SIAM J. SCI. COMPUT. © 2018 Society for Industrial and Applied Mathematics
Vol. 40, No. 5, pp. A3012–A3037

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

MULTILEVEL SPACE-TIME ADDITIVE SCHWARZ METHODS FOR PARABOLIC EQUATIONS*

SHISHUN LI†, XINPING SHAO‡, AND XIAO-CHUAN CAI§

Abstract. In this paper, we present a multilevel space-time additive Schwarz method for solving linear system of equations arising from the discretization of parabolic equations. With this method, the problem is solved in parallel on both space and time dimensions. After establishing two important properties of the space and time decomposition, i.e., a strengthened Cauchy–Schwarz-type inequality and a stable multilevel decomposition under a space-time energy norm, we develop an optimal convergence theory in R^2 and R^3 and show how the convergence rate depends on the mesh sizes, the number of subdomains, the window size, and the number of levels. Numerical experiments carried out on a parallel computer with thousands of processors for two- and three-dimensional problems confirm the theory in terms of the number of iterations, as well as the strong and weak scalabilities. Furthermore, a detailed comparison shows that the space-time method outperforms the traditional time stepping method, parallelized only in space, when the number of processors is large.

Key words. multilevel additive Schwarz method, space-time coupled method, parabolic equations, finite element discretization, parallel processing, strong/weak scalability

AMS subject classifications. 65N55, 65M55, 65M60

DOI. 10.1137/17M113808X

1. Introduction. Domain decomposition (DD) techniques are efficient parallel methods for solving time-dependent partial differential equations (PDEs) [4, 5, 6, 7, 12]. However, traditional DD methods for solving time-dependent PDEs restrict the parallelism within each time step, and are purely sequential between time steps. To solve large scale time-dependent problems more quickly and use the supercomputer more effectively, a new generation of algorithms is being developed that are parallel in both space and time, such as the waveform relaxation methods [17, 19, 23, 31], space-time multigrid methods [18, 21, 22, 32], parareal algorithms [1, 16, 20, 25, 27], and other space-time methods [8, 15, 26, 34]. In [9], some implicit space-time DD methods were introduced for solving deterministic and stochastic parabolic equations. In these methods, equations for s (window size) time steps are grouped into a single system, which is then solved by using one- and two-level overlapping Schwarz preconditioned recycling GMRES methods. The numerical results obtained on a parallel computer show that the method works very well. The optimal convergence theory for the two-level space-time additive Schwarz method was established in [24]. Under certain conditions, the convergence rate of the two-level method is bounded independently

*Submitted to the journal's Methods and Algorithms for Scientific Computing section July 14, 2017; accepted for publication (in revised form) August 7, 2018; published electronically September 18, 2018.
http://www.siam.org/journals/sisc/40-5/M113808.html
Funding: The work of the authors was partially supported by National Key R&D Program 2016YFB0200601, NSFC 11726636, 61531166003, 11701133, and Shen Zhen grant JCYJ20160331193229720.
†School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo 454003, People's Republic of China and Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, Guangdong, People's Republic of China (lss6@sina.com).
‡Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309 (xinping.shao@colorado.edu).
§Corresponding author. Department of Computer Science, University of Colorado Boulder, Boulder, CO 80309 (cai@cs.colorado.edu).

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

of the number of subdomains, the fine mesh size, and the window size. This method has also been used to solve flow control problems [33] and the inverse source problems [10, 11].

The aim of this paper is to establish a convergence theory in R^2 and R^3 for the multilevel space-time additive Schwarz method and study the numerical behavior of the method in terms of the arithmetic optimality and parallel scalability. For solving large problems, two-level algorithms may be inadequate since the coarse mesh problem may be too large to be solved without using a preconditioner of two or more levels. The main contribution of this paper is to extend the two-level method in [24] to multilevels theoretically and numerically. In the optimality analysis, we establish two important properties of the space-time decomposition, i.e., a strengthened Cauchy–Schwarz-type inequality and a stable multilevel decomposition. Then we develop an optimality theory using these two properties. The theory is complete in the sense that it shows that the convergence rate is bounded independently of the fine mesh size, the number of subdomains, the window size, and the number of levels. To confirm the theoretical estimates and demonstrate the parallel performance of the proposed method, we show some numerical experiments obtained on a supercomputer with thousands of processors.

The rest of the paper is organized as follows. In section 2, we introduce a model parabolic boundary value problem and present the multilevel space-time additive Schwarz algorithm. In section 3, we develop a convergence analysis of this algorithm by proving the strengthened Cauchy–Schwarz-type inequality and the stability of multilevel space-time decomposition. Some numerical experiments are reported to illustrate the performance of this algorithm in section 4. Finally, some concluding remarks are given in section 5.

2. Multilevel space-time additive Schwarz algorithm. We consider the following parabolic problem defined on $\Omega \times [0,T]$,

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nabla \cdot (a(x)\nabla u) + b(x)\cdot\nabla u + c(x)u = f & \text{in } \Omega\times(0,T], \\ u(x,t) = 0 & \text{on } \partial\Omega\times(0,T], \\ u(x,0) = u_0(x) & \text{in } \Omega, \end{cases} \tag{2.1}$$

where $\Omega \subset R^d$ ($d = 2$ or 3) is a bounded, open polygonal (or polyhedral) domain, $a(x) \in C^1(\overline{\Omega}, R^{d\times d})$, $b(x) \in C^1(\overline{\Omega})^d$, $c(x) \in C^1(\overline{\Omega})$, and $f(x,t) \in L^2(\Omega\times[0,T])$. Assume that $a(x) = (a_{ij}(x))_{d\times d}$ is a symmetric and uniformly positive definite matrix in Ω, i.e., there exists a positive constant m_0 such that $\xi^T a(x)\xi \geq m_0|\xi|^2$ for all $\xi \in R^d$ and $x \in \overline{\Omega}$. For brevity, we omit the variable x in the following discussion. We discretize (2.1) using finite differences in time and finite elements in space. Let $0 = t^0 < t^1 < \cdots < t^n = T$ and $\tau \equiv \triangle t^k = t^k - t^{k-1}$. Suppose u^k is the solution at time t^k. By using the backward Euler scheme for the time discretization, the variational form of (2.1) at time t^k is to find $u^k \in H_0^1(\Omega)$, $k \geq 1$, such that

$$D_\tau\left(u^k, v\right) \equiv \left(u^k, v\right) + \tau B\left(u^k, v\right) - \left(u^{k-1}, v\right) = \tau(f,v) \quad \forall v \in H_0^1(\Omega), \tag{2.2}$$

where the bilinear form is

$$B(u,v) = \int_\Omega (a\nabla u\cdot\nabla v + b\cdot\nabla u v + cuv)dx, \quad u,\ v \in H_0^1(\Omega).$$

We assume that the Gårding inequality holds [3], i.e., there exists a constant $m_1 > 0$ such that

$$B(v,v) \geq \frac{m_0}{2}\|v\|_1^2 - m_1\|v\|^2 \quad \forall v \in H_0^1(\Omega) \tag{2.3}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

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

$$B(u, v) \leq C\|u\|_1\|v\|_1 \quad \forall u,\ v \in H_0^1(\Omega), \tag{2.4}$$

where $\|\cdot\|$, $\|\cdot\|_1$ denote the L^2 and H^1 norms in Sobolev space, respectively. We split $B(u, v)$ into a second-order term and a lower-order term

$$A(u, v) = \int_\Omega a\nabla u \cdot \nabla v dx \quad \text{and} \quad S(u, v) = \int_\Omega (b \cdot \nabla uv + cuv)dx. \tag{2.5}$$

Obviously, $B(u, v) = A(u, v) + S(u, v)$ and $A(u, v)$ is self-adjoint. The following estimates come directly from the assumptions on the coefficients of (2.1):

$$|S(u, v)| \leq C\|u\|_1\|v\| \quad \text{and} \quad |S(u, v)| \leq C\|u\|\|v\|_1 \ \forall u,\ v \in H_0^1(\Omega). \tag{2.6}$$

The idea of the space-time method is to solve the following coupled system by an additive Schwarz preconditioned Krylov subspace method in parallel:

$$\begin{cases} (u^1, v^1) + \tau B(u^1, v^1) - (u^0, v^1) = \tau(f^1, v^1), \\ (u^2, v^2) + \tau B(u^2, v^2) - (u^1, v^2) = \tau(f^2, v^2), \\ \vdots \\ (u^s, v^s) + \tau B(u^s, v^s) - (u^{s-1}, v^s) = \tau(f^s, v^s), \end{cases} \tag{2.7}$$

where $s \leq n$ is often called the window size. The equivalent variational form of (2.7) is to find

$$\mathbf{u} = (u^1, u^2, \ldots, u^s) \in (H_0^1(\Omega))^s \equiv \underbrace{H_0^1(\Omega) \times H_0^1(\Omega) \times \cdots \times H_0^1(\Omega)}_{s},$$

such that

$$D_{\tau,s}(\mathbf{u}, \mathbf{v}) \equiv A_{\tau,s}(\mathbf{u}, \mathbf{v}) + S_{\tau,s}(\mathbf{u}, \mathbf{v}) + N_s(\mathbf{u}, \mathbf{v}) = (\mathbf{f}, \mathbf{v}) \quad \forall \mathbf{v} \in \left(H_0^1(\Omega)\right)^s, \tag{2.8}$$

where

$$A_{\tau,s}(\mathbf{u}, \mathbf{v}) = \tau \sum_{k=1}^{s} A\left(u^k, v^k\right) + \sum_{k=1}^{s} \left(u^k, v^k\right), \quad S_{\tau,s}(\mathbf{u}, \mathbf{v}) = \tau \sum_{k=1}^{s} S\left(u^k, v^k\right),$$

$$N_s(\mathbf{u}, \mathbf{v}) = -\sum_{k=1}^{s-1} \left(u^k, v^{k+1}\right), \quad (\mathbf{f}, \mathbf{v}) = \left(u^0, v^1\right) + \tau \sum_{k=1}^{s} \left(f^k, v^k\right).$$

For any $\mathbf{v} \in (H_0^1(\Omega))^s$, we introduce the (τ, s)-norm of $\mathbf{v}$,

$$\|\mathbf{v}\|_{\tau,s}^2 = \|\mathbf{v}\|^2 + \tau|\mathbf{v}|_1^2,$$

where $|\cdot|_1$ denotes the H^1 seminorm in Sobolev space. It is clear that the (τ, s)-norm is equivalent to the norm defined by $A_{\tau,s}(\cdot, \cdot)$. The following lemma provides some basic estimates for the bilinear forms involved in $D_{\tau,s}(\cdot, \cdot)$ [24].

LEMMA 2.1. *There exist $C > 0$ and $c_0 > 0$ independent of h, τ, and s such that*
(1) $|N_s(\boldsymbol{u}, \boldsymbol{u})| \leq \cos\left(\pi/(s+1)\right) \|\boldsymbol{u}\|^2 \quad \forall \boldsymbol{u} \in (H_0^1(\Omega))^s$,
(2) $|S_{\tau,s}(\boldsymbol{u}, \boldsymbol{v})| \leq C\tau\|\boldsymbol{u}\|_1\|\boldsymbol{v}\| \quad \forall \boldsymbol{u}, \boldsymbol{v} \in (H_0^1(\Omega))^s$,
(3) $|D_{\tau,s}(\boldsymbol{u}, \boldsymbol{v})| \leq C\|\boldsymbol{u}\|_{\tau,s}\|\boldsymbol{v}\|_{\tau,s} \quad \forall \boldsymbol{u}, \boldsymbol{v} \in (H_0^1(\Omega))^s$,

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(4) $|D_{\tau,s}(\boldsymbol{u},\boldsymbol{u})| \geq c_0\|\boldsymbol{u}\|_{\tau,s}^2 - c_1\|\boldsymbol{u}\|^2 \quad \forall \boldsymbol{u} \in (H_0^1(\Omega))^s$,
where $c_0 = \min\{m_0^2/2, 1\}$ *and* $c_1 = \cos(\pi/(s+1)) + m_1\tau$.

