



RESEARCH ARTICLE

Multilevel Additive Schwarz Preconditioners for High-Order Block Implicit Methods for Parabolic Equations

Shishun Li¹  | Lei Xu² | Xiao-Chuan Cai³ 

¹School of Mathematics and Statistics, Xinyang Normal University, Henan, P. R. China | ²Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, P. R. China | ³Department of Mathematics, University of Macau, Macau, P. R. China

Correspondence: Xiao-Chuan Cai (xccai@um.edu.mo)

Received: 16 July 2024 | **Revised:** 27 February 2025 | **Accepted:** 7 March 2025

Funding: FDCT 0141/2020/A3, 0079/2021/AFJ; MYRG-GRG 2023-00102-FST-UMDF; NSFC 12101588.

Keywords: additive Schwarz preconditioner | block implicit method | convergence rate | parabolic equations | scalability

ABSTRACT

Block implicit method (BIM) is a family of time discretization schemes with desirable stability properties, high order of accuracy, and requires a single initial value. Moreover, BIM is also time-parallel since the solutions at several time steps can be computed simultaneously at a time. However, its resulting linear system is often large and highly ill-conditioned. In this paper, we present a multilevel additive Schwarz preconditioner for solving the large sparse system of algebraic equations arising from the discretization by BIM in time and finite element in space. Under some mild assumption, we prove an optimal convergence theory showing that the convergence rate is bounded independently of the spatial mesh sizes, the time step size, the number of subdomains, and the number of levels. Numerical experiments carried out on a parallel computer with thousands of processors confirm the optimality and scalability of the method, which performs better than the low-order schemes in terms of the compute time.

1 | Introduction

Additive Schwarz (AS) preconditioners have been widely used for solving a large sparse linear system of algebraic equations arising from the implicit discretization of time-dependent partial differential equations (PDEs) on parallel computers. As examples, we mention the parabolic convection–diffusion equations [1], linear and nonlinear elasticity equations [2], Navier–Stokes equations [3, 4], Bidomain model of the heart [5, 6], magnetic reconnection problem [7], and shallow water equations [8, 9]. We note that all the applications of AS are for time-dependent problems discretized by low-order schemes like backward Euler and Crank–Nicolson. High-order implicit schemes, such as linear multistep methods, are seldom used in practice because they require multiple initial values and are subject to the Dahlquist order stability barrier [10]. Another example is the fully implicit Runge–Kutta (IRK) schemes that have all the desirable stability

properties, but at each step, a large, coupled linear system of equations needs to be solved. The diagonal IRK (DIRK) scheme is widely used in practice due to its low computational cost per time step, but it suffers from the reduction of the order of accuracy and the maximum stage order is only two [11]. For these implicit methods to be useful for practical applications, it is important to have an efficient and robust preconditioner, and AS is a good candidate with a provable optimality theory when it is used with low-order methods such as backward Euler and Crank–Nicolson schemes. Unfortunately, the optimality theory cannot be extended to IRK.

In this paper, we introduce and study a multilevel AS algorithm for solving parabolic equations discretized by a family of block implicit method (BIM) in time and finite element in space. BIM was first proposed in [12] and further studied in [13–16]. Recently, some *A*-stable BIM with up to order 10 was proposed

and studied in [17], where the scheme with block size r is described by a tableau

$$\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{a}^T & \mathbf{b}^T \end{array}$$

where $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{r \times r}$ are BIM matrices and $\mathbf{a} = -\mathbf{A}(1, 1, \dots, 1)^T$, $\mathbf{b} = \mathbf{A}(1, 2, \dots, r)^T - \mathbf{B}(1, 1, \dots, 1)^T$. Similar to the IRK, BIM results in a large and ill-conditioned algebraic system of equations when applied to parabolic equations, and some preconditioning techniques have been proposed to solve such linear systems [18–24]. Different from these preconditioners, we present a parallel multilevel AS preconditioner in the tensor form with a theoretically guaranteed optimal convergence rate. Since BIM matrices \mathbf{A} and \mathbf{B} are positive definite and positive diagonal, respectively, the corresponding variational form is also positive and satisfies the conditions required by the Lax–Milgram theorem; the optimal convergence theory of the multilevel AS algorithm can be established in the framework of the Galerkin finite element method. To obtain the optimal convergence, it is necessary to assume that $H/\sqrt{\tau}$ is sufficiently small, where H and τ denote the coarse spatial mesh size and time step size, respectively. It seems reasonable since high-order BIM allows larger time step size. Note that the optimality theory does not hold for fully IRK since the Runge–Kutta matrix is only positive stable but not positive definite [20].

One important merit of BIM is that it is time-parallel since the solutions at several time steps are computed simultaneously at a time. The multilevel AS algorithm based on BIM increases the space–time concurrency and is suitable for large-scale parallel computers. In the past few years, parallel-in-time (PinT) algorithms have attracted a lot of attention for applications on high-performance parallel computers; see, for examples, [25–27] and the references therein. In [28, 29], the authors presented some space–time AS algorithms with provable optimal convergence for parabolic equations discretized by backward Euler in time and finite elements in space. Unfortunately, the optimal convergence theory cannot be extended to higher-order time discretization methods such as Crank–Nicolson and IRK schemes. In this paper, we develop an optimal convergence theory for BIM, which outperforms AS for low-order schemes numerically.

The rest of the paper is organized as follows. In Section 2, we first review the AS algorithm for parabolic equations discretized by the backward Euler scheme and introduce the multilevel AS algorithms based on BIM. In Section 3, a detailed convergence analysis of the proposed multilevel AS algorithm is given. Some numerical results are reported in Section 4 to illustrate the performance of the proposed algorithm. Finally, some concluding remarks are given in Section 5.

2 | Block Implicit Additive Schwarz Algorithms

We consider a parabolic problem

$$\begin{cases} u_t - \nabla \cdot (\alpha(\mathbf{x}) \nabla u) + \beta(\mathbf{x}) \cdot \nabla u + \gamma(\mathbf{x})u = f(\mathbf{x}, t), & \text{in } \Omega \times (0, T], \\ u(\mathbf{x}, t) = 0, & \text{on } \partial\Omega \times (0, T], \\ u(\mathbf{x}, 0) = u_0(\mathbf{x}), & \text{in } \Omega \end{cases} \quad (1)$$

where $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) is a bounded, open polygonal (or polyhedra) domain, $(0, T]$ is the temporal interval, $f(\mathbf{x}, t) \in L^2(\Omega \times [0, T])$, and $u_0(\mathbf{x})$ is the initial value. Assume all the coefficients are sufficiently smooth, $0 < \alpha_0 \leq \alpha(\mathbf{x}) \leq \alpha_1 < +\infty$ and $\gamma(\mathbf{x}) - \frac{1}{2} \nabla \cdot \beta(\mathbf{x}) \geq 0$. For brevity, we omit the variable \mathbf{x} in the following discussion. In this paper, $C > 0$ denotes a generic constant, which is independent of any variables and may take different values at different occurrences.

2.1 | Multilevel Additive Schwarz Preconditioning for the Backward Euler Scheme

Let $0 = t_0 < t_1 < \dots < t_M = T$ be a uniform partition of $[0, T]$ and set $\tau = t_k - t_{k-1}$. Suppose u^k is the solution at time t_k . By using the backward Euler scheme for the time discretization, the weak form of the semi-discretized problem at each time step is

$$D_\tau(u, v) \equiv (u, v) + \tau D(u, v) = g(v) \quad (2)$$

where $u = u^k$, $g(v) = (u^{k-1}, v) + \tau(f^k, v)$, $f^k = f(\mathbf{x}, t_k)$ and

$$D(u, v) = \int_{\Omega} (\alpha \nabla u \cdot \nabla v + \beta \cdot \nabla uv + \gamma uv) d\mathbf{x} \quad \forall u, v \in H_0^1(\Omega)$$

Define $A(u, v) = \int_{\Omega} \alpha \nabla u \cdot \nabla v d\mathbf{x}$ and $S(u, v) = \int_{\Omega} (\beta \cdot \nabla uv + \gamma uv) d\mathbf{x}$, it is clear that $D(u, v) = A(u, v) + S(u, v)$. From the assumptions on the coefficients of (1), there exist constants $C > 0$ and $m_0 > 0$ such that

$$D(v, v) \geq m_0 \|v\|_1^2 \quad \forall v \in H_0^1(\Omega) \quad (3)$$

and

$$D(u, v) \leq C \|u\|_1 \|v\|_1 \quad \forall u, v \in H_0^1(\Omega) \quad (4)$$

where $\|v\|_1 = \left(\int_{\Omega} (v^2 + \nabla v \cdot \nabla v) d\mathbf{x} \right)^{\frac{1}{2}}$ is the $H_0^1(\Omega)$ norm in the Sobolev space.

To describe the finite element discretization and the domain decomposition method, we define a sequence of nested conforming meshes $\mathcal{T}_l = \{\mathcal{K}_i^l\}_{i=1}^{N_l}$ ($l = 1, 2, \dots, L$) obtained by the uniform refinement of an initial mesh \mathcal{T}_0 . Let $h_l = \max_i \text{diam}(\mathcal{K}_i^l)$. Now, we introduce an overlapping decomposition $\Omega = \bigcup_{i=1}^{N_l} \Omega_i^l$ on each level l ($l \geq 1$), where Ω_i^l is extended from the nonoverlapping subdomains Ω_i by adding several layers of fine mesh elements, and the corresponding overlap is denoted by δ_l ; that is, $\delta_l = O(h_l)$. We assume the decomposition satisfies Assumption 2.1 in [30]. Denote the diameter of Ω_i^l by H_l such that $H_l = O(h_{l-1})$. Then, we denote $V_l = V_{h_l}$ ($l = 0, 1, \dots, L$) as the space of continuous piecewise linear functions associated with the partition \mathcal{T}_l . For brevity, let $H = h_0$ and $h = h_L$ be the mesh sizes of the coarsest space V_0 and the finest space $V_h = V_{h_L}$, respectively. Hence, V_h can be represented as

$$V_h = V_0 + \sum_{l=1}^L V_l = V_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} V_i^l$$

where $V_i^l = V_l \cap H_0^1(\Omega_i^l)$ is the finite element subspace defined on Ω_i^l . The standard Galerkin approximation of (2) is to find $u_h^* \in V_h$ such that

$$D_\tau(u_h^*, v_h) = g(v_h), \quad \forall v_h \in V_h \quad (5)$$

ALGORITHM 1 | (MAS-BE).

