MULTIGRID METHODS FOR SADDLE POINT PROBLEMS: OPTIMALITY SYSTEMS

SUSANNE C. BRENNER, SIJING LIU, AND LI-YENG SUNG

Abstract. We develop multigrid methods for an elliptic distributed optimal control problem on convex domains that are robust with respect to a regularization parameter. We prove the uniform convergence of the W-cycle algorithm and demonstrate the performance of V-cycle and W-cycle algorithms through numerical experiments.

1. Introduction

Let Ω be a bounded convex polygonal/polyhedral domain in $\mathbb{R}^d$ $(d = 2, 3)$, $y_d \in L_2(\Omega)$, $\beta \in (0, 1]$ be a constant and $(\cdot, \cdot)_{L_2(\Omega)}$ be the inner product of $L_2(\Omega)$ (or $[L_2(\Omega)]^d$). The optimal control problem is to find

$$(\bar{y}, \bar{u}) = \underset{(y,u)\in K}{\operatorname{argmin}} \left[\frac{1}{2}\|y - y_d\|^2_{L_2(\Omega)} + \frac{\beta}{2}\|u\|^2_{L_2(\Omega)}\right], \tag{1.1}$$

where (y, u) belongs to $K \subset H^1_0(\Omega) \times L_2(\Omega)$ if and only if

$$a(y, z) = (u, z)_{L_2(\Omega)} \quad \forall z \in H^1_0(\Omega), \tag{1.2}$$

and the bilinear form $a(\cdot, \cdot)$ is given by

$$a(y, z) = \int_\Omega \nabla y \cdot \nabla z \, dx + \int_\Omega \big[(\boldsymbol{\zeta} \cdot \nabla y)z - (\boldsymbol{\zeta} \cdot \nabla z)y\big] \, dx + \int_\Omega \gamma yz \, dx. \tag{1.3}$$

Here the vector field $\boldsymbol{\zeta}$ belongs to $[W^{1,\infty}(\Omega)]^d$ and the function $\gamma \in L_\infty(\Omega)$ is nonnegative.

Remark 1.1. Throughout the paper we will follow the standard notation for differential operators, function spaces and norms that can be found for example in [13, 10].

Remark 1.2. The partial differential equation (PDE) constraint (1.2) is the weak form of a second order elliptic boundary value problem with an advective/convective term. The bilinear form $a(\cdot, \cdot)$ is nonsymmetric (unless $\boldsymbol{\zeta} = \mathbf{0}$) and it is definite because

$$a(y, y) = \int_\Omega \big(|\nabla y|^2 + \gamma|y|^2\big) \, dx. \tag{1.4}$$

Date: February 10, 2020.

1991 *Mathematics Subject Classification.* 49J20, 65N30, 65N55.

Key words and phrases. multigrid methods, elliptic distributed optimal control problems, saddle point problems.

This work was supported in part by the National Science Foundation under Grant Nos. DMS-16-20273 and DMS-19-13035.

The optimal control problem (1.1)–(1.2) has a unique solution characterized by the following first order optimality system (cf. [22, 28, 19]):

$$a(q,\bar{p}) = (q,\bar{y}-y_d)_{L_2(\Omega)} \qquad \forall\, q \in H_0^1(\Omega), \tag{1.5a}$$

$$\beta\bar{u} + \bar{p} = 0, \tag{1.5b}$$

$$a(\bar{y},z) = (\bar{u},z)_{L_2(\Omega)} \qquad \forall\, z \in H_0^1(\Omega), \tag{1.5c}$$

where $\bar{p}$ is the adjoint state. After eliminating $\bar{u}$, we arrive at a saddle point problem:

$$a(q,\bar{p}) - (q,\bar{y})_{L_2(\Omega)} = -(q,y_d)_{L_2(\Omega)} \qquad \forall\, q \in H_0^1(\Omega), \tag{1.6a}$$

$$-(\bar{p},z)_{L_2(\Omega)} - \beta a(\bar{y},z) = 0 \qquad \forall\, z \in H_0^1(\Omega). \tag{1.6b}$$

Note that the system (1.6) is unbalanced with respect to β since it only appears in (1.6b). This can be remedied by the following change of variables:

$$\bar{p} = \beta^{\frac{1}{4}}\tilde{p} \quad \text{and} \quad \bar{y} = \beta^{-\frac{1}{4}}\tilde{y}. \tag{1.7}$$

The resulting problem is

$$\beta^{\frac{1}{2}}a(q,\tilde{p}) - (q,\tilde{y})_{L_2(\Omega)} = -\beta^{\frac{1}{4}}(q,y_d)_{L_2(\Omega)} \qquad \forall\, q \in H_0^1(\Omega), \tag{1.8a}$$

$$-(\tilde{p},z)_{L_2(\Omega)} - \beta^{\frac{1}{2}}a(\tilde{y},z) = 0 \qquad \forall\, z \in H_0^1(\Omega). \tag{1.8b}$$

The saddle point problem (1.8) can be discretized by a P_1 finite element method (cf. Section 2). Our goal is to design multigrid methods for the resulting discrete saddle point problem whose performance is independent of the regularization parameter β. The key idea is to use a post-smoother that can be interpreted as a Richardson iteration for a symmetric positive definite (SPD) problem that has the same solution as the saddle point problem. Consequently we can exploit the well-known multigrid theory for SPD problems [18, 23, 6] in our convergence analysis. This idea has previously been applied to other saddle point problems in [7, 8, 9].

Our multigrid methods belong to the class of all-at-once methods where all the unknowns in (1.6) are solved simultaneously (cf. [4, 16, 26, 5, 27] and the references therein). As far as we know, the multigrid methods in this paper are the first ones that are provably robust with respect to the regularization parameter β when the elliptic PDE constraint (1.2) involves an advection/convection term.

In the case where $\boldsymbol{\zeta} = \mathbf{0}$, multigrid methods that are robust with respect to β can also be found in the papers [26, 27]. The differences are in the construction of the smoothers and in the norms that measure the convergence of the multigrid algorithms. The smoothing steps in [26, 27] are computationally less expensive than the one in the current paper, which requires solving (approximately) a reaction-diffusion problem (which however does not affect the $O(n)$ complexity). The trade-off is that the convergence of the multigrid algorithm in this paper is expressed in terms of the natural energy norm for the continuous problem, while the norms in [26, 27] are different from the energy norm. A related consequence is that the W-cycle multigrid algorithms in [26, 27] cannot take advantage of post-smoothing and hence their contraction numbers decay at the rate of $O(m^{-1/2})$, where m is the number of pre-smoothing steps, while the contraction number for our symmetric W-cycle multigrid algorithm decays at the rate of $O(m^{-1})$, where m is the number of pre-smoothing and post-smoothing steps. Moreover numerical results indicate that our V-cycle and W-cycle algorithms converge uniformly for $m = 1$.

The rest of the paper is organized as follows. We analyze the saddle point problem (1.8) and the P_1 finite element method in Section 2 and introduce the multigrid algorithms in Section 3. We derive smoothing and approximation properties in Section 4 that are the key ingredients for the convergence analysis of the W-cycle algorithm in Section 5. Numerical results are presented in Section 6 and we end with some concluding remarks in Section 7.

Throughout this paper, we use C (with or without subscripts) to denote a generic positive constant that is independent of β and any mesh parameter. Also to avoid the proliferation of constants, we use the notation $A \lesssim B$ (or $A \gtrsim B$) to represent $A \leq (\text{constant})B$, where the (hidden) positive constant is independent of β and any mesh parameter, but may depend on $\boldsymbol{\zeta}$. The notation $A \approx B$ is equivalent to $A \lesssim B$ and $B \lesssim A$.

2. P_1 Finite Element Methods

We can express (1.8) concisely as

$$\mathcal{B}((\tilde{p}, \tilde{y}), (q, z)) = -\beta^{\frac{1}{4}}(q, y_d)_{L_2(\Omega)} \qquad \forall\, (q, z) \in H_0^1(\Omega) \times H_0^1(\Omega), \tag{2.1}$$

where

$$\mathcal{B}\big((p, y), (q, z)\big) = \beta^{\frac{1}{2}} a(q, p) - (q, y)_{L_2(\Omega)} - (p, z)_{L_2(\Omega)} - \beta^{\frac{1}{2}} a(y, z). \tag{2.2}$$

2.1. **Properties of $\mathcal{B}$.** We will analyze the bilinear form $\mathcal{B}(\cdot, \cdot)$ in terms of the energy norm $\|\cdot\|_{H_\beta^1(\Omega)}$ defined by

$$\|v\|_{H_\beta^1(\Omega)}^2 = \|v\|_{L_2(\Omega)}^2 + \beta^{\frac{1}{2}} |v|_{H^1(\Omega)}^2 \qquad \forall\, v \in H^1(\Omega). \tag{2.3}$$

Let $(p, y), (q, z) \in H_0^1(\Omega) \times H_0^1(\Omega)$ be arbitrary. It follows immediately from (1.3), (2.2), (2.3) and the Cauchy-Schwarz inequality that

$$\mathcal{B}((p, y), (q, z)) \lesssim (\|p\|_{H_\beta^1(\Omega)}^2 + \|y\|_{H_\beta^1(\Omega)}^2)^{\frac{1}{2}} (\|q\|_{H_\beta^1(\Omega)}^2 + \|z\|_{H_\beta^1(\Omega)}^2)^{\frac{1}{2}}. \tag{2.4}$$

Moreover, a direct calculation using (1.4) and (2.2) shows that

$$\begin{aligned} \mathcal{B}((p, y), (p - y, -y - p)) &= \beta^{\frac{1}{2}} a(p, p) + (p, p)_{L_2(\Omega)} + \beta^{\frac{1}{2}} a(y, y) + (y, y)_{L_2(\Omega)} \\ &\geq \|p\|_{H_\beta^1(\Omega)}^2 + \|y\|_{H_\beta^1(\Omega)}^2, \end{aligned} \tag{2.5}$$

and we also have

$$\|p - y\|_{H_\beta^1(\Omega)}^2 + \|-y - p\|_{H_\beta^1(\Omega)}^2 = 2(\|p\|_{H_\beta^1(\Omega)}^2 + \|y\|_{H_\beta^1(\Omega)}^2) \tag{2.6}$$

by the parallelogram law.

It follows from (2.4)–(2.6) that

$$\|p\|_{H_\beta^1(\Omega)} + \|y\|_{H_\beta^1(\Omega)} \approx \sup_{(q,z) \in H_0^1(\Omega) \times H_0^1(\Omega)} \frac{\mathcal{B}((p, y), (q, z))}{\|q\|_{H_\beta^1(\Omega)} + \|z\|_{H_\beta^1(\Omega)}} \tag{2.7}$$

for all $(p, y) \in H_0^1(\Omega) \times H_0^1(\Omega)$.

Similarly, we have

$$\|p\|_{H^1_\beta(\Omega)} + \|y\|_{H^1_\beta(\Omega)} \approx \sup_{(q,z)\in H^1_0(\Omega)\times H^1_0(\Omega)} \frac{\mathcal{B}((q,z),(p,y))}{\|q\|_{H^1_\beta(\Omega)} + \|z\|_{H^1_\beta(\Omega)}} \tag{2.8}$$

for all $(p,y) \in H^1_0(\Omega) \times H^1_0(\Omega)$.