We next introduce a finite element discretization and an overlapping Schwarz preconditioner for (2.8). Let $\mathcal{T}_l = \{K_i^l\}_{i=1}^{N^l}$ $(l = 0, 1, \dots, L)$ be a shape regular family of nested conforming meshes covering Ω. The initial mesh $\mathcal{T}_0$ is quasi-uniform and $\mathcal{T}_l$ $(l \geq 1)$ is a refinement of $\mathcal{T}_{l-1}$. Set $h_l = \max_i \operatorname{diam}(K_i^l)$. We assume that there exist positive constants c, C, and $\gamma_1 < 1$, such that if an element K_i^{l+n} of level $n+l$ is contained in an element K_i^l of level l, then,

$$c\gamma_1^n \leq \frac{\operatorname{diam}(K_i^{l+n})}{\operatorname{diam}(K_i^l)} \leq C\gamma_1^n.$$

If $\mathcal{T}_{l+1}$ is refined uniformly from $\mathcal{T}_l$ through regular subdivision of each element K_i^l into k^d elements, we see that $\gamma_1 = 1/k$.

On each level, let $\{\Omega_{li}\}_{i=1}^{N_l}$ be a set of nonoverlapping subdomains such that

$$\Omega = \bigcup_{i=1}^{N_l} \Omega_{li}.$$

Let $V_l = V_{h_l}$, $l = 0, 1, \dots, L$, be the space of continuous piecewise linear functions associated with the partition $\mathcal{T}_l$. For simplicity, we set $V_H = V_0$ with mesh size $H = h_0$ and $V_h = V_L$ with mesh size $h = h_L$. Similarly, on each level l (≥ 0), we divide $I = [0, T] = [0, t_l^{s_l}]$ uniformly such that $0 = t_l^0 < t_l^1 < \cdots < t_l^{s_l}$ and $\tau_l \equiv \triangle t_l^k = t_l^k - t_l^{k-1}$ $(k = 1, 2, \dots, s_l)$. Denote $\gamma_2 = \tau_{l+1}/\tau_l$. For $l \geq 1$, let $\{I_{lj}\}_{j=1}^{M_l}$ be a nonoverlapping partition of I such that $I = \bigcup_{j=1}^{M_l} I_{lj}$. We denote the number of time steps on each subdomain I_{lj} by $\widehat{s}_{lj}$, hence, $\sum_{j=1}^{M_l} \widehat{s}_{lj} = s_l$, where s_l is the window size on level l.

On each level, we extend the subdomains $\{\Omega_{li}\}$ and $\{I_{lj}\}$ to larger domains $\{\Omega'_{li}\}$ and $\{I'_{lj}\}$ by adding several layers of fine mesh elements, respectively. The corresponding overlaps are denoted by $\delta_l > 0$ and $\delta_l^\tau \geq 0$. So we have $\Omega_{li} \subset \Omega'_{li} \subset \Omega$ and $I_{lj} \subseteq I'_{lj} \subset I$. Here, denote the maximum diameters of the subdomains $\{\Omega'_{li}\}_{i=1}^{N_l}$ and $\{I'_{lj}\}_{i=1}^{M_l}$ by H_l and H_l^τ, respectively. And we assume that the diameter of I'_{lj} is the same order as τ_{l-1} (i.e., $O(\tau_{l-1})$). Let s_{lj} be the number of time steps in I'_{lj}. Clearly, $s_{lj} \geq \widehat{s}_{lj}$. A typical partition of the space-time domain on level l is presented in Figure 1.

As in [35], we make the following assumption about the subdomains $\{\Omega'_{li}\}$.

Assumption 2.1. On each level, the decomposition $\Omega = \bigcup_{i=1}^{N_l} \Omega'_{li}$ satisfies the following:

(a) $\partial\Omega'_{li}$ aligns with the boundaries of level l elements, i.e., Ω'_{li} is the union of level l elements. The diameter of Ω'_{li} is the same order as h_{l-1} (i.e., $O(h_{l-1})$).
(b) The subdomains $\{\Omega'_{li}\}_{i=1}^{N_l}$ form a finite covering of Ω, with a covering constant N_c, i.e., we can color $\{\Omega'_{li}\}_{i=1}^{N_l}$ using at most N_c colors in such a way that subdomains of the same color are disjoint.
(c) There exists a partition of unity $\{\theta_{li}\}$, associated with $\{\Omega'_{li}\}_{i=1}^{N_l}$, which satisfies

$$\sum_i \theta_{li} = 1 \text{ with } \theta_{li} \in H_0^1(\Omega_{li}) \cap C^0(\Omega_{li}), \quad 0 \leq \theta_{li} \leq 1, \quad \text{and} \quad \|\nabla\theta_{li}\|_\infty \leq C/\delta_l,$$

where C is a constant independent of δ_l and H_l.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

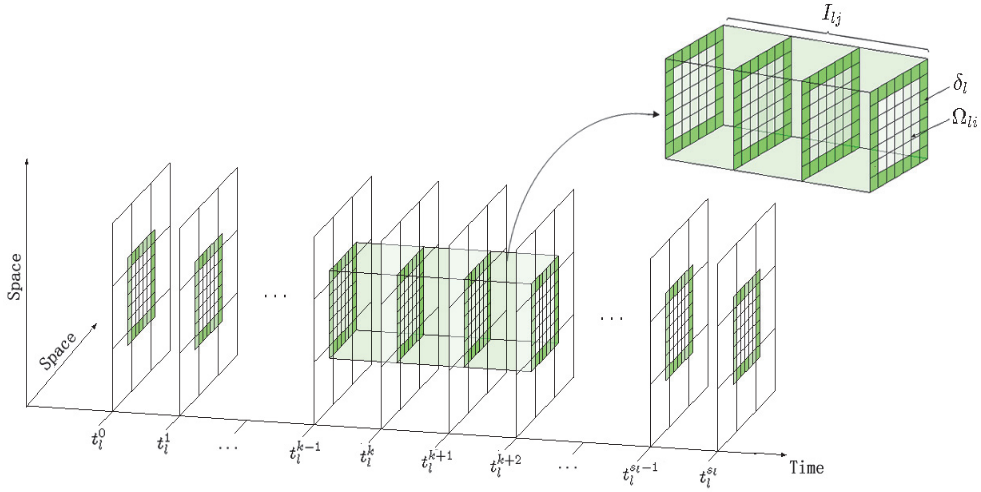

FIG. 1. *Partition of the space-time domain on level l, the top-right figure shows the space-time subdomains.*

For $l = 1, 2, \ldots, L$ and $i = 1, 2, \ldots, N_l$, the finite element space on Ω'_{li} is defined as

$$V_l^i = V_l \cap H_0^1(\Omega'_{li}).$$

Let

$$V_l^{ik} = \underbrace{0 \times \cdots \times 0 \times \overset{k}{V_l^i} \times 0 \times \cdots \times 0}_{s_l},$$

where the 0's denote the zero spaces at the other time steps and $k = 1, 2, \ldots, s_l$. Then, we can define the finite element function space as

$$\left(V_l^i\right)^{s_{lj}} = \cup_{t_l^k \in I'_{lj}} V_l^{ik} \subset (V_l)^{s_l},$$

where $(V_l)^{s_l} \subset (H_0^1(\Omega))^{s_l}$ denotes

$$\underbrace{V_l \times V_l \times \cdots \times V_l}_{s_l}.$$

The finite element solution of (2.8) is to find $\mathbf{u}_h^* \in (V_h)^s$ such that

$$D_{\tau,s}(\mathbf{u}_h^*, \mathbf{v}_h) = (\mathbf{f}, \mathbf{v}_h) \quad \forall \mathbf{v}_h \in (V_h)^s. \tag{2.9}$$

The finite element space $(V_h)^s = (V_L)^s$ can be represented as $(V_h)^s = (V_H)^{s_0} + \bigcup_{l=1}^{L}(V_l)^{s_l} = (V_H)^{s_0} + \bigcup_{l=1}^{L}\bigcup_{j=1}^{M_l}\bigcup_{i=1}^{N_l}(V_l^i)^{s_{lj}}$, where s_0 denotes the window size of the coarsest time level and

$$(V_H)^{s_0} = \underbrace{V_H \times V_H \times \cdots \times V_H}_{s_0}.$$

Define the linear operators $\mathbf{P}_0 : (V_h)^s \mapsto (V_H)^{s_0}$ and $\mathbf{P}_l^{ij} : (V_h)^s \mapsto (V_l^i)^{s_{lj}}$ such that

$$D_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{v}_H) = D_{\tau,s}(\mathbf{u}_h, \mathbf{v}_H) \ \ \forall \mathbf{u}_h \in (V_h)^s, \ \mathbf{v}_H \in (V_H)^{s_0}, \tag{2.10}$$

$$D_{\tau,s}(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{v}_l^{ij}) = D_{\tau,s}(\mathbf{u}_h, \mathbf{v}_l^{ij}) \ \ \forall \mathbf{u}_h \in (V_h)^s, \ \mathbf{v}_l^{ij} \in (V_l^i)^{s_{lj}}. \tag{2.11}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We then introduce the L-level additive Schwarz (MAS) operator $\mathbf{P}_{MAS} : (V_h)^s \mapsto (V_h)^s$ as

$$\mathbf{P}_{MAS} = \mathbf{P}_0 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}. \tag{2.12}$$

Denote $\mathbf{b} = \mathbf{P}_{MAS}\mathbf{u}_h^* = \mathbf{P}_0\mathbf{u}_h^* + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h^*$. The additive Schwarz algorithm for solving (2.9) can be written as

$$\mathbf{P}_{MAS}\mathbf{u}_h = \mathbf{b}. \tag{2.13}$$

Now we present the multilevel space-time additive Schwarz algorithm.

ALGORITHM 2.1 (L-level space-time additive Schwarz algorithm). *Find the solution of* (2.9) *by solving* (2.13) *with a Krylov subspace method.*

3. An estimate of the optimal convergence rate.

In this section, we first present a convergence theory for Algorithm 2.1 applied to problem (2.8) based on three assumptions, and then we verify the assumptions. Following [4, 14, 28], the convergence rate of the multilevel space-time additive Schwarz preconditioned GMRES method can be characterized by the two quantities

$$c_p = \inf_{\mathbf{u}_h\neq 0}\frac{A_{\tau,s}(\mathbf{P}_{MAS}\mathbf{u}_h,\mathbf{u}_h)}{A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)} \qquad \text{and} \qquad C_p = \sup_{\mathbf{u}_h\neq 0}\frac{A_{\tau,s}(\mathbf{P}_{MAS}\mathbf{u}_h,\mathbf{P}_{MAS}\mathbf{u}_h)}{A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)},$$

where $\mathbf{P}_{MAS}$ is defined by (2.12). Moreover, the residual at the kth iteration is bounded as

$$\|r_k\|_{\tau,s} \le \left(1-\frac{c_p^2}{C_p^2}\right)^{\frac{k}{2}}\|r_0\|_{\tau,s},$$

where $r_k = \mathbf{b} - \mathbf{P}_{MAS}\mathbf{u}_h^k$.

3.1. The convergence theorem.

To provide an estimate of c_p and C_p and their dependency on H, h, τ, L, and s, we first present the following three assumptions, and then give the main result of this paper based on these assumptions.

Assumption 3.1. There exists a constant $K_1 > 0$ such that

$$A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h,\mathbf{P}_0\mathbf{u}_h) + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right) \le K_1A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)\ \forall\mathbf{u}_h\in(V_h)^s, \tag{3.1}$$

where $K_1 = C(\|\Theta\|_0+1)$ and C is a positive constant independent of H, h, τ, L, and s provided that $H/\sqrt{\tau}$ is sufficiently small. $\|\Theta\|_0$ will be introduced in Assumption 3.3.

Assumption 3.2. There exists a constant $K_2 > 0$ such that

$$A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h,\mathbf{P}_0\mathbf{u}_h) + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right) \ge K_2A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)\ \forall\mathbf{u}_h\in(V_h)^s, \tag{3.2}$$

where $K_2 = cc_0(\max_{1\le l\le L}(1+h_{l-1}/\delta_l))^{-1}$, c is a positive constant independent of H, h, τ, L, and s provided that $H/\sqrt{\tau}$ is sufficiently small. The quantity c_0 is defined in Lemma 2.1.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Assumption 3.3. There exist constants $0 \leq \theta^{lk}_{jq,ip} \leq 1$ $(1 \leq i \leq N_l, 1 \leq p \leq N_k,\ 1 \leq j \leq M_l,\ 1 \leq q \leq M_k,\ 1 \leq l,k \leq L)$ independent of H, h, τ, L, and s, such that

(3.3)
$$\left|A\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right)\right| \leq \theta^{lk}_{jq,ip} A\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} A\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}} \quad \forall \mathbf{u}_l^{ij} \in \left(V_l^i\right)^{s_{lj}},\ \mathbf{u}_k^{pq} \in \left(V_k^p\right)^{s_{kq}},$$

(3.4)
$$\left|\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right)\right| \leq \theta^{lk}_{jq,ip} \left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} \left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}} \qquad \forall \mathbf{u}_l^{ij} \in \left(V_l^i\right)^{s_{lj}},\ \mathbf{u}_k^{pq} \in \left(V_k^p\right)^{s_{kq}}.$$

In the following discussion, we define the matrix $\Theta = \{\theta^{lk}_{jq,ip}\}$ and denote the l_2 norm of Θ by $\|\Theta\|_0$.