Find the finite element solution u_h^* of (5) or (9) by solving the operator equation (13) with GMRES, where the action of the preconditioner P^{-1} on the global residual $r = g_h - (M + \tau K)u_h$ is carried out as follows:

```

 $z \leftarrow R_0^T(M_0 + \tau K_0)^{-1} R_0 r$ 
For  $l = 1, 2, \dots, L$ 
     $z \leftarrow z + \sum_{i=1}^{N_l} (R_l^i)^T (M_{li} + \tau K_{li})^{-1} R_l^i r$ 
End

```

Define the linear operators $M_0, K_0 : V_0 \mapsto V_0$, $M, K : V_h \mapsto V_h$ and $M_{li}, K_{li} : V_l^i \mapsto V_l^i$ by

$$\begin{aligned} (K_0 u_0, v_0) &= D(u_0, v_0) \quad \forall v_0 \in V_0 \\ (M_0 u_0, v_0) &= (u_0, v_0) \quad \forall v_0 \in V_0 \end{aligned} \quad (6)$$

$$\begin{aligned} (K u_h, v_h) &= D(u_h, v_h) \quad \forall v_h \in V_h \\ (M u_h, v_h) &= (u_h, v_h) \quad \forall v_h \in V_h \end{aligned} \quad (7)$$

$$\begin{aligned} (K_{li} u_l^i, v_l^i) &= D(u_l^i, v_l^i) \quad \forall v_l^i \in V_l^i \\ (M_{li} u_{li}, v_{li}) &= (u_l^i, v_l^i) \quad \forall v_l^i \in V_l^i \end{aligned} \quad (8)$$

Then, (5) can be rewritten as the linear operator equation

$$(M + \tau K)u_h^* = g_h \quad (9)$$

Let $P_0 : V_h \mapsto V_0$ and $P_l^i : V_h \mapsto V_l^i$ be projections defined by

$$D_\tau(P_0 v_h, v_0) = D_\tau(v_h, v_0) \quad \forall v_h \in V_h, v_0 \in V_0 \quad (10)$$

$$D_\tau(P_l^i v_h, v_l^i) = D_\tau(v_h, v_l^i) \quad \forall v_h \in V_h, v_l^i \in V_l^i \quad (11)$$

The multilevel additive Schwarz operator $P : V_h \mapsto V_h$ can be defined by

$$P = P_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} P_l^i \quad (12)$$

Therefore, (9) is equivalent to the preconditioned operator equation

$$P u_h = b_h \quad (13)$$

where $b_h = P u_h^* = \tilde{P}^{-1}(M + \tau K)u_h^* = \tilde{P}^{-1}g_h$ and

$$\tilde{P}^{-1} = R_0^T(M_0 + \tau K_0)^{-1} R_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} (R_l^i)^T (M_{li} + \tau K_{li})^{-1} R_l^i$$

where $R_0 : V_h \mapsto V_0$ and $R_l^i : V_h \mapsto V_l^i$ denote the restriction matrices. The classical multilevel additive Schwarz preconditioner for the backward Euler (MAS-BE) scheme for solving (1) is formally described as follows [1, 31].

2.2 | Multilevel Additive Schwarz Preconditioning for Block Implicit Methods

Now, we present an additive Schwarz algorithm by replacing the backward Euler scheme in Algorithm 1 by the BIMs in the temporal discretization. The weak form of the semi-discretized problem at each time-block of size r is to find $u^i \in H_0^1(\Omega)$, $i = 1, 2, \dots, r$, such that

$$\begin{cases} \sum_{j=1}^r a_{1j}(u^j, v^1) + \tau \sum_{j=1}^r b_{1j} D(u^j, v^1) = g_1(v^1), & \forall v^1 \in H_0^1(\Omega), \\ \sum_{j=1}^r a_{2j}(u^j, v^2) + \tau \sum_{j=1}^r b_{2j} D(u^j, v^2) = g_2(v^2), & \forall v^2 \in H_0^1(\Omega), \\ \vdots \\ \sum_{j=1}^r a_{rj}(u^j, v^r) + \tau \sum_{j=1}^r b_{rj} D(u^j, v^r) = g_r(v^r), & \forall v^r \in H_0^1(\Omega) \end{cases} \quad (14)$$

where the bilinear forms (\cdot, \cdot) and $D(\cdot, \cdot)$ are defined as that in (2) and the operator

$$\begin{aligned} g_i(v^i) &= -a_{i0}(u^0, v^i) + \tau b_{i0}(f^0, v^i) - \tau b_{i0} D(u^0, v^i) \\ &\quad + \tau \sum_{j=1}^r b_{ij}(f^j, v^i), \quad \forall v^i \in H_0^1(\Omega) \end{aligned}$$

Let $\mathbf{u} = (u^1, u^2, \dots, u^r)^T \in \mathbf{H}_0^1(\Omega) \equiv (H_0^1(\Omega))^r$, $\mathbf{f} = (f^1, f^2, \dots, f^r)^T \in \mathbf{L}^2(\Omega \times [0, T]) \equiv (L^2(\Omega \times [0, T]))^r$. Set $\nabla \mathbf{u} = (\nabla u^1, \nabla u^2, \dots, \nabla u^r)^T$, $\mathbf{a} = (a_{10} \ a_{20} \ \dots \ a_{r0})^T$ and $\mathbf{b} = (b_{10} \ b_{20} \ \dots \ b_{r0})^T$. For any $\mathbf{v} \in \mathbf{H}_0^1(\Omega)$, the equivalent variational form of (14) is

$$D_\tau(\mathbf{u}, \mathbf{v}) \equiv (\mathbf{A}\mathbf{u}, \mathbf{v}) + \tau D(\mathbf{u}, \mathbf{v}) = \mathbf{g}(\mathbf{v}) \quad (15)$$

where

$$\begin{aligned} D(\mathbf{u}, \mathbf{v}) &= \int_{\Omega} (\alpha \mathbf{B} \nabla \mathbf{u} \cdot \nabla \mathbf{v} + \mathbf{B}(\beta \cdot \nabla \mathbf{u})\mathbf{v} + \gamma \mathbf{B} \mathbf{u} \mathbf{v}) d\mathbf{x} \\ \forall \mathbf{u}, \mathbf{v} &\in \mathbf{H}_0^1(\Omega) \end{aligned}$$

and

$$\begin{aligned} \mathbf{g}(\mathbf{v}) &= \tau(\mathbf{B}\mathbf{f}, \mathbf{v}) - (\mathbf{a} \otimes u^0, \mathbf{v}) + \tau(\mathbf{b} \otimes f^0, \mathbf{v}) \\ &\quad - \tau(\mathbf{b} \otimes \alpha \nabla u^0, \nabla \mathbf{v}) - \tau(\mathbf{b} \otimes (\beta \cdot \nabla u^0 + \gamma u^0), \mathbf{v}) \end{aligned}$$

Here, $\mathbf{a} \otimes u = (a_1 u, a_2 u, \dots, a_n u)^T$ denotes a vector-valued function for the vector $\mathbf{a} \in \mathbb{R}^n$ and the function u .

Define

$$\begin{aligned} \mathcal{A}_\tau(\mathbf{u}, \mathbf{v}) &= (\mathbf{u}, \mathbf{v}) + \tau(\alpha \mathbf{B} \nabla \mathbf{u}, \nabla \mathbf{v}), \quad \mathcal{N}(\mathbf{u}, \mathbf{v}) = ((\mathbf{A} - \mathbf{I})\mathbf{u}, \mathbf{v}) \\ \mathcal{S}_\tau(\mathbf{u}, \mathbf{v}) &= \tau(\mathbf{B}(\beta \cdot \nabla \mathbf{u}) + \gamma \mathbf{B} \mathbf{u}, \mathbf{v}) \quad \forall \mathbf{u}, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \end{aligned}$$

where $\mathbf{I} \in \mathbb{R}^{r \times r}$ is the identity matrix. Then, (15) is equivalent to

$$D_\tau(\mathbf{u}, \mathbf{v}) \equiv \mathcal{A}_\tau(\mathbf{u}, \mathbf{v}) + \mathcal{S}_\tau(\mathbf{u}, \mathbf{v}) + \mathcal{N}(\mathbf{u}, \mathbf{v}) = \mathbf{g}(\mathbf{v}) \quad (16)$$

Now, we consider a family of BIM with positive definite matrix \mathbf{A} and positive diagonal matrix \mathbf{B} , that is,

$$\begin{aligned} 0 < \lambda_{\min}(\mathbf{A}_H) &\leq \frac{(\mathbf{A}\mathbf{v}, \mathbf{v})}{(\mathbf{v}, \mathbf{v})} \leq \lambda_{\max}(\mathbf{A}_H) \text{ and } 0 < \min_i \{b_{ii}\} \\ &\leq \frac{(\mathbf{B}\mathbf{v}, \mathbf{v})}{(\mathbf{v}, \mathbf{v})} \leq \max_i \{b_{ii}\} \end{aligned} \quad (17)$$

where $\mathbf{A}_H = (\mathbf{A} + \mathbf{A}^T)/2$ is the Hermitian part of the matrix \mathbf{A} , $\lambda_{\min}(\mathbf{A}_H)$ and $\lambda_{\max}(\mathbf{A}_H)$ are the minimum and the maximum eigenvalues of \mathbf{A}_H , b_{ii} is the i^{th} diagonal element of \mathbf{B} . Let $\|\cdot\|$ and $|\cdot|_1$ be the $L^2(\Omega)$ norm and $H_0^1(\Omega)$ seminorm in the Sobolev space, respectively. Define the τ -norm of \mathbf{v} by $\|\mathbf{v}\|_\tau^2 = \|\mathbf{v}\|^2 + \tau|\mathbf{v}|_1^2$, it is clear that

$$\hat{c}_0 \|\mathbf{v}\|_\tau^2 \leq \mathcal{A}_\tau(\mathbf{v}, \mathbf{v}) \leq \hat{c}_1 \|\mathbf{v}\|_\tau^2 \quad (18)$$

where $\hat{c}_0 = \min\{1, \alpha_0 / \min_i \{b_{ii}\}\}$, $\hat{c}_1 = \max\{1, \alpha_1 \max_i \{b_{ii}\}\}$.

