Science in China Series A: Mathematics 65
2008 Vol. 51 No. 8 65–90
www.SciChina.com
www.springerlink.com

Cascadic Multigrid Methods for Parabolic Problems

Qiang Du[1] & Pingbing Ming[2,†]

[1]Department of Mathematics, Pennsylvania State University, University Park, PA 16802, USA
(email: qdu@math.psu.edu);
[2] LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, No. 55, Zhong-Guan-Cun East Road, Beijing, 100190, China (email: mpb@lsec.cc.ac.cn)

Abstract In this paper, we consider the cascadic multigrid method for a parabolic type equation. Backward Euler approximation in time and linear finite element approximation in space are employed. A stability result is established under some conditions on the smoother. Using new and sharper estimates for the smoothers that reflect the precise dependence on the time step and the spatial mesh parameter, these conditions are verified for a number of popular smoothers. Optimal error bounds are derived for both smooth and non-smooth data. Iteration strategies guaranteeing both the optimal accuracy and the optimal complexity are presented.

Keywords: Cascadic multigrid method, parabolic problem, finite element methods, backward Euler scheme, smoother, stability, optimal error order, optimal complexity

MSC(2000): 65N30, 65N55, 65F10

1 Introduction

The cascadic multigrid method presented by Deuflhard, Leinen and Yserentant in [1] is a *one-way* multigrid method which may be viewed as a multilevel method *without* the coarse mesh correction. The method dates back to Wachspress' pioneering work [2]. The basic idea of this method is to control the iteration number over successively refined mesh as long as the algebraic error is below the discretization error. The first algorithmic realization for two dimensional elliptic problems was given in [1] while the three dimensional realizations and convincing numerical results were reported in Bornemann, Erdmann and Kornhuber [3]. In Deuflhard [4], the use of a posteriori algorithmic control in combination with conjugate gradient method was proposed, suggesting more iterations on coarser levels to be used so as to perform less iteration on finer levels. Shaidurov [5] gave the first convergence proof that provides a theoretical justification of the numerical performance. Based on the *cascade* principle given in [1] that suggests the termination of the iteration when the discretization error dominates the algebraic error, Bornemann and Deuflhard [6] extended the results to the case when other traditional iteration methods are

Received ; accepted
DOI:
† Corresponding author
The work of Du was supported in part by NSF DMS-0409297, NSF DMR-0205232, NSF CCF-0430349 and NIH-NCI 1R01CA125707-01A1. The work of Ming was supported by National Natural Science Foundation of China under the grant 10571172, the National Basic Research Program under the grant 2005CB321704 and the Youth's Innovative Program of Chinese Academy of Sciences under the grants K7290312G9, K7502712F9.

employed as smoothers. Optimal error bounds for the cascadic solution were derived and the algorithm was shown to have the *multigrid complexity* [7]. Later, the cascadic multigrid method was applied to the elliptic problems in domains with re-entrant corners [8], Stokes problem [9], some indefinite and semi-linear problems [10], some mildly nonlinear problems [11,12], and more recently it was extended to the Mortar setting [13] and variational inequality [14]. In [15,16], the cascadic algorithm with non-conforming finite element discretization was considered, and in [17], the cascadic algorithm with finite volume discretization has been studied. We refer to [18] for the review of recent progress of this method.

Studies on the cascadic multigrid method for parabolic problems, have also been made during the last decade, see, e.g. [19, 20, 21]. With a discrete in time formulation, cascadic multigrid methods can be directly applied to the resulting elliptic problems by treating the time step size as a parameter. Though numerical experiments presented in [19] indicate that the method behaves quite well for parabolic problems, a complete mathematical analysis is not yet available. In fact, one important issue that has not been addressed is how the choice of parameters would affect the interplay between the stability of the algorithm and the iteration strategy. Moreover, it remains to be studied whether the optimal error bounds can be rigorously derived and if the algorithm is still of *multigrid complexity*. A key to the establishment of such results is a careful investigation of the stability properties of the cascadic multigrid algorithm when applied to parabolic problems with the time and space discretization. In turn, this requires improved estimates on the various smoothers that reflect the intrinsic spatial and temporal structures of the fully discrete approximations.

To put our work in a larger context, we note that there have been much interests in the study of the effect of iterative solvers on the numerical solution of parabolic equations with implicit-in-time discretizations [22]. Such studies are not only practically important but also theoretically interesting. In fact, it has been widely known that, for implicit in time discretizations, it is often possible to gain computational efficiency while preserving the order of accuracy through suitable approximations. To give an illustrative example, an earlier work of Dawson, Du and Dupont [23] proposed a coupled explicit/implicit domain decomposition algorithm as an alternative to a fully implicit discretization of parabolic equations. The domain decomposition algorithm may be seen as an approximation to the fully implicit scheme but with very different stability properties. Here, we also face the issue of establishing new stability estimates. Moreover, while the particular emphasis of our present paper is to give a comprehensive analysis of the cascadic multigrid method for parabolic equations, the framework and technical details may be useful in the study of other similar models and methods as well.

For the purpose of illustration, we focus on a linear parabolic problem in two dimensional space. We establish the stability of the cascadic algorithm under some conditions on the smoothers. We also prove an optimal error bound in the L^2 norm for the cascadic solution of the parabolic problem in spite of the fact that it is impossible to obtain such a bound when the cascadic algorithm is applied to a standard second order elliptic problem with the linear finite element discretization [24]. It is also worth mentioning that as addressed in [24], cascadic multigrid method is different from the idea of incomplete iteration proposed in [25, 22] and [Ch. 11, 26]. New techniques are used in our discussion to obtain the desired estimates. In addition, our analytical results provided here also give practical guidance on the choices of various parameters in the implementation of the cascadic algorithms for both the smooth and non-smooth initial data.

The rest of the paper is organized as follows: in § 2, we describe a Cascadic Algorithm for parabolic problems. In § 3, we study the time stability of the algorithm under some assumptions made on the smoothers. This is essential for the convergence of the cascadic algorithm when applied to the time-dependent problems. Using new estimates particularly suitable for parabolic type of problems, these assumptions are verified in § 4 for smoothers such as *Simple Jacobi, Symmetric Gauß-Seidel, and Conjugate Gradient.* Though many similar smoother estimates have been discussed in the literature, they are not directly applicable in our setting to derive the optimal results. Our improved estimates are generally sharper in their precise dependence on the mesh parameters and time steps. Error estimates are derived in § 5 for both smooth and non-smooth initial data. The iteration strategies are addressed in § 6 and some conclusion remarks are given in § 7.

Throughout this paper, C is always a generic constant and is independent of the mesh size h and the time step τ.

2 Cascadic algorithm for a parabolic problem

2.1 The model parabolic problem

We consider the following parabolic problem

$$\begin{cases} \dfrac{\partial u}{\partial t} + A\,u & = f \qquad \text{in } \ \Omega \times (0,T], \\ \quad u(x,t) & = 0 \qquad \text{on } \ \partial\Omega \times (0,T], \\ \quad u(x,0) & = u_0(x) \qquad \text{in } \ \Omega, \end{cases} \tag{2.1}$$

where Ω is a convex polygonal domain in $\mathbb{R}^2$ with boundary $\partial\Omega$, and A is an elliptic operator of the form:

$$A\,u = -\sum_{i,j=1}^{2} \frac{\partial}{\partial x_i}\Big(a_{ij}\frac{\partial u}{\partial x_j}\Big) + c(x)u.$$

A weak form of (2.1) is: *Find $u \in \mathcal{H}_0^1(\Omega)$, with $u(x,0) = u_0(x)$ in Ω and*

$$\Big(\frac{\partial u}{\partial t}, v\Big) + \mathcal{A}(u,v) = (f,v) \quad \forall v \in \mathcal{H}_0^1(\Omega),\ \forall t \in [0,T]. \tag{2.2}$$

Here, $\mathcal{H}_0^1(\Omega)$ is the standard Sobolev space and the bilinear form $\mathcal{A}$ is defined as

$$\mathcal{A}(v,w) = \sum_{i,j=1}^{2} \int_\Omega a_{ij}\frac{\partial v}{\partial x_i}\frac{\partial w}{\partial x_j}\,dx + \int_\Omega c(x)vw\,dx \quad \forall v,w \in \mathcal{H}_0^1(\Omega)\ ,$$

in particular, define $\|v\|_{\mathcal{A}}^2 := \mathcal{A}(v,v)$, and $(f,v) = \int_\Omega fv\,dx$ for $v \in \mathcal{H}_0^1(\Omega)$. The usual assumption on the bilinear form $\mathcal{A}$ reads

$$\begin{array}{lrcll} \text{(i)} & |\mathcal{A}(v,w)| & \le & C\|v\|_1\|w\|_1 & \forall v,w \in \mathcal{H}_0^1(\Omega), \\ \text{(ii)} & \mathcal{A}(v,v) & \ge & C\|v\|_1^2 & \forall v \in \mathcal{H}_0^1(\Omega). \end{array}$$

For the basic theory of parabolic equations and relevant function spaces, we refer to [27, 28]. For the application of classical multigrid methods to parabolic equations, see, for example, [29, 30, 31, 32, 33, 34] and [Ch. 11, 26].

For simplicity, we choose a Backward Euler scheme for the time discretization. Given a time interval $(0,T)$, let τ be the time step size, n the total number of time steps taken such that $n\tau = T$. The semi-discrete in time scheme is

$$\Big(\frac{u^k - u^{k-1}}{\tau}, v\Big) + \mathcal{A}(u^k, v) = (f^k, v) \quad \forall v \in \mathcal{H}_0^1(\Omega), k \geq 1, \tag{2.3}$$

with $u(x,0) = u_0(x)$ and $f^k = f(x, t^k)$.

2.2 Finite element discretization

Given a nested family of triangulation $\{\mathcal{T}_j\}_{j=0}^{\ell}$ with mesh parameter $\{h_j\}_{j=0}^{\ell}$. Throughout the paper, all triangulations are assumed to be *quasi-uniform* such that there exists a positive constant C satisfying $C^{-1} \leq 2^j h_j \leq C$. The family of continuous piecewise linear finite element spaces $X_0 \subset X_1 \subset \cdots \subset X_\ell$ are given by

$$X_j = \{\, u \in \mathcal{H}_0^1(\Omega) \mid u_{|_K} \in \mathcal{P}_1(K) \quad \forall K \in \mathcal{T}_j \,\},$$

where $\mathcal{P}_1(K)$ denotes the set of linear functions on the triangle K.

The fully discrete problem corresponding to (2.3) is defined as: Find $u_j^n \in X_j (0 \leq j \leq \ell)$ such that

$$\Big(\frac{u_j^n - u_j^{n-1}}{\tau}, v\Big) + \mathcal{A}(u_j^n, v) = (f^n, v) \qquad \forall v \in X_j. \tag{2.4}$$

Denote by $\mathcal{R}_h u \in X_\ell$ the elliptic projection with respect to $\mathcal{A}$, and $\mathcal{P}_h$ the L^2 projection on X_ℓ. Define an auxiliary bilinear form as

$$\mathcal{A}_\tau(w,v) := \tau^{-1}(w,v) + \mathcal{A}(w,v) \qquad \forall w, v \in \mathcal{H}_0^1(\Omega),$$

We define the Cascadic Algorithm for solving (2.1) as follows:

CASCADIC ALGORITHM for problem (2.1).

Step 1: For $n = 0, u_*^0 = \mathcal{P}_h u_0$.

Step 2: Once u_*^{n-1} is known, u_*^n is defined as follows: for $j = 0$, solve finite element equations

$$\mathcal{A}_\tau(w_0^n, v) = (f^n, v) - \mathcal{A}(u_*^{n-1}, v) \quad \forall v \in X_0$$

exactly, and let $w_0^{n,*} = w_0^n$.
For $j = 1, \cdots, \ell$, let $w_j^{n,*} = \mathcal{C}_{j,m_{j,n}} w_{j-1}^{n,*}$ and $w_*^n = w_\ell^{n,*}$. We then let $u_*^n = w_*^n + u_*^{n-1}$, where $\mathcal{C}_{j,m_{j,n}}$ denotes the $m_{j,n}$ steps of a basic iteration applied on level j at time step n.

Here, for simplicity, we have dropped the index ℓ for the u_*^n which always refers to the Cascadic solution at time step t_n and level ℓ.