Here, $\theta^{lk}_{jq,ip}$ characterizes the relations between the subspaces $(V_l^i)^{s_{lj}}$ and $(V_k^p)^{s_{kq}}$, i.e.,

$$\theta^{lk}_{jq,ip} = \cos\left(\left(V_l^i\right)^{s_{lj}}, \left(V_k^p\right)^{s_{kq}}\right) = \sup_{\mathbf{u}_l^{ij} \in \left(V_l^i\right)^{s_{lj}}, \mathbf{u}_k^{pq} \in \left(V_k^p\right)^{s_{kq}}} \cos\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right).$$

When $\theta^{lk}_{jq,ip} = 0$, it implies that the subspaces $(V_l^i)^{s_{lj}}$ and $(V_k^p)^{s_{kq}}$ are orthogonal while $\theta^{lk}_{jq,ip} = 1$ is the usual Cauchy–Schwarz inequality. It follows from (3.3), (3.4), and the Cauchy–Schwarz inequality that

$$\begin{aligned}
\left|A_{\tau,s}\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right)\right| &= \left|\tau A\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right) + \left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right)\right| \\
&\leq \theta^{lk}_{jq,ip} \tau A\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} A\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}} + \theta^{lk}_{jq,ip} \left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} \left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}} \\
&\leq \theta^{lk}_{jq,ip} \left(\tau A\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right) + \left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)\right)^{\frac{1}{2}} \left(\tau A\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right) + \left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)\right)^{\frac{1}{2}} \\
&= \theta^{lk}_{jq,ip} A_{\tau,s}\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} A_{\tau,s}\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}}. \qquad (3.5)
\end{aligned}$$

We now present an important lemma [24] before presenting the convergence theorem.

LEMMA 3.1. *If H is sufficiently small, i.e., there exists a constant $H_0 > 0$ such that $H < H_0$, then,*

$$\|\boldsymbol{P}_0 \boldsymbol{u}_h - \boldsymbol{u}_h\| \leq CH\sqrt{H^2 + \tau}\ \|\boldsymbol{P}_0 \boldsymbol{u}_h - \boldsymbol{u}_h\|_{\tau,s},\ \ \boldsymbol{u}_h \in (V_h)^s.$$

Furthermore,

$$\|\boldsymbol{P}_0 \boldsymbol{u}_h - \boldsymbol{u}_h\|_{\tau,s} \leq C\|\boldsymbol{u}_h\|_{\tau,s},$$

where $C > 0$ is independent of H, h, τ, and s.

THEOREM 3.2. *If Assumptions* 3.1, 3.2, *and* 3.3 *are satisfied and $H/\sqrt{\tau}$ is sufficiently small, then*

(1) *there exists a constant C_p such that*

$$A_{\tau,s}(\boldsymbol{P}_{MAS}\boldsymbol{u}_h, \boldsymbol{P}_{MAS}\boldsymbol{u}_h) \leq C_p A_{\tau,s}(\boldsymbol{u}_h, \boldsymbol{u}_h) \ \ \forall \boldsymbol{u}_h \in (V_h)^s,$$

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

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

(2) *there exists a constant* $c_p > 0$ *such that*

$$A_{\tau,s}(\boldsymbol{u}_h, \boldsymbol{P}_{MAS}\boldsymbol{u}_h) \geq c_p A_{\tau,s}(\boldsymbol{u}_h, \boldsymbol{u}_h) \quad \forall \boldsymbol{u}_h \in (V_h)^s,$$

where $c_p = cc_0(\max_{1\leq l\leq L}(1 + h_{l-1}/\delta_l))^{-1}$*; here* c *is a constant independent of* H*,* h*,* τ*,* L*, and* s*.*

Proof. (1) Using the mean value theorem, it follows from (3.5) and Assumption 3.1 that

$$\begin{aligned}
&A_{\tau,s}(\mathbf{P}_{MAS}\mathbf{u}_h, \mathbf{P}_{MAS}\mathbf{u}_h)\\
&\quad = A_{\tau,s}\left(\mathbf{P}_0\mathbf{u}_h + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h\right)\\
&\quad \leq 2A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) + 2A_{\tau,s}\left(\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h, \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h\right)\\
&\quad = 2A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) + 2\sum_{l,k=1}^{L}\sum_{j,q=1}^{M_l}\sum_{i,p=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_k^{pq}\mathbf{u}_h\right)\\
&\quad \leq 2A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) + 2\|\Theta\|_0\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right)\\
&\quad \leq 2K_1\left(1 + \|\Theta\|_0\right)A_{\tau,s}(\mathbf{u}_h, \mathbf{u}_h)\\
&\quad \leq C\left(1 + \|\Theta\|_0\right)^2 A_{\tau,s}(\mathbf{u}_h, \mathbf{u}_h).
\end{aligned}$$

(2) From the definition of $\mathbf{P}_{MAS}$, $\mathbf{P}_0$, and $\mathbf{P}_l^{ij}$, we have

$$A_{\tau,s}(\mathbf{u}_h, \mathbf{P}_{MAS}\mathbf{u}_h) = A_{\tau,s}(\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right) = \mathbf{J}_1 + \mathbf{J}_2, \tag{3.6}$$

where

$$\begin{aligned}
\mathbf{J}_1 &= A_{\tau,s}(\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h)\\
&= A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) - S_{\tau,s}(\mathbf{u}_h - \mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) - N_s(\mathbf{u}_h - \mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{J}_2 &= \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right)\\
&= \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right) - \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}S_{\tau,s}(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h)\\
&\quad - \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}N_s\left(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right).
\end{aligned}$$

Similarly to the proof of [24, Theorem 3.6], from Lemmas 2.1 and 3.1, we have

$$\mathbf{J}_1 = A_{\tau,s}(\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) \geq A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h, \mathbf{P}_0\mathbf{u}_h) - CHA_{\tau,s}(\mathbf{u}_h, \mathbf{u}_h). \tag{3.7}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Since the support of $\mathbf{P}_l^{ij}\mathbf{u}_h$ lies in a subdomain of diameter H_l, from [30, Corollary A.15] and the definition of the (τ,s)-norm, we have $\|\mathbf{P}_l^{ij}\mathbf{u}_h\| \leq CH_l|\mathbf{P}_l^{ij}\mathbf{u}_h|_1 \leq C(H/\sqrt{\tau})\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}$. Using the same techniques as in the proof of [24, Theorem 3.6], it follows from Lemma 2.1, Assumption 3.1, and (3.4) that

$$\begin{aligned}
&\left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} S_{\tau,s}(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h)\right| \\
&\quad = \left|S_{\tau,s}\left(\mathbf{u}_h, \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h\right) - \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} S_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right)\right| \\
&\quad \leq C\left(\|\Theta\|_0\right)^{\frac{1}{2}} H\|\mathbf{u}_h\|_{\tau,s}\left(\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}^2\right)^{\frac{1}{2}} + CH\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}^2 \\
&\quad \leq C\left(1 + \left(\|\Theta\|_0\right)^{\frac{1}{2}}\right) K_1 H\|\mathbf{u}_h\|_{\tau,s}^2
\end{aligned}$$

and

$$\begin{aligned}
&\left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} N_s(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h)\right| \\
&\quad \leq \left|N_s(\mathbf{u}_h, \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h)\right| + \left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} N_s(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h)\right| \\
&\quad \leq C\left(\|\Theta\|_0\right)^{\frac{1}{2}}\frac{H}{\sqrt{\tau}}\|\mathbf{u}_h\|_{\tau,s}\left(\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}^2\right)^{\frac{1}{2}} + C\frac{H^2}{\tau}\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}^2 \\
&\quad \leq C\left(1 + \left(\|\Theta\|_0\right)^{\frac{1}{2}}\right) K_1\frac{H}{\sqrt{\tau}}\|\mathbf{u}_h\|_{\tau,s}^2.
\end{aligned}$$

The above two inequalities lead to

$$\begin{aligned}
\mathbf{J}_2 &\geq \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} A_{\tau,s}(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h) \\
&\quad - \left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left(S_{\tau,s}\left(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right) + N_s\left(\mathbf{u}_h - \mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h\right)\right)\right| \\
&\geq \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} A_{\tau,s}(\mathbf{P}_l^{ij}\mathbf{u}_h, \mathbf{P}_l^{ij}\mathbf{u}_h) - C\left(1 + \left(\|\Theta\|_0\right)^{\frac{1}{2}}\right) K_1\frac{H}{\sqrt{\tau}}\|\mathbf{u}_h\|_{\tau,s}^2. \qquad (3.8)
\end{aligned}$$

If $H/\sqrt{\tau}$ is sufficiently small, combining (3.6), (3.7), (3.8), and Assumption 3.2 we arrive at

$$A_{\tau,s}(\mathbf{u}_h, \mathbf{P}\mathbf{u}_h) \geq K_2 A_{\tau,s}(\mathbf{u}_h, \mathbf{u}_h) = cc_0\left(\max_{1\leq l\leq L}\left(1 + \frac{h_{l-1}}{\delta_l}\right)\right)^{-1} A_{\tau,s}(\mathbf{u}_h, \mathbf{u}_h).$$

This completes the proof. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

3.2. Verifications of Assumptions 3.1–3.3. In this section, we focus on the verifications of the three assumptions. First, we present the verification of Assumption 3.3 and the key is to give a stronger estimate of $\theta^{lk}_{jq,ip}$ in (3.3) and (3.4), and then we obtain the upper bound of $\|\Theta\|_0$.

LEMMA 3.3. *If $l < k$, there exists a positive constant C independent of h, τ, L, and s such that*

$$\theta^{lk}_{jq,ip} \le C\gamma_1^{(k-l-1)d/2}\gamma_2^{(k-l-1)/2}.$$

Proof. For $\mathbf{u}_l^{ij} \in (V_l^i)^{s_{lj}}$, $\mathbf{u}_k^{pq} \in (V_k^p)^{s_{kq}}$, by the Cauchy–Schwarz inequality, we have

$$\begin{aligned}
&A(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq})\\
&\quad = \int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_k^{pq}dx\\
&\quad \le \left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\right)^{\frac{1}{2}}\left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_k^{pq}\cdot\nabla\mathbf{u}_k^{pq}dx\right)^{\frac{1}{2}}\\
&\quad \le \left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\right)^{\frac{1}{2}}\left(\int_{\Omega'_{kp}, I'_{kq}} \alpha(x)\nabla\mathbf{u}_k^{pq}\cdot\nabla\mathbf{u}_k^{pq}dx\right)^{\frac{1}{2}}\\
&\quad = \left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\right)^{\frac{1}{2}} A(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq})^{\frac{1}{2}}.
\end{aligned}$$

Let $K^l \subset \Omega'_{li}$ and $\tau^l \subset I'_{lj}$. From the definition of the linear finite element space of V_h, we see that $|\nabla\mathbf{u}_l^{ij}|$ is a constant. Moreover, from the assumptions on the partition of the domains Ω and I, we obtain

$$\begin{aligned}
&\int_{K^l\cap\Omega'_{kp}, \tau^l\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\\
&\quad \le C|\nabla\mathbf{u}_l^{ij}|^2\int_{K^l\cap\Omega'_{kp}, \tau^l\cap I'_{kq}} 1dx\\
&\quad \le C\frac{meas(K^l\cap\Omega'_{kp})}{meas(K^l)}\frac{meas(\tau^l\cap I'_{kq})}{meas(\tau^l)}\int_{K^l,\tau^l}\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\\
&\quad \le C\frac{h_{k-1}^d}{h_l^d}\frac{\tau_{k-1}}{\tau_l}\int_{K^l,\tau^l}\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx\\
&\quad = C\gamma_1^{d(k-l-1)}\gamma_2^{k-l-1}\int_{K^l,\tau^l}\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}dx.
\end{aligned}$$