Remark 1. Generally speaking, the stability of BIM is uniquely determined by $\mathbf{B}^{-1}\mathbf{A}$ and is independent of the individual matrices \mathbf{A} and \mathbf{B} [13, 16, 17]. Therefore, various BIMs can be constructed by choosing the BIM matrices \mathbf{A} and \mathbf{B} . Here, we only focus on some BIM with special \mathbf{A} and \mathbf{B} . For example, for a fourth-order L -stable BIM defined by the BIM matrices

$$\mathbf{A} = \begin{pmatrix} 19/6 & -3/2 & 5/6 & -1/6 \\ -56/15 & 12/5 & -8/15 & 2/15 \\ 1/4 & -3/4 & 5/12 & 1/8 \\ -1/15 & 3/20 & -1/5 & 5/48 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4/5 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & 1/20 \end{pmatrix}$$

and the BIM vectors $\mathbf{a} = -\mathbf{A}\mathbf{e}$, $\mathbf{b} = \mathbf{A}\tilde{\mathbf{x}} - \mathbf{B}\mathbf{e}$, where $\tilde{\mathbf{x}} = (1, 2, 3, 4)^T$, $\mathbf{e} = (1, 1, 1, 1)^T$, it is easy to verify that \mathbf{A} is positive definite and

$$\|\mathbf{v}\|^2 + \frac{\alpha_0}{20} \tau |\mathbf{v}|_1^2 \leq \mathcal{A}_\tau(\mathbf{v}, \mathbf{v}) \leq \|\mathbf{v}\|^2 + \alpha_1 \tau |\mathbf{v}|_1^2$$

Then, (18) holds with $\hat{c}_0 = \min\{1, \alpha_0/20\}$, $\hat{c}_1 = \max\{1, \alpha_1\}$.

Some basic estimates for the bilinear forms involved $D_\tau(\cdot, \cdot)$ are summarized as follows:

Lemma 1. *There exists a constant $C > 0$ independent of τ such that*

1. $|S_\tau(\mathbf{u}, \mathbf{v})| \leq C\tau \|\mathbf{u}\|_1 \|\mathbf{v}\| \quad \forall \mathbf{u}, \mathbf{v} \in H_0^1(\Omega),$
2. $D_\tau(\mathbf{u}, \mathbf{v}) \leq C \|\mathbf{u}\|_\tau \|\mathbf{v}\|_\tau \quad \forall \mathbf{u}, \mathbf{v} \in H_0^1(\Omega),$
3. $D_\tau(\mathbf{u}, \mathbf{u}) \geq c_0 \|\mathbf{u}\|_\tau^2 \quad \forall \mathbf{u} \in H_0^1(\Omega),$

where $c_0 = \min\{\lambda_{\min}(\mathbf{A}_H), m_0\} / \min_i \{b_{ii}\}$ is a constant.

Proof. 1. It follows from (14), (16), (17) and the Cauchy-Schwarz inequality that

$$\begin{aligned} S_\tau(\mathbf{u}, \mathbf{v}) &= \tau \sum_{i,j=1}^r b_{ij} (\beta \cdot \nabla \mathbf{u}^i + \gamma \mathbf{u}^i, \mathbf{v}^j) \leq \tau \sum_{i=1}^r (\beta \cdot \nabla \mathbf{u}^i + \gamma \mathbf{u}^i, \mathbf{v}^i) \\ &\leq C\tau \sum_{i=1}^r |\mathbf{u}^i|_1 \|\mathbf{v}^i\| \leq C\tau \|\mathbf{u}\|_1 \|\mathbf{v}\| \end{aligned}$$

2. Combining (4), (14), (17) and the Cauchy-Schwarz inequality

$$\begin{aligned} D_\tau(\mathbf{u}, \mathbf{v}) &= \sum_{i,j=1}^r a_{ij} (\mathbf{u}^i, \mathbf{v}^j) + \tau \sum_{i,j=1}^r b_{ij} D(\mathbf{u}^i, \mathbf{v}^j) \\ &\leq \max_{i,j} |a_{ij}| \sum_{i,j=1}^r (\mathbf{u}^i, \mathbf{v}^j) + \tau \max_i \{b_{ii}\} \sum_{i=1}^r D(\mathbf{u}^i, \mathbf{v}^i) \end{aligned}$$

$$\begin{aligned} &\leq C \left(\|\mathbf{u}\| \|\mathbf{v}\| + \tau \sum_{i=1}^r \|\mathbf{u}^i\|_1 \|\mathbf{v}^i\|_1 \right) \\ &\leq C \|\mathbf{u}\|_\tau \|\mathbf{v}\|_\tau \end{aligned}$$

3. From (3), (14) and (17), we have

$$\begin{aligned} D_\tau(\mathbf{u}, \mathbf{u}) &= \sum_{i,j=1}^r a_{ij} (\mathbf{u}^i, \mathbf{u}^j) + \tau \sum_{i,j=1}^r b_{ij} D(\mathbf{u}^i, \mathbf{u}^j) \\ &\geq \lambda_{\min}(\mathbf{A}_H) \|\mathbf{u}\|^2 + \frac{\tau}{\min_i \{b_{ii}\}} \sum_{i=1}^r D(\mathbf{u}^i, \mathbf{u}^i) \\ &\geq \lambda_{\min}(\mathbf{A}_H) \|\mathbf{u}\|^2 + \frac{m_0 \tau}{\min_i \{b_{ii}\}} \|\mathbf{u}\|_1^2 \\ &\geq c_0 \|\mathbf{u}\|_\tau^2 \end{aligned}$$

where $c_0 = \min\{\lambda_{\min}(\mathbf{A}_H), m_0\} / \min_i \{b_{ii}\}$. \square

Assume that the time interval $[0, T]$ is divided into M_t time blocks and there are r time steps in each block, we have $rM_t = M$ and $M\tau = T$. The partitions of the domain Ω for BIM is described in Figure 1. Define the coarse space $\mathbf{V}_0 = (V_0)^r$, subspaces $\mathbf{V}_l^i = (V_l^i)^r$, the finite element space $\mathbf{V}_h = (V_h)^r$ can be represented as

$$\mathbf{V}_h = \mathbf{V}_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathbf{V}_l^i$$

The Galerkin approximation of (15) is to find $\mathbf{u}_h^* \in \mathbf{V}_h$ such that

$$D_\tau(\mathbf{u}_h^*, \mathbf{v}_h) = \mathbf{g}(\mathbf{v}_h), \quad \forall \mathbf{v}_h \in \mathbf{V}_h \quad (19)$$

which can be rewritten as the linear operator equation in the tensor form:

$$\mathbf{D}\mathbf{u}_h^* \equiv (\mathbf{A} \otimes \mathbf{M} + \tau \mathbf{B} \otimes \mathbf{K}) \mathbf{u}_h^* = \mathbf{g}_h \quad (20)$$

where \mathbf{K} and \mathbf{M} are defined in (7). Let $\mathcal{P}_0 : \mathbf{V}_h \mapsto \mathbf{V}_0$ and $\mathcal{P}_l^i : \mathbf{V}_h \mapsto \mathbf{V}_l^i$ be projections defined by

$$D_\tau(\mathcal{P}_0 \mathbf{v}_h, \mathbf{v}_0) = D_\tau(\mathbf{v}_h, \mathbf{v}_0) \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \mathbf{v}_0 \in \mathbf{V}_0 \quad (21)$$

$$D_\tau(\mathcal{P}_l^i \mathbf{v}_h, \mathbf{v}_l^i) = D_\tau(\mathbf{v}_h, \mathbf{v}_l^i) \quad \forall \mathbf{v}_h \in \mathbf{V}_h, \mathbf{v}_l^i \in \mathbf{V}_l^i \quad (22)$$

The multilevel additive Schwarz (MAS) operator $\mathcal{P} : \mathbf{V}_h \mapsto \mathbf{V}_h$ for BIM can be defined by

$$\mathcal{P} = \mathcal{P}_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \quad (23)$$

Therefore, (20) is equivalent to the preconditioned operator equation

$$\mathcal{P}\mathbf{u}_h = \mathbf{b}_h \quad (24)$$

where $\mathbf{b}_h = \mathcal{P}\mathbf{u}_h^* = \tilde{\mathcal{P}}^{-1} \mathbf{D}\mathbf{u}_h^* = \tilde{\mathcal{P}}^{-1} \mathbf{g}_h$ and the additive Schwarz preconditioner

$$\tilde{\mathcal{P}}^{-1} = \mathcal{R}_0^T \mathcal{D}_0^{-1} \mathcal{R}_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} (\mathcal{R}_l^i)^T \mathcal{D}_l^{-1} \mathcal{R}_l^i$$

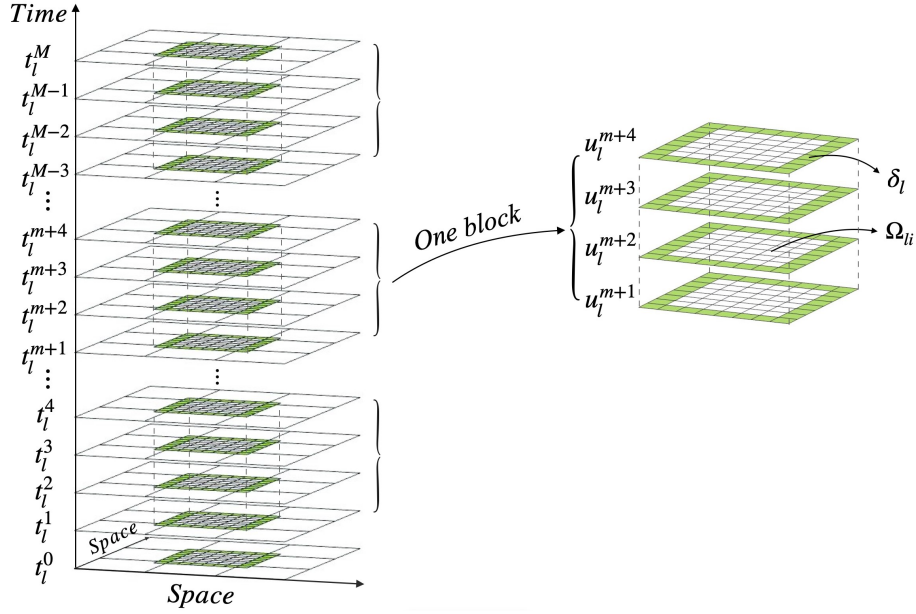


FIGURE 1 | Partition of the space–time domain and subdomains on level l . Here, Ω_{li} denotes the subdomains in space, δ_l is the overlapping size on the mesh level l , $m = 4j$ ($j = 0, 1, 2, \dots, M_l - 1$) and M_l denotes the number of blocks.