We call a cascadic multigrid algorithm *optimal* on level ℓ if the algebraic error is commensurate with the discretization error, i.e.,

$$\|u_*^n - u_\ell^n\|_\tau \approx \|u^n - u_\ell^n\|_\tau,$$

and with *multigrid complexity* if the amount of work on time step t_n is $\mathcal{O}(n_\ell)$, where $n_\ell = \dim X_\ell$.

2.3 Additional notations and technical lemmas

The following lemma gives the regularity of the resulting elliptic problem, the proof is standard (see [35]).

Lemma 2.1. *For a given $g \in \mathcal{H}^{-1}(\Omega)$, the problem*

$$\mathcal{A}_\tau(w, v) = (g, v) \qquad \forall v \in \mathcal{H}_0^1(\Omega)$$

has a unique solution $w \in \mathcal{H}_0^1(\Omega)$, and if $g \in L^2(\Omega)$, then w admits the following regularity estimate:

$$\tau^{-1/2}\|w\|_1 + \|w\|_2 \leq C_R\|g\|_0\ , \tag{2.5}$$

for some constant C_R.

Let us define the τ-norm by $\|v\|_\tau^2 := \mathcal{A}_\tau(v, v)$, the τ-inner product by $(v, w)_\tau := \mathcal{A}_\tau(v, w)$ for any $v, w \in \mathcal{H}_0^1(\Omega)$, and the orthogonal subspaces by

$$X_{j-1}^\perp := \{\, v \in X_j \mid (v, w)_\tau = 0 \quad \forall w \in X_{j-1}\,\}. \tag{2.6}$$

For $0 \leq j \leq \ell$, we define some linear operators $A_{\tau,j}: X_j \to X_j$ by

$$(A_{\tau,j}v, w) := \mathcal{A}_\tau(v, w) \quad \forall v, w \in X_j\ .$$

Note that $A_{\tau,j} = \tau^{-1}I + A_j$ is positive definite with A_j defined by

$$(A_jv, w) := \mathcal{A}(v, w) \quad \forall v, w \in X_j.$$

In particular, we let $\mathcal{A}_h = A_\ell$. Denote by $\hat{\lambda}_j$ and $\hat{\lambda}_1$ the largest and smallest eigenvalue of $A_{\tau,j}$ and by κ_j the condition number of $A_{\tau,j}$. We see that $\hat{\lambda}_j = \tau^{-1} + \lambda_j$, with λ_j the largest eigenvalue of A_j. It is well known that $\lambda_j = \mathcal{O}(h_j^{-2})$.

As a convention, we let $\|B\|$ be the matrix norm $\|B\| := \sup_{\|x\|=1} x^{\mathrm{T}}Bx$ for any matrix B, and $\rho(B)$ be its spectra radius, and $\kappa(B)$ be its condition number.

3 Stability

For the sake of clarity, we first present a new stability analysis of the Cascadic Algorithm for the parabolic equations under some assumptions on the smoothers. We assume that the smoothers satisfy: for $j = 1, \ldots, \ell$ and $k = 1, \ldots, n$,

$$\begin{aligned} \|\mathcal{C}_{j,m_{j,k}}v\|_\tau &\leq \|v\|_\tau && \forall v \in X_j, \\ \|\mathcal{C}_{j,m_{j,k}}v\|_\tau &\leq \gamma_{j,k}\|v\|_\tau && \forall v \in X_{j-1}^\perp\ . \end{aligned} \tag{3.1}$$

Detailed derivation of the above estimates are presented later for some smoothers of interests (see Theorem 4.2, Corollary 4.3 and Theorem 4.4).

Theorem 3.1. *Under the condition*

$$\sum_{j=1}^{\ell} \gamma_{j,k}^2 < 1 \qquad \textit{for}\ \ k = 1, \ldots, n, \tag{3.2}$$

the Cascadic Algorithm is stable in the sense that the solution u_*^n *satisfies*

$$\|u_*^n\|_{\mathcal{A}}^2 \le C\|u_0\|_{\mathcal{A}}^2 + \sum_{k=1}^n \tau\|f^k\|_0^2. \tag{3.3}$$

Proof. For $1 \le k \le n$ and $1 \le j \le \ell$, let $\overline{u}_\ell^k$ be the solution of

$$\mathcal{A}_\tau(\overline{u}_\ell^k, v) = (f^k, v) + \tau^{-1}(u_*^{k-1}, v) \qquad \forall v \in X_\ell\ . \tag{3.4}$$

And we define $\overline{w}_j^k \in X_j$ satisfying

$$\mathcal{A}_\tau(\overline{w}_j^k, v) = (f^k, v) - \mathcal{A}(u_*^{k-1}, v) \qquad \forall v \in X_j.$$

Comparing with the algorithm, we have $\overline{u}_\ell^k = \overline{w}_\ell^k + u_*^{k-1}$, thus $u_*^k - \overline{u}_\ell^k = w_*^k - \overline{w}_\ell^k$ and a bound on $u_*^k - \overline{u}_\ell^k$ can be found by getting a bound on $w_*^k - \overline{w}_\ell^k$.

Similar to [6], we note that for any $1 \le k \le n$ and $1 \le j \le \ell$,

$$\begin{aligned} w_j^{k,*} - \overline{w}_j^k &= \mathcal{C}_{j,m_{j,k}}(w_{j-1}^{k,*} - \overline{w}_j^k) \\ &= \mathcal{C}_{j,m_{j,k}}(w_{j-1}^{k,*} - \overline{w}_{j-1}^k) + \mathcal{C}_{j,m_{j,k}}(\overline{w}_{j-1}^k - \overline{w}_j^k). \end{aligned} \tag{3.5}$$

Invoking (3.1) as well as (3.5) yields

$$\|w_j^{k,*} - \overline{w}_j^k\|_\tau \le \|w_{j-1}^{k,*} - \overline{w}_{j-1}^k\|_\tau + \gamma_{j,k}\|\overline{w}_j^k - \overline{w}_{j-1}^k\|_\tau.$$

A recursive application of the above inequality leads to

$$\|w_*^k - \overline{w}_\ell^k\|_\tau \le \sum_{j=1}^\ell \gamma_{j,k}\, \|\overline{w}_j^k - \overline{w}_{j-1}^k\|_\tau. \tag{3.6}$$

Using Cauchy-Schwartz inequality and $\|\overline{w}_j^k - \overline{w}_{j-1}^k\|_\tau^2 = \|\overline{w}_j^k\|_\tau^2 - \|\overline{w}_{j-1}^k\|_\tau^2$, we get

$$\begin{aligned} \|w_*^k - \overline{w}_\ell^k\|_\tau &\le \Big(\sum_{j=1}^\ell \gamma_{j,k}^2\Big)^{1/2}\Big(\sum_{j=1}^\ell \|\overline{w}_j^k - \overline{w}_{j-1}^k\|_\tau^2\Big)^{1/2} \\ &= \Big(\sum_{j=1}^\ell \gamma_{j,k}^2\Big)^{1/2}\Big(\sum_{j=1}^\ell \|\overline{w}_j^k\|_\tau^2 - \|\overline{w}_{j-1}^k\|_\tau^2\Big)^{1/2} \le \Big(\sum_{j=1}^\ell \gamma_{j,k}^2\Big)^{1/2}\|\overline{w}_l^k\|_\tau\ . \end{aligned}$$

In view of the assumption (3.2), we obtain

$$\|u_*^k - \overline{u}_\ell^k\|_\tau \le \|\overline{u}_\ell^k - u_*^{k-1}\|_\tau,$$

which implies

$$\|u_*^k\|_\tau^2 \le \|u_*^{k-1}\|_\tau^2 + 2(u_*^k - u_*^{k-1}, \overline{u}_\ell^k)_\tau.$$

Notice that $\mathcal{A}_\tau$ is symmetric and using (3.4), we have

$$\begin{aligned} (u_*^k - u_*^{k-1}, \overline{u}_\ell^k)_\tau &= (\overline{u}_\ell^k, u_*^k - u_*^{k-1})_\tau = (f^k, u_*^k - u_*^{k-1}) + \tau^{-1}(u_*^{k-1}, u_*^k - u_*^{k-1}) \\ &= (f^k, u_*^k - u_*^{k-1}) + \frac{1}{2\tau}(\|u_*^k\|_0^2 - \|u_*^{k-1}\|_0^2 - \|u_*^k - u_*^{k-1}\|_0^2) \\ &\le \frac{\tau}{2}\|f^k\|_0^2 + \frac{1}{2\tau}(\|u_*^k\|_0^2 - \|u_*^{k-1}\|_0^2). \end{aligned}$$

A combination of the above two inequalities leads to

$$\|u_*^k\|_\tau^2 \le \|u_*^{k-1}\|_\tau^2 + \tau\|f^k\|_0^2 + \frac{1}{\tau}(\|u_*^k\|_0^2 - \|u_*^{k-1}\|_0^2),$$

which in turn implies

$$\|u_*^k\|_{\mathcal{A}}^2 \le \|u_*^{k-1}\|_{\mathcal{A}}^2 + \tau\|f^k\|_0^2.$$

Finally, a recursive application of the above inequality and using

$$\|u_*^0\|_{\mathcal{A}} = \|\mathcal{P}_h u_0\|_{\mathcal{A}} \le C\|u_0\|_{\mathcal{A}}$$

yields (3.3). □

Remark 3.2. *By Theorem (3.1), we see that sufficiently many smoothing operations at each time step would not affect the stability of the marching algorithm, even though the discrete solutions are only computed approximately. The condition (3.2) allows us to quantitatively characterize the properties of the smoothers to guarantee the stability in time. It will be shown later that efficient iteration strategies can be developed for several popular smoothers so that both the stability property and the optimal multigrid complexity hold simultaneously. This in turn implies the convergence of the cascadic algorithms with both optimal accuracy and optimal complexity.*

4 Smoothers

To avoid complicated notation, we focus on the smoother estimate at a particular time step. Thus, we drop the subscript k used for indexing the time steps. For example, we simply use $\mathcal{C}_{j,m_j}$ to denote the basic iterations applied m_j times on level j. As in [6], we call the basic iteration a *smoother*, if it satisfies

$$\|\mathcal{C}_{j,m_j}v\|_a \le \|v\|_a \ , \ \|\mathcal{C}_{j,m_j}v\|_a \le C\frac{h_j^{-1}}{m_j^\gamma}\|v\|_0 \ , \quad \forall v \in X_j, \tag{4.1}$$

where $\|\cdot\|_a$ is the energy norm corresponding to the basic iteration, that is, in our case, $\|\cdot\|_a = \|\cdot\|_\tau$. It is known that $\gamma = 1/2$ for *Simple Jacobi, Symmetric Gauß-Seidel, SSOR* [7] and $\gamma = 1$ for *Conjugate Gradient* iterations [6, 7, 36, 5, 8]. Notice that in practice, it is expected that an increase in iteration number should lead to a decrease of $\|\mathcal{C}_{j,m_j}v\|_a/\|v\|_a$; similarly, the smaller κ_j is, the smaller $\|\mathcal{C}_{j,m_j}v\|_a/\|v\|_a$ and $\|\mathcal{C}_{j,m_j}v\|_a/\|v\|_0$ ought to be. Unfortunately, such expected behaviors are not reflected in (4.1). In addition, the dependence on h and τ is also not explicitly revealed. In fact, the smoother estimates derived in the literature usually do not make a clear and precise distinction on the effects of h and τ in the smoothing step. We now derive some new estimates for the afore-mentioned smoothers with respect to τ-norm. Two cases are discriminated, one for the usual symmetric iteration, another for *Conjugate Gradient* iteration.

4.1 Symmetric iterations

For symmetric iterations, the iteration matrix usually takes the form $\mathcal{S} = I - W^{-1}B$, with smoother $\mathcal{S}_m = \mathcal{S}^m$, $m \in \mathbb{N}$. Here, W and B are operators (matrices) from X_j to X_j, and I is the identity operator. Denote the energy norm by $\|x\|_a := (Bx, x)$ for any $x \in X_j$.

For our discussion, we only consider the symmetric iterations satisfying the following general assumption: 1) B is symmetric and positive definite; 2)W is regular with $W = W^{\mathrm{T}}$ and 3) $W \geq B$, i. e., $W - B$ is positive definite.

The following theorems contain smoother estimates along the same spirits of those obtained in [37, 5, 20]. We omit some technical derivations but emphasize on the precise nature of the estimates particularly suitable to parabolic problems.