2.2. **Discrete Problems.** Let $\mathcal{T}_h$ be a triangulation of Ω and $V_h \subset H^1_0(\Omega)$ be the P_1 finite element space associated with $\mathcal{T}_h$. The P_1 finite element method for (2.1) is to find $(\tilde{p}_h, \tilde{y}_h) \in V_h \times V_h$ such that

$$\mathcal{B}((\tilde{p}_h, \tilde{y}_h),(q_h,z_h)) = -\beta^{\frac{1}{4}}(q_h, y_d)_{L_2(\Omega)} \qquad \forall\,(q_h,z_h) \in V_h \times V_h. \tag{2.9}$$

For the convergence analysis of the multigrid algorithms, it is necessary to consider a more general problem: Find $(p,y) \in H^1_0(\Omega) \times H^1_0(\Omega)$ such that

$$\mathcal{B}((p,y),(q,z)) = (f,q)_{L_2(\Omega)} + (g,z)_{L_2(\Omega)} \qquad \forall\,(q,z) \in H^1_0(\Omega) \times H^1_0(\Omega), \tag{2.10}$$

where $f, g \in L_2(\Omega)$, together with the following dual problem: Find $(p,y) \in H^1_0(\Omega) \times H^1_0(\Omega)$ such that

$$\mathcal{B}((q,z),(p,y)) = (f,q)_{L_2(\Omega)} + (g,z)_{L_2(\Omega)} \qquad \forall\,(q,z) \in H^1_0(\Omega) \times H^1_0(\Omega). \tag{2.11}$$

The unique solvability of (2.10) (resp., (2.11) follows immediately from (2.7) (resp., (2.8)).

The P_1 finite element method for (2.10) is to find $(p_h, y_h) \in V_h \times V_h$ such that

$$\mathcal{B}((p_h,y_h),(q_h,z_h)) = (f,q_h)_{L_2(\Omega)} + (g,z_h)_{L_2(\Omega)} \qquad \forall\,(q_h,z_h) \in V_h \times V_h, \tag{2.12}$$

and the P_1 finite element method for (2.11) is to find $(p_h, y_h) \in V_h \times V_h$ such that

$$\mathcal{B}((q_h,z_h),(p_h,y_h)) = (f,q_h)_{L_2(\Omega)} + (g,z_h)_{L_2(\Omega)} \qquad \forall\,(q_h,z_h) \in V_h \times V_h. \tag{2.13}$$

Note that (2.4)–(2.6) also yield the following analog of (2.7):

$$\|p_h\|_{H^1_\beta(\Omega)} + \|y_h\|_{H^1_\beta(\Omega)} \approx \sup_{(q_h,z_h)\in V_h\times V_h} \frac{\mathcal{B}((p_h,y_h),(q_h,z_h))}{\|q_h\|_{H^1_\beta(\Omega)} + \|z_h\|_{H^1_\beta(\Omega)}} \qquad \forall\,(p_h,y_h) \in V_h \times V_h. \tag{2.14}$$

Similarly, we also have

$$\|p_h\|_{H^1_\beta(\Omega)} + \|y_h\|_{H^1_\beta(\Omega)} \approx \sup_{(q_h,z_h)\in V_h\times V_h} \frac{\mathcal{B}((q_h,z_h),(p_h,y_h))}{\|q_h\|_{H^1_\beta(\Omega)} + \|z_h\|_{H^1_\beta(\Omega)}} \qquad \forall\,(p_h,y_h) \in V_h \times V_h. \tag{2.15}$$

Therefore the discrete problems (2.12) and (2.13) are uniquely solvable.

2.3. **Error Estimates.** From (2.4), (2.14), (2.15) and the saddle point theory [2, 11, 30], we have the following quasi-optimal error estimate.

Lemma 2.1. *Let* (p,y) (*resp.,* (p_h,y_h)) *be the solution of* (2.10) *or* (2.11) (*resp.,* (2.12) *or* (2.13)). *We have*

$$\|p-p_h\|_{H^1_\beta(\Omega)} + \|y-y_h\|_{H^1_\beta(\Omega)} \lesssim \inf_{(q_h,z_h)\in V_h\times V_h} \big(\|p-q_h\|_{H^1_\beta(\Omega)} + \|y-z_h\|_{H^1_\beta(\Omega)}\big). \tag{2.16}$$

In order to convert (2.16) into a concrete error estimate, we need the regularity of the solutions of (2.10) and (2.11)

Lemma 2.2. *The solution* (p, y) *of* (2.10) *or* (2.11) *belongs to* $H^2(\Omega) \times H^2(\Omega)$ *and we have*

$$\|\beta^{\frac{1}{2}} p\|_{H^2(\Omega)} + \|\beta^{\frac{1}{2}} y\|_{H^2(\Omega)} \lesssim \|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}. \tag{2.17}$$

Proof. We will only consider (2.10) since the arguments for (2.11) are similar. In view of (2.2), we can write (2.10) as

$$\begin{aligned} a(q, \beta^{\frac{1}{2}} p) &= (y + f, q)_{L_2(\Omega)} && \forall\, q \in H_0^1(\Omega), \\ a(\beta^{\frac{1}{2}} y, z) &= (-p - g, z)_{L_2(\Omega)} && \forall\, z \in H_0^1(\Omega), \end{aligned}$$

and hence, by the elliptic regularity for convex domains [17, 14, 19],

$$\|\beta^{\frac{1}{2}} p\|_{H^2(\Omega)} \lesssim \|y\|_{L_2(\Omega)} + \|f\|_{L_2(\Omega)}, \tag{2.18a}$$

$$\|\beta^{\frac{1}{2}} y\|_{H^2(\Omega)} \lesssim \|p\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}. \tag{2.18b}$$

From (2.3), (2.7) and (2.10) we also have

$$\|p\|_{L_2(\Omega)} + \|y\|_{L_2(\Omega)} \lesssim \|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}. \tag{2.19}$$

The estimate (2.17) follows from (2.18) and (2.19). □

We can now derive concrete error estimates for the P_1 finite element methods.

Theorem 2.3. *Let* (p, y) *(resp.,* (p_h, y_h)*) be the solution of* (2.10) *or* (2.11) *(resp.,* (2.12) *or* (2.13)*). We have*

$$\|p - p_h\|_{H^1_\beta(\Omega)} + \|y - y_h\|_{H^1_\beta(\Omega)} \le C(1 + \beta^{\frac{1}{2}} h^{-2})^{\frac{1}{2}} \beta^{-\frac{1}{2}} h^2 (\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}), \tag{2.20}$$

$$\|p - p_h\|_{L_2(\Omega)} + \|y - y_h\|_{L_2(\Omega)} \le C(1 + \beta^{\frac{1}{2}} h^{-2}) \beta^{-1} h^4 (\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}), \tag{2.21}$$

where the positive constant C *is independent of* β *and* h.

Proof. We will only consider the case that involves (2.10) and (2.12). Let $\Pi_h : H^2(\Omega) \cap H_0^1(\Omega) \longrightarrow V_h$ be the nodal interpolation operator. We have the following standard interpolation error estimate [13, 10]:

$$\|\zeta - \Pi_h \zeta\|_{L_2(\Omega)} + h|\zeta - \Pi_h \zeta|_{H^1(\Omega)} \le C h^2 |\zeta|_{H^2(\Omega)} \qquad \forall\, \zeta \in H^2(\Omega) \cap H_0^1(\Omega), \tag{2.22}$$

where the positive constant C only depends on the shape regularity of $\mathcal{T}_h$.

The estimate (2.20) follows from (2.3), (2.16), (2.17) and (2.22):

$$\begin{aligned} \|p - p_h\|^2_{H^1_\beta(\Omega)} + \|y - y_h\|^2_{H^1_\beta(\Omega)} &\lesssim \|p - \Pi_h p\|^2_{H^1_\beta(\Omega)} + \|y - \Pi_h y\|^2_{H^1_\beta(\Omega)} \\ &= \|p - \Pi_h p\|^2_{L_2(\Omega)} + \beta^{\frac{1}{2}} |p - \Pi_h p|^2_{H^1(\Omega)} + \|y - \Pi_h y\|^2_{L_2(\Omega)} + \beta^{\frac{1}{2}} |y - \Pi_h y|^2_{H^1(\Omega)} \\ &\lesssim (\beta^{-1} h^4 + \beta^{-\frac{1}{2}} h^2)(\|f\|^2_{L_2(\Omega)} + \|g\|^2_{L_2(\Omega)}) \\ &= (1 + \beta^{\frac{1}{2}} h^{-2}) \beta^{-1} h^4 (\|f\|^2_{L_2(\Omega)} + \|g\|^2_{L_2(\Omega)}). \end{aligned}$$

The estimate (2.21) is established by a duality argument. Let $(\xi, \theta) \in H_0^1(\Omega) \times H_0^1(\Omega)$ be defined by

$$\mathcal{B}((q, z), (\xi, \theta)) = (q, p - p_h)_{L_2(\Omega)} + (z, y - y_h)_{L_2(\Omega)} \qquad \forall\, (q, z) \in H_0^1(\Omega) \times H_0^1(\Omega). \tag{2.23}$$

We have, by (2.4), Lemma 2.2 (applied to (2.23)), (2.22), (2.23) and Galerkin orthogonality,

$$\begin{aligned}
\|p-p_h\|^2_{L_2(\Omega)}+\|y-y_h\|^2_{L_2(\Omega)} &= \mathcal{B}((p-p_h,y-y_h),(\xi,\theta))\\
&= \mathcal{B}((p-p_h,y-y_h),(\xi-\Pi_h\xi,\theta-\Pi_h\theta))\\
&\lesssim (\|\xi-\Pi_h\xi\|^2_{H^1_\beta(\Omega)}+\|\theta-\Pi_h\theta\|^2_{H^1_\beta(\Omega)})^{\frac12}(\|p-p_h\|^2_{H^1_\beta(\Omega)}+\|y-y_h\|^2_{H^1_\beta(\Omega)})^{\frac12}\\
&\lesssim (1+\beta^{\frac12}h^{-2})^{\frac12}\beta^{-\frac12}h^2(\|p-p_h\|^2_{L_2(\Omega)}+\|y-y_h\|^2_{L_2(\Omega)})^{\frac12}\\
&\qquad\times(\|p-p_h\|^2_{H^1_\beta(\Omega)}+\|y-y_h\|^2_{H^1_\beta(\Omega)})^{\frac12},
\end{aligned}$$

which together with (2.20) implies (2.21). □

The performance of the P_1 finite element method for (2.10) is illustrated in the following numerical example.

Example 2.4. We solve (2.10) on $\Omega=(0,1)\times(0,1)$ with $\beta=1$, $\boldsymbol{\zeta}=\frac12[1\ \ 0]^t$, $\gamma=0$ and exact solution

$$(p,y)=\big(\sin(2\pi x_1)\sin(2\pi x_2),x_1(1-x_1)x_2(1-x_2)\big).$$

There are 10 degrees of freedom (dofs) for the P_1 finite element space associated with the initial mesh ($k=1$). After 7 uniform mesh refinements, the P_1 finite element space associated with the final mesh ($k=8$) has 261122 dofs. The relative errors are displayed in Table 2.1. We observe $O(h)$ convergence in $|\cdot|_{H^1(\Omega)}$ and $O(h^2)$ convergence in $\|\cdot\|_{L_2(\Omega)}$, which agrees with Theorem 2.3.

k	$\frac{\lvert p-p_h\rvert_{H^1(\Omega)}}{\lvert p\rvert_{H^1(\Omega)}}$	Order	$\frac{\|p-p_h\|_{L_2(\Omega)}}{\|p\|_{L_2(\Omega)}}$	Order	$\frac{\lvert y-y_h\rvert_{H^1(\Omega)}}{\lvert y\rvert_{H^1(\Omega)}}$	Order	$\frac{\|y-y_h\|_{L_2(\Omega)}}{\|y\|_{L_2(\Omega)}}$	Order
1	1.60e-01	-	1.45e-01	-	2.77e-01	-	1.96e-01	-
2	1.92e-01	-0.27	3.64e-01	0.09	1.33e-01	1.06	6.97e-02	1.49
3	9.54e-02	1.01	4.20e-02	1.70	5.83e-02	1.19	2.01e-02	1.79
4	4.67e-02	1.03	1.10e-02	1.93	2.67e-02	1.13	5.31e-03	1.92
5	2.32e-02	1.01	2.79e-03	1.97	1.27e-02	1.07	1.36e-03	1.92
6	1.16e-02	1.00	7.00e-04	2.00	6.24e-03	1.03	3.45e-04	2.02
7	5.79e-03	1.00	1.75e-04	2.00	3.09e-03	1.00	8.67e-05	1.99
8	2.89e-03	1.00	4.38e-05	2.00	1.53e-03	1.05	2.17e-05	2.00

TABLE 2.1. Relative errors for the P_1 finite element method for (2.10)

2.4. **A P_1 Finite Element Method for** (1.6)**.** The P_1 finite element method for (1.6) is to find $(\bar p_h,\bar y_h)\in V_h\times V_h$ such that

$$a(q_h,\bar p_h)-(q_h,\bar y_h)_{L_2(\Omega)}=-(q_h,y_d)_{L_2(\Omega)}\qquad\forall\,q_h\in V_h, \tag{2.24a}$$

$$-(\bar p_h,z_h)_{L_2(\Omega)}-\beta a(\bar y_h,z_h)=0\qquad\forall\,z_h\in V_h, \tag{2.24b}$$

which is equivalent to (2.9) under the change of variables

$$\bar{p}_h = \beta^{\frac{1}{4}}\tilde{p}_h \quad \text{and} \quad \bar{y}_h = \beta^{-\frac{1}{4}}\tilde{y}_h. \tag{2.25}$$

Applying the results in Section 2.3 to (2.1) and (2.9), we arrive at the following error estimates through the change of variables (1.7) and (2.25).