Here $meas(\,)$ is the area (in R^2) and volume (in R^3) of the domain. Summing over

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

$K^l \subset \Omega'_{li}$ and $\tau^l \subset I'_{lj}$ and combining the above inequalities, we have

$$
\begin{aligned}
&A\left(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}\right) \\
&\quad \leq \left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \alpha(x)\nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}\,dx\right)^{\frac{1}{2}} A\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}} \\
&\quad \leq C\gamma_1^{d(k-l-1)/2}\gamma_2^{(k-l-1)/2} A(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq})^{\frac{1}{2}}\left(\int_{\Omega'_{li}, I'_{lj}} \nabla\mathbf{u}_l^{ij}\cdot\nabla\mathbf{u}_l^{ij}\,dx\right)^{\frac{1}{2}} \\
&\quad = C\gamma_1^{d(k-l-1)/2}\gamma_2^{(k-l-1)/2} A\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} A\left(\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq}\right)^{\frac{1}{2}}.
\end{aligned}
$$

Let $K^l \subset \Omega'_{li}$ have vertices $y_1, y_2, \ldots, y_{n_l}$, and let $x_1, x_2, \ldots, x_{n_k}$ be all the nodal points of $\mathcal{T}_k$ in Ω'_{kp}. It can be shown that $n_k \leq C(h_{k-1}/h_k)^d$ [6, Lemma 3.3]. Since $\mathbf{u}_l^{ij}$ is linear in $K^l \cap \Omega'_{kp}$, we obtain

$$
\begin{aligned}
\int_{K^l\cap\Omega'_{kp}, \tau^l\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx &= \int_{K^l\cap\Omega'_{kp}} \widetilde{\mathbf{u}}_l^{ij}\cdot\widetilde{\mathbf{u}}_l^{ij}\,dx \leq Ch_k^d\left(\sum_{1\leq\sigma\leq n_k, x_\sigma\in K^l}\left(\widetilde{\mathbf{u}}_l^{ij}(x_\sigma)\right)^2\right) \\
&\leq Cn_k h_k^d \max_{x_\sigma\in K^l}\left\{\left(\widetilde{\mathbf{u}}_l^{ij}(x_\sigma)\right)^2\right\} \leq Cn_k h_k^d \sum_{\xi=1}^{n_l}\left(\widetilde{\mathbf{u}}_l^{ij}(y_\xi)\right)^2 \\
&\leq Cn_k \frac{h_k^d}{h_l^d}\int_{K^l} \widetilde{\mathbf{u}}_l^{ij}\cdot\widetilde{\mathbf{u}}_l^{ij}\,dx \leq C\frac{h_{k-1}^d}{h_l^d}\int_{K^l, \tau^l\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx \\
&= C\gamma_1^{d(k-l-1)/2}\int_{K^l, \tau^l\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx,
\end{aligned}
$$

where $\widetilde{\mathbf{u}}_l^{ij}$ is defined in (3.9) and (3.10). Analogously, we can also have

$$
\int_{K^l, \tau^l\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx \leq C\gamma_2^{(k-l-1)/2}\int_{K^l, \tau^l} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx.
$$

Combining the above two inequalities and the Cauchy–Schwarz inequality we have

$$
\begin{aligned}
(\mathbf{u}_l^{ij}, \mathbf{u}_k^{pq}) &= \int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_k^{pq}\,dx \\
&\leq C\left(\int_{\Omega'_{li}\cap\Omega'_{kp}, I'_{lj}\cap I'_{kq}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx\right)^{\frac{1}{2}} (\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq})^{\frac{1}{2}} \\
&\leq C\gamma_1^{d(k-l-1)/2}\gamma_2^{(k-l-1)/2}\left(\int_{\Omega'_{li}, I'_{lj}} \mathbf{u}_l^{ij}\cdot\mathbf{u}_l^{ij}\,dx\right)^{\frac{1}{2}} (\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq})^{\frac{1}{2}} \\
&= C\gamma_1^{d(k-l-1)/2}\gamma_2^{(k-l-1)/2}\left(\mathbf{u}_l^{ij}, \mathbf{u}_l^{ij}\right)^{\frac{1}{2}} (\mathbf{u}_k^{pq}, \mathbf{u}_k^{pq})^{\frac{1}{2}}.
\end{aligned}
$$

This completes the proof. ∎

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Remark 3.1. For simplicity, in the proof of Lemma 3.3, we define

$$A(\mathbf{u},\mathbf{v}) = \int_{\Omega_1\cap\Omega_2, I_1\cap I_2} \alpha(x)\nabla\mathbf{u}\cdot\nabla\mathbf{v}dx = \int_{\Omega_1\cap\Omega_2} \alpha(x)\nabla\widetilde{\mathbf{u}}\cdot\nabla\widetilde{\mathbf{v}}dx = \sum_{k=1}^{m}\int_{\Omega_1\cap\Omega_2} \alpha(x)\nabla u^{i_k}\cdot\nabla v^{i_k}dx \tag{3.9}$$

and

$$(\mathbf{u},\mathbf{v}) = \int_{\Omega_1\cap\Omega_2, I_1\cap I_2} \mathbf{u}\cdot\mathbf{v}dx = \int_{\Omega_1\cap\Omega_2} \widetilde{\mathbf{u}}\cdot\widetilde{\mathbf{v}}dx = \sum_{k=1}^{m}\int_{\Omega_1\cap\Omega_2} u^{i_k}\cdot v^{i_k}dx, \tag{3.10}$$

where $\widetilde{\mathbf{u}} = (u^{i_1}, u^{i_2}, \ldots, u^{i_m})$, $\{i_1, i_2, \ldots, i_m\}$ is a subset of $\{1, 2, \ldots, s\}$, and m denotes the number of time steps contained in both the time domains I_1 and I_2. From Lemma 3.3, we see that the farther apart the two levels are, the more orthogonal are the associated subspaces.

Now we prove the upper bound of $\|\Theta\|_0$ in Assumption 3.3. The following lemma [29, Lemma 9] is necessary in the analysis.

LEMMA 3.4. *Let $A \in R^{n\times n}$ be a matrix with at most n_c entries per column. Then*

$$\|A\|_0 \le \sqrt{n_c}\max_i\left(\sum_j A_{ij}^2\right)^{1/2}.$$

The matrix defined in Assumption 3.3,

$$\Theta = \{\theta^{lk}_{jq,ip}\}_{l,k\le L, j\le M_l, q\le M_k, i\le N_l, p\le N_k}, \tag{3.11}$$

can be partitioned into an $L\times L$ block matrix

$$\Theta = \{\theta^{lk}\}_{1\le l,k\le L},$$

where $\theta^{lk} = \{\theta^{lk}_{jq}\}_{1\le j\le M_l, 1\le q\le M_k}$ is also an $M_l\times M_k$ block matrix and $\theta^{lk}_{jq} = \{\theta^{lk}_{jq,ip}\}_{1\le i\le N_l, 1\le p\le N_k}$ is an $N_l\times N_k$ submatrix. Hence, θ^{lk} is an $M_l\times N_l$ by $M_k\times N_k$ matrix.

LEMMA 3.5. *For $1\le l\le k\le L$, there exists a constant C independent of h, τ, L, and s such that*

$$\left\|\theta^{lk}\right\|_0 \le CN_c\gamma_2^{(k-l-1)/2}.$$

Proof. We consider first the case $l = k$. From the assumptions on the partition of the time domain I, it is clear that $I'_{lj}\cap I'_{lq} \ne \emptyset$ only if $|j-q|\le 1$, i.e., θ^{ll}_{jq} is a zero matrix except $q = j-1$, $q = j$, and $q = j+1$, so θ^{ll} is actually a block triangular matrix. Moreover, from Assumption 2.1, we see that $\{\Omega'_{li}\}_{i=1}^{N_l}$ can be colored with at most N_c colors, so the number of nonzero $\theta^{ll}_{jq,ip}$ in each row of the submatrix θ^{ll}_{jq} is N_c. Therefore, the number of nonzero $\theta^{ll}_{jq,ip}$ in each row of the matrix θ^{ll} is $3N_c$. Since θ^{ll} is symmetric, we have

$$\left\|\theta^{ll}\right\|_0 \le \left\|\theta^{ll}\right\|_\infty \le 3N_c, \tag{3.12}$$

where $\|\cdot\|_\infty$ denotes the l_∞ norm of the matrix.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

We consider next the case $l < k$. It is clear that θ_{jq}^{lk} is a zero matrix if $I'_{lj} \cap I'_{kq} = \emptyset$. Further, from (3.9) and (3.10), we see that the number of time steps is zero if I'_{kq} is completely contained in a single coarse mesh element of I'_{lj}. Hence, θ_{jq}^{lk} is also a zero matrix when $I'_{kq} \subset \tau^l \subset I'_{lj}$. It implies that most submatrices θ_{jq}^{lk} are zero matrices, and the number of nonzero θ_{jq}^{lk} per row (fixed j) of the matrix θ^{lk} is bounded by $\widetilde{C}$, the number of nonzero θ_{jq}^{lk} per column (fixed q) is no more than 3, where the constant $\widetilde{C}$ is independent of the mesh sizes. Moreover, it is clear that $\theta_{jq,ip}^{lk} = 0$ when $\Omega'_{li} \cap \Omega'_{kp} = \emptyset$. From (3.9), we can obtain $\theta_{jq,ip}^{lk} = 0$ if $\Omega'_{kp} \subset K^l \subset \Omega'_{li}$ since $\mathbf{u}_l^{ij} \in (V_l^i)^{s_j}$ is linear and $\mathbf{v}_k^{pq} \in (V_k^p)^{s_q}$ vanishes on ∂K^l. Unfortunately, from (3.10), we cannot obtain $\theta_{jq,ip}^{lk} = 0$ when $\Omega'_{kp} \subset K^l \subset \Omega'_{li}$. Therefore, the number of nonzero $\theta_{jq,ip}^{lk}$ per row (fixed i) of the submatrix θ_{jq}^{lk} is bounded by $\widehat{C}\gamma_1^{-(k-l-1)d}$ and the number of nonzero $\theta_{jq,ip}^{lk}$ per column (fixed p) of θ_{jq}^{lk} is bounded by N_c, where the constant $\widehat{C}$ is only dependent on the number of finitely many edges/faces of $K^l \subset \Omega'_{li}$. The above analysis implies that, in the case of $l < k$, the number of nonzero $\theta_{jq,ip}^{lk}$ per row of θ^{lk} is bounded by $\widetilde{C}\widehat{C}\gamma_1^{-(k-l-1)d}$ and the number of nonzero $\theta_{jq,ip}^{lk}$ per column of θ^{lk} is bounded by $3N_c$. Combining Lemmas 3.3 and 3.4 yields

(3.13)
$$\left\|\theta^{lk}\right\|_0 \le (3N_c)^{\frac{1}{2}} \left(\widetilde{C}\widehat{C}\gamma_1^{-(k-l-1)d}\left(C\gamma_1^{|k-l-1|d/2}\gamma_2^{|k-l-1|/2}\right)^2\right)^{\frac{1}{2}} \le CN_c\gamma_2^{(k-l-1)/2}.$$

The proof is completed from (3.12) and (3.13). ∎

Lemma 3.6. *For the matrix* Θ *defined in* (3.11)*, we have*

$$\|\Theta\|_0 \le CN_c\frac{1}{1-\sqrt{\gamma_2}},$$

where $C > 0$ *is independent of* h*,* τ*,* L*, and* s*.*

Proof. Define
$$\widetilde{\Theta} = \left\{\left\|\theta^{lk}\right\|_0\right\}_{1\le l,k\le L}.$$

Since $\widetilde{\Theta}$ is symmetric, from [35, Lemma 3.3] and Lemma 3.5, we have

$$\|\Theta\|_0 \le \|\widetilde{\Theta}\|_0 \le \|\widetilde{\Theta}\|_\infty \le CN_c\frac{1}{1-\sqrt{\gamma_2}}.$$

This completes the proof. ∎

From Lemma 3.6, we see that the upper bound of $\|\Theta\|_0$ is independent of h, τ, L, and s, which completes the proof of Assumptions 3.3. In the following, we verify Assumptions 3.1 and 3.2.