Theorem 4.1. *Under the above assumptions, we have that for any $v \in X_j$,*

$$\begin{aligned} \|\mathcal{S}_m v\|_a &\leq \frac{\rho(I - W^{-1}B)^i}{\sqrt{2(m-i)}} \|W - B\|^{1/2} \|v\|_0 \quad i \in [0, m), \\ \|\mathcal{S}_m v\|_a &\leq \frac{\rho(I - W^{-1}B)^i}{\sqrt{2(m-i)+1}} \|W\|^{1/2} \|v\|_0 \qquad i \in [0, m] . \end{aligned} \tag{4.2}$$

Proof. Let $C = W^{-1/2} B W^{-1/2}$, we have $0 \leq C \leq I$. Since $I - W^{-1}B = W^{-1/2}(I - C)W^{1/2}$, we get $\mathcal{S}_m = (I - W^{-1}B)^m = W^{-1/2}(I - C)^m W^{1/2}$ and

$$\|\mathcal{S}_m v\|_a^2 = (B\mathcal{S}_m v, \mathcal{S}_m v) = \big(C(I-C)^{2m} w, w\big)$$

with $w = W^{1/2} v$, Then for $i \in [0, m)$,

$$\begin{aligned} \|\mathcal{S}_m v\|_a^2 &\leq \rho(I-C)^{2i} \big(C(I-C)^{2(m-i)} w, w\big) \\ &= \Big((I-C)^{2m-2i} w - (I-C)^{2m-2i+1} w, w\Big) \rho(I-C)^{2i} \\ &\leq \frac{\rho(I-C)^{2i}}{2(m-i)} \Big(\sum_{k=1}^{2m-2i} (I-C)^k w - (I-C)^{k+1} w, w\Big) \\ &\leq \frac{\rho(I-C)^{2i}}{2(m-i)} (w - Cw, w) \leq \frac{\rho(I - W^{-1}B)^{2i}}{2(m-i)} \|W - B\| \, \|v\|_0^2. \end{aligned}$$

This gives $(4.2)_1$. For $(4.2)_2$, we note that for $i \in [0, m]$,

$$\|\mathcal{S}_m v\|_a^2 \leq \frac{\rho(I-C)^{2i}}{2(m-i)+1} (w, w) \leq \frac{\rho(I - W^{-1}B)^{2i}}{2(m-i)+1} \|W\| \, \|v\|_0^2 \, .$$

□

Applying $(4.2)_2$ to the *Simple Jacobi* iteration gives

$$\|\mathcal{C}_{j,m_j} v\|_\tau \leq \Big(\frac{\hat{\lambda}_j}{2m_j - 2i + 1}\Big)^{1/2} \Big(\frac{\lambda_j}{\tau^{-1} + \lambda_j}\Big)^i \|v\|_0 \,, \quad \forall v \in X_j \,,\ \forall i \in [0, m_j] \,. \tag{4.3}$$

For the *Symmetric Gauß-Seidel*, the following lemma is given as a remark in [38]. A slightly weaker form valid for more general matrices and norm is given in [39].

Lemma 4.2. *For any real $n \times n$, $m-$band symmetric positive definite matrix B with $\lambda_{\max}(B)$ and the $\lambda_{\min}(B)$ being the largest and smallest eigenvalues and L being its lower triangular part, we have for some constant C and $C_L = C \log 2m$ that*

$$\|L\| \leq C_L [\lambda_{\max}(B) - \lambda_{\min}(B)] \, . \tag{4.4}$$

It is easy to see that (4.4) can be rewritten as

$$\|L\| \le C_L \lambda_{\max}(B)\big(1 - 1/\kappa(B)\big). \tag{4.5}$$

The iteration matrix for the *Gauß-Seidel* is $M_{GS} = -(D_j + L_j)^{-1} L_j^T$ with $A_{\tau,j} = D_j - L_j - L_j^{\mathrm{T}}$, where D_j and L_j are the diagonal part and the lower triangular part of $A_{\tau,j}$. By [40], the *Gauß-Seidel* iteration admits the bound:

$$\|M_{GS}\|_\tau^2 = 1 - \|A_{\tau,j}^{-1/2}(D_j + L_j)D_j^{-1/2}\|^{-2}. \tag{4.6}$$

We now have the following theorem for the *Symmetric Gauß-Seidel* iteration.

Theorem 4.3. *Assume the diagonal part of* $A_{\tau,j}$ *admits the following estimate*

$$\|D_j^{-1/2}\| \le C_D \hat{\lambda}_j^{-1/2}, \tag{4.7}$$

then for any $v \in X_j$, *the* Symmetric Gauß-Seidel *iteration satisfies, for* $i \in [0, m_j)$,

$$\|\mathcal{S}_{j,m_j} v\|_\tau \le C_D C_L \Big(\frac{\lambda_j}{2m_j - 2i}\Big)^{1/2} \Big(\frac{\lambda_j}{C_{GS}\tau^{-1} + \lambda_j}\Big)^i \|v\|_0, \tag{4.8}$$

and for $i \in [0, m_j]$,

$$\|\mathcal{S}_{j,m_j} v\|_\tau \le (1 + C_D^2 C_L^2)^{1/2} \Big(\frac{\hat{\lambda}_j}{2m_j - 2i + 1}\Big)^{1/2} \Big(\frac{\lambda_j}{C_{GS}\tau^{-1} + \lambda_j}\Big)^i \|v\|_0 \tag{4.9}$$

with $C_{GS} = 1/(1 + C_D C_L)^2$.

Proof. By Theorem 4.1, we only need to estimate terms like $\|W_j - A_{\tau,j}\|$, $\|W_j\|$ and $\rho(I - W_j^{-1} A_{\tau,j})$. Note that $W_j - A_{\tau,j} = L_j D_j^{-1} L_j^T$, by Lemma 4.2, we get

$$\|W_j - A_{\tau,j}\| \le C_D^2 \|L_j\|^2 / \hat{\lambda}_j \le C_D^2 C_L^2 (\hat{\lambda}_j - \hat{\lambda}_1)^2 / \hat{\lambda}_j \le C_D^2 C_L^2 \lambda_j, \tag{4.10}$$

which together with the triangle inequality leads to

$$\|W_j\| \le (1 + C_D^2 C_L^2) \hat{\lambda}_j \ . \tag{4.11}$$

We now turn to (4.8). Resorting to Lemma 4.2 once again, we obtain

$$\begin{aligned} \|(D_j + L_j) D_j^{-1/2}\| &\le \|D_j^{1/2}\| + \|L_j\| \, \|D_j^{-1/2}\| \le \hat{\lambda}_j^{1/2} + C_D C_L (\hat{\lambda}_j - \hat{\lambda}_1) \hat{\lambda}_j^{-1/2} \\ &\le \hat{\lambda}_j^{1/2} \big(1 + C_D C_L (1 - 1/\kappa_j)\big). \end{aligned} \tag{4.12}$$

A combination of (4.12) and (4.6) gives

$$\|M_{GS}\|_\tau^2 \le 1 - \frac{1}{\kappa_j \big(1 + C_D C_L (1 - 1/\kappa_j)\big)^2}. \tag{4.13}$$

A simple calculation yields

$$\begin{aligned} \kappa_j (1 + C_D C_L (1 - 1/\kappa_j))^2 &\le \kappa_j + 2 C_D C_L (\kappa_j - 1) + C_D^2 C_L^2 (\kappa_j - 1) \\ &= 1 + (1 + C_D C_L)^2 (\kappa_j - 1). \end{aligned} \tag{4.14}$$

With $C_{GS} = 1/(1 + C_D C_L)^2$, it follows from (4.13) and (4.14) that

$$\|M_{GS}\|_\tau^2 \le \frac{\kappa_j - 1}{\kappa_j - 1 + C_{GS}}.$$

Note that $\kappa_j = (\tau^{-1} + \lambda_j)(\tau^{-1} + \lambda_1)^{-1}$, we thus have

$$\|M_{GS}\|_\tau^2 \le \frac{\lambda_j - \lambda_1}{C_{GS}(\tau^{-1} + \lambda_1) + \lambda_j - \lambda_1} \le \frac{\lambda_j}{C_{GS}\tau^{-1} + \lambda_j}.$$

By [40, Theorem 4.8.10], the spectral radius of the *Symmetric Gauß-Seidel* iteration $\rho(S_{GS}) = \|M_{GS}\|_\tau^2$, using (4.10) and (4.11), we get (4.8) and (4.9), respectively. □

4.2 Conjugate Gradient iterations

We now give an estimate for *Conjugate Gradient* (CG) iterations. The classical approach for estimating the convergence rate of the CG-iteration is to exploit *dominated polynomials* that may yield different bounds. Let $\hat{\mathcal{Q}}_k$ be the scaled Chebyshev polynomial defined as

$$\hat{\mathcal{Q}}_k(x) = \mathcal{C}_k(x/d)/\mathcal{C}_k(1/d) \qquad \text{for } x \in [0, d].$$

Here, $\mathcal{C}_k(x) = \cos(k \arccos(x))$ for $x \in [-1, 1]$ is the kth degree Chebyshev polynomial of the first kind. Let $p_k = \sqrt{d}/(2k+1)$, the Lanczos polynomial [42] is defined as

$$\sqrt{x}\mathcal{Q}_k(x) = (-1)^k p_k \cos\big((2k+1)\arccos(\sqrt{x/d})\big) \qquad \text{for } x \in [0, d].$$

For any $i \in [0, k]$, define $\mathcal{S}_k^i(x) := \hat{\mathcal{Q}}_i(x)\mathcal{Q}_{k-i}(x)$. $\mathcal{Q}_k$ and $\mathcal{S}_k^i$ satisfy (see [41], [§4.1, 36], [Lemma 3.1, 8] and [7, 43]):

Lemma 4.4. *For interval* $[0, b]$, $a \in [0, b]$, *integers* k, *and* $i \in [0, k]$,

1. *For any* k,

$$\max_{0 \le x \le b} |\mathcal{Q}_k(x)| \le 1 \qquad \textit{and} \qquad \max_{0 \le x \le b} |\sqrt{x}\mathcal{Q}_k(x)| \le \sqrt{b}/(2k+1)\ .$$

2. $\mathcal{S}_k^i(0) = 1$,

$$\max_{0 \le x \le b} |\mathcal{S}_k^i(x)| \le 1, \qquad \textit{and} \qquad \max_{a \le x \le b} |\mathcal{S}_k^i(x)| \le 2\Big(\frac{\sqrt{b} - \sqrt{a}}{\sqrt{b} + \sqrt{a}}\Big)^i\ .$$

3. *For weight* $\sqrt{x}$,

$$\max_{0 \le x \le b} |\sqrt{x}\mathcal{S}_k^i(x)| \le \frac{2\sqrt{b}}{2(k-i)+1}\Big(\frac{\sqrt{b} - \sqrt{a}}{\sqrt{b} + \sqrt{a}}\Big)^i\ .$$

We now define a family of auxiliary operators by

$$\mathcal{S}_{j,m_j}^i := \mathcal{S}_{m_j}^i(A_{\tau,j}) \quad \forall i \in [0, m_j], \tag{4.15}$$

which dominate the error reduction operator $\mathcal{C}_{j,m_j}$ for the CG-method and they are smoothers in the sense of (4.1).

Theorem 4.5. *Define $\mathcal{S}^i_{j,m_j}$ as in (4.15), then for any $v \in X_j$, there holds*

$$\|\mathcal{S}^i_{j,m_j} v\|_\tau \le \|v\|_\tau,$$
$$\|\mathcal{S}^i_{j,m_j} v\|_\tau \le \frac{2\hat{\lambda}_j^{1/2}}{2(m_j - i) + 1}\Big(\frac{\lambda_j}{4\tau^{-1} + \lambda_j}\Big)^i \|v\|_0.$$

The proof of the above theorem is standard (see [8]) and we omit the details.

4.3 Smoother estimates on orthogonal subspaces

It is known that the smoother on the level j actually only damps out the error components in some subspaces rather than the entire space. To be more precise, we will translate our previous estimate for the smoother $\mathcal{S}_{j,m_j}$ into one confined to the subspace $X^\perp_{j-1}$ instead of X_j, here $X^\perp_{j-1}$ is defined as in (2.6). Such kind of refined estimate is crucial for the convergence study of classical multigrid method [37, 44], while it is not yet exploited in the present setting. We start from the following lemma which is actually a dual estimate for the parabolic problem.