Lemma 2.5. *Let* $(\bar{p},\bar{y})$ *(resp.,* $(\bar{p}_h,\bar{y}_h)$*) be the solution of* (1.6) *(resp.,* (2.24)*). We have*

$$\|\bar{p}-\bar{p}_h\|_{H^1_\beta(\Omega)} + \beta^{\frac{1}{2}}\|\bar{y}-\bar{y}_h\|_{H^1_\beta(\Omega)} \le C(1+\beta^{\frac{1}{2}}h^{-2})^{\frac{1}{2}}h^2\|y_d\|_{L_2(\Omega)},$$

$$\|\bar{p}-\bar{p}_h\|_{L_2(\Omega)} + \beta^{\frac{1}{2}}\|\bar{y}-\bar{y}_h\|_{L_2(\Omega)} \le C(1+\beta^{\frac{1}{2}}h^{-2})\beta^{-\frac{1}{2}}h^4\|y_d\|_{L_2(\Omega)},$$

where the positive constant C *is independent of* β *and* h*.*

Remark 2.6. According to Lemma 2.5, the performance of the P_1 finite element method with respect to the norms of $H^1(\Omega)$ and $L_2(\Omega)$ will deteriorate as $\beta \downarrow 0$. Therefore it is necessary to use very fine mesh when β is small in which case it is crucial to have an efficient iterative solver.

Remark 2.7. We can approximate the optimal control $\bar{u}$ in (1.1) by $\bar{u}_h = -\beta^{-1}\bar{p}_h$. It then follows from (1.5b) that the relative error for $\bar{u}_h$ is identical to the relative error for $\bar{p}_h$.

3. Multigrid Algorithms

Let $\mathcal{T}_0$ be a triangulation of Ω and the triangulations $\mathcal{T}_1, \mathcal{T}_2, \ldots$ be generated from $\mathcal{T}_0$ through a refinement process so that $h_k = h_{k-1}/2$ and the shape regularity is maintained. The P_1 finite element subspace of $H^1_0(\Omega)$ associated with $\mathcal{T}_k$ is denoted by V_k.

We want to design multigrid methods for the problem of finding $(p,y) \in V_k \times V_k$ such that

$$\mathcal{B}((p,y),(q,z)) = F(q) + G(z) \qquad \forall\,(q,z) \in V_k \times V_k, \tag{3.1}$$

where $F, G \in V_k'$, and for the dual problem of finding $(p,q) \in V_k \times V_k$ such that

$$\mathcal{B}((q,z),(p,y)) = F(q) + G(z) \qquad \forall\,(q,z) \in V_k \times V_k. \tag{3.2}$$

3.1. A Mesh-Dependent Inner Product.

It is convenient to use a mesh-dependent inner product on $V_k \times V_k$ to rewrite (3.1) and (3.2) in terms of an operator that maps $V_k \times V_k$ to $V_k \times V_k$. First we introduce a mesh-dependent inner product on V_k:

$$(v,w)_k = h_k^d \sum_{x\in\mathcal{V}_k} v(x)w(x) \qquad \forall\, v,w \in V_k, \tag{3.3}$$

where $\mathcal{V}_k$ is the set of the interior vertices of $\mathcal{T}_k$. We have

$$(v,v)_k \approx \|v\|^2_{L_2(\Omega)} \qquad \forall\, v \in V_k \tag{3.4}$$

by a standard scaling argument [13, 10], where the hidden constants only depend on the shape regularity of $\mathcal{T}_0$.

We then define the mesh-dependent inner product $[\cdot,\cdot]_k$ on $V_k \times V_k$ by

$$[(p,y),(q,z)]_k = (p,q)_k + (y,z)_k. \tag{3.5}$$

Let the operator $\mathfrak{B}_k : V_k \times V_k \longrightarrow V_k \times V_k$ be defined by

$$[\mathfrak{B}_k(p,y),(q,z)]_k = \mathcal{B}((p,y),(q,z)) \qquad \forall\,(p,y),(q,z) \in V_k \times V_k. \tag{3.6}$$

We can then rewrite (3.1) in the form

$$\mathfrak{B}_k(p,y) = (f,g), \tag{3.7}$$

where $(f,g) \in V_k \times V_k$ is defined by

$$[(f,g),(q,z)]_k = F(q) + G(z) \qquad \forall\,(q,z) \in V_k \times V_k,$$

and (3.2) becomes

$$\mathfrak{B}_k^t(p,y) = (f,g), \tag{3.8}$$

where

$$[\mathfrak{B}_k^t(p,y),(q,z)]_k = [(p,y),\mathfrak{B}_k(q,z)]_k = \mathcal{B}((q,z),(p,y)) \qquad \forall\,(p,y),(q,z) \in V_k \times V_k. \tag{3.9}$$

We take the coarse-to-fine operator $I_{k-1}^k : V_{k-1} \times V_{k-1} \longrightarrow V_k \times V_k$ to be the natural injection and define the fine-to-coarse operator $I_k^{k-1} : V_k \times V_k \longrightarrow V_{k-1} \times V_{k-1}$ to be the transpose of I_{k-1}^k with respect to the mesh-dependent inner products, i.e.,

$$[I_k^{k-1}(p,y),(q,z)]_{k-1} = [(p,y),I_{k-1}^k(q,z)]_k \quad \forall\,(p,y) \in V_k \times V_k, (q,z) \in V_{k-1} \times V_{k-1}.$$

3.2. **A Block-Diagonal Preconditioner.** Let $L_k : V_k \longrightarrow V_k$ be a linear operator symmetric with respect to the inner product $(\cdot,\cdot)_k$ on V_k such that

$$(L_k v, v)_k \approx \|v\|^2_{H^1_\beta(\Omega)} = \|v\|^2_{L_2(\Omega)} + \beta^{\frac12}|v|^2_{H^1(\Omega)} \qquad \forall\, v \in V_k. \tag{3.10}$$

Then the operator $\mathfrak{C}_k : V_k \times V_k \longrightarrow V_k \times V_k$ defined by

$$\mathfrak{C}_k(p,y) = (L_k p, L_k y) \tag{3.11}$$

is SPD with respect to $[\cdot,\cdot]_k$ and we have

$$[\mathfrak{C}_k(p,y),(p,y)]_k \approx \|p\|^2_{H^1_\beta(\Omega)} + \|y\|^2_{H^1_\beta(\Omega)} \qquad \forall\,(p,y) \in V_k \times V_k, \tag{3.12}$$

where the hidden constants are independent of k and β.

Remark 3.1. We will use $\mathfrak{C}_k^{-1}$ as a preconditioner in the constructions of the smoothing operators. In practice we can take L_k^{-1} to be an approximate solve of the P_1 finite element discretization of the following boundary value problem:

$$-\beta^{\frac12}\Delta u + u = \phi \quad \text{in } \Omega \quad \text{and} \quad u = 0 \quad \text{on } \partial\Omega. \tag{3.13}$$

The multigrid algorithms in Section 3 are $O(n)$ algorithms as long as L_k^{-1} is also an $O(n)$ algorithm. We refer to [24, 15] for the general construction of block diagonal preconditioners for saddle point problems arising from the discretization of PDEs.

Lemma 3.2. *We have*

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k \approx \|p\|^2_{H^1_\beta(\Omega)} + \|y\|^2_{H^1_\beta(\Omega)} \qquad \forall\,(p,y) \in V_k \times V_k, \tag{3.14}$$

$$[\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t(p,y),(p,y)]_k \approx \|p\|^2_{H^1_\beta(\Omega)} + \|y\|^2_{H^1_\beta(\Omega)} \qquad \forall\,(p,y) \in V_k \times V_k, \tag{3.15}$$

where the hidden constants are independent of k and β.

Proof. Let $(p,y)\in V_k\times V_k$ be arbitrary and $(r,x)=\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y)$. Using (2.14), (3.6), (3.12) and duality, we derive (3.14) as follows:

$$\begin{aligned}[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k &= [\mathfrak{C}_k(\mathfrak{C}_k^{-1}\mathfrak{B}_k)(p,y),\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y)]_k\\ &= [\mathfrak{C}_k(r,x),(r,x)]_k\\ &= \sup_{(q,z)\in V_k\times V_k}\frac{[\mathfrak{C}_k(r,x),(q,z)]_k^2}{[\mathfrak{C}_k(q,z),(q,z)]_k}\\ &\approx \sup_{(q,z)\in V_k\times V_k}\frac{[\mathfrak{B}_k(p,y),(q,z)]_k^2}{\|q\|^2_{H^1_\beta(\Omega)}+\|z\|^2_{H^1_\beta(\Omega)}}\\ &= \sup_{(q,z)\in V_k\times V_k}\frac{\big[\mathcal{B}((p,y),(q,z))\big]^2}{\|q\|^2_{H^1_\beta(\Omega)}+\|z\|^2_{H^1_\beta(\Omega)}}\approx \|p\|^2_{H^1_\beta(\Omega)}+\|y\|^2_{H^1_\beta(\Omega)}.\end{aligned}$$

The derivation of (3.15) is analogous, with (2.14) (resp., (3.6)) replaced by (2.15) (resp., (3.9)). □

Lemma 3.3. *The minimum and maximum eigenvalues of* $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$ *and* $\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t$ *satisfy the following bounds*:

$$\lambda_{\min}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k),\ \lambda_{\min}(\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t)\geq C_{\min}, \tag{3.16}$$

$$\lambda_{\max}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k),\ \lambda_{\max}(\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t)\leq C_{\max}(1+\beta^{\frac12}h_k^{-2}), \tag{3.17}$$

where the positive constants $C_{\min}$ *and* $C_{\max}$ *are independent of* k *and* β.

Proof. We will only derive the estimates for $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$ since the derivation for $\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t$ is similar. We have, from (3.4) and (3.5),

$$[(p,y),(p,y)]_k\approx\|p\|^2_{L_2(\Omega)}+\|y\|^2_{L_2(\Omega)}\qquad\forall\,(p,y)\in V_k\times V_k, \tag{3.18}$$

where the hidden constants only depend on the shape regularity of $\mathcal{T}_0$. It follows from (2.3), (3.14) and (3.18) that

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k\geq C_{\min}[(p,y),(p,y)]_k\qquad\forall\,(p,y)\in V_k\times V_k, \tag{3.19}$$

which then implies (3.16) by the Rayleigh quotient formula.