Define the projection operators $\widehat{\mathbf{P}}_l : (V_h)^s \to (V_l)^{s_l}$ such that

$$A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h, \mathbf{v}_l) = A_{\tau,s}(\mathbf{u}_h, \mathbf{v}_l) \quad \forall \mathbf{u}_h \in (V_h)^s,\ \mathbf{v}_l \in (V_l)^{s_l}. \tag{3.14}$$

Then, we consider the decomposition of any function $\mathbf{u}_h \in (V_h)^s$ as follows:

$$\mathbf{u}_h = \sum_{l=0}^{L}\mathbf{u}_l, \quad \mathbf{u}_0 = \widehat{\mathbf{P}}_0\mathbf{u}_h, \quad \mathbf{u}_l = (\widehat{\mathbf{P}}_l - \widehat{\mathbf{P}}_{l-1})\mathbf{u}_h, \quad l = 1, 2, \ldots, L. \tag{3.15}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Further, $\mathbf{u}_l$ is decomposed by

$$\mathbf{u}_l = \sum_{j=1}^{M_l}\sum_{i=1}^{N_l} \mathbf{u}_l^{ij}, \quad \mathbf{u}_l^{ij} \in \left(V_l^i\right)^{s_{lj}}.$$

Similarly to [24, Lemma 3.4], using Nitsche's trick, we obtain

$$\|\widehat{\mathbf{P}}_l\mathbf{u}_h - \mathbf{u}_h\| \le Ch_l\sqrt{h_l^2+\tau}\ \|\widehat{\mathbf{P}}_l\mathbf{u}_h - \mathbf{u}_h\|_{\tau,s},\ \mathbf{u}_h \in (V_h)^s,$$

and

$$\|\widehat{\mathbf{P}}_l\mathbf{u}_h - \mathbf{u}_h\|_{\tau,s} \le C\|\mathbf{u}_h\|_{\tau,s},$$

where $C > 0$ is independent of the mesh size h, the time step size τ, and the window size s. Since $\widehat{\mathbf{P}}_l$ is a projection, i.e., $\widehat{\mathbf{P}}_l^2 = \widehat{\mathbf{P}}_l$, we have $\widehat{\mathbf{P}}_{l-1}\mathbf{u}_l = \widehat{\mathbf{P}}_{l-1}\widehat{\mathbf{P}}_l\mathbf{u}_h - \widehat{\mathbf{P}}_{l-1}^2\mathbf{u}_h = 0$. Combining this with the above two inequalities yields

$$\|\mathbf{u}_l\| = \|(\mathbf{I}-\widehat{\mathbf{P}}_{l-1})\mathbf{u}_l\| \le Ch_{l-1}\sqrt{h_{l-1}^2+\tau}\|\mathbf{u}_l\|_{\tau,s}. \tag{3.16}$$

Since $\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h \in (V_{l-1})^{s_{l-1}} \subset (V_l)^{s_l}$, we have

$$A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h, \widehat{\mathbf{P}}_{l-1}\mathbf{u}_h) = A_{\tau,s}(\mathbf{u}_h, \widehat{\mathbf{P}}_{l-1}\mathbf{u}_h) = A_{\tau,s}(\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h, \widehat{\mathbf{P}}_{l-1}\mathbf{u}_h).$$

Hence, it follows from the above equality, (3.14), and (3.15) that

$$\begin{aligned}
&A_{\tau,s}(\mathbf{u}_l,\mathbf{u}_l)\\
&= A_{\tau,s}\left((\widehat{\mathbf{P}}_l-\widehat{\mathbf{P}}_{l-1})\mathbf{u}_h, (\widehat{\mathbf{P}}_l-\widehat{\mathbf{P}}_{l-1})\mathbf{u}_h\right)\\
&= A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h,\widehat{\mathbf{P}}_l\mathbf{u}_h) - 2A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h,\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h) + A_{\tau,s}(\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h,\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h)\\
&= A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h,\widehat{\mathbf{P}}_l\mathbf{u}_h) - A_{\tau,s}(\widehat{\mathbf{P}}_l\mathbf{u}_h,\widehat{\mathbf{P}}_{l-1}\mathbf{u}_h)\\
&= A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_l).
\end{aligned} \tag{3.17}$$

Now we present a bounded decomposition lemma.

LEMMA 3.7. *For any $\boldsymbol{v}_h \in (V_h)^s$, there exist $\boldsymbol{v}_0 = \boldsymbol{v}_H \in (V_H)^{s_0}$ and $\boldsymbol{v}_l^{ij} \in (V_l^i)^{s_{lj}}$ such that $\boldsymbol{v}_h = \boldsymbol{v}_0 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\boldsymbol{v}_l^{ij}$ and*

$$\|\boldsymbol{v}_0\|_{\tau,s}^2 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left\|\boldsymbol{v}_l^{ij}\right\|_{\tau,s}^2 \le C\max_{1\le l\le L}\left(1+\frac{h_{l-1}}{\delta_l}\right)\|\boldsymbol{v}_h\|_{\tau,s}^2, \tag{3.18}$$

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

Proof. For each j, we define $(V_l)^{s_{lj}} = \bigcup_{i=1}^{N_l}(V_l^i)^{s_{lj}}$ and let

$$\mathbf{v}_l^{ij} = \widehat{R}_{ij}^l\left(\widehat{I}^l\left(\theta_{li}\mathbf{v}_l^j\right)\right) \in \left(V_l^i\right)^{s_{lj}},$$

where $\mathbf{v}_l^j \in (V_l)^{s_{lj}}$, and θ_{li} denotes a partition of unity as in Assumption 2.1, $\widehat{I}^l$ is the linear interpolation operator onto $(V_l)^{s_{lj}}$, and $\widehat{R}_{ij}^l : (V_l)^{s_{lj}} \to (V_l^i)^{s_{lj}}$ is the restriction operator.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Using the techniques in [30, Lemma 3.12] and according to the definition of the (τ, s)-norm, we have

$$\begin{aligned} A_{\tau,s}&\left(\mathbf{v}_l^{ij}, \mathbf{v}_l^{ij}\right) \\ &= \tau \left|\widehat{I}^l\left(\theta_{li}\mathbf{v}_l^j\right)\right|^2_{1,\Omega'_{li}} + \left\|\widehat{I}^l\left(\theta_{li}\mathbf{v}_l^j\right)\right\|^2_{\Omega'_{li}} \\ &\le \tau\left(\left\|\theta_{li}\nabla\mathbf{v}_l^j\right\|^2_{\Omega'_{li}} + \left\|\nabla\theta_{li}\mathbf{v}_l^j\right\|^2_{\Omega'_{li}}\right) + \left\|\theta_{li}\mathbf{v}_l^j\right\|^2_{\Omega'_{li}} \\ &\le C\left(\tau\left|\mathbf{v}_l^j\right|^2_{1,\Omega'_{li}} + \left\|\mathbf{v}_l^j\right\|^2_{\Omega'_{li}}\right) + \tau\left\|\nabla\theta_{li}\mathbf{v}_l^j\right\|^2_{\Omega'_{li}}, \end{aligned} \tag{3.19}$$

where $|\cdot|_{1,\Omega'_{li}}$ and $\|\cdot\|_{\Omega'_{li}}$ denote the H^1 seminorm and L^2 norm on the subdomain Ω'_{li}, respectively. We note that $\nabla\theta_{li}$ is zero except in a strip Ω'_{li,δ_l} of width δ_l in the vicinity of $\partial\Omega'_{li}$. From Assumption 2.1 and [13, Lemma 3.1] we have

$$\left\|\nabla\theta_{li}\mathbf{v}_l^j\right\|^2_{\Omega'_{li}} \le C\frac{1}{\delta_l^2}\left\|\mathbf{v}_l^j\right\|^2_{\Omega'_{li,\delta_l}} \le C\left(1+\frac{h_{l-1}}{\delta_l}\right)\left\|\mathbf{v}_l^j\right\|^2_{1,\Omega'_{li}} + C\frac{1}{h_{l-1}\delta_l}\left\|\mathbf{v}_l^j\right\|^2_{\Omega'_{li}}, \tag{3.20}$$

where $\|\cdot\|_{\Omega'_{li,\delta_l}}$ denotes the L^2 norm in the strip Ω'_{li,δ_l}.

Combining the finite covering property of Ω'_{li}, (3.19), (3.20), and summing over i and j yields

$$\begin{aligned} \sum_{j=1}^{M_l}\sum_{i=1}^{N_l} &A_{\tau,s}\left(\mathbf{v}_l^{ij}, \mathbf{v}_l^{ij}\right) \\ &\le C\sum_{j=1}^{M_l}\left(\tau\left(1+\frac{h_{l-1}}{\delta_l}\right)\left|\mathbf{v}_l^j\right|_1^2 + \left\|\mathbf{v}_l^j\right\|^2\right) + C\tau\sum_{j=1}^{M_l}\frac{1}{h_{l-1}\delta_l}\left\|\mathbf{v}_l^j\right\|^2 \\ &\le C\left(\tau\left(1+\frac{h_{l-1}}{\delta_l}\right)|\mathbf{v}_l|_1^2 + \|\mathbf{v}_l\|^2\right) + C\frac{\tau}{h_{l-1}\delta_l}\|\mathbf{v}_l\|^2 \\ &\le C\left(1+\frac{h_{l-1}}{\delta_l}\right)A_{\tau,s}(\mathbf{v}_l, \mathbf{v}_l). \end{aligned} \tag{3.21}$$

The second to the last inequality of (3.21) is satisfied since the time subdomain I'_{lj} is extended from I_{lj} by adding several layers of the fine mesh elements, and the last inequality holds because of (3.16).

From (3.15), (3.17), and (3.21), summing over l, we have

$$\begin{aligned} A_{\tau,s}(\mathbf{v}_0, \mathbf{v}_0) &+ \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} A_{\tau,s}\left(\mathbf{v}_l^{ij}, \mathbf{v}_l^{ij}\right) \\ &\le A_{\tau,s}(\mathbf{v}_0, \mathbf{v}_0) + C\sum_{l=1}^{L}\left(1+\frac{h_{l-1}}{\delta_l}\right)A_{\tau,s}(\mathbf{v}_l, \mathbf{v}_l) \\ &\le C\max_{1\le l\le L}\left(1+\frac{h_{l-1}}{\delta_l}\right)\sum_{l=0}^{L}A_{\tau,s}(\mathbf{v}_h, \mathbf{v}_l) \\ &= C\max_{1\le l\le L}\left(1+\frac{h_{l-1}}{\delta_l}\right)A_{\tau,s}(\mathbf{v}_h, \mathbf{v}_h), \end{aligned}$$

which completes the proof. ∎

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

LEMMA 3.8. *If $H/\sqrt{\tau}$ is sufficiently small, then Assumptions* 3.1 *and* 3.2 *are satisfied.*

Proof. (1) Verification of Assumption 3.1. Since $\|\mathbf{P}_l^{ij}\mathbf{u}_h\| \le C\left(H/\sqrt{\tau}\right)\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}$, it follows from Lemma 2.1 that

$$\begin{aligned}&\left|D_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right|\\ &\quad=\left|A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)+S_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)+N_s\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right|\\ &\quad\ge\left|A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)-\cos\left(\frac{\pi}{s+1}\right)\left\|\mathbf{P}_l^{ij}\mathbf{u}_h\right\|^2-C\tau\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_1\|\mathbf{P}_l^{ij}\mathbf{u}_h\|\right.\\ &\quad\ge\left(1-CH-C\frac{H^2}{\tau}\right)A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right).\end{aligned}$$

If $H/\sqrt{\tau}$ is sufficiently small, the above inequality implies that there exists a constant $\widehat{C} > 1$ such that

$$A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\le\widehat{C}\left|D_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right|.$$

Then, from (3.5) and Lemma 2.1, we have

$$\begin{aligned}&\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\\ &\qquad\le\widehat{C}\left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}D_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right|=\widehat{C}\left|\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}D_{\tau,s}\left(\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right|\\ &\qquad\le CA_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)^{\frac12}A_{\tau,s}\left(\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h,\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{P}_l^{ij}\mathbf{u}_h\right)^{\frac12}\\ &\qquad\le C(\|\Theta\|_0)^{\frac12}A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)^{\frac12}\left(\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\right)^{\frac12}.\end{aligned}\tag{3.22}$$

Obviously, (3.22) implies

$$\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\le C\|\Theta\|_0A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)\quad\forall\mathbf{u}_h\in(V_h)^s.\tag{3.23}$$

It follows from Lemma 3.1 that

$$A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h,\mathbf{P}_0\mathbf{u}_h)\le CA_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)\quad\forall\mathbf{u}_h\in(V_h)^s.\tag{3.24}$$