Lemma 4.6. *Let $u_j \in X_j$ satisfy the following finite element approximation*

$$\mathcal{A}_\tau(u_j, v) = 0 \qquad \forall v \in X_{j-1}. \tag{4.16}$$

Let C_I be a constant in the following estimate

$$\inf_{v \in X_{j-1}} \|u - v\|_\tau \le C_I \lambda_j^{-1/2}(\tau^{-1/2}\|u\|_1 + \|u\|_2), \tag{4.17}$$

and C_R be defined in (2.5). We have for $C_B = max(1, C_I C_R)$ that,

$$\|u_j\|_0 \le C_B \hat{\lambda}_j^{-1/2} \|u_j\|_\tau. \tag{4.18}$$

Proof. Resorting to the Aubin-Nitsche trick, we let $w \in \mathcal{H}^1_0(\Omega)$ satisfy

$$\mathcal{A}_\tau(v, w) = (u_j, v) \qquad \forall v \in \mathcal{H}^1_0(\Omega). \tag{4.19}$$

By virtue of (2.5), we have

$$\tau^{-1/2}\|w\|_1 + \|w\|_2 \le C_R \|u_j\|_0. \tag{4.20}$$

Take $v = u_j$ on the right-hand side of (4.19), let $\Pi w \in X_{j-1}$ be the Clément interpolant of w [45], using (4.16), (4.17) and (4.20), we have

$$\begin{aligned}\|u_j\|_0^2 &= \mathcal{A}_\tau(u_j, w) = \mathcal{A}_\tau(u_j, w - \Pi w)\\ &\le \|u_j\|_\tau \|w - \Pi w\|_\tau \le C_I C_R \lambda_j^{-1/2} \|u_j\|_\tau \|u_j\|_0,\end{aligned}$$

so $\|u_j\|_0 \le C_I C_R \lambda_j^{-1/2} \|u_j\|_\tau$. Together with the bound $\|u_j\|_0 \le \tau^{1/2}\|u_j\|_\tau$, we get

$$(\tau^{-1} + \lambda_j)\|u_j\|_0^2 \le \max(1, C_I^2 C_R^2)\|u_j\|_\tau^2.$$

This in turn implies (4.18). □

Combining Theorem 4.3, Theorem 4.5 and Lemma 4.6, we have:

Theorem 4.7. *The Symmetric smoothers and the CG smoother $\mathcal{C}_{j,m_j}$ satisfy*

$$\|\mathcal{C}_{j,m_j} v\|_\tau \le \gamma_j(i)\|v\|_\tau \quad \forall v \in X_{j-1}^\perp , \tag{4.21}$$

where

$$\gamma_j(i) = \frac{C_S C_B}{\big(2(m_j - i) + 1\big)^\gamma}\Big(\frac{\lambda_j}{C_*\tau^{-1} + \lambda_j}\Big)^i \quad i \in [0, m_j],$$

with $\gamma = 1/2$ for the Symmetric smoothers and $\gamma = 1$ for the CG smoother; C_B is defined in Lemma 4.6; C_S and C_* are constants depending on the smoother, defined as in previous theorems.

Remark 4.8. *Note that in practice, we may allow m_j to vary not only with the spatial level j, but also with the temporal step k. Thus, in such case, m_j and γ_j should be replaced by $m_{j,k}$ and $\gamma_{j,k}$ just like that in the previous section .*

5 Convergence analysis

We now present the error estimate for our algorithm. Discussions of convergence of other multigrid methods for parabolic problems have been given, for example, in [33].

In simple matrix terms, the Backward Euler method is given by:

$$(I + \tau\mathcal{B})U^n = U^{n-1} + \tau f^n, \quad \text{for } n \ge 1, \quad \text{with } U^0 = v, \tag{5.1}$$

where $\mathcal{B}$ is a positive definite self-adjoint operator in the Hilbert space $\mathcal{H}$. Let $|v| = \|(I + \tau\mathcal{B})^{1/2}v\|$ where $\|\cdot\|$ is the norm in $\mathcal{H}$. The corresponding dual norm and the associated s-norms are defined by

$$|v|_* = \|(I + \tau\mathcal{B})^{-1/2}v\| \ , \ \ |v|_s = \|\mathcal{B}^{s/2}v\| \ \text{ and } |v|_{*,s} = |\mathcal{B}^{s/2}v|_*. \tag{5.2}$$

In case of $\mathcal{B} = \mathcal{A}_h$, we use instead the notation $|\chi|_{-s,h} = \|\mathcal{A}_h^{-s/2}\chi\|_0$.

First, we state some stability estimates for the Backward Euler scheme:

Lemma 5.1. *Let U^n be the solutions of (5.1), $\overline{\partial}U^n = \tau^{-1}(U^n - U^{n-1})$, and $p \ge 0$. Then for $n \ge 1$ and $t_n = n\tau$,*

$$t_n^p\|U^n\|^2 + \tau\sum_{k=1}^n t_k^p|U^k|_1^2 \le C(|v|_{-p}^2 + \tau^p\|v\|^2) + C\tau\sum_{k=1}^n(|f^k|_{-p-1}^2 + t_k^p|f^k|_{-1}^2) \ , \tag{5.3}$$

$$\tau\sum_{k=1}^n t_k^p|\overline{\partial}U^k|^2 \le C(\tau^{p-1}|v|^2 + |v|_{*,-p+1}^2) + C\tau\sum_{k=1}^n(t_k^p|f^k|_*^2 + |f^k|_{*,-p}^2) \ , \tag{5.4}$$

$$t_n^p|U^n|_1^2 \le C(\tau^{p-1}|v|^2 + |v|_{*,-p+1}^2) + C\tau\sum_{k=1}^n(t_k^p|f^k|_*^2 + |f^k|_{*,-p}^2) \ . \tag{5.5}$$

Proof. The estimates (5.3) and (5.4) are derived in [Lemma 10.3, 26] and [Lemma 11.1, 26], respectively. To prove (5.5), by eigen-decomposition, it suffices to consider the scalar case with $\mathcal{B} = \mu > 0$. For such a case, (5.5) reduces to

$$\tau^p n^p \mu(U^n)^2 \le C(\tau^{p-1}(1+\tau\mu) + \mu^{-p+1}(1+\tau\mu)^{-1})v^2 + C\tau\sum_{k=1}^n(1+\tau\mu)^{-1}|f^k|^2(k^p\tau^p + \mu^{-p}) \ .$$

Replacing $\tau\mu$ by λ and τf^k by g^k, we have

$$n^p(U^n)^2 \le C(1+1/\lambda+\lambda^{-p}(1+\lambda)^{-1})v^2 + \frac{C}{\lambda(1+\lambda)}\sum_{j=1}^{n} g^{2j}(j^p+\lambda^{-p}) \ , \tag{5.6}$$

The proof of the above inequality can be made in two cases, first for $g^j = 0$ with $j \ge 1$ and $v = 1$, then for $v = 0$. The final results follow from the linearity of the equation.

In the first case we have by the defining equation, $U^n = (1+\lambda)^{-n}$ for $n \ge 0$. It is easy to see that there exists a constant $C > 0$ such that

$$n^p(1+\lambda)^{-2n} \le C(\lambda^{-p}(1+\lambda)^{-1}+1+1/\lambda) \ ,$$

for any n, which implies (5.6).

In the second case we have

$$U^n = \sum_{j=1}^{n}(1+\lambda)^{-(n+1-j)}g^j \qquad \text{for } \ j \ge 1.$$

Using the inequality

$$n^p \le C(p)(j^p+(n-j)^p) \quad \text{with } C(p) = \max(2^{p-1},1),$$

we obtain that

$$\begin{aligned} n^p(U^n)^2 &\le \frac{C(p)}{(1+\lambda)^2}\Big(\sum_{j=0}^{n-1}(1+\lambda)^{-j}(j^{p/2}+(n-j)^{p/2})g^{n-j}\Big)^2 \\ &\le \frac{C(p)}{(1+\lambda)^2}\sum_{j=0}^{n-1}(1+\lambda)^{-2j}\sum_{j=1}^{n} j^p g^{2j} + \frac{C(p)}{(1+\lambda)^2}\sum_{j=0}^{n-1}(1+\lambda)^{-2j}j^p\sum_{j=1}^{n} g^{2j} \\ &=: I_1 + I_2. \end{aligned}$$

I_1 can be easily bounded as

$$|I_1| \le \frac{C(p)}{(1+\lambda)^2}\frac{1}{1-(1+\lambda)^{-2}}\sum_{j=1}^{n} j^p g^{2j} \le \frac{C(p)}{\lambda(1+\lambda)}\sum_{j=1}^{n} j^p g^{2j}. \tag{5.7}$$

Using the inequality

$$\sum_{j=1}^{\infty} j^p x^j \le Cx(1-x)^{-p-1} \qquad \text{for } \ 0 \le x < 1,$$

we have

$$\begin{aligned} \frac{1}{(1+\lambda)^2}\sum_{j=0}^{n-1}(1+\lambda)^{-2j}j^p &\le C(1+\lambda)^{-4}\big(1-(1+\lambda)^{-2}\big)^{-p-1} \\ &= C\lambda^{-p-1}(1+\lambda)^{2p-2}(\lambda+2)^{-p-1}. \end{aligned}$$

If $\lambda \ge 1$, we have

$$\frac{(1+\lambda)^{2p-2}}{\lambda^{p+1}(\lambda+2)^{p+1}} = \frac{1}{\lambda(1+\lambda)}\Big(1+\frac{1}{\lambda(2+\lambda)}\Big)^p \le \Big(\frac{4}{3}\Big)^p\frac{1}{6\lambda(1+\lambda)}.$$

If $0 < \lambda < 1$, we have

$$\frac{(1+\lambda)^{2p-2}}{\lambda^{p+1}(\lambda+2)^{p+1}} = \frac{1}{\lambda^{p+1}(1+\lambda)}\frac{(1+\lambda)^{2p}}{(1+\lambda)(2+\lambda)^{p+1}} \le 2^{p-1}\lambda^{-p-1}(1+\lambda)^{-1}.$$

Combining the above two inequalities leads to

$$|I_2| \le \frac{C_1(p)}{\lambda(1+\lambda)}\sum_{j=1}^{n} g^{2j}(1+\lambda^{-p}) \le \frac{C_1(p)}{\lambda(1+\lambda)}\sum_{j=1}^{n} g^{2j}(j^p+\lambda^{-p}), \tag{5.8}$$

with $C_1(p) = C(p)\max((4/3)^p/6, 2^{p-1})$. A combination of (5.7) and (5.8) gives (5.6) and thus (5.5). □

5.1 Convergence for the smooth data

Note that for smooth data, whenever the Backward Euler scheme is applicable in the time discretization, it is customary to have $\tau \ge Ch_\ell^2$ with some positive constant C. Thus, a simple calculation shows that there exists $j_0 \in [1,\ell]$ such that $\lambda_{j_0} \le \tau^{-1} < \lambda_{j_0+1}$. We express γ_j as follows,

$$\gamma_j = \begin{cases} C_S C_B\Big(\frac{\lambda_j}{C_*\tau^{-1}+\lambda_j}\Big)^{m_j} & \text{if } j \in [0, j_0-1]; \\ \frac{C_S C_B}{(2(m_j-1)+1)^\gamma}\frac{\lambda_j}{C_*\tau^{-1}+\lambda_j} & \text{if } j \in [j_0,\ell]\ , \end{cases} \tag{5.9}$$

where $C_B = max(1, C_I C_R)$ as in the Lemma 4.6. The choice of constants C_*, C_S and γ depends on the particular smoother, such constants for several smoothers are listed in the following table.

Table 1: Constants in the estimate of smoother (5.9)

Smoother	C_*	C_S	γ
Simple Jacobi	1	1	1/2
S-GS	C_{GS}	$(1+C_D^2C_L^2)^{1/2}$	1/2
CG	4	2	1

We note that there is a mild dependence of C_S on j (or ℓ) in our theoretical estimates (due to the dependence on the bandwidth as in Lemma 4.2) for the *Symmetric Gauß-Seidel* smoother.