ALGORITHM 2 | (MAS-BIM).

Find the finite element solution \mathbf{u}_h^* of (20) by solving the operator equation (24) with GMRES, where the action of the preconditioner $\tilde{\mathcal{P}}^{-1}$ on the global residual $\mathbf{r} = \mathbf{g}_h - \mathbf{D}\mathbf{u}_h$ is carried out as follows:

```

 $\mathbf{z} \leftarrow \mathcal{R}_0^T \mathbf{D}_0^{-1} \mathcal{R}_0 \mathbf{r}$ 
For  $l = 1, 2, \dots, L$ 
   $\mathbf{z} \leftarrow \mathbf{z} + \sum_{i=1}^{N_l} (\mathcal{R}_l^i)^T \mathbf{D}_{li}^{-1} \mathcal{R}_l^i \mathbf{r}$ 
End

```

Here, $\mathbf{D}_0 = \mathbf{A} \otimes \mathbf{M}_0 + \tau \mathbf{B} \otimes \mathbf{K}_0$, $\mathbf{D}_{li} = \mathbf{A} \otimes \mathbf{M}_{li} + \tau \mathbf{B} \otimes \mathbf{K}_{li}$, \mathbf{M}_0 , \mathbf{K}_0 , \mathbf{M}_{li} and \mathbf{K}_{li} are defined in (6) and (8). The block restriction matrices $\mathcal{R}_0 : \mathbf{V}_h \rightarrow \mathbf{V}_0$ and $\mathcal{R}_l^i : \mathbf{V}_h \rightarrow \mathbf{V}_l^i$ are defined as $\mathcal{R}_0 = \mathbf{I} \otimes \mathbf{R}_0$ and $\mathcal{R}_l^i = \mathbf{I} \otimes \mathbf{R}_l^i$.

Now, we present the multilevel additive Schwarz algorithm for BIM.

Assume the solution of the elliptic problem $D(\mathbf{v}, \mathbf{u}) = \mathbf{g}(\mathbf{v})$ is $H^2(\Omega)$ -regular; that is $\|\mathbf{u}\|_2 \leq \|\mathbf{g}\|$, we have the following lemma:

Lemma 2. For any $\mathbf{g} \in L^2(\Omega)$, the problem

$$D_\tau(\mathbf{v}, \mathbf{w}) = \mathbf{g}(\mathbf{v}), \quad \mathbf{v} \in \mathbf{H}_0^1(\Omega)$$

has a solution such that $\|\mathbf{w}\|_2 \leq C\|\mathbf{g}\|/\tau$.

Proof. From (15), we have

$$(\mathbf{A}\mathbf{w}, \mathbf{w}) + \tau D(\mathbf{w}, \mathbf{w}) = (\mathbf{g}, \mathbf{w})$$

It follows from (3) and the positive definiteness of \mathbf{A} that $\|\mathbf{w}\| \leq \|\mathbf{g}\|/\lambda_{\min}(\mathbf{A}_H)$, which implies that

$$\|\mathbf{g} - \mathbf{A}\mathbf{w}\| \leq \|\mathbf{g}\| + \|\mathbf{A}\mathbf{w}\| \leq \left(1 + \frac{\lambda_{\max}(\mathbf{A}_H)}{\lambda_{\min}(\mathbf{A}_H)}\right) \|\mathbf{g}\| \leq C\|\mathbf{g}\|$$

Therefore, we have $D(\mathbf{v}, \mathbf{w}) = ((\mathbf{g} - \mathbf{A}\mathbf{w})/\tau, \mathbf{v})$, combining the regularity assumption yields

$$\|\mathbf{w}\|_2 \leq \|\mathbf{g} - \mathbf{A}\mathbf{w}\|/\tau \leq C\|\mathbf{g}\|/\tau$$

This completes the proof. \square

Then, we obtain the following error estimate:

Lemma 3. There exists a positive constant C independent of H , h and τ , such that

$$\|\mathcal{P}_0 \mathbf{u}_h - \mathbf{u}_h\| \leq C \frac{H}{\sqrt{\tau}} \sqrt{\frac{H^2}{\tau} + 1} \|\mathcal{P}_0 \mathbf{u}_h - \mathbf{u}_h\|_\tau$$

and

$$\|\mathcal{P}_0 \mathbf{u}_h - \mathbf{u}_h\|_\tau \leq C\|\mathbf{u}_h\|_\tau$$

Proof. Consider an auxiliary problem: Find $\mathbf{w} \in \mathbf{H}_0^1(\Omega)$, such that

$$D_\tau(\mathbf{v}, \mathbf{w}) = (\mathbf{u}_h - \mathcal{P}_0 \mathbf{u}_h, \mathbf{v}) \quad \mathbf{v} \in \mathbf{H}_0^1(\Omega)$$

Let $\mathbf{w}_H \in \mathbf{V}_0$ be the finite element approximation of \mathbf{w} , it follows from (21) and Lemma 1 that

$$\begin{aligned} \|\mathbf{u}_h - \mathcal{P}_0 \mathbf{u}_h\|^2 &= D_\tau(\mathbf{u}_h - \mathcal{P}_0 \mathbf{u}_h, \mathbf{w}) = D_\tau(\mathbf{u}_h - \mathcal{P}_0 \mathbf{u}_h, \mathbf{w} - \mathbf{w}_H) \\ &\leq C\|\mathbf{u}_h - \mathcal{P}_0 \mathbf{u}_h\|_\tau \|\mathbf{w} - \mathbf{w}_H\|_\tau \end{aligned} \quad (25)$$

From the definition of the τ -norm and the a priori error estimates for the classical elements, we obtain

$$\begin{aligned} \|\mathbf{w} - \mathbf{w}_H\|_\tau &= \sqrt{\|\mathbf{w} - \mathbf{w}_H\|^2 + \tau \|\mathbf{w} - \mathbf{w}_H\|_1^2} \\ &\leq \sqrt{CH^4 \|\mathbf{w}\|_2^2 + C\tau H^2 \|\mathbf{w}\|_2^2} \leq CH\sqrt{H^2 + \tau} \|\mathbf{w}\|_2 \end{aligned}$$

Combining the above inequality with (25) and Lemma 2 implies

$$\begin{aligned}\|u_h - P_0 u_h\|^2 &\leq CH\sqrt{H^2 + \tau}\|u_h - P_0 u_h\|_\tau \|w\|_2 \\ &\leq C\frac{H}{\sqrt{\tau}}\sqrt{\frac{H^2}{\tau} + 1}\|u_h - P_0 u_h\|_\tau \|u_h - P_0 u_h\|\end{aligned}$$

which yields

$$\|u_h - P_0 u_h\| \leq C\frac{H}{\sqrt{\tau}}\sqrt{\frac{H^2}{\tau} + 1}\|u_h - P_0 u_h\|_\tau$$

From (21) and Lemma 1, we obtain

$$\begin{aligned}c_0\|u_h - P_0 u_h\|_\tau^2 &= D_\tau(u_h - P_0 u_h, u_h - P_0 u_h) \\ &= D_\tau(u_h - P_0 u_h, u_h) \leq C\|u_h - P_0 u_h\|_\tau \|u_h\|_\tau\end{aligned}$$

Hence, $\|P_0 u_h - u_h\|_\tau \leq C\|u_h\|_\tau$. \square

3 | Optimality of the Multilevel Additive Schwarz Algorithm

In this section, we present an optimal convergence theory for the multilevel additive Schwarz algorithm for parabolic equations discretized by BIM. We first prove the following two important properties of the decomposition of V_h .

Lemma 4. (Stable decomposition) For any $v_h \in V_h$, there exist $v_0 = v_0 \in V_0$ and $v_l^i \in V_l^i$ such that $v_h = v_0 + \sum_{l=1}^L \sum_{i=1}^{N_l} v_l^i$ and

$$\|v_0\|_\tau^2 + \sum_{l=1}^L \sum_{i=1}^{N_l} \|v_l^i\|_\tau^2 \leq C \max_{1 \leq l \leq L} (1 + h_{l-1}/\delta_l) \|v_h\|_\tau^2 \quad (26)$$

where $C > 0$ is a constant independent of H, h, τ , and L .

Remark 2. The upper bound for the multilevel stable decomposition in $H^1(\Omega)$ -norm and $L^2(\Omega)$ -norm is established in [30] and [31], respectively. And the results can be extended easily to $\|\cdot\|_\tau$ norm. For more details, we refer to [5, 29].

