**Research Article**

Sijing Liu*

# Robust Multigrid Methods for Discontinuous Galerkin Discretizations of an Elliptic Optimal Control Problem

https://doi.org/10.1515/cmam-2023-0132
Received May 16, 2023; revised January 12, 2024; accepted January 17, 2024

**Abstract:** We consider discontinuous Galerkin methods for an elliptic distributed optimal control problem, and we propose multigrid methods to solve the discretized system. We prove that the $W$-cycle algorithm is uniformly convergent in the energy norm and is robust with respect to a regularization parameter on convex domains. Numerical results are shown for both $W$-cycle and $V$-cycle algorithms.

**Keywords:** Elliptic Distributed Optimal Control Problems, General State Equations, Multigrid Methods, Discontinuous Galerkin Methods

**MSC 2020:** 49J20, 49M41, 65N30, 65N55

## 1 Introduction

In this paper, we consider the following elliptic optimal control problem. Let $\Omega$ be a bounded convex polygonal/polyhedral domain in $\mathbb{R}^n$ ($n = 2, 3$), $y_d \in L_2(\Omega)$ and let $\beta$ be a positive constant. Find

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

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

$$a(y, v) = \int_\Omega uv\,dx \quad \text{for all } v \in H^1_0(\Omega). \tag{1.2}$$

Here the bilinear form $a(\cdot\,,\cdot)$ is defined as

$$a(y, v) = \int_\Omega \nabla y \cdot \nabla v\,dx + \int_\Omega (\boldsymbol{\zeta} \cdot \nabla y)v\,dx + \int_\Omega \gamma y v\,dx, \tag{1.3}$$

where vector field $\boldsymbol{\zeta} \in [W^{1,\infty}(\Omega)]^2$ and the function $\gamma \in L_\infty(\Omega)$ is nonnegative. If $\boldsymbol{\zeta} \neq 0$, then constraint (1.2) is the weak form of a general second-order PDE with an advective/convective term. We assume

$$\gamma - \frac{1}{2}\nabla \cdot \boldsymbol{\zeta} \geq \gamma_0 > 0 \tag{1.4}$$

such that problem (1.2) is well-posed.

**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 [15, 22].

***Corresponding author: Sijing Liu,** Department of Mathematics, University Of Connecticut, Storrs, CT; and current address: The Institute for Computational and Experimental Research in Mathematics, Brown University, Providence, RI, USA, e-mail: sijing_liu@brown.edu

It is well known [35, 42] that the solution of (1.1)–(1.2) is characterized by

$$a(q,\bar{p}) = (\bar{y}-y_d, q)_{L_2(\Omega)} \quad \text{for all } q \in H_0^1(\Omega), \tag{1.5a}$$
$$\bar{p} + \beta\bar{u} = 0, \tag{1.5b}$$
$$a(\bar{y}, z) = (\bar{u}, z)_{L_2(\Omega)} \quad \text{for all } z \in H_0^1(\Omega), \tag{1.5c}$$

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

$$\begin{aligned} a(q,\bar{p}) - (\bar{y}, q)_{L_2(\Omega)} &= -(y_d, q)_{L_2(\Omega)} && \text{for all } q \in H_0^1(\Omega),\\ -(\bar{p}, z)_{L_2(\Omega)} - \beta a(\bar{y}, z) &= 0 && \text{for all } z \in H_0^1(\Omega). \end{aligned} \tag{1.6}$$

Notice that $\beta$ is only in the second equation (1.6). In order to make the equations more balanced, we perform a change of variables. Let

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

then system (1.6) becomes

$$\begin{aligned} \beta^{\frac{1}{2}} a(q,\tilde{p}) - (\tilde{y}, q)_{L_2(\Omega)} &= -\beta^{\frac{1}{4}}(y_d, q)_{L_2(\Omega)} && \text{for all } q \in H_0^1(\Omega),\\ -(\tilde{p}, z)_{L_2(\Omega)} - \beta^{\frac{1}{2}} a(\tilde{y}, z) &= 0 && \text{for all } z \in H_0^1(\Omega). \end{aligned} \tag{1.8}$$

The scaling technique (1.7) can also be found in [13, 26, 27].

In this paper, we employ a discontinuous Galerkin (DG) method [2, 3, 37] to discretize the saddle point problem (1.8). There are several advantages to use DG methods; for example, DG methods are more flexible regarding choices of meshes and more suitable for convection-dominated problems. It is well known that DG methods can capture sharp gradients in the solutions such that spurious oscillations can be avoided. Our main goal is to design multigrid methods for solving the discretized system resulting from the DG discretization that are robust with respect to the regularization parameter $\beta$. Note that we reformulate the optimal control problem into a saddle point problem by eliminating the control variable $\bar{u}$ (cf. [30]). Therefore, we do not discretize the control explicitly but recover the discrete control using the discrete analog of relation (1.5b). This is a well-known strategy to discretize the optimal control problem.