Let $\mathcal{K} = \sum_{j=1}^{\ell}\gamma_j$ with γ_j defined in (5.9). By Theorem 3.1, the Cascadic Algorithm is stable if $\mathcal{K} < 1$. Obviously, we have $\mathcal{K} < \beta$ by the expression of γ_j where

$$\beta := C_S C_B\Big[\sum_{j=1}^{j_0-1}\Big(\frac{\lambda_j}{C_*\tau^{-1}+\lambda_j}\Big)^{m_j-1} + \sum_{j=j_0}^{\ell}\frac{1}{(2m_j-1)^\gamma}\Big]\ . \tag{5.10}$$

Theorem 5.2. *Let u_*^n be the solution of the Cascadic Algorithm, u is a smooth solution of (2.1). If $u_*^1 = \overline{u}_\ell^1$, then there exists a positive constant $\delta < 1$ such that for $\mathcal{K} \le \delta$,*

$$\|u_*^n - u(x,t_n)\|_0 \le C(T,u)(h_\ell^2 + \beta\tau), \quad \textit{for } t_n \le T\ . \tag{5.11}$$

Proof. With $u^n = u(x, t_n)$, we have that

$$e^n = u_*^n - u^n = u_*^n - \mathcal{R}_h\, u^n + \mathcal{R}_h\, u^n - u^n = \theta^n + \rho^n,$$

The estimate for ρ^n is standard, i. e.,

$$\|\rho^n\|_0 \le C(u) h_\ell^2. \tag{5.12}$$

Define $\overline{\partial}\theta^n = \tau^{-1}(\theta^n - \theta^{n-1})$ and let $\overline{u}_\ell^n$ be defined by (3.4) and $\omega^n = \tau^{-1}(u_*^n - \overline{u}_\ell^n)$. Notice that $\mathcal{A}_h\mathcal{R}_h = \mathcal{P}_h A$, where $\mathcal{P}_h$ is the L^2 projection onto X_ℓ, we get

$$\begin{aligned}\overline{\partial}\theta^n + \mathcal{A}_h\theta^n &= \overline{\partial}u_*^n + \mathcal{A}_h u_*^n - (\mathcal{A}_h\mathcal{R}_h u^n + \mathcal{R}_h\overline{\partial}u^n)\\ &= \tau^{-1}(u_*^n - u_*^{n-1}) + \mathcal{A}_h(u_*^n - \overline{u}_\ell^n) + \mathcal{A}_h\overline{u}_\ell^n - \mathcal{P}_h A u^n - \overline{\partial}\mathcal{R}_h u^n\\ &= \tau^{-1}(u_*^n - \overline{u}_\ell^n) + \mathcal{A}_h(u_*^n - \overline{u}_\ell^n) + \mathcal{P}_h(f^n - Au^n) - \overline{\partial}\mathcal{R}_h u^n\\ &= \mathcal{P}_h\partial_t u^n - \overline{\partial}\mathcal{R}_h u^n + (I + \tau\mathcal{A}_h)\omega^n =: \sigma_1^n + \sigma_2^n\ ,\end{aligned}$$

for $n \ge 1$. Since $\|\theta^0\|_0 \le Ch_\ell^2$, by (5.3) with $p = 0$, we have

$$\|\theta^n\|_0^2 \le Ch_\ell^2 + C\tau\sum_{k=1}^{n}(|\sigma_1^k|_{-1,h}^2 + |\sigma_2^k|_{-1,h}^2)\,. \tag{5.13}$$

Obviously, using standard techniques, we have

$$\begin{aligned}|\sigma_1^k|_{-1,h} &\le C\|\sigma_1^k\|_0 \le C\|\partial_t u^k - \overline{\partial}u^k\|_0 + C\|(\mathcal{P}_h - \mathcal{R}_h)\overline{\partial}u^k\|_0\\ &\le C\tau^{1/2}\Big(\int_{t_{k-1}}^{t_k}\|u_{tt}\|_0^2\,dt\Big)^{1/2} + Ch_\ell^2\tau^{-1/2}\Big(\int_{t_{k-1}}^{t_k}\|u_t\|_2^2\,dt\Big)^{1/2}\,.\end{aligned}$$

We also have $|\sigma_2^k|_{-1,h} \le C\tau^{1/2}\|\omega^k\|_\tau$ and using $u_*^1 = \overline{u}_\ell^1$, so we get the bound on the right-hand side of (5.13):

$$\|\theta^n\|_0^2 \le C(u)(h_\ell^2 + \tau)^2 + C\tau^2\sum_{k=2}^{n}\|\omega^k\|_\tau^2. \tag{5.14}$$

It remains to estimate $\tau^2\sum_{k=2}^n\|\omega^k\|_\tau^2 = \sum_{k=2}^n\|u_*^k - \overline{u}_\ell^k\|_\tau^2$. By (3.6) and using

$$\|\overline{w}_j^k - \overline{w}_{j-1}^k\|_\tau \le \inf_{v\in X_{j-1}}\|\overline{w}_\ell^k - v\|_\tau\ ,$$

we have

$$\|u_*^k - \overline{u}_\ell^k\|_\tau \le \sum_{j=1}^{\ell}\gamma_j\inf_{v\in X_{j-1}}\|\overline{w}_\ell^k - v\|_\tau\ .$$

Taking $v = \mathcal{P}_{j-1}(u_*^{k-1} - u_*^{k-2})$ in the above inequality, with $\mathcal{P}_{j-1}$ defined by

$$\mathcal{A}_\tau(\mathcal{P}_{j-1}u, v) = \mathcal{A}_\tau(u, v) \qquad \forall v \in X_{j-1},$$

and using the obvious decomposition,

$$\begin{aligned}&\overline{w}_\ell^k - \mathcal{P}_{j-1}(u_*^{k-1} - u_*^{k-2}) = \overline{u}_\ell^k - u_*^{k-1} - \mathcal{P}_{j-1}(u_*^{k-1} - u_*^{k-2})\\ &\ = \overline{u}_\ell^k - u_*^k + (u_*^k - 2u_*^{k-1} + u_*^{k-2}) + (I - \mathcal{P}_{j-1})(u_*^{k-1} - u_*^{k-2}),\end{aligned}$$

we have

$$\|u_*^k - \overline{u}_\ell^k\|_\tau \le \sum_{j=1}^{\ell}\gamma_j\|u_*^k - \overline{u}_\ell^k\|_\tau + \sum_{j=1}^{\ell}\gamma_j\|u_*^k - 2u_*^{k-1} + u_*^{k-2}\|_\tau$$
$$+ \sum_{j=1}^{\ell}\gamma_j\|(I-\mathcal{P}_{j-1})(u_*^{k-1} - u_*^{k-2})\|_\tau =: I_1 + I_2 + I_3. \tag{5.15}$$

By the definition of $\mathcal{K}$, $I_1 = \mathcal{K}\|u_*^k - \overline{u}_\ell^k\|_\tau$. Notice that

$$\begin{aligned} u_*^k - 2u_*^{k-1} + u_*^{k-2} &= \tau^2(\overline{\partial}^2\theta^k + \mathcal{R}_h\overline{\partial}^2 u^k) \\ &= \tau(\overline{\partial}\theta^k - \overline{\partial}\theta^{k-1}) + \tau^2\mathcal{R}_h\overline{\partial}^2 u^k, \end{aligned} \tag{5.16}$$

and

$$\begin{aligned} \mathcal{K}\tau^2\|\mathcal{R}_h\overline{\partial}^2 u^k\|_\tau &\le C\mathcal{K}\tau^{\frac{3}{2}}\|\overline{\partial}^2 u^k\|_1 \le C\mathcal{K}\tau^{\frac{3}{2}}\|\overline{\partial}^2\int_{t_{k-2}}^{t_k}(t_k - s)u_{tt}(s)\,ds\|_1 \\ &\le C\mathcal{K}\tau\Big(\int_{t_{k-2}}^{t_k}\|u_{tt}\|_1^2\,ds\Big)^{1/2}, \end{aligned}$$

so we bound I_2 by

$$|I_2| \le \mathcal{K}\tau(\|\overline{\partial}\theta^k\|_\tau + \|\overline{\partial}\theta^{k-1}\|_\tau) + C\mathcal{K}\tau\Big(\int_{t_{k-2}}^{t_k}\|u_{tt}\|_1^2\,ds\Big)^{1/2}. \tag{5.17}$$

Note that $(I-\mathcal{P}_{j-1})\mathcal{R}_h = (I-\mathcal{P}_{j-1})(\mathcal{R}_h - I) + I - \mathcal{P}_{j-1}$, we decompose I_3 into

$$\begin{aligned} I_3 &= \sum_{j=1}^{\ell}\gamma_j\Big(\tau\|(I-\mathcal{P}_{j-1})\overline{\partial}\theta^{k-1}\|_\tau + \tau\|(I-\mathcal{P}_{j-1})\mathcal{R}_h\overline{\partial}u^{k-1}\|_\tau\Big) \\ &\le \mathcal{K}\tau\|\overline{\partial}\theta^{k-1}\|_\tau + \sum_{j=1}^{\ell}\gamma_j\tau\|(I-\mathcal{R}_h)\overline{\partial}u^{k-1}\|_\tau + \sum_{j=1}^{\ell}\gamma_j\tau\|(I-\mathcal{P}_{j-1})\overline{\partial}u^{k-1}\|_\tau \\ &=: I_{31} + I_{32} + I_{33}. \end{aligned}$$

where $\|(I-\mathcal{P}_{j-1})u\|_\tau \le \|u\|_\tau$ is used in deriving the last inequality.

The standard estimate for the Galerkin projection $\mathcal{R}_h$ gives us an bound on I_{32}:

$$|I_{32}| \le C\mathcal{K}(h_\ell^2 + h_\ell\tau^{\frac{1}{2}})\tau^{\frac{1}{2}}\|\overline{\partial}u^{k-1}\|_2 \le C\mathcal{K}(h_\ell^2 + h_\ell\tau^{\frac{1}{2}})\Big(\int_{t_{k-1}}^{t_k}\|u_t\|_2^2\,dt\Big)^{1/2}.$$

I_{33} can thus be estimated by

$$|I_{33}| \le \sum_{j=1}^{\ell}\gamma_j h_{j-1}^2\tau^{\frac{1}{2}}\|\overline{\partial}u^{k-1}\|_2 + \sum_{j=1}^{\ell}\gamma_j h_{j-1}\tau\|\overline{\partial}u^{k-1}\|_2 =: B_1 + B_2\,.$$

From the construction of γ_j, we see that B_1 can be further decomposed into

$$B_1 = \Big(\sum_{j=1}^{j_0-1}\gamma_j h_{j-1}^2 + \sum_{j=j_0}^{\ell}\gamma_j h_{j-1}^2\Big)\tau^{\frac{1}{2}}\|\overline{\partial}u^{k-1}\|_2.$$

Moreover, using $\lambda_j h_{j-1}^2 \le C$, we have

$$\sum_{j=1}^{j_0-1} \gamma_j h_{j-1}^2 \le \frac{C_S C_B}{C_*} \sum_{j=1}^{j_0-1} \Big(\frac{\lambda_j}{C_*\tau^{-1} + \lambda_j} \Big)^{m_j - 1} \tau \lambda_j h_{j-1}^2$$
$$\le C\, C_S C_B \tau \sum_{j=1}^{j_0-1} \Big(\frac{\lambda_j}{C_*\tau^{-1} + \lambda_j} \Big)^{m_j - 1},$$

and

$$\sum_{j=j_0}^{\ell} \gamma_j h_{j-1}^2 \le C \sum_{j=j_0}^{\ell} \frac{C_S C_B}{(2m_j - 1)^\gamma} \tau \lambda_j h_{j-1}^2 = C\tau \sum_{j=j_0}^{\ell} \frac{C_S C_B}{(2m_j - 1)^\gamma} \,.$$

Combining the above two, and using (5.10), we get a bound on B_1:

$$|B_1| \le C\beta\tau \Big(\int_{t_{k-1}}^{t_k} \|u_t\|_2^2 \, dt \Big)^{1/2} . \tag{5.18}$$

Repeating the above procedure and using $\lambda_j \le C_*\tau^{-1} + \lambda_j$, we bound B_2 as

$$|B_2| \le C\beta\tau \Big(\int_{t_{k-1}}^{t_k} \|u_t\|_2^2 \, dt \Big)^{1/2},$$

which, in combination with (5.18), leads to a bound on I_{33}:

$$|I_{33}| \le C\beta\tau \Big(\int_{t_{k-1}}^{t_k} \|u_t\|_2^2 \, dt \Big)^{1/2} .$$

Combining the estimates for I_1, I_2 and I_3 together, we get

$$\sum_{k=2}^{n} \|u_*^k - \overline{u}_\ell^k\|_\tau^2 \le C \Big(\frac{2\mathcal{K}}{1-\mathcal{K}} \Big)^2 \tau^2 \sum_{k=2}^{n} \|\overline{\partial}\theta^k\|_\tau^2 + C(h_\ell^4 + \beta^2\tau^2) \int_0^T \|\partial_t u\|_2^2 \, dt \,. \tag{5.19}$$