Lemma 5. (Strengthened Cauchy-Schwarz inequalities) There exist constants $0 \leq \theta_{ij}^{lk} \leq 1$ ($1 \leq i \leq N_l, 1 \leq j \leq N_k, 1 \leq l, k \leq L$) independent of H, h, τ , and L , such that

$$(u_l^i, u_k^j) \leq \theta_{ij}^{lk} (u_l^i, u_l^i)^{\frac{1}{2}} (u_k^j, u_k^j)^{\frac{1}{2}} \quad \forall u_l^i \in V_l^i, u_k^j \in V_k^j \quad (27)$$

$$\mathcal{A}_\tau(u_l^i, u_k^j) \leq \theta_{ij}^{lk} \mathcal{A}_\tau(u_l^i, u_l^i)^{\frac{1}{2}} \mathcal{A}_\tau(u_k^j, u_k^j)^{\frac{1}{2}} \quad \forall u_l^i \in V_l^i, u_k^j \in V_k^j \quad (28)$$

Proof. The inequality (27) can be obtained directly from Lemma 3.3 of [31]. Since B is positive diagonal, it follows from Lemma 3.2 of [30] and Cauchy-Schwarz inequality that

$$\begin{aligned}\mathcal{A}_\tau(u_l^i, u_k^j) &= (u_l^i, u_k^j) + \tau(\alpha B \nabla u_l^i, \nabla u_k^j) \\ &\leq \theta_{ij}^{lk} (u_l^i, u_l^i)^{\frac{1}{2}} (u_k^j, u_k^j)^{\frac{1}{2}} + \tau \theta_{ij}^{lk} (\alpha B u_l^i, u_l^i)^{\frac{1}{2}} (\alpha B u_k^j, u_k^j)^{\frac{1}{2}} \\ &\leq \theta_{ij}^{lk} \mathcal{A}_\tau(u_l^i, u_l^i)^{\frac{1}{2}} \mathcal{A}_\tau(u_k^j, u_k^j)^{\frac{1}{2}}\end{aligned}$$

This completes the proof. \square

It has been proven that $\theta_{ij}^{lk} \leq C\eta^{d|k-l-1|/2}$ in [30], where $C > 0$ is a constant and η denotes the ratio of the spatial mesh sizes from one level to the next coarser level. Set $\Theta = \{\Theta_{ij}^{lk}\}_{1 \leq l, k \leq L}$ and $\tilde{\Theta} = \{\|\Theta_{ij}^{lk}\|_0\}_{1 \leq l, k \leq L}$, where $\Theta_{ij}^{lk} = \{\theta_{ij}^{lk}\}_{1 \leq i \leq N_l, 1 \leq j \leq N_k}$. It shows that

$$\|\Theta\|_0 \leq \|\tilde{\Theta}\|_0 \leq CN_c \frac{1}{1 - \sqrt{\eta}} \quad (29)$$

Here, $\|\cdot\|_0$ denotes the l_2 -norm (the Euclidean norm), N_c is the number of colors of the decompositions $\{\Omega'_{ij}\}$, that is, if $x \in \Omega$, then it belongs to at most N_c subdomains in $\{\Omega'_{ij}\}$.

In Algorithm 2, the preconditioned operator equation is solved by GMRES. Following [32, 33], the convergence rate of the multilevel additive Schwarz preconditioned GMRES method can be estimated by the ratio of minimal eigenvalue of the symmetric part of the operator to the norm of the operator, and the residual at the k^{th} iteration is bounded as

$$\|r_k\|_\tau \leq \left(1 - \frac{c_p^2}{C_p^2}\right)^{\frac{k}{2}} \|r_0\|_\tau \quad (30)$$

where $r_k = b_h - \mathcal{P}u_h^{(k)}$, and the two quantities are defined by

$$c_p = \inf_{u_h \neq 0} \frac{\mathcal{A}_\tau(\mathcal{P}u_h, u_h)}{\mathcal{A}_\tau(u_h, u_h)} \quad \text{and} \quad C_p = \sup_{u_h \neq 0} \frac{\mathcal{A}_\tau(\mathcal{P}u_h, \mathcal{P}u_h)^{\frac{1}{2}}}{\mathcal{A}_\tau(u_h, u_h)^{\frac{1}{2}}} \quad (31)$$

where \mathcal{P} is defined by (23).

Remark 3. The convergence bounds (30) for preconditioned GMRES are originally derived using the Euclidean inner product and the associated norm [32, 34]. For the additive Schwarz preconditioned GMRES method applied to the linear systems arising from discretizations of elliptic problems, the optimal bounds are obtained in the energy norm (or the equivalent Sobolev $H^1(\Omega)$ norm) [35, 36]. However, some numerical examples are constructed to show that the bounds in l_2 -norm may not be optimal [37], that is, the quantities appearing in the bounds may depend on the mesh size h . In [33], an asymptotic optimality of the above method in l_2 -norm was obtained. By using an a posteriori bound for GMRES in [38], the authors prove that the additive Schwarz preconditioned GMRES without coarse grid correction is either optimal or close to be optimal; see [39]. Following the idea of [1, 31], we prove below that the bounds in (30) are optimal in the $\|\cdot\|_\tau$ norm.

Assumption 1. Let \mathcal{P}_0 and \mathcal{P}_l^i be defined as in (21) and (22), respectively, for any $u_h \in V_h$, there exist constants K_1 and K_2 such that

$$\begin{aligned}K_2 \mathcal{A}_\tau(u_h, u_h) &\leq \mathcal{A}_\tau(\mathcal{P}_0 u_h, \mathcal{P}_0 u_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i u_h, \mathcal{P}_l^i u_h) \\ &\leq K_1 \mathcal{A}_\tau(u_h, u_h)\end{aligned} \quad (32)$$

where $K_1 = C\|\Theta\|_0 \hat{c}_1^2 / (c_0 \hat{c}_0)^2$, $K_2 = C\hat{c}_0 \max_{1 \leq l \leq L} (1 + h_{l-1}/\delta_l)^{-1} / \hat{c}_1$, C is a constant independent of H, h, τ and L .

Theorem 1. Let \mathcal{P} be defined as in (23), if Assumption 1 holds, we have

1. there exists a constant C_p such that

$$\mathcal{A}_\tau(\mathcal{P}\mathbf{u}_h, \mathcal{P}\mathbf{u}_h) \leq C_p \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \quad \forall \mathbf{u}_h \in \mathbf{V}_h$$

where $C_p^2 = 2K_1(1 + \|\Theta\|_0)$ and C is a constant independent of H, h, τ and L ;

2. there exists a constant $c_p > 0$ such that

$$\mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P}\mathbf{u}_h) \geq c_p \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \quad \forall \mathbf{u}_h \in \mathbf{V}_h$$

where $c_p = (K_2 - C(\sqrt{\|\Theta\|_0} + 1)K_1 H \sqrt{H^2/\tau + 1}/(\hat{c}_0 \sqrt{\tau}))$ and C is a constant independent of H, h, τ and L .

Proof. 1. It follows from (23), (28) and (32) that

$$\begin{aligned} \mathcal{A}_\tau(\mathcal{P}\mathbf{u}_h, \mathcal{P}\mathbf{u}_h) &= \mathcal{A}_\tau\left(\mathcal{P}_0\mathbf{u}_h + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h\right) \\ &\leq 2\mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) + 2\mathcal{A}_\tau\left(\sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h, \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h\right) \\ &= 2\mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) + 2 \sum_{l,k=1}^L \sum_{i,j=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_k^j \mathbf{u}_h) \\ &\leq 2\mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &\quad + 2 \sum_{l,k=1}^L \sum_{i,j=1}^{N_l} \theta_{ij}^{lk} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^j \mathbf{u}_h)^{\frac{1}{2}} \mathcal{A}_\tau(\mathcal{P}_k^j \mathbf{u}_h, \mathcal{P}_k^i \mathbf{u}_h)^{\frac{1}{2}} \\ &\leq 2\mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) + 2\|\Theta\|_0 \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\ &\leq 2K_1(1 + \|\Theta\|_0) \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \end{aligned}$$

the first inequality holds because $2xy \leq x^2 + y^2$ for $x, y \in \mathbb{R}$.

2. Since $\sqrt{\tau}\|\mathbf{u}_h\|_1 \leq C\|\mathbf{u}_h\|_\tau$, it follows from the inequality (1) in Lemmas 1 and 3, we have

$$\begin{aligned} |S_\tau(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h)| &\leq C\tau\|\mathcal{P}_0\mathbf{u}_h\|_1\|\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h\| \\ &\leq CH\sqrt{H^2/\tau + 1}\|\mathcal{P}_0\mathbf{u}_h\|_\tau\|\mathbf{u}_h\|_\tau \\ &\leq CH\sqrt{H^2/\tau + 1}\|\mathbf{u}_h\|_\tau^2 \end{aligned} \quad (33)$$

Similarly,

$$\begin{aligned} |\mathcal{N}(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h)| &= ((\mathbf{A} - \mathbf{I})\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &\leq C\|\mathcal{P}_0\mathbf{u}_h\|\|\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h\| \\ &\leq C\frac{H}{\sqrt{\tau}}\sqrt{\frac{H^2}{\tau} + 1}\|\mathbf{u}_h\|_\tau^2 \end{aligned} \quad (34)$$

Combining (18), (21), (33) and (34) implies

$$\begin{aligned} \mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) &= \mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) + \mathcal{A}_\tau(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &= \mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) + \mathcal{D}_\tau(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &\quad - S_\tau(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) - \mathcal{N}(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \end{aligned}$$

$$\begin{aligned} &= \mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) - S_\tau(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &\quad - \mathcal{N}(\mathbf{u}_h - \mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) \\ &\geq \mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) - C\frac{H}{\sqrt{\tau}}\sqrt{\frac{H^2}{\tau} + 1}\|\mathbf{u}_h\|_\tau^2 \\ &\geq \mathcal{A}_\tau(\mathcal{P}_0\mathbf{u}_h, \mathcal{P}_0\mathbf{u}_h) - C\frac{H}{\hat{c}_0\sqrt{\tau}}\sqrt{\frac{H^2}{\tau} + 1}\mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \end{aligned} \quad (35)$$

Since the support of $\mathcal{P}_l^i \mathbf{u}_h$ lies in a region of diameter H_l , it follows from Friedrichs' inequality and the definition of the norm $\|\cdot\|_\tau$ that

$$\|\mathcal{P}_l^i \mathbf{u}_h\| \leq CH_l \|\mathcal{P}_l^i \mathbf{u}_h\|_1 \leq CH_l / \sqrt{\tau} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau \quad (36)$$