Combining (3.23) and (3.24) yields

$$A_{\tau,s}(\mathbf{P}_0\mathbf{u}_h,\mathbf{P}_0\mathbf{u}_h)+\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}A_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{P}_l^{ij}\mathbf{u}_h\right)\le C(\|\Theta\|_0+1)A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h).$$

Therefore, Assumption 3.1 holds with $K_1 = C(\|\Theta\|_0 + 1)$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

(2) Verification of Assumption 3.2. It follows from Lemmas 2.1, 3.7 and the Cauchy–Schwarz inequality that

$$
\begin{aligned}
D_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h) &= D_{\tau,s}(\mathbf{u}_h,\mathbf{u}_0) + D_{\tau,s}\left(\mathbf{u}_h, \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\mathbf{u}_l^{ij}\right)\\
&= D_{\tau,s}(\mathbf{P}_0\mathbf{u}_h,\mathbf{u}_0) + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l} D_{\tau,s}\left(\mathbf{P}_l^{ij}\mathbf{u}_h,\mathbf{u}_l^{ij}\right)\\
&\le C\|\mathbf{P}_0\mathbf{u}_h\|_{\tau,s}\|\mathbf{u}_0\|_{\tau,s} + C\sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left\|\mathbf{P}_l^{ij}\mathbf{u}_h\right\|_{\tau,s}\left\|\mathbf{u}_l^{ij}\right\|_{\tau,s}\\
&\le C\left(\|\mathbf{P}_0\mathbf{u}_h\|_{\tau,s}^2 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\|\mathbf{P}_l^{ij}\mathbf{u}_h\|_{\tau,s}^2\right)^{\frac{1}{2}}\left(\|\mathbf{u}_0\|_{\tau,s}^2 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left\|\mathbf{u}_l^{ij}\right\|_{\tau,s}^2\right)^{\frac{1}{2}}\\
&\le C\max_{1\le l\le L}\left(1+\frac{h_{l-1}}{\delta_l}\right)^{\frac{1}{2}}\|\mathbf{u}_h\|_{\tau,s}\left(\|\mathbf{P}_0\mathbf{u}_h\|_{\tau,s}^2 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left\|\mathbf{P}_l^{ij}\mathbf{u}_h\right\|_{\tau,s}^2\right)^{\frac{1}{2}}.
\end{aligned}
\tag{3.25}
$$

From Lemmas 2.1 and 3.1, we obtain

$$
\begin{aligned}
|D_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h)| &\ge c_0\|\mathbf{u}_h\|_{\tau,s}^2 - c_1\|\mathbf{u}_h\|^2\\
&\ge c_0\|\mathbf{u}_h\|_{\tau,s}^2 - \left(\cos\left(\frac{\pi}{s+1}\right)+m_1\tau\right)\|\mathbf{P}_0\mathbf{u}_h-\mathbf{u}_h\|^2\\
&\quad - \left(\cos\left(\frac{\pi}{s+1}\right)+m_1\tau\right)\|\mathbf{P}_0\mathbf{u}_h\|^2\\
&\ge \left(c_0 - CH^2\right)\|\mathbf{u}_h\|_{\tau,s}^2 - (1+m_1\tau)\|\mathbf{P}_0\mathbf{u}_h\|_{\tau,s}\|\mathbf{u}_h\|_{\tau,s}.
\end{aligned}
\tag{3.26}
$$

Combining (3.25) and (3.26) and rearranging the terms, we have

$$
\|\mathbf{P}_0\mathbf{u}_h\|_{\tau,s}^2 + \sum_{l=1}^{L}\sum_{j=1}^{M_l}\sum_{i=1}^{N_l}\left\|\mathbf{P}_l^{ij}\mathbf{u}_h\right\|_{\tau,s}^2 \ge cc_0\left(\max_{1\le l\le L}\left(1+\frac{h_{l-1}}{\delta_l}\right)\right)^{-1}A_{\tau,s}(\mathbf{u}_h,\mathbf{u}_h).
$$

Hence, Assumption 3.2 holds with $K_2 = cc_0(\max_{1\le l\le L}(1+h_{l-1}/\delta_l))^{-1}$. ∎

4. Numerical results. In this section, we report some numerical experiments to illustrate the optimal convergence and the parallel performance of the proposed algorithm. The software is implemented on the top of PETSc [2]. The computational domain Ω is initially divided into a coarse mesh of size h_1. The fine meshes are then obtained by refining the coarsest mesh uniformly. Moreover, the space-time domain $\Omega\times[0,t^s]$ is divided into N_p subdomains and each subdomain is assigned to one processor. The problem is solved by a preconditioned GMRES(30) method with stopping conditions $rtol = 1.0\times10^{-6}$ (the relative convergence tolerance) and $atol = 1.0\times10^{-15}$ (the absolute convergence tolerance). In the following, the overlap is denoted by "ovlp". The quantities τ_l and h_l denote the time step size and the mesh size on the mesh level l, respectively. For simplicity, the window size, and the time step size on the finest mesh are denoted by s and τ, respectively.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 1
The number of GMRES iterations for Example 4.1 *by Algorithm* 2.1 *(three level) with* $h_1 = \tau_1 = 1/32$, $h_2 = \tau_2 = 1/64$, *and* $h_3 = \tau_3 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	50	45	46	56	33	30	49	44	42
32	51	34	30	56	38	34	51	39	33
64	40	35	30	46	39	33	57	43	37
96	40	37	30	44	40	33	58	42	35
128	41	38	30	41	39	32	54	45	34

TABLE 2
The number of GMRES iterations for Example 4.1 *by Algorithm* 2.1 *(two level) with* $h_1 = \tau_1 = 1/32$ *and* $h_2 = \tau_2 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	46	39	43	50	61	40	55	44	54
32	53	70	36	59	56	42	54	40	45
64	44	40	33	54	73	42	64	80	45
96	45	41	34	56	57	43	53	45	49
128	45	42	36	46	47	35	56	49	49

4.1. Numerical studies of the convergence rate. In this subsection, we study the convergence rate of Algorithm 2.1, including the dependence on the overlap size, the number of the subdomains, the mesh sizes, the window size, and the number of levels for several test problems. All of these properties of the algorithm are machine independent. The machine dependent issues will be investigated in the next two subsections.

Example 4.1. Consider a heat equation

$$\begin{cases} u_t - \triangle u = f & \textit{in } \Omega \times (0, T], \\ u(x, y, t) = 0 & \textit{on } \partial\Omega \times (0, T], \\ u(x, y, 0) = \sin(\pi x)\sin(\pi y) & \textit{in } \Omega, \end{cases}$$

where f is chosen such that $u = \sin(\pi x)\sin(\pi y)e^t$ is the exact solution, T is chosen as t^s since we only compute the solution for one time window consisting of s time steps. The process can be repeated to cover a longer time interval.

In the first example we set $h_1 = 1/32$, $h_2 = 1/64$, $h_3 = 1/128$, and $\tau_i = h_i$ ($i = 1, 2, 3$). In the experiment, we vary the window size, the subdomain partition, and the overlap size. All the subdomain problems are solved using LU. The numerical results in Table 1 show that increasing the overlap reduces the number of iterations, and the number of iterations is bounded and independent of the window size and the number of subdomains. Some numerical results for two-level space-time algorithm are also given. From Table 2, it is clear that the number of iterations decreases when increasing the overlap and is bounded independently of s and N_p. We see that the three-level algorithm is more expensive than the two-level algorithm in terms of the cost per iteration. But the number of iterations for the three-level algorithm is smaller.

In the next set of experiments, we fix the mesh size $h_1 = 1/16$, $h_2 = 1/32$, $h_3 = 1/64$, the terminal time $T = 1.0$, and change the time step size $\tau_i = h_i/k$ ($i = 1, 2, 3, k = 1, 2, 3, 4$). In this case, the window size s increases with the decrease of

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

TABLE 3
The number of GMRES iterations for Example 4.1 *by Algorithm* 2.1 *(three level) with* $T = 1.0$, $h_1 = 1/16$, $h_2 = 1/32$, $h_3 = 1/64$, $\tau_1 = h_1/k$, $\tau_2 = h_2/k$, *and* $\tau_3 = h_3/k$ $(k = 1, 2, 3, 4)$.

	$N_p = 64$	$N_p = 128$	$N_p = 256$
$s = 64$	40	42	48
$s = 128$	34	36	36
$s = 256$	30	31	32
$s = 512$	28	29	29

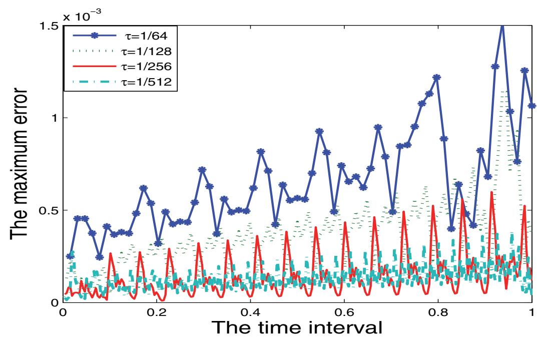

FIG. 2. *The maximum norm of the error for Example* 4.1 *by Algorithm* 2.1 *(three level) with* $T = 1.0$, $N_p = 256$, $h_1 = 1/16$, $h_2 = 1/32$, $h_3 = 1/64$, $\tau_1 = h_1/k$, $\tau_2 = h_2/k$, *and* $\tau_3 = h_3/k$ $(k = 1, 2, 3, 4)$.

τ since $s\tau = 1.0$. The numerical results listed in Table 3 show that the number of iterations decreases when increasing the window size s and is bounded independently of the number of subdomains. From Figure 2, we see that the error in the maximum norm of the solution goes down when the time step size τ decreases from $1/64$ to $1/512$.

Example 4.2. Consider a convection diffusion equation

$$\begin{cases} u_t - \varepsilon\triangle u + \mathbf{b}\cdot\nabla u = f & in\ \Omega\times(0,T], \\ u(x,y,0) = \frac{1}{0.8\varepsilon}e^{-(x^2+y^2)/(0.8\varepsilon)} & in\ \Omega, \end{cases}$$

where $\mathbf{b} = (1,1)$, T is chosen as t^s, $f = -\frac{x+y}{8\varepsilon^2(t+0.2)^2}e^{-(x^2+y^2)/(4\varepsilon(t+0.2))}$, and the analytical solution is

$$u(x,y,t) = \frac{1}{4\varepsilon(t+0.2)}e^{-(x^2+y^2)/(4\varepsilon(t+0.2))}$$

with consistent Dirichlet boundary conditions.

Similarly to Example 4.1, we first set $h_1 = 1/32$, $h_2 = 1/64$, $h_3 = 1/128$, and $\tau_i = h_i$ $(i = 1, 2, 3)$, and vary the window size, the subdomain partition, and the overlap size. From Table 4, we observe that the number of iterations decreases when increasing the overlap and is bounded independently of the window size and the number of subdomains. Comparing the numerical results in Tables 4 and 5, we see that the number of iterations for the three-level algorithm is fewer than that of the two-level algorithm when the overlap size is zero, and it becomes almost the same when we increase the overlap size.

Next, we fix the mesh size $h_1 = 1/16$, $h_2 = 1/32$, $h_3 = 1/64$, the terminal time $T = 1.0$, and vary the time step sizes. The numerical results listed in Table 6 show

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 4
The number of GMRES iterations for Example 4.2 *by Algorithm* 2.1 *(three level) with* $h_1 = \tau_1 = 1/32$, $h_2 = \tau_2 = 1/64$, *and* $h_3 = \tau_3 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	36	29	26	41	29	26	41	38	32
32	34	26	23	37	29	26	37	29	27
64	29	23	21	33	26	23	36	29	26
96	28	23	21	33	26	23	32	27	24
128	28	22	21	28	24	21	31	26	24

TABLE 5
The number of GMRES iterations for Example 4.2 *by Algorithm* 2.1 *(two level) with* $h_1 = \tau_1 = 1/32$ *and* $h_2 = \tau_2 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	41	28	35	45	28	26	51	36	30
32	44	26	24	47	28	27	47	27	26
64	37	23	22	43	26	24	46	28	26
96	36	23	22	43	27	24	43	26	25
128	35	23	21	37	23	22	41	26	24

TABLE 6
The number of GMRES iterations for Example 4.2 *by Algorithm* 2.1 *(three level) with* $T = 1.0$, $h_1 = 1/32$, $h_2 = 1/64$, $h_3 = 1/128$, $\tau_1 = h_1/k$, $\tau_2 = h_2/k$, *and* $\tau_3 = h_3/k$ $(k = 1, 2, 3, 4)$.