Now, let $\mathcal{K} \le \delta$, for some δ to be specified later and let $\mathcal{K}/(1-\mathcal{K}) \le \delta/(1-\delta) = \epsilon$, then (5.14) and (5.19) yield

$$\|\theta^n\|_0^2 \le C(u)(h_\ell^2 + \tau)^2 + C\epsilon^2\tau^2 \sum_{k=2}^{n} \|\overline{\partial}\theta^k\|_\tau^2 + C(u)\beta^2\tau^2 \,. \tag{5.20}$$

Applying (5.4) with $p = 0$, we are led to

$$\tau^2 \sum_{k=1}^{n} \|\overline{\partial}\theta^k\|_\tau^2 \le C\tau \sum_{k=1}^{n} (|\sigma_1^k|_*^2 + |\sigma_2^k|_*^2) \,.$$

As above, we can get

$$\tau \sum_{k=1}^{n} (|\sigma_1^k|_*^2 + |\sigma_2^k|_*^2) \le C(u)(h_\ell^2 + \tau)^2 + C\tau^2 \sum_{k=1}^{n} \|\omega^k\|_\tau^2$$
$$\le C(u)(h_\ell^2 + \tau)^2 + C\epsilon^2\tau^2 \sum_{k=1}^{n} \|\overline{\partial}\theta^k\|_\tau^2 + C(u)\beta^2\tau^2$$
$$\le C(u)(h_\ell^2 + \tau)^2 + C\epsilon^2\tau \sum_{k=1}^{n} (|\sigma_1^k|_*^2 + |\sigma_2^k|_*^2) + C(u)\beta^2\tau^2 \,. \tag{5.21}$$

Taking ϵ suitably small (thus, δ suitably small), we have

$$\tau^2 \sum_{k=1}^{n} \|\overline{\partial}\theta^k\|_\tau^2 \le C\tau \sum_{k=1}^{n} |\sigma^k|_*^2 \le C(u)(h_\ell^2+\tau)^2 + C(u)\beta^2\tau^2 \ . \tag{5.22}$$

A combination of (5.20) and (5.22) gives (5.11). □

An error bound in the energy norm is given below:

Theorem 5.3. *Under the same assumption of Theorem 5.2, we have*

$$\|u_*^n - u(x,t_n)\|_1 \le C(u,T)(h_\ell + \beta\tau). \tag{5.23}$$

Proof. Following the argument given in Theorem 5.2, we have $e^n = \theta^n + \rho^n$ with

$$\|\rho^n\|_1 \le Ch_\ell \|u^n\|_2 \ . \tag{5.24}$$

To estimate $\|\theta^n\|_1$, since $\|\theta^0\|_1 \le Ch_\ell$, instead of (5.3), we have by (5.5) with $p = 0$ that

$$\|\theta^n\|_1^2 \le Ch_\ell^2 + C\tau \sum_{k=1}^{n} |\sigma_1^k|_*^2 + C\tau \sum_{k=1}^{n} |\sigma_2^k|_*^2.$$

In view of (5.21), we have

$$\|\theta^n\|_1 \le C(u,T)(h_\ell + \beta\tau),$$

which, together with (5.24), yields (5.23). □

5.2 Convergence for nonsmooth data

In the remaining part of this section, we consider the homogeneous equation with nonsmooth initial data. Recall that the Backward-Euler satisfies

$$\|u_l^n - u(x,t^n)\|_0 \le C(h_\ell^2 + \tau)t_n^{-1}\|u_0\|_0 \ . \tag{5.25}$$

We show that our Cascadic Algorithm can be designed so that the above type of error bound remains valid.

We define the semi-discrete in space approximation by:

$$u_{h,t} + \mathcal{A}_h u_h = 0, \quad \text{for } t > 0, \quad \text{with } u_h(0) = \mathcal{P}_h u_0, \tag{5.26}$$

then the solution of (5.26) satisfies

$$\|u_h(t) - u(x,t)\|_0 \le Ch_\ell^2 t^{-1}\|u_0\|_0, \quad \text{for } t > 0. \tag{5.27}$$

By virtue of [Theorem 3.4, 26], we have

$$\|\partial_t\big(u_h(t) - u(x,t)\big)\|_0 \le Ch_\ell^2 t^{-2}\|u_0\|_0, \tag{5.28}$$

the above estimate together with the inverse inequality [46] leads to

$$\begin{aligned}\|\partial_t\big(u_h(t) - u(x,t)\big)\|_1 &\le Ch_\ell^{-1}\|\partial_t\big(u_h(t) - \mathcal{R}_h u\big)\|_0 + \|\partial_t\big(\mathcal{R}_h u - u(x,t)\big)\|_1 \\ &\le Ch_\ell^{-1}\Big(\|\partial_t\big(u_h(t) - u(x,t)\big)\|_0 + \|\partial_t\big(\mathcal{R}_h u - u(x,t)\big)\|_0\Big) \\ &\quad + C\|\partial_t\big(\mathcal{R}_h u - u(x,t)\big)\|_1 \\ &\le Ch_\ell t^{-2}\|u_0\|_0 + Ch_\ell\|\partial_t u\|_2 \le Ch_\ell t^{-2}\|u_0\|_0 \ . \end{aligned} \tag{5.29}$$

Here, we have used $\|\partial_s u\|_2 \le Cs^{-2}\|u_0\|_0$ in the last step [26].

To more effectively resolving the initial layer, we allow the iteration strategies to vary with respect to time. Thus, to emphasize on the dependence on the time steps, we introduce the subscript k for the time step t_k and define $\alpha_k := \sum_{j=1}^{\ell} \gamma_{j,k}$ and

$$\beta_k := C_S C_B \Big[\sum_{j=1}^{j_0-1}\Big(\frac{\lambda_{j,k}}{C_*\tau^{-1}+\lambda_{j,k}}\Big)^{m_{j,k}-1} + \sum_{j=j_0}^{\ell}\frac{1}{(2m_{j,k}-1)^{\gamma}}\Big], \tag{5.30}$$

where $\gamma_{j,k}$'s are the constants in the smoother estimates $m_{j,k}$ are the iteration number used in the smoothers.

Theorem 5.4. *For the fully discrete method (2.4) with $f = 0, j = \ell$ and $u_{0,h} = \mathcal{P}_h u_0$, let $u_*^k = \overline{u}_\ell^k$ for $k = 1, 2$, and let $m_{k,j}$ be the iteration number on the j-th level at the time step t_k. If for some suitably small constant $\epsilon \in (0, 1)$, we have that*

$$\frac{\beta_k}{1-\beta_k} \le \epsilon \min(t_k^2, 1), \tag{5.31}$$

then there exists a constant $C > 0$ such that

$$\|u_*^n - u(x, t_n)\|_0 \le C(h_\ell^2 + \tau)t_n^{-1}\|u_0\|_0 \quad \text{for } n \ge 3 \text{ and } t_n \le T. \tag{5.32}$$

Proof. With $\omega^n = (u_*^n - \overline{u}_\ell^n)/\tau$, $\vartheta^n = \overline{\partial} u_h(t_n) - u_{h,t}(t_n)$, and $e^n = u_*^n - u_h(t_n)$, we have, as in Theorem 5.2, the error equation

$$\overline{\partial} e^n + \mathcal{A}_h e^n = -\vartheta^n + (I + \tau\mathcal{A}_h)\omega^n =: \sigma^n . \tag{5.33}$$

since $e^0 = 0$, an application of (5.3) with $p = 2$ gives

$$t_n^2\|e^n\|_0^2 \le C\tau\sum_{k=1}^{n}(t_k^2|\sigma^k|_{-1,h}^2 + |\sigma^k|_{-3,h}^2).$$

Since $\mathcal{A}_h$ is positive definite and t_n is bounded,

$$\begin{aligned} t_k^2|(I+\tau\mathcal{A}_h)\omega^k|_{-1,h}^2 + |(I+\tau\mathcal{A}_h)\omega^k|_{-3,h}^2 &\le C|(I+\tau\mathcal{A}_h)\omega^k|_{-1,h}^2 \\ &\le C\|(I+\tau\mathcal{A}_h)^{\frac{1}{2}}\omega^k\|_0^2 = C\tau\|\omega^k\|_\tau^2 . \end{aligned} \tag{5.34}$$

Since $\omega^1 = \omega^2 = 0$ by assumption, we thus have

$$t_n^2\|e^n\|_0^2 \le C\tau\sum_{k=1}^{n}(t_k^2|\vartheta^k|_{-1,h}^2 + |\vartheta^k|_{-3,h}^2) + C\tau^2\sum_{k=3}^{n}\|\omega^k\|_\tau^2. \tag{5.35}$$

The next step is to show that

$$\tau\sum_{k=1}^{n}(t_k^2|\vartheta^k|_{-1,h}^2 + |\vartheta^k|_{-3,h}^2) \le C\tau^2\|u_0\|_0^2 . \tag{5.36}$$

Let $s = 1$ or 3. By the definition of ϑ^k, we get

$$|\vartheta^k|_{-s,h}^2 \le C\tau\int_{t_{k-1}}^{t_k}|u_{h,tt}(y)|_{-s,h}^2\,dy.$$

Then, for $k > 1$ when $s = 1$ and $k \geq 1$ when $s = 3$, we have

$$\tau t_k^{3-s} |\vartheta^k|_{-s,h}^2 \leq C\tau^2 \int_{t_{k-1}}^{t_k} y^{3-s} |u_{h,tt}(y)|_{-s,h}^2 \, dy.$$

By the eigen-decomposition of the operator $\mathcal{A}_h$, we have

$$\begin{aligned}\int_0^\infty y^{3-s} |u_{h,tt}(y)|_{-s,h}^2 \, dy &\leq \int_0^\infty y^{3-s} \sum_{m=1}^{n_j} \lambda_m^{4-s} \exp^{-2\lambda_m y} (\mathcal{P}_h u_0, \phi_l)^2 \, dy \\ &\leq C \sum_{m=1}^{n_j} (\mathcal{P}_h u_0, \phi_l)^2 = C\|\mathcal{P}_h u_0\|_0^2 \leq C\|u_0\|_0^2. \end{aligned} \tag{5.37}$$

Consequently, we obtain (5.36) except for the terms related to $k = 1$ and $s = 1$. For these terms we have

$$\begin{aligned}\tau\, t_1^2 |\vartheta^1|_{-1,h}^2 &= \tau^3 |\overline{\partial} u_h(t_1) - u_{h,t}(t_1)|_{-1,h}^2 \leq C\tau^3 (|\overline{\partial} u_h(t_1)|_{-1}^2 + |u_{h,t}(t_1)|_{-1}^2) \\ &\leq C\tau^2 \int_0^\tau |u_{h,t}|_{-1}^2 \, dt + C\tau^3 |u_h(\tau)|_1^2 \leq C\tau^2 \|\mathcal{P}_h u_0\|_0^2 \leq C\tau^2 \|u_0\|_0^2. \end{aligned} \tag{5.38}$$

So, (5.35) together with (5.36) gives

$$t_n^2 \|e^n\|_0^2 \leq C\tau^2 \|u_0\|_0^2 + C\tau^2 \sum_{k=3}^n \|\omega^k\|_\tau^2. \tag{5.39}$$

As in the proof of Theorem 5.2, we can bound the second term in the above sum as

$$\begin{aligned}\tau\|\omega^k\|_\tau = \|u_*^k - \overline{u}_\ell^k\|_\tau &\leq \sum_{j=1}^\ell \gamma_{j,k} \big(\|u_*^k - \overline{u}_\ell^k\|_\tau + \|u_*^k - 2u_*^{k-1} + u_*^{k-2}\|_\tau \\ &\quad + \|(I - \mathcal{P}_{j-1})(u_*^{k-1} - u_*^{k-2})\|_\tau \big) =: J_1 + J_2 + J_3. \end{aligned} \tag{5.40}$$

By using similar estimates on J_1 and J_2 as that in Theorem 5.2 and

$$\frac{\alpha_k}{1-\alpha_k} \leq \frac{\beta_k}{1-\beta_k} \leq \epsilon t_k^2, \tag{5.41}$$

we may recast (5.40) as

$$\begin{aligned}\sum_{k=3}^n \|u_*^k - \overline{u}_\ell^k\|_\tau^2 &\leq 8\epsilon^2\tau^2 \sum_{k=3}^n t_k^4 \|\overline{\partial} e^k\|_\tau^2 + C\tau^4 \sum_{k=3}^n t_k^4 \|\overline{\partial}^2 u_h(t_k)\|_\tau^2 \\ &\quad + C\tau^2 \sum_{k=3}^n \frac{1}{(1-\alpha_k)^2} \Big(\sum_{j=1}^\ell \gamma_{j,k} \|(I - \mathcal{P}_{j-1}) \overline{\partial} u_h(t_k)\|_\tau \Big)^2 \\ &= :I_1 + I_2 + I_3. \end{aligned} \tag{5.42}$$