Combining (27) and (36) implies

$$\begin{aligned} \left\| \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h \right\|^2 &= \sum_{l,k=1}^L \sum_{i,j=1}^{N_l} (\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_k^j \mathbf{u}_h) \\ &\leq \sum_{l,k=1}^L \sum_{i,j=1}^{N_l} \theta_{ij}^{lk} (\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^j \mathbf{u}_h)^{\frac{1}{2}} (\mathcal{P}_k^j \mathbf{u}_h, \mathcal{P}_k^i \mathbf{u}_h)^{\frac{1}{2}} \\ &\leq C\|\Theta\|_0 \frac{H^2}{\tau} \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \end{aligned} \quad (37)$$

It follows from Lemmas 1, (18), (32), (36), and (37) that

$$\begin{aligned} &\sum_{l=1}^L \sum_{i=1}^{N_l} S_\tau(\mathbf{u}_h - \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\ &= S_\tau\left(\mathbf{u}_h, \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h\right) - \sum_{l=1}^L \sum_{i=1}^{N_l} S_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\ &\leq C\tau\|\mathbf{u}_h\|_1 \left\| \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h \right\| + C\tau \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_1 \|\mathcal{P}_l^i \mathbf{u}_h\| \\ &\leq C\sqrt{\|\Theta\|_0} H \|\mathbf{u}_h\|_\tau \left(\sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \right)^{\frac{1}{2}} \\ &\quad + CH \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \\ &\leq C(\sqrt{\|\Theta\|_0} + 1)K_1 H / \hat{c}_0 \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \end{aligned} \quad (38)$$

Similarly, we have

$$\begin{aligned} &\sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{N}(\mathbf{u}_h - \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\ &= \mathcal{N}\left(\mathbf{u}_h, \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h\right) - \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{N}(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\ &\leq C\|\mathbf{u}_h\| \left\| \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h \right\| + C \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|^2 \\ &\leq C\sqrt{\|\Theta\|_0} \frac{H}{\sqrt{\tau}} \|\mathbf{u}_h\|_\tau \left(\sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \right)^{\frac{1}{2}} \\ &\quad + C\frac{H^2}{\tau} \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \\ &\leq C(\sqrt{\|\Theta\|_0} + 1)K_1 \frac{H}{\hat{c}_0 \sqrt{\tau}} \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \end{aligned} \quad (39)$$

From (22), (38), and (39), we have

$$\begin{aligned}
 & \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &= \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathbf{u}_h - \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &= \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) - \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{D}_\tau(\mathbf{u}_h - \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\quad - \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{N}(\mathbf{u}_h - \mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\geq \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\quad - C(\sqrt{\|\Theta\|_0} + 1)K_1 \frac{H}{\hat{c}_0 \sqrt{\tau}} \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h)
 \end{aligned} \tag{40}$$

Combining (32), (35), and (40) implies

$$\begin{aligned}
 \mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P} \mathbf{u}_h) &= \mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P}_0 \mathbf{u}_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\geq \mathcal{A}_\tau(\mathcal{P}_0 \mathbf{u}_h, \mathcal{P}_0 \mathbf{u}_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\quad - C(\sqrt{\|\Theta\|_0} + 1)K_1 \frac{H}{\hat{c}_0 \sqrt{\tau}} \sqrt{\frac{H^2}{\tau} + 1} \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \\
 &\geq \left(K_2 - C(\sqrt{\|\Theta\|_0} + 1)K_1 \frac{H}{\hat{c}_0 \sqrt{\tau}} \sqrt{\frac{H^2}{\tau} + 1} \right) \\
 &\quad \times \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h)
 \end{aligned}$$

This completes the proof. \square

Next, we shall verify Assumption 1.

Lemma 6. Let \mathcal{P} be defined as in (23), then Assumption 1 is satisfied with $K_1 = C\|\Theta\|_0 \hat{c}_1^2 / (c_0 \hat{c}_0)^2$ and $K_2 = C\hat{c}_0 \max_{1 \leq l \leq L} (1 + h_{l-1} / \delta_l)^{-1} / \hat{c}_1$.

Proof. Upper bound. From the inequality (3) of Lemmas 1 and (18), we have

$$\mathcal{D}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \geq c_0 \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \geq c_0 / \hat{c}_1 \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \tag{41}$$

It follows from the inequality (3) of Lemmas 1, (22), (28), and (41) that

$$\begin{aligned}
 & \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \leq \hat{c}_1 / c_0 \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{D}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &= \hat{c}_1 / c_0 \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{D}_\tau(\mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \leq C\hat{c}_1 / c_0 \|\mathbf{u}_h\|_\tau \left\| \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{P}_l^i \mathbf{u}_h \right\|_\tau \\
 &\leq C\hat{c}_1 \sqrt{\|\Theta\|_0} / (c_0 \hat{c}_0) \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h)^{\frac{1}{2}} \left(\sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \right)^{\frac{1}{2}}
 \end{aligned}$$

It is clear that

$$\sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \leq C\|\Theta\|_0 \hat{c}_1^2 / (c_0 \hat{c}_0)^2 \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \tag{42}$$

By Lemmas (3) and (18), we have

$$\mathcal{A}_\tau(\mathcal{P}_0 \mathbf{u}_h, \mathcal{P}_0 \mathbf{u}_h) \leq C\hat{c}_1 / \hat{c}_0 \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h) \tag{43}$$

Then, combining (42) and (43) implies

$$\begin{aligned}
 & \mathcal{A}_\tau(\mathcal{P}_0 \mathbf{u}_h, \mathcal{P}_0 \mathbf{u}_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\leq C\|\Theta\|_0 \hat{c}_1^2 / (c_0 \hat{c}_0)^2 \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h)
 \end{aligned}$$

Lower bound. Using the Cauchy-Schwarz inequality, it follows from Lemmas 1 and 4 that

$$\begin{aligned}
 \mathcal{D}_\tau(\mathbf{u}_h, \mathbf{u}_h) &= \mathcal{D}_\tau(\mathbf{u}_h, \mathbf{u}_0) + \mathcal{D}_\tau\left(\mathbf{u}_h, \sum_{l=1}^L \sum_{i=1}^{N_l} \mathbf{u}_l^i\right) \\
 &= \mathcal{D}_\tau(\mathcal{P}_0 \mathbf{u}_h, \mathbf{u}_0) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{D}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathbf{u}_l^i) \\
 &\leq C\|\mathcal{P}_0 \mathbf{u}_h\|_\tau \|\mathbf{u}_0\|_\tau + C \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau \|\mathbf{u}_l^i\|_\tau \\
 &\leq C \left(\|\mathcal{P}_0 \mathbf{u}_h\|_\tau^2 + \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \right)^{\frac{1}{2}} \\
 &\quad \times \left(\|\mathbf{u}_0\|_\tau^2 + \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathbf{u}_l^i\|_\tau^2 \right)^{\frac{1}{2}} \\
 &\leq C \max_{1 \leq l \leq L} (1 + h_{l-1} / \delta_l)^{\frac{1}{2}} \|\mathbf{u}_h\|_\tau \\
 &\quad \times \left(\|\mathcal{P}_0 \mathbf{u}_h\|_\tau^2 + \sum_{l=1}^L \sum_{i=1}^{N_l} \|\mathcal{P}_l^i \mathbf{u}_h\|_\tau^2 \right)^{\frac{1}{2}}
 \end{aligned} \tag{44}$$

From the inequality (3) of Lemma 1, it is clear that $\mathcal{D}_\tau(\mathbf{u}_h, \mathbf{u}_h) \geq c_0 \|\mathbf{u}_h\|_\tau^2$. Then, combining (18) and (44) yields

$$\begin{aligned}
 & \mathcal{A}_\tau(\mathcal{P}_0 \mathbf{u}_h, \mathcal{P}_0 \mathbf{u}_h) + \sum_{l=1}^L \sum_{i=1}^{N_l} \mathcal{A}_\tau(\mathcal{P}_l^i \mathbf{u}_h, \mathcal{P}_l^i \mathbf{u}_h) \\
 &\geq C\hat{c}_0 \max_{1 \leq l \leq L} (1 + h_{l-1} / \delta_l)^{-1} / \hat{c}_1 \mathcal{A}_\tau(\mathbf{u}_h, \mathbf{u}_h)
 \end{aligned}$$

This completes the proof. \square

4 | Numerical Experiments

In this section, some numerical results are reported to illustrate the optimal convergence and the parallel performance of the proposed algorithm. The algorithm is implemented using PETSc [40], and all the examples are tested on a supercomputer with 16,000 compute nodes, each comprising two Intel Ivy Bridge Xeon processors and 88 gigabytes of memory. We only use the CPU cores in our tests. In the experiments, the parabolic equations are discretized by the fourth-order BIM (presented in

TABLE 1 | The numerical results for Example 1 obtained with Algorithm 2 (two-level) with $N_p = 64$, $h_2 = 1/1024$ and $\tau = \sqrt{h_2}$.