By a standard inverse estimate [13, 10], we have

$$\|v\|^2_{H^1_\beta(\Omega)}=\|v\|^2_{L_2(\Omega)}+\beta^{\frac12}|v|^2_{H^1(\Omega)}\leq(1+C\beta^{\frac12}h_k^{-2})\|v\|^2_{L_2(\Omega)}\qquad\forall\,v\in V_k,$$

where the positive constant C depends only on the shape regularity of $\mathcal{T}_0$. It then follows from (2.3), (3.14) and (3.18) that

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k\leq C_{\max}(1+\beta^{\frac12}h_k^{-2})[(p,y),(p,y)]_k\qquad\forall\,(p,y)\in V_k\times V_k,$$

and hence (3.17) holds because of the Rayleigh quotient formula. □

Remark 3.4. It follows from (3.16) and (3.17) that the operators $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$ and $\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t$ are well-conditioned when $\beta^{\frac12}h_k^{-2}=O(1)$.

3.3. **A W-Cycle Multigrid Algorithm for (3.7).** Let the output of the W-cycle algorithm for (3.7) with initial guess (p_0, y_0) and m_1 (resp., m_2) pre-smoothing (resp., post-smoothing) steps be denoted by $MG_W(k, (f, g), (p_0, y_0), m_1, m_2)$.

We use a direct solve for $k = 0$, i.e., we take $MG_W(0, (f, g), (p_0, y_0), m_1, m_2)$ to be $\mathfrak{B}_0^{-1}(f, g)$. For $k \geq 1$, we obtain $MG_W(k, (f, g), (p_0, y_0), m_1, m_2)$ in three steps.

Pre-Smoothing We compute $(p_1, y_1), \ldots, (p_{m_1}, y_{m_1})$ recursively by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{C}_k^{-1} \mathfrak{B}_k^t ((f, g) - \mathfrak{B}_k(p_{j-1}, y_{j-1})) \tag{3.20}$$

for $1 \leq j \leq m_1$. The choice of the damping factor λ_k will be given below in (3.23) and (3.24).

Coarse Grid Correction Let $(f', g') = I_k^{k-1}((f, g) - \mathfrak{B}_k(p_{m_1}, y_{m_1}))$ be the transferred residual of (p_{m_1}, y_{m_1}) and let $(p_1', y_1'), (p_2', y_2') \in V_{k-1} \times V_{k-1}$ be computed by

$$(p_1', y_1') = MG_W(k-1, (f', g'), (0, 0), m_1, m_2), \tag{3.21a}$$

$$(p_2', y_2') = MG_W(k-1, (f', g'), (p_1', y_1'), m_1, m_2). \tag{3.21b}$$

We then take (p_{m_1+1}, y_{m_1+1}) to be $(p_{m_1}, y_{m_1}) + I_{k-1}^k (p_2', y_2')$.

Post-Smoothing We compute $(p_{m_1+2}, y_{m_1+2}), \ldots, (p_{m_1+m_2+1}, y_{m_1+m_2+1})$ recursively by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{B}_k^t \mathfrak{C}_k^{-1} ((f, g) - \mathfrak{B}_k(p_{j-1}, y_{j-1})) \tag{3.22}$$

for $m_1 + 2 \leq j \leq m_1 + m_2 + 1$.

The final output is $MG_W(k, (f, g), (p_0, y_0), m_1, m_2) = (p_{m_1+m_2+1}, y_{m_1+m_2+1})$.

To complete the description of the algorithm, we choose the damping factor λ_k as follows:

$$\lambda_k = \frac{2}{\lambda_{\min}(\mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k) + \lambda_{\max}(\mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k)} \qquad \text{if } \beta^{\frac{1}{2}} h_k^{-2} < 1, \tag{3.23}$$

and

$$\lambda_k = [C_\dagger(1 + \beta^{\frac{1}{2}} h_k^{-2})]^{-1} \qquad \text{if } \beta^{\frac{1}{2}} h_k^{-2} \geq 1, \tag{3.24}$$

where $C_\dagger$ is greater than or equal to the constant $C_{\max}$ in (3.17).

Remark 3.5. Note that the post-smoothing step is exactly the Richardson iteration for the equation

$$\mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k(p, y) = \mathfrak{B}_k^t \mathfrak{C}_k^{-1}(f, g),$$

which is equivalent to (3.7).

Remark 3.6. In the case where $\beta^{\frac{1}{2}} h_k^{-2} < 1$, the choice of λ_k is motivated by the well-conditioning of $\mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k$ (cf. Remark 3.4) and the optimal choice of damping factor for the Richardson iteration [25, p. 114]. In practice the relation (3.23) only holds approximately, but it affects neither the analysis nor the performance of the W-cycle algorithm. In the case where $\beta^{\frac{1}{2}} h_k^{-2} \geq 1$, the choice of λ_k is motivated by the condition $\lambda_{\max}(\lambda_k \mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k) \leq 1$ (cf. (3.17)) that will ensure the highly oscillatory part of the error is damped out when Richardson iteration is used as a smoother for an ill-conditioned system (cf. Lemma 4.2).

3.4. **A V-Cycle Multigrid Algorithm for** (3.7)**.** Let the output of the V-cycle algorithm for (3.7) with initial guess (p_0, y_0) and m_1 (resp., m_2) pre-smoothing (resp., post-smoothing) steps be denoted by $MG_V(k,(f,g),(p_0,y_0),m_1,m_2)$. The difference between the computations of $MG_V(k,(f,g),(p_0,y_0),m_1,m_2)$ and $MG_W(k,(f,g),(p_0,y_0),m_1,m_2)$ is only in the coarse grid correction step, where we compute

$$(p_1', y_1') = MG_V(k-1,(f',g'),(0,0),m_1,m_2)$$

and take (p_{m_1+1}, y_{m_1+1}) to be $(p_{m_1}, y_{m_1}) + I_{k-1}^k(p_1', y_1')$.

Remark 3.7. We will focus on the analysis of the W-cycle algorithm in this paper. But numerical results (cf. Section 6) indicate that the performance of the V-cycle algorithm is also robust respect to k and β.

3.5. **Multigrid Algorithms for** (3.8)**.** We can define W-cycle and V-cycle algorithms for (3.8) by simply interchanging the operators $\mathfrak{B}_k$ and $\mathfrak{B}_k^t$ in Section 3.3 and Section 3.4. In particular, the pre-smoothing step is given by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{C}_k^{-1}\mathfrak{B}_k((f,g) - \mathfrak{B}_k^t(p_{j-1}, y_{j-1})), \tag{3.25}$$

and the post-smoothing step is given by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{B}_k\mathfrak{C}_k^{-1}((f,g) - \mathfrak{B}_k^t(p_{j-1}, y_{j-1})). \tag{3.26}$$

4. Smoothing and Approximation Properties

We will develop in this section two key ingredients for the convergence analysis of the W-cycle algorithm, namely, the smoothing and approximation properties. They will be expressed in terms of two scales of mesh-dependent norms defined by

$$|||(p,y)|||_{s,k} = [(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)^s(p,y),(p,y)]_k^{\frac{1}{2}} \qquad \forall (p,y) \in V_k \times V_k, \tag{4.1}$$

$$|||(p,y)|||_{s,k}^{\sim} = [(\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t)^s(p,y),(p,y)]_k^{\frac{1}{2}} \qquad \forall (p,y) \in V_k \times V_k. \tag{4.2}$$

Note that

$$|||(p,y)|||_{0,k}^2 \approx \|p\|_{L_2(\Omega)}^2 + \|y\|_{L_2(\Omega)}^2 \approx (|||(p,y)|||_{0,k}^{\sim})^2 \qquad \forall\,(p,y) \in V_k \times V_k \tag{4.3}$$

by (3.18), and

$$|||(p,y)|||_{1,k}^2 \approx \|p\|_{H_\beta^1(\Omega)}^2 + \|y\|_{H_\beta^1(\Omega)}^2 \approx (|||(p,y)|||_{1,k}^{\sim})^2 \qquad \forall\,(p,y) \in V_k \times V_k \tag{4.4}$$

by (3.14) and (3.15).

4.1. **Post-Smoothing Properties.** The error propagation operator for one post-smoothing step defined by (3.22) is given by

$$R_k = Id_k - \lambda_k \mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k, \tag{4.5}$$

where Id_k is the identity operator on $V_k \times V_k$.

Similarly, the error propagation operator for one post-smoothing step defined by (3.26) is given by

$$\tilde{R}_k = Id_k - \lambda_k \mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t. \tag{4.6}$$

Lemma 4.1. *In the case where* $\beta^{\frac{1}{2}}h_k^{-2} < 1$*, we have*

$$|||R_k(p,y)|||_{1,k} \le \tau |||(p,y)|||_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k, \tag{4.7}$$

$$|||\tilde{R}_k(p,y)|||^{\sim}_{1,k} \le \tau |||(p,y)|||^{\sim}_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k, \tag{4.8}$$

where the constant $\tau\in(0,1)$ *is independent of* k *and* β*.*

Proof. In this case λ_k given by (3.23) is the optimal damping parameter for the Richardson iteration and we have

$$C_{\min} \le \lambda_{\min}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k) \le \lambda_{\max}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)) < 2C_{\max}$$

by Lemma 3.3. It follows that (cf. [25, p. 114])

$$\begin{aligned}
|||R_k(p,y)|||_{1,k} &= [\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_kR_k(p,y),R_k(p,y)]_k^{\frac{1}{2}}\\
&\le \Big(\frac{\lambda_{\max}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)-\lambda_{\min}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)}{\lambda_{\max}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)+\lambda_{\min}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)}\Big)[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k^{\frac{1}{2}}\\
&\le \Big(\frac{2C_{\max}-C_{\min}}{2C_{\max}+C_{\min}}\Big)|||(p,y)|||_{1,k}.
\end{aligned}$$

Therefore (4.7) holds for $\tau=(2C_{\max}-C_{\min})/(2C_{\max}+C_{\min})$.

The derivation of (4.8) is identical. □

Lemma 4.2. *In the case where* $\beta^{\frac{1}{2}}h_k^{-2}\ge 1$*, we have, for* $0\le s\le 1$*,*

$$|||R_k^m(p,y)|||_{1,k} \le C(1+\beta^{\frac{1}{2}}h_k^{-2})^{s/2}m^{-s/2}|||(p,y)|||_{1-s,k} \quad \forall(p,y)\in V_k\times V_k, \tag{4.9}$$

$$|||\tilde{R}_k^m(p,y)|||^{\sim}_{1,k} \le C(1+\beta^{\frac{1}{2}}h_k^{-2})^{s/2}m^{-s/2}|||(p,y)|||^{\sim}_{1-s,k} \quad \forall(p,y)\in V_k\times V_k, \tag{4.10}$$

where the positive constant C *is independent of* k *and* β*.*

Proof. In this case λ_k is given by (3.24) and $\lambda_{\max}(\lambda_k\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)\le 1$. It follows from (3.24), (4.1), (4.5), calculus and the spectral theorem that

$$\begin{aligned}
|||R_k^m(p,y)|||_{1,k}^2 &= [\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_kR_k^m(p,y),R_k^m(p,y)]_k\\
&= \lambda_k^{-s}[(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)^{1-s}(\lambda_k\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)^sR_k^m(p,y),R_k^m(p,y)]_k\\
&\le C_\dagger^{-s}(1+\beta^{\frac{1}{2}}h_k^{-2})^s\max_{0\le x\le 1}[(1-x)^{2m}x^s][(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k)^{1-s}(p,y),(p,y)]_k\\
&\le C(1+\beta^{\frac{1}{2}}h_k^{-2})^sm^{-s}|||(p,y)|||_{1-s,k}^2.
\end{aligned}$$

The proof for (4.10) is identical. □

Remark 4.3. In the special case where $s=0$, the calculation in the proof of Lemma 4.2 shows that

$$|||R_k(p,y)|||_{1,k}\le|||(p,y)|||_{1,k} \quad\text{and}\quad |||\tilde{R}_k(p,y)|||^{\sim}_{1,k}\le|||(p,y)|||^{\sim}_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k.$$