We estimate I_2 and I_3 firstly. For $k \geq 3$, we have

$$t_k^3 \|\overline{\partial}^2 u_h(t_k)\|_0^2 \leq C t_k^3 \|\overline{\partial}^2 \big(\int_{t_{k-2}}^t (t-s) u_{h,tt}(s)\, ds \big)_{t=t_k} \|_0^2 \leq C\tau^{-1} \int_{t_{k-2}}^{t_k} s^3 \|u_{h,tt}(s)\|_0^2 \, ds,$$

$$t_k^4 \|\overline{\partial}^2 u_h(t_k)\|_1^2 \leq C t_k^4 \mid \big(\overline{\partial}^2 \int_{t_{k-2}}^t (t-s) u_{h,tt}(s)\, ds \big)_{t=t_k} |_1^2 \leq C\tau^{-1} \int_{t_{k-2}}^{t_k} s^4 |u_{h,tt}(s)|_1^2 \, ds.$$

For bounded t_n, using the above two inequalities, we may bound I_2 as

$$\begin{aligned}|I_2| &\le C\sum_{k=3}^{n}\left(\tau^3 t_k^3\|\overline{\partial}^2 u_h(t_k)\|_0 + \tau^4 t_k^4\|\overline{\partial}^2 u_h(t_k)\|_1^2\right)\\ &\le C\tau^2\int_0^\infty s^3\|u_{h,tt}(s)\|_0^2\,ds + C\tau^3\int_0^\infty s^4|u_{h,tt}(s)|_1^2\,ds\ .\end{aligned}$$

As given in (5.37), the above inequality is estimated as

$$|I_2| \le C\tau^2\|u_0\|_0^2.$$

Note that I_3 can be further decomposed into two terms

$$\begin{aligned}I_3 \le\ & C\sum_{k=3}^{n}\frac{\tau^2}{(1-\alpha_k)^2}\Big(\sum_{j=1}^{\ell}\gamma_{j,k}\|(I-\mathcal{P}_{j-1})\overline{\partial}(u_h(t_k)-u(x,t_k))\|_\tau\Big)^2\\ &+C\sum_{k=3}^{n}\frac{\tau}{(1-\alpha_k)^2}\Big(\sum_{j=1}^{\ell}\gamma_{j,k}\|(I-\mathcal{P}_{j-1})\overline{\partial}u(x,t_k)\|_\tau\Big)^2 =: I_{31}+I_{32}.\end{aligned}$$

In view of (5.28) and (5.29),

$$\|\overline{\partial}(u_h(t_k)-u(x,t_k))\|_\tau^2 \le C\Big((h_\ell^4/\tau^2+h_\ell^2/\tau)\int_{t_{k-1}}^{t_k}\frac{ds}{s^4}\Big)\|u_0\|_0^2.$$

Notice that $\|I-\mathcal{P}_{j-1}\|_\tau \le 1$ and (5.41), we see that I_{31} is bounded:

$$\begin{aligned}|I_{31}| &\le C\sum_{k=3}^{n}\frac{\tau^2}{(1-\alpha_k)^2}\Big(\sum_{j=1}^{\ell}\gamma_{j,k}\|\overline{\partial}(u_h(t_k)-u(x,t_k))\|_\tau\Big)^2\\ &\le C\epsilon^2(h_\ell^4+h_\ell^2\tau)\sum_{k=3}^{n}\int_{t_{k-1}}^{t_k}\frac{t_k^4}{t^4}\,dt\,\|u_0\|_0^2 \le C\epsilon^2(h_\ell^4+h_\ell^2\tau)\|u_0\|_0^2.\end{aligned} \tag{5.43}$$

Similar to the estimation on I_{33} in Theorem 5.1, we have

$$\sum_{j=1}^{\ell}\gamma_{j,k}\|(I-\mathcal{P}_{j-1})\overline{\partial}u(t_k)\|_\tau \le C\beta_k\tau^{\frac{1}{2}}\|\overline{\partial}u(t_k)\|_2\ . \tag{5.44}$$

In view of (5.44) and (5.41), we bound I_{32} as

$$\begin{aligned}|I_{32}| &\le C\tau^2\sum_{k=3}^{n}\frac{\beta_k^2}{(1-\alpha_k)^2}\int_{t_{k-1}}^{t_k}\|\partial_s u\|_2^2\,ds \le C\tau^2\sum_{k=3}^{n}\frac{\beta_k^2}{(1-\beta_k)^2}\int_{t_{k-1}}^{t_k}\|\partial_s u\|_2^2\,ds\\ &\le C\epsilon^2\tau^2\sum_{k=3}^{n}t_k^4\int_{t_{k-1}}^{t_k}\|\partial_s u\|_2^2\,ds \le C\epsilon^2\tau^2\sum_{k=3}^{n}\int_{t_{k-1}}^{t_k}\frac{t_k^4}{s^4}ds\|u_0\|_0^2 \le C\epsilon^2\tau^2\,\|u_0\|_0^2.\end{aligned}$$

Summing up the estimate for I_2 and I_3, notice that t_n is bounded, we conclude that for any $\epsilon > 0$, there holds

$$\tau^2\sum_{k=3}^{n}\|\omega^k\|_\tau^2 \le C(h_\ell^2+\tau)^2\|u_0\|_0^2 + C\epsilon\tau\sum_{k=3}^{n}t_k^3|\overline{\partial}e^k|^2\ .$$

Invoking Lemma 5.1 once again, we obtain

$$\begin{aligned}\tau\sum_{k=3}^n t_k^3|\overline{\partial}e^k|^2 &\le C\tau\sum_{k=3}^n(t_k^3|\vartheta^k|_*^2+|\vartheta^k|_{*,3}^2)+C\tau^2\sum_{k=3}^n\|\omega^k\|_\tau^2\\ &\le C\tau^2\|u\|_0^2+C\tau^2\sum_{k=3}^n\|\omega^k\|_\tau^2.\end{aligned}$$

Combining the above two and choosing a sufficiently small ϵ gives

$$\tau^2\sum_{k=3}^n\|\omega^k\|_\tau^2\le C(h_\ell^2+\tau)^2\|u_0\|_0^2, \tag{5.45}$$

which together with (5.39) implies

$$\|u_*^n-u_h(t_n)\|_0\le C(h_\ell^2+\tau)t_n^{-1}\|u_0\|_0. \tag{5.46}$$

Combining with (5.27), the classical error bound for u_h, we get

$$\|u_h(t_n)-u(x,t_n)\|_0\le Ch_\ell^2t_n^{-1}\|u_0\|_0\ , \tag{5.47}$$

we get the desired result (5.32). □

An error bound in the energy norm is given below:

Theorem 5.5. *Under the same assumption of Theorem 5.4, we have*

$$\|u_*^n-u(x,t_n)\|_1\le C(u)(h_\ell t_n^{-1}+(h_\ell^2+\tau)t_n^{-3/2})\|u_0\|_0. \tag{5.48}$$

Proof. Following the argument given in Theorem 5.4, we still have the error equation (5.33). Since $e^0=0$, instead of (5.4) we have by (5.5) with $p=3$ that

$$t_n^3\|u_*^n-u_h(t_n)\|_1^2\le C\tau\sum_{k=1}^n(t_k^3|\sigma^k|_*^2+|\sigma^k|_{*,-3}^2),$$

with $\sigma^k=\vartheta^k+(I+\tau\mathcal{A}_h)\omega^k$. As in (5.36), we have

$$C\tau\sum_{k=1}^n(t_k^3|\vartheta^k|_*^2+|\vartheta^k|_{*,-3}^2)\le C\tau^2\|u_0\|_0^2.$$

And as in (5.34), we get

$$t_k^3|(I+\tau\mathcal{A}_h)\omega^k|_*^2+|(I+\tau\mathcal{A}_h)\omega^k|_{*,-3}^2\le C\tau\|\omega^k\|_\tau^2.$$

Combining the above three estimates and (5.45) leads to

$$\|u_*^n-u_h(t_n)\|_1\le C(h_\ell^2+\tau)t_n^{-3/2}\|u_0\|_0.$$

As that in (5.29), we have

$$\|u_h(t_n)-u(x,t_n)\|_1\le Ch_\ell\|u(x,t_n)\|_2\le Ch_\ell t_n^{-1}\|u_0\|_0.$$

A combination of the above two gives (5.48). □

Remark 5.6. *Notice that if we assume $u_0 \in \mathcal{H}_0^1(\Omega)$, the error estimate in (5.48) can be improved to $\mathcal{O}(h_\ell + \tau)/t_n$ since we may use (5.5) with $p = 2$ in such a case.*

Remark 5.7. *We require that $u_0 \in \mathcal{H}_0^1(\Omega)$ in the stability estimate (cf. Theorem 3.1), which is not realistic for the nonsmooth initial data. However, we assume that $u_*^1 = \overline{u}_\ell^1$ in Theorem 5.4, hence for the case when $u_0 \in \mathcal{L}^2(\Omega)$ and $f = 0$, the stability estimate can be modified as*

$$\|u_*^n\|_{\mathcal{A}}^2 \le \|u_*^1\|_{\mathcal{A}}^2 \le \frac{1}{2\tau}\|u_0\|_0^2.$$

6 Iteration Strategy

For achieving good performance for the Cascadic Algorithms in practice, parameter tuning is an important issue in their actual implementation. The theoretical analysis of the Cascadic Algorithm made in this paper can be useful in practice as a guide for assigning values to the various parameters used in the algorithm. We now make some discussions on this issue.

Since the constraint on the iteration number for achieving the optimal error bounds is generally tighter than that for stability, we only consider how the iteration number is selected so as to give the optimal error bounds.

In view of Theorem 5.2, 5.3, 5.4 and Theorem 5.5, the following three conditions are required for the Cascadic Algorithm to be of optimal complexity for parabolic equations: for each k,

1. $\beta < 1$ (or $\beta_k < 1$).
2. $\beta/(1-\beta)$, or $\beta_k/(1-\beta_k)$, is sufficiently small.
3. The overall computing cost (complexity) is of the order $\mathcal{O}(n_\ell)$, i.e.,

$$\sum_{j=1}^{\ell} m_j n_j \approx \mathcal{O}(n_\ell)\ .$$

To achieve the optimal complexity for smooth data, we have the following choice for the iteration number m_j.

$$m_j = \begin{cases} m_{j_0} & 0 \le j \le j_0, \\ \lfloor (m_l - \frac{1}{2}) 2^{\frac{2(l-j)}{\gamma+1}} + \frac{1}{2} \rfloor & j = j_0+1, \cdots, \ell. \end{cases}$$

Define $d_j\colon = \lambda_j/(C_*\tau^{-1} + \lambda_j)$. Noting $\lambda_j < \lambda_{j+1}$, we thus define $\hat{c}\colon = \max_{1\le j\le j_0} \lambda_j/\lambda_{j+1}$, which in turn implies that for any $1 \le j \le j_0 - 1$:

$$\begin{aligned} \frac{d_j}{d_{j+1}} &= \frac{(\lambda_j/\lambda_{j+1})C_*\tau^{-1} + \lambda_j}{C_*\tau^{-1} + \lambda_j} \le \frac{\hat{c}C_*\tau^{-1} + \lambda_j}{C_*\tau^{-1} + \lambda_j} \\ &\le \frac{\hat{c}C_* + 1}{C_* + 1} = : \underline{c} < 1, \end{aligned} \tag{6.1}$$

where we have used $\tau\lambda_j \le \tau\lambda_{j_0} \le 1$. In view of (6.1), we obtain

$$d_j = d_{j_0} \prod_{k=j}^{j_0-1} \frac{d_k}{d_{k+1}} \le \underline{c}^{j_0-j} d_{j_0} \le \frac{\underline{c}^{j_0-j}}{C_*+1}\ ,$$

since $d_{j_0} \le 1/(C_* + 1)$. We then get

$$\beta \le \frac{C_S C_B}{(1 - \underline{c}^{m_{j_0}-1})}\Big(\frac{\underline{c}}{1+C_*}\Big)^{m_{j_0}-1} + \frac{C_S C_B}{(2m_\ell - 1)^\gamma}\frac{1}{1 - 2^{-2\gamma/(\gamma+1)}} \le \epsilon\ ,$$

which can be smaller than some suitable constant ϵ. It is easy to verify that β is bounded uniformly for such m_j.