		1/16			1/32			1/64		
h_1		$\delta = 0$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 1$	$\delta = 2$	$\delta = 0$	$\delta = 1$	$\delta = 2$
ILU(0)	IT	12	11	10	8	7	6	7	6	5
	Time (s)	2.32	2.15	2.05	1.86	1.81	1.82	1.84	1.80	1.78
ILU(1)	IT	11	9	8	8	8	8	7	6	5
	Time (s)	2.52	2.39	2.29	2.25	2.10	2.13	2.27	2.22	2.16
ILU(2)	IT	10	8	7	8	6	6	8	6	6
	Time (s)	2.93	2.56	2.51	2.55	2.39	2.45	2.52	2.48	2.54
LU	IT	11	8	8	9	7	7	8	7	7
	Time (s)	7.91	6.32	6.45	7.51	6.16	6.28	7.68	6.22	6.40

TABLE 2 | The number of GMRES iterations for Example 1 with $\delta = 2$, $N_p = 64$ and $\tau = \sqrt{h_L}$.

h_L	1/64		1/128			1/256				1/512					1/1024						
L	2	3	2	3	4	2	3	4	5	2	3	4	5	6	2	3	4	5	6	7	
ILU (0)	10	12	9	11	13	7	10	12	13	7	9	10	11	13	6	8	9	11	12	12	
ILU (1)	9	12	8	11	13	8	10	11	13	7	9	10	12	13	6	8	9	10	11	12	
ILU (2)	10	12	8	11	13	8	10	11	13	7	9	10	12	12	6	8	9	11	11	12	
LU	10	12	9	11	13	8	10	12	13	8	10	11	12	14	7	9	10	12	13	14	

Remark 1) in time and piecewise linear conforming finite elements in space, the spatial domain Ω is initially partitioned into a coarse mesh of size h_1 , and the fine meshes are obtained by refining the coarsest mesh uniformly. We divide Ω into N_p subdomains on each level and assign each subdomain to one processor. The resulting linear system is solved by a preconditioned GMRES(30) method, and the stopping condition is set to be $rtol = 1.0 \times 10^{-8}$ (the relative convergence tolerance). The subdomain problems are solved by LU factorization or ILU(k). For simplicity, the notations “ $\delta := \delta_l/h_l$ ”, “IT”, and “Time(s)” denote the overlapping size, the number of iterations, and the compute time (seconds), respectively.

Example 1. We consider an advection–diffusion equation

$$\begin{cases} u_t - \Delta u + \beta(x, y) \cdot \nabla u = f, & \text{in } \Omega \times (0, T], \\ u(x, y, t) = 0, & \text{on } \partial\Omega \times (0, T], \\ u(x, y, 0) = x(1-x)y(1-y), & \text{in } \Omega \end{cases}$$

where $\beta(x, y) = (y(1-x^2), -x(1-y^2))$, $\Omega = [0, 1] \times [0, 1]$, and f is chosen such that the exact solution is $u(x, y, t) = x(1-x)y(1-y)e^{-t}$.

In the first test, we fix the number of subdomains to $N_p = 64$, the fine spatial mesh size $h_2 = 1/1024$, the time step size $\tau = \sqrt{h_2}$, and vary the coarse spatial mesh size h_1 , the overlapping size δ as well as the subdomain solver. The numerical results in Table 1 show that both the iteration counts and the compute time decrease if we vary the overlapping size from $\delta = 0$ to $\delta = 1$. And the change of the iteration counts is not sensitive to different coarse mesh sizes. Additionally, if we increase $\delta = 1$ to $\delta = 2$,

the iteration counts do not change for some cases, but the total compute time grows since the communication time increases. It is clear that the iteration counts decrease slightly, but the compute time increases as the level of fill-ins increases for the case $h_1 = 1/16$. The iteration counts stay the same for different fill-in levels k if we choose smaller coarse mesh sizes. Moreover, the ILU(k) performs better than LU in terms of the compute time for this test example.

Next, we show how the convergence rate depends on the number of levels and the subdomain solver. The number of subdomains is $N_p = 64$, and the coarsening factor is fixed as 2. The results listed in Table 2 show that the number of iterations is not sensitive to the subdomain solvers LU and ILU with different fill-ins. For the fixed finest mesh size, the iteration counts increase a little with the number of levels since the coarsest mesh sizes are different. However, the iteration counts are almost the same with different mesh levels if we fix the coarse mesh size; see the last column of $h_L = 1/64$, $h_L = 1/128$, $h_L = 1/256$, $h_L = 1/512$ and $h_L = 1/1024$. This illustrates that the convergence rate of MAS-BIM is independent of the finest mesh size and the number of levels.

Finally, we investigate the strong parallel scalability for Example 1. In the experiment, the finest spatial mesh size is fixed as $h_3 = 1/8192$ and $h_3 = 1/10,240$, respectively, the number of levels and the coarsening factor are $L = 3$ and 4, the time step size and the overlapping size are fixed as $\tau = 1/256$ and $\delta = 1$, and the subdomain solver is set ILU(1). We increase the number of processors from 256 to 4096; Table 3 shows that the number of iterations changes very little, and the total compute

time decreases. Moreover, Figure 2 also shows that Algorithm (2) has a nearly linear speedup with up to 4096 processors.

Example 2. Consider a convection dominated parabolic equation

$$\begin{cases} u_t - \varepsilon \Delta u + \beta(x, y, z) \cdot \nabla u = f, & \text{in } \Omega \times (0, 1], \\ u(x, y, z, t) = 0, & \text{on } \partial\Omega \times (0, 1], \\ u(x, y, z, 0) = \sin(\pi x) & \text{in } \Omega, \\ \sin(\pi y) \sin(\pi z) \end{cases}$$

where $\varepsilon > 0$ is a constant, $\beta(x, y, z) = (1, 1, 1)$, $\Omega = [0, 1] \times [0, 1] \times [0, 1]$, f is chosen such that the exact solution is $u(x, y, z, t) = \sin(\pi x) \sin(\pi y) \sin(\pi z) \cos pt$, where p is a constant to be specified later.

TABLE 3 | Strong scaling results of Algorithm 2 (three-level) on Tianhe-2 for Example 1 with $\delta = 1$ and $\tau = 1/256$.

N_p	$h_3 = 1/8192$		$h_3 = 1/10,240$	
	IT	Time (s)	IT	Time (s)
256	8	34.8	8	52.8
512	9	21.7	8	31.6
1024	9	13.3	9	20.7
2048	9	8.4	9	12.6
4096	9	5.8	9	8.9

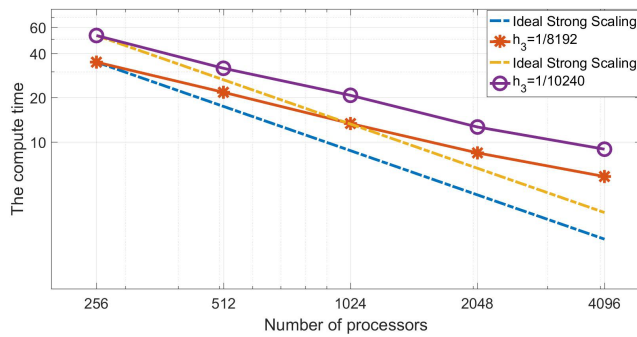


FIGURE 2 | Strong scaling results of Algorithm 2 (three-level) on Tianhe-2 for Example 1.

TABLE 4 | Numerical results for Example 2 obtained by MAS-BE, MAS-CN, MAS-BIM with $\varepsilon = 0.1$, $N_p = 512$, $\delta = 1$, $h_1 = 1/32$ and $h_2 = 1/128$.

		$p = 2.5\pi$			$p = 5.5\pi$			$p = 11\pi$		
		IT	Time (s)	$\ u - u_h\ $	IT	Time (s)	$\ u - u_h\ $	IT	Time (s)	$\ u - u_h\ $
MAS-BE	1/4096	2.1	273.71	1.80e-4	2.2	291.40	2.25e-5	2.4	301.31	3.80e-4
	1/2048	2.8	167.03	5.71e-4	2.9	175.80	5.37e-5	3.0	199.92	5.26e-4
MAS-CN	1/256	6.5	45.39	3.64e-5	6.9	47.46	4.20e-5	7.4	49.74	5.23e-4
	1/128	7.6	27.13	4.52e-5	8.2	28.69	1.60e-4	8.6	30.57	2.17e-3
MAS-BIM	1/64	7.7	15.20	4.65e-5	8.1	15.78	5.33e-5	9.0	17.15	5.14e-4
	1/32	8.1	9.40	5.30e-5	8.6	9.80	9.51e-5	8.9	10.11	8.32e-3

We test a three-dimensional problem to compare the performance of BIM with the backward Euler and the Crank–Nicolson schemes. In the experiments, we set $\varepsilon = 0.1$ and fix the number of levels $L = 2$, the coarse mesh size $h_1 = 1/32$, the fine mesh size $h_2 = 1/128$, and the number of subdomains on each level $N_p = 512$. In [29], we showed that space–time additive Schwarz algorithm based on the backward Euler scheme is more efficient than Algorithm 1 when the number of processors is large. To make a fair comparison, we compare Algorithm 2 and the two-level space–time additive Schwarz algorithm based on the backward Euler scheme as well as Crank–Nicolson scheme for the case when the window size is 4. So, the solutions at four time steps are solved simultaneously for these three algorithms denoted by “MAS-BE”, “MAS-CN”, and “MAS-BIM”, respectively.

Table 4 shows that the average number of GMRES iterations and the total compute time of all three algorithms increase with the increase of the parameter p . Moreover, the average number of iterations for MAS-BE is smaller than MAS-CN and MAS-BIM. However, MAS-BE requires a smaller time step size and larger number of blocks to achieve the same error at the final time. It is clear that the total compute time of MAS-BE is much more than MAS-CN and MAS-BIM. Since a larger time step size can be used for higher-order schemes, the increase of the iteration counts is subordinate to the decrease in the number of blocks. Therefore, we see that the higher-order methods are superior to lower-order methods when the solution has the required regularity. Finally, we choose $\varepsilon = 10^{-4}$ and list the numerical results in Table 5. Compared with Tables 4 and 5, we see that both the average number of iterations and the compute time of MAS-BE and MAS-CN decrease with smaller ε , but the errors increase with the same stopping condition. On the contrary, the average number of iterations and the compute time of MAS-BIM grow, which means that we should choose smaller τ for this case. However, MAS-BIM still takes less compute time than MAS-BE and MAS-CN when achieving the same error.

5 | Conclusion

We introduced and studied a multilevel additive Schwarz algorithm for parabolic equations discretized by a family of BIM in time and finite element in space. An optimal convergence theory is proved for the multilevel additive Schwarz method for BIM with positive definite matrix A and positive diagonal matrix B .

TABLE 5 | Numerical results for Example 2 obtained by MAS-BE, MAS-CN, and MAS-BIM with $\varepsilon = 10^{-4}$, $N_p = 512$, $\delta = 1$, $h_1 = 1/32$ and $h_2 = 1/128$.