4.2. **Approximation Properties.** We define two Ritz projection operators $P_k^{k-1} : V_k \times V_k \to V_{k-1} \times V_{k-1}$ and $\tilde{P}_k^{k-1} : V_k \times V_k \to V_{k-1} \times V_{k-1}$ in terms of the bilinear form $\mathcal{B}(\cdot,\cdot)$ and the natural injection $I_{k-1}^k : V_{k-1} \times V_{k-1} \longrightarrow V_k \times V_k$ as follows. For any $(p,y) \in V_k \times V_k$ and $(q,z) \in V_{k-1} \times V_{k-1}$,

$$\mathcal{B}(P_k^{k-1}(p,y),(q,z)) = \mathcal{B}((p,y), I_{k-1}^k(q,z)) = \mathcal{B}((p,y),(q,z)), \tag{4.11}$$

$$\mathcal{B}((q,z),\tilde{P}_k^{k-1}(p,y)) = \mathcal{B}(I_{k-1}^k(q,z),(q,y)) = \mathcal{B}((q,z),(p,y)). \tag{4.12}$$

It follows that

$$P_k^{k-1} I_{k-1}^k = Id_{k-1} = \tilde{P}_k^{k-1} I_{k-1}^k$$

and hence

$$(I_{k-1}^k P_k^{k-1})^2 = I_{k-1}^k P_k^{k-1} \quad \text{and} \quad (Id_k - I_{k-1}^k P_k^{k-1})^2 = Id_k - I_{k-1}^k P_k^{k-1}, \tag{4.13}$$

$$(I_{k-1}^k \tilde{P}_k^{k-1})^2 = I_{k-1}^k \tilde{P}_k^{k-1} \quad \text{and} \quad (Id_k - I_{k-1}^k \tilde{P}_k^{k-1})^2 = Id_k - I_{k-1}^k \tilde{P}_k^{k-1}. \tag{4.14}$$

Moreover we have the following Galerkin orthogonality relations:

$$\mathcal{B}((Id_k - I_{k-1}^k P_k^{k-1})(p,y), I_{k-1}^k(q,z)) = 0 \quad \forall\,(p,y) \in V_k \times V_k, (q,z) \in V_{k-1} \times V_{k-1}, \tag{4.15}$$

$$\mathcal{B}(I_{k-1}^k(q,z),(Id_k - I_{k-1}^k \tilde{P}_k^{k-1})(p,y)) = 0 \quad \forall\,(p,y) \in V_k \times V_k, (q,z) \in V_{k-1} \times V_{k-1}. \tag{4.16}$$

The effects of the operators $Id_k - I_{k-1}^k P_k^{k-1}$ and $Id_k - I_{k-1}^k \tilde{P}_k^{k-1}$ are measured by the following approximation properties.

Lemma 4.4. *There exists a positive constant C independent of k and β such that*

$$|||(Id_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{0,k} \le C(1+\beta^{\frac12} h_k^{-2})^{\frac12} \beta^{-\frac12} h_k^2 |||(p,y)|||_{1,k} \quad \forall\,(p,y) \in V_k \times V_k, \tag{4.17}$$

$$|||(Id_k - I_{k-1}^k \tilde{P}_k^{k-1})(p,y)|||_{0,k}^{\sim} \le C(1+\beta^{\frac12} h_k^{-2})^{\frac12} \beta^{-\frac12} h_k^2 |||(p,y)|||_{1,k}^{\sim} \quad \forall\,(p,y) \in V_k \times V_k. \tag{4.18}$$

Proof. We will only present the detailed arguments for (4.17). Let $(p,y) \in V_k \times V_k$ be arbitrary and

$$(\zeta,\mu) = (Id_k - I_{k-1}^k P_k^{k-1})(p,y). \tag{4.19}$$

In view of (4.3), it suffices to establish the estimate

$$\|\zeta\|_{L_2(\Omega)} + \|\mu\|_{L_2(\Omega)} \lesssim (1+\beta^{\frac12} h_k^{-2})^{\frac12} \beta^{-\frac12} h_k^2 |||(p,y)|||_{1,k} \tag{4.20}$$

by a duality argument.

Let $(\xi,\theta) \in H_0^1(\Omega) \times H_0^1(\Omega)$ be defined by

$$\mathcal{B}((q,z),(\xi,\theta)) = (\zeta,q)_{L_2(\Omega)} + (\mu,z)_{L_2(\Omega)} \qquad \forall\,(q,z) \in H_0^1(\Omega) \times H_0^1(\Omega), \tag{4.21}$$

and $(\xi_{k-1},\theta_{k-1}) \in V_{k-1} \times V_{k-1}$ be defined by

$$\mathcal{B}((q,z),(\xi_{k-1},\theta_{k-1})) = (\zeta,q)_{L_2(\Omega)} + (\mu,z)_{L_2(\Omega)} \qquad \forall\,(q,z) \in V_{k-1} \times V_{k-1}. \tag{4.22}$$

Since $h_{k-1} = 2h_k$, we have, according to Theorem 2.3,

$$\|\xi - \xi_{k-1}\|_{H_\beta^1(\Omega)} + \|\theta - \theta_{k-1}\|_{H_\beta^1(\Omega)} \lesssim (1+\beta^{\frac12} h_k^{-2})^{\frac12} \beta^{-\frac12} h_k^2 (\|\zeta\|_{L_2(\Omega)} + \|\mu\|_{L_2(\Omega)}). \tag{4.23}$$

Putting (2.4), (4.4), (4.15), (4.19) and (4.21)–(4.23) together, we find

$$\begin{aligned}\|\zeta\|^2_{L_2(\Omega)}+\|\mu\|^2_{L_2(\Omega)} &= \mathcal{B}((\zeta,\mu),(\xi,\theta))\\ &= \mathcal{B}((Id_k - I^k_{k-1}P^{k-1}_k)(p,y),(\xi,\theta))\\ &= \mathcal{B}((Id_k - I^k_{k-1}P^{k-1}_k)(p,y),(\xi,\theta)-(\xi_{k-1},\theta_{k-1}))\\ &= \mathcal{B}((p,y),(\xi,\theta)-(\xi_{k-1},\theta_{k-1}))\\ &\lesssim (\|\xi-\xi_{k-1}\|^2_{H^1_\beta(\Omega)}+\|\theta-\theta_{k-1}\|^2_{H^1_\beta(\Omega)})^{\frac12}(\|p\|^2_{H^1_\beta(\Omega)}+\|y\|^2_{H^1_\beta(\Omega)})^{\frac12}\\ &\lesssim (1+\beta^{\frac12}h_k^{-2})^{\frac12}\beta^{-\frac12}h_k^2(\|\zeta\|_{L_2(\Omega)}+\|\mu\|_{L_2(\Omega)})|||(p,y)|||_{1,k},\end{aligned}$$

which implies (4.20).

The estimate (4.18) is established by similar arguments based on (4.16). □

We will also need the following stability estimates.

Lemma 4.5. *We have*

$$|||I^k_{k-1}(q,z)|||_{1,k} \approx |||(q,z)|||_{1,k-1} \qquad \forall\,(q,z)\in V_{k-1}\times V_{k-1}, \tag{4.24}$$

$$|||P^{k-1}_k(p,y)|||_{1,k-1} \lesssim |||(p,y)|||_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k, \tag{4.25}$$

$$|||\tilde{P}^{k-1}_k(p,y)|||^{\sim}_{1,k-1} \lesssim |||(p,y)|||^{\sim}_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k, \tag{4.26}$$

where the hidden constants are independent of k and β.

Proof. The estimate (4.24) follows from (4.4) and the fact that I^k_{k-1} is the natural injection. The estimate (4.25) then follows from (2.14), (4.4), (4.11) and (4.24) :

$$\begin{aligned}|||P^{k-1}_k(p,y)|||_{1,k-1} &\approx \sup_{(q,z)\in V_{k-1}\times V_{k-1}} \frac{\mathcal{B}\big(P^{k-1}_k(p,y),(q,z)\big)}{|||(q,z)|||_{1,k-1}}\\ &= \sup_{(q,z)\in V_{k-1}\times V_{k-1}} \frac{\mathcal{B}\big((p,y),I^k_{k-1}(q,z)\big)}{|||(q,z)|||_{1,k-1}} \lesssim |||(p,y)|||_{1,k}.\end{aligned}$$

Similarly we obtain (4.26) by using (2.15), (4.4), (4.12) and (4.24). □

5. Convergence Analysis of the W-Cycle Algorithms

Let $E_k : V_k\times V_k \longrightarrow V_k\times V_k$ be the error propagation operator for the k-th level W-cycle algorithm for (3.7). We have the following well-known recursive relation (cf. [18, 23, 6]):

$$E_k = R_k^{m_2}(Id_k - I^k_{k-1}P^{k-1}_k + I^k_{k-1}E^2_{k-1}P^{k-1}_k)S_k^{m_1}, \tag{5.1}$$

where R_k is given by (4.5) and

$$S_k = Id_k - \lambda_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t\mathfrak{B}_k \tag{5.2}$$

is the error propagation operator for one pre-smoothing step (cf. (3.20)).

Note that S_k is the transpose of $\tilde{R}_k$ (the error propagation operator of one post-smoothing step for the dual problem (3.8)) with respect to the variational form $\mathcal{B}(\cdot,\cdot)$. Indeed we have, by (3.6), (3.9) and (4.6),

$$\begin{aligned}\mathcal{B}(S_k(p,y),(q,z)) &= [\mathfrak{B}_k(Id_k - \lambda_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t\mathfrak{B}_k)(p,y),(q,z)]_k \\ &= [\mathfrak{B}_k(p,y),(Id_k - \lambda_k\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t)(q,z)]_k \qquad (5.3)\\ &= \mathcal{B}((p,y),\tilde{R}_k(q,z)) \qquad \forall\,(p,y),(q,z)\in V_k\times V_k.\end{aligned}$$

Remark 5.1. The duality between S_k and $\tilde{R}_k$ is the reason why we consider multigrid algorithms for (3.7) and (3.8) simultaneously.

The relations (4.11) and (5.3) lead to the following useful result.

Lemma 5.2. *We have*

$$\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\| \approx \|\tilde{R}_k^m(Id_k - I_{k-1}^k\tilde{P}_k^{k-1})\|, \qquad (5.4)$$

where $\|\cdot\|$ *denotes the operator norm with respect to* $|||\cdot|||_{1,k}$ *and the hidden constants are independent of* k *and* β.

Proof. It follows from (2.14), (4.4), (4.11), (4.12) and (5.3) that

$$\begin{aligned}&|||(Id_k - I_{k-1}^k P_k^{k-1})S_k^m(p,y)|||_{1,k}\\ &\qquad\approx \sup_{(q,z)\in V_k\times V_k}\frac{\mathcal{B}\big((Id_k - I_{k-1}^k P_k^{k-1})S_k^m(p,y),(q,z)\big)}{|||(q,z)|||_{1,k}}\\ &\qquad= \sup_{(q,z)\in V_k\times V_k}\frac{\mathcal{B}\big((p,y),\tilde{R}_k^m(Id_k - I_{k-1}^k\tilde{P}_k^{k-1})(q,z)\big)}{|||(q,z)|||_{1,k}}\\ &\qquad\lesssim |||(p,y)|||_{1,k}\|\tilde{R}_k^m(Id_k - I_{k-1}^k\tilde{P}_k^{k-1})\|\end{aligned}$$

and hence

$$\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\| \lesssim \|\tilde{R}_k^m(Id_k - I_{k-1}^k\tilde{P}_k^{k-1})\|.$$

The estimate in the other direction is established by a similar argument that uses (2.15) instead of (2.14). □

5.1. **Convergence of the Two-Grid Algorithm for** (3.7)**.** In the two-grid algorithm the coarse grid residual equation is solved exactly. By setting $E_{k-1} = 0$ in (5.1), we obtain the error propagation operator $R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}$ for the two-grid algorithm with m_1 (resp., m_2) pre-smoothing (resp., post-smoothing) steps.

We will separate the convergence analysis into two cases.

The case where $\beta^{\frac{1}{2}}h_k^{-2} < 1$**.** Here we can apply Lemma 4.1 which states that R_k (resp., $\tilde{R}_k$) is a contraction with respect to $|||\cdot|||_{1,k}$ (resp., $|||\cdot|||_{\widetilde{1,k}}$) and the contraction number τ is independent of k and β.