	$N_p = 64$	$N_p = 128$	$N_p = 256$
$s = 64$	27	28	32
$s = 128$	22	23	25
$s = 256$	23	23	23
$s = 512$	23	24	24

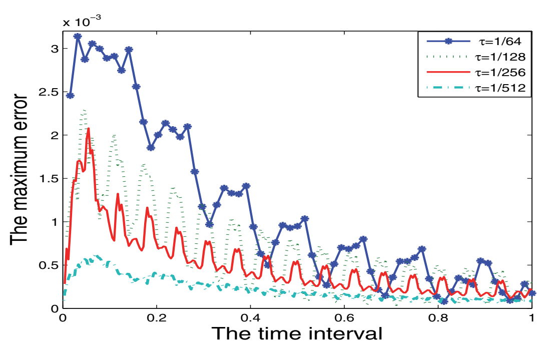

FIG. 3. *The maximum norm of the error for Example* 4.2 *by Algorithm* 2.1 *with* $T = 1.0$, $N_p = 256$, $h_1 = 1/16$, $h_2 = 1/32$, $h_3 = 1/64$, $\tau_1 = h_1/k$, $\tau_2 = h_2/k$, *and* $\tau_3 = h_3/k (k = 1, 2, 3, 4)$.

that the number of iterations does not change much with the number of subdomains and is bounded independently of the window size. Moreover, from Figure 3, it is clear that the error in the maximum norm of the solution decreases when choosing smaller time step size.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 7

The number of GMRES iterations for Example 4.3 *by Algorithm* 2.1 *(three level) with* $h_1 = \tau_1 = 1/32$, $h_2 = \tau_2 = 1/64$, *and* $h_3 = \tau_3 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	36	27	25	36	28	25	37	32	29
32	37	25	25	36	27	26	36	28	25
64	45	25	24	37	25	25	36	27	26
96	45	24	23	37	25	24	37	26	25
128	45	24	23	45	25	24	37	25	25

TABLE 8

The number of GMRES iterations for Example 4.3 *by Algorithm* 2.1 *(two level) with* $h_1 = \tau_1 = 1/32$ *and* $h_2 = \tau_2 = 1/128$.

N_p	64			128			256		
s \ ovlp	0	1	2	0	1	2	0	1	2
16	42	25	24	43	27	25	46	31	28
32	43	26	24	43	27	24	42	27	25
64	43	26	23	43	26	24	43	27	24
96	43	26	23	43	26	24	42	26	24
128	43	26	23	43	26	23	43	26	24

TABLE 9

The number of GMRES iterations for Example 4.3 *by Algorithm* 2.1 *(three level) with* $T = 1.0$, $h_1 = 1/32$, $h_2 = 1/64$, $h_3 = 1/128$, $\tau_1 = h_1/k$, $\tau_2 = h_2/k$, *and* $\tau_3 = h_3/k (k = 1, 2, 3, 4)$.

	$N_p = 64$	$N_p = 128$	$N_p = 256$
$s = 64$	29	30	33
$s = 128$	27	27	27
$s = 256$	27	27	27
$s = 512$	27	28	28

Example 4.3. Consider an advection diffusion reaction equation

$$\begin{cases} u_t - \triangle u + \mathbf{b} \cdot \nabla u + cu = 0 & \text{in } \Omega \times (0, T], \\ u(x, y, t) = 0 & \text{on } \partial\Omega \times (0, T], \\ u(x, y, 0) = e^{-10\sqrt{(x-0.5)^2 + (y-0.5)^2}} & \text{in } \Omega, \end{cases}$$

where $\mathbf{b} = (1, 1)$, $c = 1$, and T is chosen as t^s.

We test the third example which doesn't have an analytic solution. The numerical results of the three-level and two-level algorithms are given in Tables 7 and 8, respectively. We observe that the number of iterations decreases when the overlap size increases and is bounded independently of the window size and the number of subdomains. Moreover, the results in Table 9 also show that the number of iterations does not change much with the window size and the number of subdomains. This confirms that optimal convergence of Algorithm 2.1 is as indicated by the theory.

4.2. Parallel performance of the algorithm. In this subsection, we investigate the parallel performance of the algorithm in terms of the total compute time, and the strong/weak scalability. All experiments are carried out on the Tianhe-2 supercomputer located at the National Supercomputer Center in Guangzhou. The machine ranks #2 in the current Top500 list. Only CPUs are used in the calculations, without the accelerators.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 10
Strong scaling results of Algorithm 2.1 (three level) on Tianhe-2 for Examples 4.1–4.3 with $h_1 = 1/128$, $h_2 = 1/256$, $h_3 = 1/512$, $\tau_1 = 1/64$, $\tau_2 = 1/128$, *and* $\tau_3 = 1/256$.

N_p	Example 4.1		Example 4.2		Example 4.3	
	GMRES	Time (s)	GMRES	Time (s)	GMRES	Time (s)
512	26	180.5	24	178.1	22	176.3
1024	25	34.0	23	34.2	22	32.8
2048	29	29.4	29	29.9	25	28.7
3072	31	14.9	29	14.5	25	14.0
4096	32	10.2	28	8.7	25	8.7

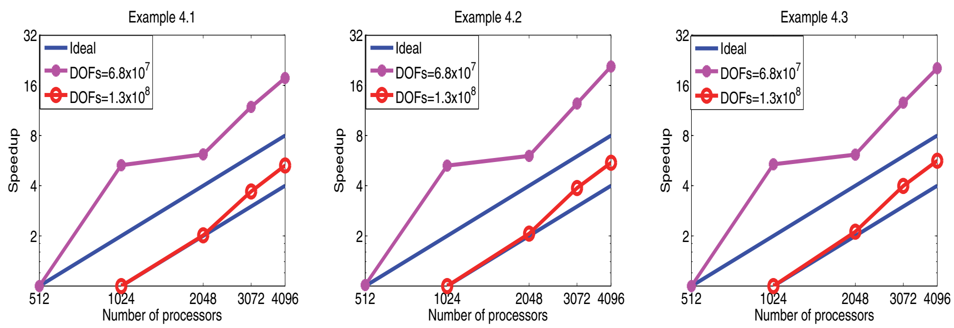

FIG. 4. *Strong scaling results of Algorithm 2.1 (three level) on Tianhe-2 for Examples 4.1–4.3.*

TABLE 11
Strong scaling results of Algorithm 2.1 (three level) on Tianhe-2 for Examples 4.1–4.3 with $h_1 = \tau_1 = 1/128$, $h_2 = \tau_2 = 1/256$, *and* $h_3 = \tau_3 = 1/512$.

N_p	Example 4.1		Example 4.2		Example 4.3	
	GMRES	Time (s)	GMRES	Time (s)	GMRES	Time (s)
512	-	-	-	-	-	-
1024	27	183.1	24	174.3	22	176.1
2048	36	91.0	28	85.9	24	83.1
3072	36	49.6	29	45.9	25	44.2
4096	36	34.6	29	32.4	25	31.2

In the strong scaling test, we first fix the terminal time $T = 1.0$ and set the finest mesh size $h = 1/512$ and $\tau = 1/256$ (the system has about 68 million unknowns) and increase the number of processors from 512 to 4096. The results for Examples 4.1–4.3 are listed in Table 10. It shows that the number of iterations changes slightly with the increase of the number of processors. And the total compute time decreases a lot when we increase the number of processors from 512 to 1024, but the speedup is not as good when we go from 1024 to 2048, etc. In Figure 4, "DOFs" denotes the total degrees of freedom. It shows that the proposed algorithm has superlinear speedup up to 4096 processors when $DOFs = 6.8 \times 10^7$. Next, we increase the problem size by setting the finest mesh size $h = 1/512$ and $\tau = 1/512$, so the DOFs is about 134 million. From Table 11, we observe that the number of iterations for Example 4.1 increases greatly when the number of processors changes from 1024 to 2048, but it does not change for later cases. For Examples 4.2 and 4.3, it is clear that the number of iterations changes only slightly with the increase of the number of processors. Moreover, Figure 4 also shows that Algorithm 2.1 has superlinear speedup up to 4096 processors for solving Examples 4.1–4.3.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Table 12
Weak scaling results of Algorithm 2.1 *(three level) on Tianhe*-2 *for Examples* 4.1–4.3 *with* $h_1 = \tau_1 = 1/128$*,* $h_2 = \tau_2 = 1/256$*, and* $h_3 = \tau_3 = 1/512$*.*

			Example 4.1		Example 4.2		Example 4.3	
N_p	s	$DOFs$	GMRES	Time (s)	GMRES	Time (s)	GMRES	Time (s)
256	32	8388608	30	28.5	31	29.0	30	28.6
512	64	16777216	26	26.9	29	27.9	25	26.6
1024	128	33554432	27	27.7	29	28.3	25	27.2
2048	256	67108864	29	29.4	29	29.6	25	28.7
4096	512	134217728	36	34.6	29	32.0	25	30.9

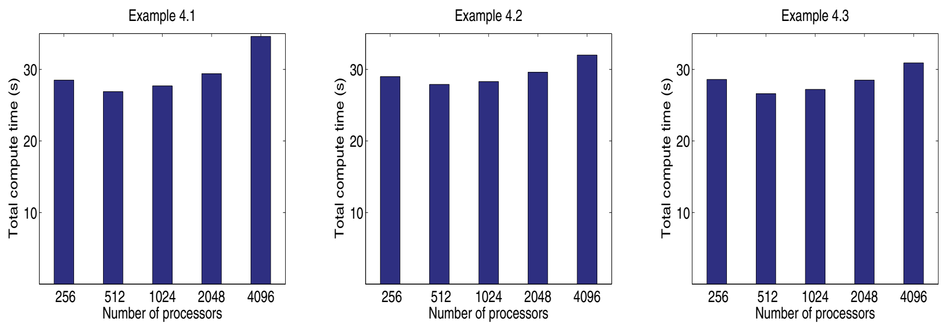

Fig. 5. *Weak scaling results of Algorithm* 2.1 *(three level) on Tianhe*-2 *for Examples* 4.1–4.3*.*

For problems requiring a large number of time steps, the window size s may be large. Hence, we consider the weak scaling of Algorithm 2.1 in term of s, i.e., we fix the mesh size and the time step size, and vary the window size and the number of the processors, with a fixed problem size on each processor. The results for Examples 4.1–4.3 are presented in Table 12 and Figure 5. It is clear that the number of iterations of Example 4.1 changes a little but is bounded independently of the number of the processors and the window size, and there is almost no change for Examples 4.2 and 4.3. Moreover, the total compute time changes only slightly with the increase of N_p and s. This indicates that the proposed algorithm has a very good weak scalability.

4.3. Comparison with a traditional parallel time stepping method in three dimensions (3D). In this subsection, we investigate the performance of the proposed algorithm for a three-dimensional problem and also compare the algorithm with a traditional time stepping method parallelized only in space.

Example 4.4. This is an extension of Example 4.1 to 3D, where $\varepsilon = 0.1$, $\Omega = [0,1] \times [0,1] \times [0,1]$, T is chosen as t^s, $f = -\sin(\pi x)\sin(\pi y)\sin(\pi z)(\sin t - 3\varepsilon\pi^2 \cos t)$, and the analytical solution is $u(x,y,t) = \sin(\pi x)\sin(\pi y)\sin(\pi z)\cos t$.