		$p = 2.5\pi$			$p = 5.5\pi$			$p = 11\pi$		
	τ	IT	Time (s)	$\ u - u_h\ $	IT	Time (s)	$\ u - u_h\ $	IT	Time (s)	$\ u - u_h\ $
MAS-BE	1/4096	2.0	221.41	5.50e-4	2.0	213.53	8.85e-4	2.0	208.15	3.10e-4
	1/2048	2.0	105.69	1.09e-3	2.0	123.49	1.76e-3	2.0	113.77	6.13e-4
MAS-CN	1/256	2.9	25.49	3.14e-4	3.1	25.98	1.02e-4	3.3	26.82	5.41e-4
	1/128	4.6	18.67	1.66e-4	4.9	19.26	3.05e-4	5.2	20.52	2.20e-3
MAS-BIM	1/128	4.0	16.25	1.71e-4	4.5	18.03	2.22e-4	5.0	19.50	2.98e-4
	1/64	8.6	16.37	1.43e-4	9.3	17.40	1.85e-4	10.3	18.96	5.09e-4

The numerical results for a fourth-order BIM with L -stability confirm the theory in terms of the convergence rate and accuracy. The linear speedup is obtained on a parallel computer using up to 4096 processor cores. The numerical comparisons also demonstrate that our algorithm outperforms the multilevel additive Schwarz algorithm for low-order schemes in terms of the compute time.

Acknowledgments

The work was supported in part by FDCT 0141/2020/A3, 0079/2021/AFJ, MYRG-GRG 2023-00102-FST-UMDF, and the National Natural Science Foundation of China (No. 12101588).

Data Availability Statement

The authors have nothing to report.

References

1. X.-C. Cai, “Additive Schwarz Algorithms for Parabolic Convection-Diffusion Equations,” *Numerische Mathematik* 60 (1991): 41–61.

2. F. Kong and X.-C. Cai, “A Highly Scalable Multilevel Schwarz Method With Boundary Geometry Preserving Coarse Spaces for 3D Elasticity Problems on Domains With Complex Geometry,” *SIAM Journal on Scientific Computing* 38 (2016): C73–C95.

3. P. F. Fischer, “An Overlapping Schwarz Method for Spectral Element Solution of the Incompressible Navier-Stokes Equations,” *Journal of Computational Physics* 133 (1997): 84–101.

4. F.-N. Hwang and X.-C. Cai, “A Parallel Nonlinear Additive Schwarz Preconditioned Inexact Newton Algorithm for Incompressible Navier-Stokes Equations,” *Journal of Computational Physics* 204 (2005): 666–691.

5. L. F. Pavarino and S. Scacchi, “Multilevel Additive Schwarz Preconditioners for the Bidomain Reaction-Diffusion System,” *SIAM Journal on Scientific Computing* 31 (2008): 420–445.

6. S. Scacchi, “A Hybrid Multilevel Schwarz Method for the Bidomain Model,” *Computer Methods in Applied Mechanics and Engineering* 197 (2008): 4051–4061.

7. S. Ovtchinnikov, F. Dobrian, X.-C. Cai, and D. E. Keyes, “Additive Schwarz-Based Fully Coupled Implicit Methods for Resistive Hall Magnetohydrodynamic Problems,” *Journal of Computational Physics* 225 (2006): 1919–1936.

8. C. Yang and X.-C. Cai, “Parallel Multilevel Methods for Implicit Solution of Shallow Water Equations With Nonsmooth Topography on the Cubed-Sphere,” *Journal of Computational Physics* 230 (2011): 2523–2539.

9. C. Yang, J. Cai, and X.-C. Cai, “A Fully Implicit Domain Decomposition Algorithm for Shallow Water Equations on the Cubed-Sphere,” *SIAM Journal on Scientific Computing* 32 (2010): 418–438.

10. G. Dahlquist, “A Special Stability Problem for Linear Multistep Methods,” *BIT* 3 (1963): 27–43.

11. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations. II: Stiff and Differential-Algebraic Problems* (Springer, 1996).

12. W. E. Milne, *Numerical Solution of Differential Equations* (Wiley, 1953).

13. L. Brugnano and D. Trigiante, “Block Implicit Methods for ODEs,” in *Recent Trends in Numerical Analysis*. Advances in Computational Mathematics and Modelling, vol. 3 (Nova Science Publishers, 2001), 81–105.

14. L. F. Shampine and H. A. Watts, “Block Implicit One-Step Methods,” *Mathematics of Computation* 23 (1969): 731–740.

15. H. A. Watts, “A Stable Block Implicit One-Step Methods. PhD Thesis, The University of New Mexico,” 1971.

16. H. A. Watts and L. F. Shampine, “Ai-Stable Block One-Step Methods,” *BIT* 12 (1972): 252–266.

17. S. Li, J.-Y. Wang, and X.-C. Cai, “A-Stable High Order Block Implicit Methods for Parabolic Equations,” *SIAM Journal on Numerical Analysis* 61 (2021): 1858–1884.

18. O. Axelsson, I. Dravins, and M. Neytcheva, “Stage-Parallel Preconditioners for Implicit Runge–Kutta Methods of Arbitrarily High Order, Linear Problems,” *Numerical Linear Algebra with Applications* 31 (2024): e2532.

19. X. Jiao, X. Wang, and Q. Chen, “Optimal and Low-Memory Near-Optimal Preconditioning of Fully Implicit Runge-Kutta Schemes for Parabolic PDEs,” *SIAM Journal on Scientific Computing* 43 (2021): A3527–A3551.

20. K.-A. Mardal, T. K. Nilssen, and G. A. Staff, “Order-Optimal Preconditioners for Implicit Runge-Kutta Schemes Applied to Parabolic PDEs,” *SIAM Journal on Scientific Computing* 29 (2007): 361–375.

21. P. Munch, I. Dravins, M. Kronbichler, and M. Neytcheva, “Stage-Parallel Fully Implicit Runge–Kutta Implementations With Optimal Multilevel Preconditioners at the Scaling Limit,” *SIAM Journal on Scientific Computing* 46 (2023): S71–S96.

22. M. M. Rana, V. E. Howle, K. Long, A. Meek, and W. Milestone, “A New Block Preconditioner for Implicit Runge-Kutta Methods for Parabolic PDE Problems,” *SIAM Journal on Scientific Computing* 43 (2021): S475–S495.

23. B. S. Southworth, O. Krzysik, W. Pazner, and H. D. Sterck, "Fast Solution of Fully Implicit Runge-Kutta and Discontinuous Galerkin in Time for Numerical PDEs, Part I: The Linear Setting," *SIAM Journal on Scientific Computing* 44 (2022): A416–A443.
24. G. A. Staff, K.-A. Mardal, and T. K. Nilssen, "Preconditioning of Fully Implicit Runge-Kutta Schemes for Parabolic PDEs," *Modelling Identification and Control* 27 (2006): 109–123.
25. M. J. Gander, "50 Years of Time Parallel Time Integration. Multiple Shooting and Time Domain Decomposition Methods," in *Contributions in Mathematical and Computational Sciences*, vol. 9 (Springer, 2015), 69–113.
26. S. Li, X. Shao, and X.-C. Cai, "Highly Parallel Space-Time Domain Decomposition Methods for Parabolic Problems," *CCF Transactions on High Performance Computing* 1 (2019): 25–34.
27. B. W. Ong and J. B. Schroder, "Applications of Time Parallelization," *Computing and Visualization in Science* 23 (2020): 11.
28. S. Li and X.-C. Cai, "Convergence Analysis of Two-Level Space-Time Additive Schwarz Method for Parabolic Equations," *SIAM Journal on Numerical Analysis* 53 (2015): 2727–2751.
29. S. Li, X. Shao, and X.-C. Cai, "Multilevel Space-Time Additive Schwarz Methods for Parabolic Equation," *SIAM Journal on Scientific Computing* 40 (2018): A3012–A3037.
30. X. Zhang, "Multilevel Schwarz Methods," *Numerische Mathematik* 63 (1992): 521–539.
31. X.-C. Cai, "Multiplicative Schwarz Methods for Parabolic Problems," *SIAM Journal on Scientific Computing* 15 (1994): 587–603.
32. S. C. Eisenstat, H. C. Elman, and M. H. Schultz, "Variational Iterative Methods for Nonsymmetric Systems of Linear Equations," *SIAM Journal on Numerical Analysis* 20 (1983): 345–357.
33. M. Sarkis and D. B. Szyld, "Optimal Left and Right Additive Schwarz Preconditioning for Minimal Residual Methods With Euclidean and Energy Norms," *Computer Methods in Applied Mechanics and Engineering* 196 (2003): 1612–1621.
34. H. C. Elman, "Iterative Methods for Sparse, Nonsymmetric Systems of Linear Equations"(PhD thesis, Yale University, Department of Computer Science, 1982).
35. X.-C. Cai and O. B. Widlund, "Domain Decomposition Algorithms for Indefinite Elliptic Problems," *SIAM Journal on Scientific and Statistical Computing* 13 (1992): 243–258.
36. X.-C. Cai and O. B. Widlund, "Multiplicative Schwarz Algorithms for Some Nonsymmetric and Indefinite Problems," *SIAM Journal on Numerical Analysis* 30 (1993): 936–952.
37. X.-C. Cai and J. Zou, "Some Observations on the l^2 Convergence of the Additive Schwarz Preconditioned GMRES Method," *Numerical Linear Algebra with Applications* 9 (2002): 379–397.
38. D. B. Szyld and V. Simoncini, "On the Occurrence of Superlinear Convergence of Exact and Inexact Krylov Subspace Methods," *SIAM Review* 47 (2005): 247–272.
39. X. Du and D. B. Szyld, "A Note on the Mesh Independence of Convergence Bounds for Additive Schwarz Preconditioned GMRES," *Numerical Linear Algebra with Applications* 15 (2008): 547–557.
40. S. Balay, S. Abhyankar, M. F. Adams, et al., *PETSc Users Manual* (Argonne National Laboratory, 2024).