Lemma 5.3. *In the case where* $\beta^{\frac{1}{2}}h_k^{-2} < 1$, *there exists a positive constant* $C_\sharp$ *independent of* k *and* β *such that*

$$\|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| \le C_\sharp\tau^{m_1+m_2}, \qquad (5.5)$$

where $\|\cdot\|$ *is the operator norm with respect to* $|||\cdot|||_{1,k}$.

Proof. We have, from (4.7) and Lemma 4.5,

$$\begin{aligned}&|||R_k^m(Id_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{1,k}\\ &\qquad \le \tau^m |||(Id_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{1,k} \lesssim \tau^m |||(p,y)|||_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k,\end{aligned}$$

and hence

$$\|R_k^m(Id_k - I_{k-1}^k P_k^{k-1})\| \lesssim \tau^m. \tag{5.6}$$

Similarly, we also have, by (4.4), (4.8) and Lemma 4.5,

$$\|\tilde{R}_k^m(Id_k - I_{k-1}^k \tilde{P}_k^{k-1})\| \lesssim \tau^m,$$

which together with Lemma 5.2 implies

$$\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\| \lesssim \tau^m. \tag{5.7}$$

Finally we establish (5.5) by combining (4.13), (5.6) and (5.7):

$$\begin{aligned}&\|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\|\\ &\qquad = \|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\|\\ &\qquad \le \|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})\|\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| \lesssim \tau^{m_1+m_2}.\end{aligned}$$

□

The case where $\beta^{\frac{1}{2}}h_k^{-2} \ge 1$. Here we can apply Lemma 4.2.

Lemma 5.4. *In the case where $\beta^{\frac{1}{2}}h_k^{-2} \ge 1$, there exists a positive constant $C_\flat$ independent of k and β such that*

$$\|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| \le C_\flat[\max(1,m_1)\max(1,m_2)]^{-\frac{1}{2}}, \tag{5.8}$$

where $\|\cdot\|$ is the operator norm with respect to $|||\cdot|||_{1,k}$.

Proof. Let m be any positive integer. We have, from (4.9) and (4.17),

$$\begin{aligned}&|||R_k^m(Id_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{1,k}\\ &\qquad \lesssim (1+\beta^{\frac{1}{2}}h_k^{-2})^{\frac{1}{2}} m^{-\frac{1}{2}} |||(Id_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{0,k}\\ &\qquad \lesssim (1+\beta^{\frac{1}{2}}h_k^{-2})^{\frac{1}{2}} m^{-\frac{1}{2}} (1+\beta^{\frac{1}{2}}h_k^{-2})^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 |||(p,y)|||_{1,k}\\ &\qquad = m^{-\frac{1}{2}}(\beta^{-\frac{1}{2}}h_k^2+1)|||(p,y)|||_{1,k}\\ &\qquad \le 2m^{-\frac{1}{2}}|||(p,y)|||_{1,k} \qquad \forall\,(p,y)\in V_k\times V_k,\end{aligned}$$

and hence

$$\|R_k^m(Id_k - I_{k-1}^k P_k^{k-1})\| \lesssim m^{-\frac{1}{2}}. \tag{5.9}$$

Similarly, we also have, by (4.4), (4.10) and (4.18),

$$\|\tilde{R}_k^m(Id_k - I_{k-1}^k \tilde{P}_k^{k-1})\| \lesssim m^{-\frac{1}{2}}.$$

It then follows from Lemma 5.2 that

$$\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\| \lesssim m^{-\frac{1}{2}}. \tag{5.10}$$

Combining (4.13), (5.9) and (5.10), we obtain for $m_1, m_2 \geq 1$,

$$\begin{aligned}\|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| &= \|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| \\ &\leq \|R_k^{m_2}(Id_k - I_{k-1}^k P_k^{k-1})\|\|(Id_k - I_{k-1}^k P_k^{k-1})S_k^{m_1}\| \\ &\lesssim (m_1 m_2)^{-\frac{1}{2}}.\end{aligned}$$

The cases where $m_1 = 0$ or $m_2 = 0$ follow directly from (5.9) and (5.10). □

5.2. **Convergence of the W-Cycle Algorithm for** (3.7)**.** We will derive error estimates for the W-cycle algorithm through (5.1) and the results for the two-grid algorithm in Section 5.1. For simplicity we will focus on the symmetric W-cycle algorithm where $m_1 = m_2 = m \geq 1$.

According to (4.4) and Remark 4.3, there exists a positive constant C_1 independent of k and m such that

$$\|R_k^m\|, \|\tilde{R}_k^m\| \leq C_1, \tag{5.11}$$

where $\|\cdot\|$ is the operator norm with respect to $\|\!|\cdot|\!\|_{1,k}$. Moreover it follows from (2.14), (4.4) and (5.3) that

$$\begin{aligned}\|\!|S_k^m(p,y)|\!\|_{1,k} &\approx \sup_{(q,z)\in V_k\times V_k} \frac{\mathcal{B}\big(S_k^m(p,y),(q,z)\big)}{\|\!|(q,z)|\!\|_{1,k}} \\ &= \sup_{(q,z)\in V_k\times V_k} \frac{\mathcal{B}\big((p,y),\tilde{R}_k^m(q,z)\big)}{\|\!|(q,z)|\!\|_{1,k}} \lesssim \|\!|(p,y)|\!\|_{1,k}\|\tilde{R}_k^m\| \qquad \forall\,(p,y)\in V_k\times V_k,\end{aligned}$$

and hence, by (5.11),

$$\|S_k^m\| \leq C_2, \tag{5.12}$$

where the positive constant C_2 is also independent of k and m.

Putting Lemma 4.5, (5.1), (5.11) and (5.12) together, we obtain the recursive estimate

$$\|E_k\| \leq \|R_k^m(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\| + C_*\|E_{k-1}\|^2 \qquad \text{for } k \geq 1, \tag{5.13}$$

where the positive constant C_* is independent of k and β. The behavior of $\|E_k\|$ is therefore determined by (5.13), the behavior of $\|R_k^m(Id_k - I_{k-1}^k P_k^{k-1})S_k^m\|$, and the initial condition

$$\|E_0\| = 0. \tag{5.14}$$

Specifically, for $\beta^{\frac{1}{2}}h_k^{-2} < 1$, we have

$$\|E_k\| \leq C_\sharp \tau^{2m} + C_*\|E_{k-1}\|^2 \tag{5.15}$$

by Lemma 5.3, and for $\beta^{\frac{1}{2}}h_k^{-2} \geq 1$, we have

$$\|E_k\| \leq C_\flat m^{-1} + C_*\|E_{k-1}\|^2 \tag{5.16}$$

by Lemma 5.4.

The following result is useful for the analysis of (5.14)–(5.16).

Lemma 5.5. *Let α_k $(k = 0, 1, 2, \ldots)$ be a sequence of nonnegative numbers such that*

$$\alpha_k \le 1 + \delta\alpha_{k-1}^2 \qquad \text{for} \quad k \ge 1, \tag{5.17}$$

where the positive constant δ satisfies

$$\delta \le \frac{1}{4(1+\alpha_0)}. \tag{5.18}$$

Then we have

$$\alpha_k \le 2 + 4^{1-2^k}\alpha_0 \qquad \text{for} \quad k \ge 0. \tag{5.19}$$

Proof. The bound (5.19) holds trivially for $k = 0$. Suppose it holds for $k \ge 0$. We have, by (5.17) and (5.18),

$$\begin{aligned} \alpha_{k+1} \le 1 + \delta\alpha_k^2 &\le 1 + \delta(2 + 4^{1-2^k}\alpha_0)^2 \\ &= 1 + \delta(4 + 4^{1-2^k}4\alpha_0) + (\delta\alpha_0)4^{2-2^{k+1}}\alpha_0 \\ &\le 1 + \delta(4 + 4\alpha_0) + \Big(\frac{1}{4}\Big)4^{2-2^{k+1}}\alpha_0 \le 2 + 4^{1-2^{k+1}}\alpha_0. \end{aligned}$$

Therefore the bound (5.19) holds for $k \ge 0$ by mathematical induction. □

Theorem 5.6. *There exists a positive integer m_* independent of k and β such that $m \ge m_*$ implies*

$$\|E_k\| \le 2C_\sharp\tau^{2m} \qquad \forall\, 1 \le k \le k_*, \tag{5.20}$$

$$\|E_k\| \le 2C_\flat m^{-1} + 4^{1-2^{k-k_*}}(2C_\sharp\tau^{2m}) \qquad \forall\, k \ge k_* + 1, \tag{5.21}$$

where $\|\cdot\|$ is the operator norm with respect to $\|\!|\cdot|\!\|_{1,k}$ and k_* is the largest positive integer such that $\beta^{\frac{1}{2}}h_{k_*}^{-2} < 1$.

Proof. For $1 \le k \le k_*$, we take $\alpha_k = \|E_k\|/(C_\sharp\tau^{2m})$ and observe that

$$\alpha_k \le 1 + (C_* C_\sharp\tau^{2m})\alpha_{k-1}^2$$

by (5.15). It then follows from (5.14) and Lemma 5.5 that $\alpha_k \le 2$, or equivalently

$$\|E_k\| \le 2C_\sharp\tau^{2m},$$

provided that

$$C_* C_\sharp\tau^{2m} \le \frac{1}{4}. \tag{5.22}$$

We now define $\mu_k = \|E_{k_*+k}\|/(C_\flat m^{-1})$ and observe that

$$\mu_k \le 1 + (C_* C_\flat m^{-1})\mu_{k-1}^2 \qquad \text{for} \quad k \ge 1$$

by (5.16). It then follows from Lemma 5.5 that

$$\mu_k \le 2 + 4^{1-2^k}\mu_0 \qquad \text{for} \quad k \ge 1,$$

or equivalently

$$\|E_k\| \le 2C_\flat m^{-1} + 4^{1-2^{k-k_*}}\|E_{k_*}\| \qquad \text{for} \quad k \ge k_* + 1,$$

provided that

$$C_* C_\flat m^{-1} \le \frac{1}{4(1+\|E_{k_*}\|/(C_\flat m^{-1}))},$$

or equivalently

$$C_* C_\flat m^{-1} + C_*\|E_{k_*}\| \le \frac{1}{4}. \tag{5.23}$$

Finally we observe that if we choose m_* so that

$$C_* C_\flat m_*^{-1} + 2C_* C_\sharp \tau^{2m_*} \le \frac{1}{4},$$

then (5.22) and (5.23) are satisfied for $m \ge m_*$. □

Remark 5.7. According to (4.4) and Theorem 5.6, the k-th level symmetric W-cycle algorithm for (3.7) is a contraction in the energy norm $\|\cdot\|_{H^1_\beta(\Omega)}$ if the number of smoothing steps is sufficiently large and the contraction number is bounded away from 1 uniformly in k and β. Moreover, for the coarser levels where $\beta^{\frac{1}{2}} h_k^{-2} < 1$, the contraction number of the symmetric W-cycle algorithm will decrease exponentially with respect to the number of smoothing steps m. After a few transition levels the dominant term on the right-hand side of (5.21) becomes $2C_\flat m^{-1}$ and the contraction number will decrease at the rate of m^{-1} for the finer levels where $\beta^{\frac{1}{2}} h_k^{-2} \ge 1$.