It remains to estimate the overall computing cost on each time level. Notice that $4^j/c_* \le \dim X_j \le c_* 4^j$, a simple calculation yields that

$$\sum_{j=1}^{\ell} m_j n_j \le c_*^2\Big(m_{j_0} n_\ell (2^{2(j_0-\ell)} - 2^{-2\ell})/3 + (m_\ell - 1/2) n_\ell \frac{1 - 2^{\frac{2\gamma}{\gamma+1}(j_0-\ell-1)}}{1 - 2^{-2\gamma/(\gamma+1)}} + \frac{2}{3} n_\ell (1 - 2^{2(j_0-\ell-1)})\Big).$$

Notice that $m_{j_0} \le m_{j_0+1}$, we thus have

$$m_{j_0}(2^{2(j_0-\ell)} - 2^{-2\ell}) \le (m_\ell - 1/2)\big(2^{\frac{2\gamma}{\gamma+1}(j_0-\ell) - \frac{2}{\gamma+1}} - 2^{\frac{-2(\gamma\ell+1)}{\gamma+1}}\big) + 2^{2(j_0-\ell)-1} - 2^{-2\ell-1}.$$

A combination of the above two estimates leads to

$$\sum_{j=1}^{\ell} m_j n_j \le c_*^2/3\Big((m_\ell - 1/2) n_\ell (2^{\frac{2\gamma}{\gamma+1} j_0} - 1) 2^{\frac{-2(\gamma\ell+1)}{\gamma+1}} + 3(m_\ell - 1/2) n_\ell \frac{1 - 2^{\frac{2\gamma}{\gamma+1}(j_0-\ell-1)}}{1 - 2^{-2\gamma/(\gamma+1)}} + n_\ell\Big). \tag{6.2}$$

As to the nonsmooth data, the strategy is basically the same, except when k is small. For the initial transient period, i. e., small k, we let m_j depend on the index k, that is, $m_j = m_{k,j}$ so that it becomes large for small k. The rationale behind the choice is due to the fact that, in this case, we need

$$\beta_k \le \frac{C_S C_B}{(1 - \underline{c}_k^{m_{j_0}-1})}\Big(\frac{\underline{c}_k}{1+C_*}\Big)^{m_{j_0}-1} + \frac{C_S C_B}{(2m_{k,\ell} - 1)^\gamma}\frac{1}{1 - 2^{-2\gamma/(\gamma+1)}} \le \epsilon t_k^2\ ,$$

for some suitably small constant ϵ, where

$$\underline{c}_k := \frac{\hat{c}_k C_* + 1}{C_* + 1} \quad \text{with} \ \ \hat{c}_k := \max_{1\le j\le j_0} \lambda_{k,j}/\lambda_{k,j+1}.$$

Such a scenario is as expected when an initial transient layer needs to be resolved.

As above, the overall computing cost on time level k is

$$\sum_{j=1}^{\ell} m_{k,j} n_j \le c_*^2/3\Big((m_{k,\ell} - 1/2) n_\ell (2^{\frac{2\gamma}{\gamma+1} j_0} - 1) 2^{\frac{-2\gamma\ell}{\gamma+1} - \frac{2}{\gamma+1}} + 3(m_{k,\ell} - 1/2) n_\ell \frac{1 - 2^{\frac{2\gamma}{\gamma+1}(j_0-\ell-1)}}{1 - 2^{-2\gamma/(\gamma+1)}} + n_\ell\Big). \tag{6.3}$$

In the cases of *Jacobi* smoother and the *CG* iteration, $m_j(m_{k,j})$'s are taken to be suitably large but independent of j and ℓ and we thus have optimal multigrid complexity. In the case of the symmetric *Gauß-Seidel* smoother, we may need to let m_ℓ be proportional to some (say, quadratic) power of $\log(2m)$ (m being the bandwidth). For most equations and discretizations considered in this paper here, we typically expect that $\log(2m)$ is on the order of the level index ℓ, thus the complexity of the Cascadic Algorithm is nearly optimal in the sense that the total work is on the order of $O(n_\ell \log^2(n_\ell))$.

7 Conclusion

In this paper, a comprehensive analysis of a cascadic multigrid algorithm for an implicit in time discretization of some parabolic equations is presented. New and sharper estimates on smoothers are established to reflect the spatial and temporal structure of the discrete approximation to the parabolic equations. The stability of the algorithm is established based on these smoother estimates. Complete error estimates for both smooth and nonsmooth data are provided. We also combine with a complexity analysis to provide guidance on some optimal choices of various parameter values. Moreover, the general framework and the technical derivations provide a basis for studying the applications of cascadic multigrid algorithms to other time dependent equations.

References

[1] Deuflhard P, Leinen P, Yserentant H. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Engrg.*, **1**: 3–35 (1989)

[2] Wachspress E L. Iterative Solution of Elliptic Systems, and Applications to the Neutron Diffusion Equations of Reactor Physics. Prentice-Hall, Engle-wood Cliffs, N.J, (1966)

[3] Bornemann F A, Erdmann B, Kornhuber R. Adaptive multilevel methods in three space dimensions. *Internat. J. Numer. Methods Engrg.*, **36**: 3187–3203 (1993)

[4] Deuflhard P. Cascadic conjugate gradient methods for elliptic partial differential equations: algorithm and numerical results, in Proceedings of the 7th international conference on domain decomposition methods 1993, Keyes D, Xu J eds., AMS, Providence, pp. 29-42.

[5] Shaidurov V. Some estimates of the rate of convergence for the cascadic conjugate-gradient method. *Comput. Math. Appl.*, **31**: 161–171 (1996)

[6] Bornemann F A, Deuflhard P. The cascadic multigrid method for elliptic problems, *Numer. Math.*, **75**: 135–152 (1996)

[7] Hackbusch W. Multi-Grid Methods and Applications. Springer-Verlag, Berlin-Heidelberg-New York, (1985)

[8] Shaidurov V, Tobiska L. The convergence of the cascadic conjugate-gradient method applied to elliptic problem in domains with re-entrant corners. *Math. Comp.*, **69**: 501–520 (2000)

[9] Braess D, Dahmen W. A cascadic multigrid algorithm for the Stokes equations. *Numer. Math.*, **82**: 179–191 (1999)

[10] Shaidurov V, Timmermann G. A cascadic multigrid algorithm for semilinear indefinite elliptic problem. *Computing.*, **64**: 349–366 (2000)

[11] Timmermann G. A cascadic multigrid algorithm for semilinear elliptic problems. *Numer. Math.*, **86**: 717–731 (2000)

[12] Huang Y Q, Shi Z-C, Tang T, Xue W-M. A mulilevel successive iteration method for nonlinear elliptic problem. *Math. Comp.*, **73**: 525–539 (2004)

[13] Braess D, Deuflhard P, Lipnikov K. A subspace cascadic multigrid method for mortar elements. *Computing*, **69**: 205–225 (2002)

[14] Blum H, Braess D, Suttmeier F T. A cascadic multigrid algorithm for variational inequalities. *Comput. Vis. Sci.*, **7**: 153–157 (2004)

[15] Shi Z-C, Xu X. Cascadic multigrid method for ellptic problems. *East-West J. Numer. Math.*, **7**: 199–209

(1999)
[16] Stevenson R. Nonconforming finite elements and the cascadic multi-grid method. *Numer. Math.*, **91**: 351–387 (2002)
[17] Shi Z-C, Xu X, Man H. Cascadic multigrid for finite volume methods for elliptic problems. *J. Comput. Math.*, **22**: 905–920 (2004)
[18] Shi Z.-C, Xu X, Huang Y Q. Economical cascadic multigrid methods. *Sci. China Ser. A*, **50**: 1765–1780 (2007)
[19] Bornemann F A. An adaptive multilevel approach to parabolic equations II: variable order time discretization based on a multiplicative error correction. *IMPACT Comput. Sci. Engrg.*, **3**: 93–122 (1991)
[20] Shi Z.-C, Xu X. Cascadic multigrid for parabolic problems. *J. Comput. Math.*, **18**: 551–560 (2000)
[21] Xu X, Chen J. Multigrid for the mortar finite element for parabolic problem. *J. Comput. Math.*, **21**: 411–420 (2003)
[22] Douglas J, Dupont T, Ewing R. Incomplete iteration for time-stepping: a Galerkin method for a quasilinear parabolic problem. *SIAN J. Numer. Anal.*, **16**: 503–524 (1979)
[23] Dawson C, Du Q, Dupont T. A finite difference domain decomposition algorithm for numerical solution of the heat equation. *Math. Comp.*, **57**: 63–71 (1991)
[24] Bornemann F A, Krause R. Classical and cascadic multigrid-a methodological comparision, in Domain Decompose Methods in Scientific and Engineering Computing, P. Bjorstad, M. Espedal and D. Keys, eds., Domain Decomposition Press, Bergen, 1998, pp. 64–71.
[25] Bramble J H, Pasciak J E, Sammon P H, Thomée V. Incomplete iterations in multistep backward difference methods for parabolic problems with smooth and nonsmooth data. *Math. Comp.*, **52**: 339–367 (1989)
[26] Thomée V. Galerkin Finite Element Methods for Parabolic Problem. Springer-Verlag, (1997)
[27] Admas R A. Sobolev Spaces. Academic Press, (1975)
[28] Lions J L, Magenes E. Probléms aux non homogénes et applications. tome I. Dunod, Paris, (1968)
[29] Bank R E, Dupont T. An optimal order process for solving elliptic finite element equations. *Math. Comp.*, **36**: 35–51 (1981)
[30] Brandt A, Greenwald J. Parabolic multigrid revisited, in Multigrid III, Hackbusch W, Trottenberg U eds., International Series of Numerical Mathematics, Vol. 98, Birkhäuser Verlak, Basel, 1986, pp. 143–154.
[31] Cai X C. Additive Schwarz algorithms for parabolic convection-diffusion equations. *Numer. Math.*, **60**: 41–61 (1991)
[32] Hackbusch W. Parabolic multi-grid methods, in Computing Methods in Applied Sciences and Engineering, VI, Glowinski R, Lions J L eds., North-Holland, Amerstdam, 1984, pp. 189–197.
[33] Lubich Ch, Ostermann A. Multigrid dynamic iteration for parabolic equations. *BIT.*, **27**: 216–234 (1987)
[34] Yserentant H. Hierarchical bases in the numerical solution to parabolic problems, in Large Scale Scientific Computing (Oberwolfach), Proc. Sci. Comput, Vol 7. Birkhäuser Boston, Deuflhard P, Engquist B eds. 1985, pp. 22–36.
[35] Grisvard P. Elliptic Problems in Nonsmooth Domain. Boston, Pitman, (1985)
[36] Shaidurov V. Multigrid Methods for Finite Elements. Kluwer Academic Publishers, Dordrecht, (1995)
[37 Bank R E, Douglas C C. Sharp estimates for multigrid rates of convergence with general smoothing and acceleration. *SIAM J. Numer. Anal.*, **22**: 617–633 (1985)
[38] Oswald P. On the convergence rate of SOR: a worst case estimate. *Computing.*, **52**: 245–255 (1994)
[39] Kwapien S, Pelzcynski A. The main triangle projection in matrix spaces and its applications. *Studia. Math.*, **34**: 43–68 (1970)
[40] Hackbusch W. Iterative Solution of Large Sparse Systems of Equations. Springer-Verlag. Berlin-Heidelberg-New York, (1994)
[41] Lanczos C. Solutions of systems of linear equations by minimized iterations. *J. Research. Nat. Bur. Standards.*, **49**: 33–53 (1952)
[42] II'IN B.-P. Some estimates for conjugate gradient methods. *USSR Comp. Math and Math. Phys.*, **16**: 22–30. (1976)(firstly published in *Zh. vȳchisl, Mat. mat. Fiz.*, **16**: 847–856 (1976))
[43] Saad Y. Iterative Methods for Sparse Linear Systems. PWS Publishing Company: Boston, (1996)
[44] Braess D, Hackbusch W. A new convergence proof for the multigrid method including the V-cycle. *SIAM J. Numer. Anal.*, **20**: 967–975 (1983)
[45] Clément P. Approximation by finite element functions using local regularization. *Rev. Française Automat. Infomat. Recherche Opérationnelle Sér. Rouge Anal. Numér.*, **9**(R-2): 77–84 (1975)
[46] Ciarlet P G. The Finite Element Method for Elliptic Problems. North-Holland, Amsterdam, (1978)