We test this example on a $65 \times 65 \times 65 \times 2048$ space-time mesh. We fix $h_1 = 1/32$, $h_2 = 1/64$, and $\tau_i = h_i$ $(i = 1, 2)$, and increase the number of processors from 8 to 8192. The space-time domain $[0,1]^3 \times [0,32]$ is divided into $N_p = P_x \times P_y \times P_z \times P_t$ subdomains and each subdomain is assigned to one processor. All the subdomain problems are solved using ILU except that LU is used on the coarse mesh. Because the processor distribution impacts the ratio of computation and communication, different arrangement of the processors may lead to different parallel performance. In the experiment, we distribute the spatial domain evenly such that each processor holds approximately a cube in space, i.e., $P_x = P_y = P_z$. Otherwise, we omit some cases

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

TABLE 13

The averaged number of iterations and the total compute time for solving Example 4.4 *on a* $65 \times 65 \times 65 \times 2048$ *space-time mesh using Algorithm* 2.1 *(corresponding to rows with* $s > 1$*) and the traditional time stepping method parallelized only in space with a two-level additive Schwarz method with* $h_1 = \tau_1 = 1/32$ *and* $h_2 = \tau_2 = 1/64$ *(corresponding to the row with* $s = 1$*).*

N_p		8	16	32	64	128	256	512	1024	2048	4096	8192
$s = 1$	GMRES	4.6	-	-	4.7	-	-	6.7	-	-	4.7	3.7
	Time(s)	340.8	-	-	65.7	-	-	43.3	-	-	51.6	51.7
$s = 2$	GMRES	6.0	-	-	5.5	-	-	5.5	-	-	5.3	4.8
	Time(s)	422.5	-	-	67.3	-	-	42.5	-	-	46.2	47.8
$s = 4$	GMRES	6.3	-	-	6.2	-	-	6.6	-	-	6.5	7.2
	Time(s)	457.2	-	-	67.9	-	-	27.6	-	-	31.4	40.5
$s = 8$	GMRES	*	6.5	6.5	6.5	6.6	6.5	6.7	7.0	7.0	6.6	6.7
	Time(s)	*	247.2	121.6	69.2	40.9	25.5	22.9	21.6	20.0	23.2	29.7
$s = 32$	GMRES	*	6.8	7.0	7.2	7.3	6.8	7.2	6.9	7.2	7.3	7.3
	Time(s)	*	284.5	146.7	77.1	42.7	26.8	20.2	17.3	16.8	13.4	16.6
$s = 64$	GMRES	*	7.3	7.3	7.4	7.4	7.4	7.2	7.4	7.2	7.4	7.5
	Time(s)	*	341.7	171.9	88.1	47.9	26.7	20.2	15.0	11.0	10.8	15.2
$s = 256$	GMRES	*	*	7.6	7.6	7.7	7.8	7.8	7.8	7.7	7.8	7.7
	Time(s)	*	*	311.4	156.7	84.1	45.1	29.1	16.2	13.9	10.2	13.9
$s = 2048$	GMRES	*	*	*	*	*	8.0	8.0	8.0	9.0	9.0	9.0
	Time(s)	*	*	*	*	*	220.5	117.7	66.7	37.5	20.8	12.2

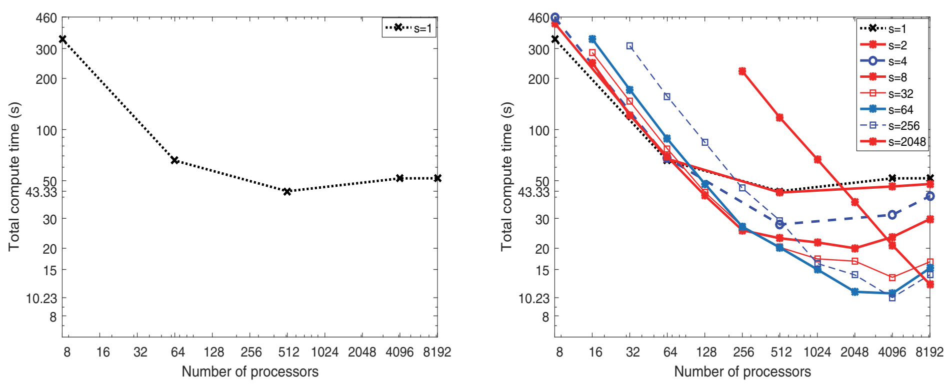

FIG. 6. *The total compute time for solving Example* 4.4 *using sequential time stepping (left) and Algorithm* 2.1 *with different window size* s *(right).*

and denote it by "-" in Table 13. "*" indicates that the number of processors is too small to solve the problem.

In Table 13, we present the compute time and the number of iterations. Note that the number of iterations is averaged, i.e., we sum all the iteration counts in each window and divide by the number of windows ($2048/s$). The sequential time stepping algorithm corresponds to the case $s = 1$ in Algorithm 2.1. From Table 13, we see that the average number of iterations is bounded independently of the number of processors for each window size s. The total compute time of the traditional time stepping method ($s = 1$) first goes down fast and then increases a little with the number of processors. It indicates that the cost of communication becomes dominate when the number of processors is large. The strong scalability of the sequential time stepping method is not good when the number of processors is large. As shown in Figure 6, for the cases $s = 2, 4, 8, 32, 64, 256$, the total compute time curves are

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

similar to the case of $s = 1$, but the total compute time is less than that required by the sequential time stepping method. The compute time curves also show that the crossover point for which it becomes beneficial to use Algorithm 2.1, compared with the sequential time stepping method. For examples, the crossover point is at about 64 processors for the cases $s = 2, 4, 8, 32$, and 128 processors for $s = 64$ and 256 processors for $s = 256$. For the case of $s = 2048$, i.e., solve the global problem all at once, we observe that the crossover point is at about 2048 processors. In this case, the compute time decreases very fast and the speedup is linear.

5. Concluding remarks. We introduced and studied a multilevel space-time additive Schwarz method for solving parabolic equations. Under certain conditions, we proved that the convergence of the method is only dependent on the ratio between the subdomain size and the overlap size on each level, but is bounded independently of the mesh sizes, the number of subdomains, the window size, and the number of levels. The numerical experiments confirm the convergence estimates. Moreover, experiments carried out on a supercomputer show that our method is robust and scalable, measured in total compute time, with respect to the number of processors, mesh sizes, and the window size. Our experiments also show that the new space-time additive Schwarz method outperforms the sequential time stepping method parallelized only in space when the number of processors is large.

Acknowledgments. We would like to thank the anonymous referees for their suggestions.

REFERENCES

[1] G. Bal and Y. Maday, *A "parareal" time discretization for non-linear PDE's with application to the pricing of an American put*, in Recent Developments in Domain Decomposition Methods, Lect. Notes Comput. Sci. Eng. 23, Springer, Berlin, 2001, pp. 189–202.

[2] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune. K. Buschelman, L. Dalcin, V. Eijkhout, D. Kaushik, M. G. Knepley, L. C. McInnes, W. D. Gropp, K. Rupp, B. F. Smith, S. Zampini, and H. Zhang, *PETSc Users Manual*, Technical report ANL 95/11 - Revision 3.7, Argonne National Laboratory, Argonne, IL, 2017.

[3] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts Appl. Math. 15, Springer, New York, 2008.

[4] X.-C. Cai, *Some Domain Decomposition Algorithms for Nonselfadjoint Elliptic and Parabolic Partial Differential Equations*, Ph.D. thesis, Courant Institute, New York, 1989.

[5] X.-C. Cai, *Additive Schwarz algorithms for parabolic convection-diffusion equations*, Numer. Math., 60 (1991), pp. 41–61.

[6] X.-C. Cai, *Multiplicative Schwarz methods for parabolic problems*, SIAM J. Sci. Comput., 15 (1994), pp. 587–603.

[7] Y. Cao, M. Gunzburger, X. He, and X. Wang, *Parallel, non-iterative, multi-physics domain decomposition methods for time-dependent Stokes-Darcy systems*, Math. Comp., 83 (2014), pp. 1617–1644.

[8] A. J. Christlieb, R. D. Haynes, and B. W. Ong, *A parallel space-time algorithm*, SIAM J. Sci. Comput., 34 (2012), pp. C233–C248.

[9] C. Cong, X.-C. Cai, and K. Gustafson, *Implicit space-time domain decomposition methods for stochastic parabolic partial differential equations*, SIAM J. Sci. Comput., 36 (2014), pp. C1–C24.

[10] X. Deng, X.-C. Cai, and J. Zou, *A parallel space-time domain decomposition method for some unsteady source inversion problems*, Inverse Probl. Imaging, 9 (2015), pp. 1069–1091.

[11] X. Deng, X.-C. Cai, and J. Zou, *Two-level space-time domain decomposition methods for unsteady inverse problems*, J. Sci. Comput., 67 (2016), pp. 860–882.

[12] M. Dryja and X. M. Tu, *A domain decomposition discretization of parabolic problems*, Numer. Math., 107 (2007), pp. 625–640.

[13] M. Dryja and O. B. Widlund, *Domain decomposition algorithms with small overlap*, SIAM J. Sci. Comput., 15 (1994), pp. 604–620.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 10/24/18 to 128.138.244.88. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

[14] S. C. Eisenstat, H. C. Elman, and M. H. Schultz, *Variational iterative methods for nonsymmetric systems of linear equations*, SIAM J. Numer. Anal., 20 (1983), pp. 345–357.

[15] R. D. Falgout, S. Friedhoff, Tz. V. Kolev, S. P. MacLachlan, and J. B. Schroder, *Parallel time integration with multigrid*, SIAM J. Sci. Comput., 36 (2014), pp. C635–C661.

[16] C. Farhat and M. Chandesris, *Time-decomposed parallel time-integrators: Theory and feasibility studies for fluid, structure, and fluid-structure applications*, Internat. J. Numer. Methods Engrg., 58 (2003), pp. 1397–1434.

[17] M. J. Gander and L. Halpern, *Optimized Schwarz waveform relaxation methods for advection reaction diffusion problems*, SIAM J. Numer. Anal., 45 (2007), pp. 666–697.

[18] M. J. Gander and M. Neumüller, *Analysis of a new space-time parallel multigrid algorithm for parabolic problems*, SIAM J. Sci. Comput., 38 (2016), pp. A2173–A2208.

[19] M. J. Gander and A. M. Stuart, *Space-time continuous analysis of waveform relaxation for the heat equation*, SIAM J. Sci. Comput., 19 (1998), pp. 2014–2031.

[20] M. J. Gander and S. Vandewalle, *Analysis of the parareal time-parallel time-integration method*, SIAM J. Sci. Comput., 29 (2007), pp. 556–578.

[21] G. Horton, *The time-parallel multigrid method*, Comm. Appl. Numer. Methods, 8 (1992), pp. 585–595.

[22] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16 (1995), pp. 848–864.

[23] E. Lelarasmee, A. E. Ruehli, and A. L. Sangiovanni-Vincentelli, *The waveform relaxation method for time-domain analysis of large-scale integrated circuits*, IEEE Trans. Comput-Aided Des., CAD-1 (1982), pp. 131–145.

[24] S. Li and X.-C. Cai, *Convergence analysis of two-level space-time additive Schwarz method for parabolic equations*, SIAM J. Numer. Anal., 53 (2015), pp. 2727–2751.

[25] J.-L. Lions, Y. Maday, and G. Turinici, *Résolution d'EDP par un schéma en temps pararéel*, C. R. Math. Acad. Sci. Paris Sér. I, 332 (2001), pp. 661–668.

[26] M. L. Minion, R. Speck, M. Bolten, M. Emmett, and D. Ruprecht, *Interweaving PFASST and parallel multigrid*, SIAM J. Sci. Comput., 37 (2015), pp. S244–S263.

[27] D. Ruprecht and R. krause, *Explicit parallel-in-time integration of linear acoustic-advection system*, Comput. & Fluids, 59 (2012), pp. 72–83.

[28] M. Sarkis and D. B. Szyld, *Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 1612–1621.

[29] B. Smith, P. Bjørstad, and W. Gropp, *Domain Decomposition: Parallel Multilevel Methods for Elliptic Partial Differential Equations*, Cambridge University Press, Cambridge, 1996.

[30] A. Toselli and O. B. Widlund, *Domain Decomposition Methods-Algorithms and Theory*, Springer-Verlag, Berlin, 2005.

[31] M.-B. Tran, *Parallel Schwarz waveform relaxation algorithm for an n-dimensional semilinear heat equation*, ESAIM Math. Model. Numer. Anal., 48 (2014), pp. 795–813.

[32] T. Weinzierl and T. Koppl, *A geometric space-time multigrid algorithm for the heat equation*, Numer. Math. Theory Methods Appl., 5 (2012), pp. 110–130.

[33] H. Yang and X.-C. Cai, *Two-level space-time domain decomposition methods for flow control problems*, J. Sci. Comput., 70 (2017), pp. 717–743.

[34] H. Yu, *A local space-time adaptive scheme in solving two-dimensional parabolic problems based on domain decomposition methods*, SIAM J. Sci. Comput., 23 (2001), pp. 304–322.

[35] X. Zhang, *Multilevel Schwarz methods*, Numer. Math., 63 (1992), pp. 521–539.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.