Remark 5.8. For the nonsymmetric W-cycle algorithm with m_1 (resp., m_2) pre-smoothing (resp., post-smoothing) steps, the estimates (5.20) and (5.21) are replaced by

$$\begin{aligned} &\|E_k\| \le 2C_\sharp \tau^{m_1+m_2} && \forall\, 1 \le k \le k_*, \\ &\|E_k\| \le 2C_\flat [\max(1,m_1)\max(1,m_2)]^{-\frac{1}{2}} + 4^{1-2^{k-k_*}}(2C_\sharp \tau^{m_1+m_2}) && \forall\, k \ge k_*+1. \end{aligned}$$

5.3. **Convergence of the *W*-cycle Multigrid Algorithms for** (3.8)**.** The error propagation operator $\tilde{E}_k : V_k \times V_k \longrightarrow V_k \times V_k$ for the W-cycle algorithm for (3.8) satisfies the following analog of (5.1):

$$\tilde{E}_k = \tilde{R}_k^{m_2}(Id_k - I_{k-1}^k \tilde{P}_k^{k-1} + I_{k-1}^k \tilde{E}_{k-1}^2 \tilde{P}_k^{k-1})\tilde{S}_k^{m_1},$$

where $\tilde{R}_k$ is given by (4.6) and $\tilde{S}_k = Id_k - \lambda_k \mathfrak{C}_k^{-1}\mathfrak{B}_k\mathfrak{B}_k^t$ is the error propagation operator for one pre-smoothing step (cf. (3.25)), and we have the relations

$$\begin{aligned} \mathcal{B}((p,y), \tilde{S}_k(q,z)) &= \mathcal{B}(R_k(p,y),(q,z)) && \forall\,(p,y),(q,z) \in V_k \times V_k, \\ \|(Id_k - I_{k-1}^k \tilde{P}_k^{k-1})\tilde{S}_k^m\| &\approx \|R_k^m(Id_k - I_{k-1}^k P_k^{k-1})\|, \end{aligned}$$

that are the analogs of (5.3) and (5.4). The results for E_k in Section 5.2 also holds for $\tilde{E}_k$ by essentially identical arguments based on Lemma 4.1, Lemma 4.2, (4.14), Lemma 4.4 and Lemma 4.5.

6. Numerical Results

In this section we report numerical results of the symmetric W-cycle and V-cycle algorithms for (3.7) on two and three dimensional convex domains, where the preconditioner $\mathfrak{C}_k^{-1}$ is based on a $V(4,4)$ multigrid solve for (3.13). We employed the MATLAB/C++ toolbox FELICITY [29] in our computations.

Example 6.1. (Unit Square)

The domain Ω for this example is the unit square $(0,1)^2$. We take $\boldsymbol{\zeta} = \frac{1}{2}[1\ \ 0]^t$ and $\gamma = 0$ in (1.3), and $C_\dagger = 5$ in (3.24). The initial triangulation $\mathcal{T}_0$ is generated by the two diagonals of Ω, and the triangulations $\mathcal{T}_1, \mathcal{T}_2, \ldots$ are generated by uniform subdivisions.

The contraction numbers of the k-th level symmetric W-cycle algorithm in the energy norm with $\beta = 10^{-2}$ (resp., $\beta = 10^{-4}$ and $\beta = 10^{-6}$) are presented in Table 6.1 (resp., Table 6.2 and Table 6.3), where the number m of pre-smoothing and post-smoothing steps increases from 2^0 to 2^5.

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	2.9e-01	8.8e-02	7.8e-03	6.1e-05	6.4e-08	9.1e-17
2	6.0e-01	3.9e-01	1.9e-01	4.9e-02	1.8e-02	2.7e-03
3	4.5e-01	2.4e-01	1.0e-01	3.5e-02	1.6e-02	7.4e-03
4	3.8e-01	2.2e-01	8.7e-02	3.7e-02	2.0e-02	7.5e-03
5	3.7e-01	2.1e-01	8.2e-02	4.0e-02	2.0e-02	9.6e-03
6	3.7e-01	2.1e-01	8.1e-02	4.0e-02	2.0e-02	1.0e-02

Table 6.1. The contraction numbers of the k-th level symmetric W-cycle algorithm with m smoothing steps for $\beta = 10^{-2}$ (unit square)

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	1.2e-01	1.5e-02	2.2e-04	5.1e-08	1.0e-15	1.4e-17
2	2.3e-01	7.5e-02	5.9e-03	3.6e-05	3.9e-07	1.4e-16
3	4.9e-01	2.6e-01	7.1e-02	7.0e-03	2.9e-04	5.5e-07
4	5.5e-01	3.2e-01	1.7e-01	6.0e-02	2.4e-02	7.1e-03
5	4.1e-01	2.4e-01	1.0e-01	4.8e-02	2.4e-02	1.2e-02
6	3.8e-01	2.2e-01	8.6e-02	4.2e-02	2.2e-02	1.1e-02
7	3.7e-01	2.1e-01	8.2e-02	4.0e-02	2.1e-02	1.1e-02

Table 6.2. The contraction numbers of the k-th level symmetric W-cycle algorithm with m smoothing steps for $\beta = 10^{-4}$ (unit square)

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	2.6e-01	6.8e-02	4.6e-03	2.1e-05	6.5e-11	1.3e-16
2	3.9e-01	1.7e-01	3.0e-02	8.9e-04	9.8e-07	8.2e-14
3	2.4e-01	5.6e-02	3.1e-03	8.9e-06	2.8e-11	1.3e-16
4	3.8e-01	1.4e-01	2.3e-02	1.0e-03	2.4e-06	3.3e-12
5	7.0e-01	4.9e-01	2.8e-01	1.3e-01	3.3e-02	6.0e-03
6	4.9e-01	2.9e-01	1.4e-01	6.0e-02	2.8e-02	1.1e-02
7	4.0e-01	2.3e-01	9.4e-02	4.5e-02	2.4e-02	1.2e-02
8	3.7e-01	2.1e-01	8.4e-02	4.1e-02	2.1e-02	1.1e-02

Table 6.3. The contraction numbers of the k-th level symmetric W-cycle algorithmwith m smoothing steps for $\beta = 10^{-6}$ (unit square)

We observe that the symmetric W-cycle algorithm is a contraction with $m = 1$ for all three choices of β, and the behavior of the contraction numbers as k and m vary agree with Remark 5.7. The robustness with respect to β and k is also clearly observed.

The times for one iteration of the symmetric W-cycle algorithm at level 7 (where there are roughly 6×10^4 dofs) are reported in Table 6.4. They are proportional to the number of smoothing steps, which confirms that this is an $O(n)$ algorithm.

m	2^0	2^1	2^2	2^3	2^4	2^5
Times (s)	3.0e-1	5.4e-1	1.0e+0	2.0e+0	4.0e+0	7.9e+0

Table 6.4. The times for one iteration of the symmetric W-cycle algorithm with m smoothing steps at level 7 (unit square)

We have also computed the contraction numbers for the k-th level symmetric V-cycle algorithm, which are similar to those of the W-cycle algorithm. For brevity we only present the results for $k = 1, \ldots, 7$, $\beta = 10^{-2}, 10^{-4}, 10^{-6}$ and $m = 2^0, 2^1, 2^2$ in Table 6.5. Again we observe that the V-cycle algorithm is a contraction for $m = 1$ and the contraction numbers are robust with respect to both β and k.

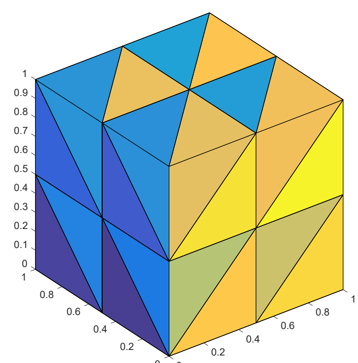
Example 6.2. (Unit Cube)

The domain for this example is the unit cube $(0,1)^3$. We take $\boldsymbol{\zeta} = \frac{1}{2}[1\ \ 1\ \ 1]^t$ and $\gamma = 0$ in (1.3), and $C_\dagger = 4$ in (3.24). The triangulations $\mathcal{T}_0$ and $\mathcal{T}_1$ are depicted in Figure 6.1. The number of grid points in all directions are doubled in each refinement and the triangulations inside the cubic subdomains at all levels are similar to one another.

The contraction numbers of the k-th level symmetric W-cycle algorithm in the energy norm with $\beta = 10^{-2}$ (resp., $\beta = 10^{-4}$ and $\beta = 10^{-6}$) are displayed in Table 6.6 (resp., Table 6.7 and Table 6.8), where the number m of pre-smoothing and post-smoothing steps

m \ k	1	2	3	4	5	6	7	Time (s)
$\beta = 10^{-2}$								
2^0	2.94e-01	6.01e-01	5.58e-01	5.38e-01	5.33e-01	5.28e-01	5.12e-01	7.01e-02
2^1	8.84e-02	3.87e-01	3.44e-01	3.31e-01	3.01e-01	2.93e-01	2.76e-01	1.29e-01
2^2	7.81e-03	1.86e-01	1.67e-01	1.55e-01	1.33e-01	1.31e-01	1.29e-01	2.44e-01
$\beta = 10^{-4}$								
2^0	1.21e-01	2.31e-01	4.88e-01	5.46e-01	4.94e-01	4.86e-01	4.85e-01	7.11e-02
2^1	1.47e-02	7.59e-02	2.55e-01	3.20e-01	3.18e-01	3.17e-01	3.16e-01	1.30e-01
2^2	2.17e-04	5.73e-03	7.18e-02	1.68e-01	1.73e-01	1.75e-01	1.75e-01	2.51e-01
$\beta = 10^{-6}$								
2^0	2.56e-01	3.91e-01	2.36e-01	3.71e-01	7.03e-01	6.31e-01	6.03e-01	7.14e-02
2^1	6.79e-02	1.68e-01	5.61e-02	1.42e-01	4.93e-01	4.12e-01	4.03e-01	1.30e-01
2^2	4.61e-03	3.09e-02	3.13e-03	2.35e-02	2.82e-01	2.54e-01	2.48e-01	2.51e-01

TABLE 6.5. The contraction numbers of the k-th level symmetric V-cycle algorithm with $\beta = 10^{-2}, 10^{-4}, 10^{-6}$ and $m = 2^0, 2^1, 2^2$, together with the times for one iteration of the V-cycle algorithm at level 7 (unit square)



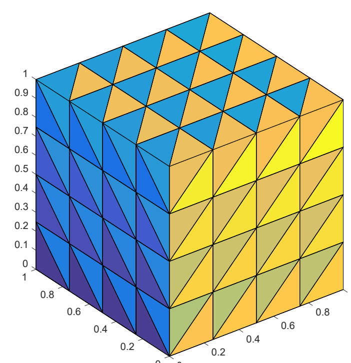


FIGURE 6.1. Triangulations $\mathcal{T}_0$ and $\mathcal{T}_1$ for the unit cube

increases from 2^0 to 2^5. We observe that the symmetric W-cycle algorithm is a contraction for $m = 1$. The behavior of the contraction numbers agree with Remark 5.7, and the contraction numbers are robust with respect to both β and k. The times for one iteration of the W-cycle algorithm at level 5 (where there are roughly 5×10^5 dofs) are reported in Table 6.9. They are proportional to m, which confirms the $O(n)$ complexity of the algorithm.

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	4.7e-01	2.2e-01	5.2e-02	5.0e-03	4.6e-05	7.2e-08
2	6.7e-01	4.7e-01	2.8e-01	1.3e-01	5.9e-02	1.9e-02
3	6.0e-01	4.2e-01	2.5e-01	1.5e-01	7.1e-02	2.8e-02
4	5.6e-01	4.0e-01	2.5e-01	1.4e-01	7.5e-02	3.4e-02
5	5.6e-01	3.9e-01	2.5e-01	1.4e-01	7.7e-02	3.7e-02

TABLE 6.6. The contraction numbers of the k-th level symmetric W-cycle algorithm with m smoothing steps for $\beta = 10^{-2}$ (unit cube)

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	2.3e-01	5.4e-02	2.9e-03	8.3e-06	5.8e-12	1.3e-16
2	4.8e-01	2.6e-01	9.3e-02	1.8e-02	4.7e-04	4.9e-07
3	4.9e-01	3.3e-01	1.9e-01	8.3e-02	2.3e-02	3.0e-03
4	6.5e-01	4.8e-01	3.1e-01	1.9e-01	9.6e-02	4.3e-02
5	5.9e-01	4.2e-01	2.7e-01	1.6e-01	9.1e-02	4.2e-02
6	5.6e-01	4.0e-01	2.6e-01	1.5e-01	8.2e-02	4.3e-02

TABLE 6.7. The contraction numbers of the k-th level symmetric W-cycle algorithm with m smoothing steps for $\beta = 10^{-4}$ (unit cube)

k \ m	2^0	2^1	2^2	2^3	2^4	2^5
1	2.9e-01	8.5e-02	7.4e-03	5.3e-05	4.5e-07	1.6e-16
2	2.7e-01	7.2e-02	5.2e-03	2.3e-05	2.1e-11	1.9e-16
3	4.7e-01	1.9e-01	4.4e-02	2.0e-03	5.3e-06	1.4e-12
4	5.2e-01	3.4e-01	1.7e-01	5.1e-02	5.6e-03	8.6e-05
5	7.6e-01	6.0e-01	4.3e-01	2.7e-01	1.4e-01	5.9e-02
6	6.7e-01	4.9e-01	3.2e-01	1.9e-01	1.1e-01	5.5e-02
7	5.8e-01	4.1e-01	2.6e-01	1.6e-01	8.9e-01	4.7e-02

TABLE 6.8. The contraction numbers of the k-th level symmetric W-cycle algorithm with m smoothing steps for $\beta = 10^{-6}$ (unit cube)

The performance of the symmetric V-cycle algorithm is similar and we only present the numerical results for $m = 2^0$, 2^1 and 2^2 in Table 6.10. Again the symmetric V-cycle algorithm is a contraction for $m = 1$.

m	2^0	2^1	2^2	2^3	2^4	2^5
Times (s)	8.3e-1	1.5e+0	2.8e+0	5.4e+0	1.1e+1	2.1e+1

TABLE 6.9. The times for one iteration of the symmetric W-cycle algorithm with m smoothing steps at level 5 (unit cube)

m \ k	1	2	3	4	5	Time (s)
$\beta = 10^{-2}$						
2^0	4.74e-01	6.71e-01	7.03e-01	7.11e-01	7.13e-01	7.13e-01
2^1	2.25e-01	4.76e-01	5.23e-01	5.39e-01	5.42e-01	1.28e+00
2^2	5.15e-02	2.76e-01	3.36e-01	3.58e-01	3.65e-01	2.39e+00
$\beta = 10^{-4}$						
2^0	2.32e-01	4.85e-01	5.54e-01	6.52e-01	7.00e-01	7.62e-01
2^1	5.41e-02	2.59e-01	3.61e-01	4.77e-01	5.37e-01	1.29e+00
2^2	2.92e-03	9.33e-02	2.00e-01	3.23e-01	3.72e-01	2.41e+00
$\beta = 10^{-6}$						
2^0	2.91e-01	2.65e-01	4.35e-01	5.38e-01	7.64e-01	7.22e-01
2^1	8.51e-02	7.09e-02	1.97e-01	3.49e-01	5.97e-01	1.30e+00
2^2	4.29e-03	5.23e-03	4.37e-02	1.74e-01	4.31e-01	2.44e+00

TABLE 6.10. The contraction numbers of the k-th level symmetric V-cycle algorithm with $\beta = 10^{-2}, 10^{-4}, 10^{-6}$ and $m = 2^0, 2^1, 2^2$, together with the times for one iteration of the V-cycle algorithm at level 5 (unit cube)

7. Concluding Remarks

In this paper we developed multigrid algorithms for the first order optimality system of a model linear-quadratic elliptic optimal control problem where the state equation contains a convective/advective term, and proved that for convex domains the W-cycle algorithm with a sufficiently large number of smoothing steps is uniformly convergent with respect to mesh refinements and a regularizing parameter. The theoretical estimates and the performance of the algorithms are demonstrated by numerical results.

Numerical results also indicate that our multigrid algorithms are robust for nonconvex domains. For the L-shaped domain $\Omega = (0,1)^2 \setminus [0.5,1)^2$ with $\boldsymbol{\zeta} = \frac{1}{2}[1\ \ 0]^t$, $\gamma = 0$ and $C_\dagger = 5$, the contraction numbers for the symmetric V-cycle (resp., W-cycle) algorithm in the energy norm with 1 pre-smoothing step and 1 post-smoothing step can be found in Table 7.1 (resp., Table 7.2), where the preconditioner is based on a $V(1,1)$ solve for (3.13).

The times for one iteration of the multigrid algorithms at level 6 (where there are roughly 5×10^4 dofs) are also included in Table 7.1 and Table 7.2.

β \ k	1	2	3	4	5	6	Time
10^{-2}	7.97e-01	7.85e-01	7.89e-01	7.93e-01	7.96e-01	7.99e-01	4.70e-02
10^{-4}	2.18e-01	4.67e-01	7.56e-01	7.57e-01	7.64e-01	7.71e-01	4.73e-02
10^{-6}	4.02e-01	1.62e-01	4.20e-01	8.62e-01	8.40e-01	8.36e-01	4.74e-02

Table 7.1. The contraction numbers of the symmetric V-cycle algorithm with $m = 1$, together with the time (in seconds) for one iteration of the V-cycle algorithm at level 6 (L-shaped domain)

β \ k	1	2	3	4	5	6	Time
10^{-2}	7.97e-01	7.04e-01	6.32e-01	6.07e-01	6.01e-01	5.92e-01	1.56e-01
10^{-4}	2.18e-01	4.64e-01	7.54e-01	6.68e-01	6.18e-01	5.91e-01	1.57e-01
10^{-6}	4.02e-01	1.63e-01	4.06e-01	8.61e-01	7.67e-01	6.57e-01	1.59e-01

Table 7.2. The contraction numbers of the symmetric W-cycle algorithm with $m = 1$, together with the time (in seconds) for one iteration of the W-cycle algorithm at level 6 (L-shaped domain)

The extensions of our analysis to V-cycle algorithms and to nonconvex domains are ongoing projects. Another direction is to develop multigrid algorithms for optimal control problems with advection/convection dominated PDE constraints [3, 31, 32, 19, 33, 21, 1, 12, 20].

References

[1] A. Allendes, E. Hernández, and E. Otárola. A robust numerical method for a control problem involving singularly perturbed equations. *Comput. Math. Appl.*, 72:974–991, 2016.

[2] I. Babuška. The finite element method with Lagrange multipliers. *Numer. Math.*, 20:179–192, 1973.

[3] R. Becker and B. Vexler. Optimal control of the convection-diffusion equation using stabilized finite element methods. *Numer. Math.*, 106:349–367, 2007.

[4] A. Borzì and V. Schulz. Multigrid methods for PDE optimization. *SIAM Rev.*, 51:361–395, 2009.

[5] A. Borzì and V. Schulz. *Computational Optimization of Systems Governed by Partial Differential Equations.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012.

[6] J.H. Bramble and X. Zhang. The Analysis of Multigrid Methods. In P.G. Ciarlet and J.L. Lions, editors, *Handbook of Numerical Analysis, VII*, pages 173–415. North-Holland, Amsterdam, 2000.

[7] S.C. Brenner, H. Li, and L.-Y. Sung. Multigrid methods for saddle point problems: Stokes and Lamé systems. *Numer. Math.*, 128:193–216, 2014.

[8] S.C. Brenner, H. Li, and L.-Y. Sung. Multigrid methods for saddle point problems: Oseen system. *Comput. Math. Appl.*, 74:2056–2067, 2017.

[9] S.C. Brenner, D.-S. Oh, and L.-Y. Sung. Multigrid methods for saddle point problems: Darcy systems. *Numer. Math.*, 138:437–471, 2018.

[10] S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods* (*Third Edition*). Springer-Verlag, New York, 2008.

[11] F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers. *RAIRO Anal. Numér.*, 8:129–151, 1974.

[12] G. Chen, W. Hu, J. Shen, J.R. Singler, Y. Zhang, and X. Zheng. An HDG method for distributed control of convection diffusion PDEs. *J. Comput. Appl. Math.*, 343:643–661, 2018.

[13] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1978.

[14] M. Dauge. *Elliptic Boundary Value Problems on Corner Domains,* Lecture Notes in Mathematics 1341. Springer-Verlag, Berlin-Heidelberg, 1988.

[15] H.C. Elman, D.J. Silvester, and A.J. Wathen. *Finite Elements and Fast Iterative Solvers: with Applications in Incompressible Fluid Dynamics*. Oxford University Press, Oxford, second edition, 2014.

[16] M. Engel and M. Griebel. A multigrid method for constrained optimal control problems. *J. Comput. Appl. Math.*, 235:4368–4388, 2011.

[17] P. Grisvard. *Elliptic Problems in Non Smooth Domains*. Pitman, Boston, 1985.

[18] W. Hackbusch. *Multi-grid Methods and Applications*. Springer-Verlag, Berlin-Heidelberg-New York-Tokyo, 1985.

[19] M. Heinkenschloss and D. Leykekhman. Local error estimates for SUPG solutions of advection-dominated elliptic linear-quadratic optimal control problems. *SIAM J. Numer. Anal.*, 47:4607–4638, 2010.

[20] W. Hu, J. Shen, J.R. Singler, Y. Zhang, and X. Zheng. A superconvergent HDG method for distributed control of convection diffusion PDEs. *J. Sci. Comput.*, 76:1436–1457, 2018.

[21] D. Leykekhman and M. Heinkenschloss. Local error analysis of discontinuous Galerkin methods for advection-dominated elliptic linear-quadratic optimal control problems. *SIAM J. Numer. Anal.*, 50:2012–2038, 2012.

[22] J.-L. Lions. *Optimal Control of Systems Governed by Partial Differential Equations*. Springer-Verlag, New York, 1971.

[23] J. Mandel, S. McCormick, and R. Bank. Variational Multigrid Theory. In S. McCormick, editor, *Multigrid Methods, Frontiers In Applied Mathematics 3*, pages 131–177. SIAM, Philadelphia, 1987.

[24] K.-A. Mardal and R. Winther. Preconditioning discretizations of systems of partial differential equations. *Numer. Linear Algebra Appl.*, 18:1–40, 2011.

[25] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2003.

[26] J. Schöberl, R. Simon, and W. Zulehner. A robust multigrid method for elliptic optimal control problems. *SIAM J. Numer. Anal.*, 49:1482–1503, 2011.

[27] S. Takacs and W. Zulehner. Convergence analysis of all-at-once multigrid methods for elliptic control problems under partial elliptic regularity. *SIAM J. Numer. Anal.*, 51:1853–1874, 2013.

[28] F. Tröltzsch. *Optimal Control of Partial Differential Equations*. American Mathematical Society, Providence, RI, 2010.

[29] S.W. Walker. FELICITY: A MATLAB/C++ toolbox for developing finite element methods and simulation modeling. *SIAM J. Sci. Comput.*, 40:C234–C257, 2018.

[30] J. Xu and L. Zikatanov. Some observations on Babuška and Brezzi theories. *Numer. Math.*, 94:195–202, 2003.

[31] N. Yan and Z. Zhou. A priori and a posteriori error estimates of streamline diffusion finite element method for optimal control problem governed by convection dominated diffusion equation. *Numer. Math. Theory Methods Appl.*, 1:297–320, 2008.

[32] N. Yan and Z. Zhou. A priori and a posteriori error analysis of edge stabilization Galerkin method for the optimal control problem governed by convection-dominated diffusion equation. *J. Comput. Appl. Math.*, 223:198–217, 2009.

[33] Z. Zhou and N. Yan. The local discontinuous Galerkin method for optimal control problem governed by convection diffusion equations. *Int. J. Numer. Anal. Model.*, 7:681–699, 2010.

Susanne C. Brenner, Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
E-mail address: brenner@math.lsu.edu

Sijing Liu, Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
E-mail address: sliu42@lsu.edu

Li-yeng Sung, Department of Mathematics and Center for Computation and Technology, Louisiana State University, Baton Rouge, LA 70803, USA
E-mail address: sung@math.lsu.edu