Multigrid methods for (1.8) based on continuous Galerkin methods are intensively studied in the literature, for example, in [6, 13, 36, 39–41] and the references therein. In [13], based on the approaches in [11, 12, 14], the authors developed multigrid methods for (1.8) using a continuous $P_1$ finite element method. Besides the robustness of the multigrid methods with respect to $\beta$, the estimates in [13] are established in a natural energy norm, and the multigrid methods have a standard $\mathcal{O}(m^{-1})$ performance, where $m$ is the number of smoothing steps. In this paper, we extend the results in [13] to DG methods where the diffusion term in (1.3) is discretized by a symmetric interior penalty (SIP) method and the convection term in (1.3) is discretized by an unstabilized/centered-fluxes DG method [18, 23]. Multigrid methods based on DG methods are investigated in [1, 9, 10, 16, 28, 31, 32] and the references therein. However, not much work has been done towards the multigrid methods based on DG discretizations for optimal control problems with provable results. The general idea in this paper is to construct a block-diagonal preconditioner and convert the saddle point problem (1.8) into an equivalent symmetric positive definite (SPD) problem using the preconditioner. Therefore, well-known multigrid theories for SPD system [7, 29] can be utilized. The preconditioner requires solving a reaction-diffusion equation (approximately) based on a SIP discretization. This, however, does not affect the overall optimal computational complexity of our multigrid methods since the preconditioner itself can be constructed by multigrid methods (cf. [16]).

Note that the extension of the analysis in [13] to DG methods is not trivial. First, DG methods are nonconforming methods in the sense that $V_h \not\subset V$. Consequently, the discrete bilinear form $\mathcal{B}_k$ is now different from the continuous bilinear form $\mathcal{B}$ (see Section 4). Therefore, all the definitions at the discrete level are different from the ones in [13], for example, the inter-grid transfer operators and the projection operators. Secondly, in order to establish $L_2$ error estimates for the DG methods, we need to make sure that the adjoint consistency (cf. [3]) of the DG methods hold; then a duality argument can be utilized (see Appendix B). Additionally, we need multigrid methods for DG discretization of a singularly perturbed reaction-diffusion equation to construct the crucial preconditioner $\mathfrak{C}_k^{-1}$ (cf. [16, 28]). At last, and most importantly, the proof of the approximation property (Lemma 5.3) cannot rely on the continuous problem anymore. Instead, a connection between level $k$ and level $k-1$ is exploited, as well as the connection between primal and dual problems, to prove the approximation property.

The rest of the paper is organized as follows. In Section 2, we gather some known results regarding the continuous problem. In Section 3, we discretize the optimal control problem with a DG method and establish concrete error estimates. A crucial block-diagonal preconditioner is introduced and the multigrid methods for the discretized system are described in Section 4. In Sections 5 and 6, we establish the main theorem, which is based on the smoothing and approximation properties of the multigrid methods. Finally, we provide some numerical results in Section 7 and end with some concluding remarks in Section 8. Some technical proofs are provided in Appendices A and B.

Throughout this paper, we use $C$ (with or without subscripts) to denote a generic positive constant that is independent of any mesh parameters and the regularization parameter $\beta$. Also, to avoid the proliferation of constants, we use the notation $A \lesssim B$ (or $A \gtrsim B$) to represent $A \leq (\text{constant})B$. The notation $A \approx B$ is equivalent to $A \lesssim B$ and $B \lesssim A$. Note that we do not consider the convection-dominated case in this paper; hence the constants might depend on $\boldsymbol{\zeta}$ and $\gamma$.

## 2 Continuous Problem

We rewrite (1.8) in a concise form

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

where

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

Let $\|p\|_{H^1_\beta(\Omega)}$ be defined by

$$\|p\|^2_{H^1_\beta(\Omega)} = \beta^{\frac{1}{2}}|p|^2_{H^1(\Omega)} + \|p\|^2_{L_2(\Omega)}.$$

We have the following lemmas (cf. [13]) regarding the bilinear form $\mathcal{B}$ with respect to the norm $\|\cdot\|_{H^1_\beta(\Omega)}$.

**Lemma 2.1.** *We have*

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

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

**Remark 2.2.** Lemma 2.1 guarantees the well-posedness of (1.8) by the standard theory [5, 17].

**Remark 2.3.** We also have the similar stability estimate

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

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

We also need the following regularity results (cf. [13]) on convex domains. Let $(p,y) \in H_0^1(\Omega)\times H_0^1(\Omega)$ satisfy

$$\mathcal{B}((p,y),(q,z)) = (f,q)_{L_2(\Omega)} + (g,z)_{L_2(\Omega)} \quad \text{for all } (q,z) \in H_0^1(\Omega)\times H_0^1(\Omega), \tag{2.3}$$

where $(f,g) \in L_2(\Omega)\times L_2(\Omega)$ and $\mathcal{B}$ is defined in (2.2).

**Lemma 2.4.** *The solution* $(p,y)$ *of* (2.3) *belongs to* $H^2(\Omega)\times H^2(\Omega)$, *and we have*

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

**Remark 2.5.** Lemma 2.4 is also valid for the following dual problem. Find $(p,y) \in H_0^1(\Omega)\times H_0^1(\Omega)$ such that

$$\mathcal{B}((q,z),(p,y)) = (f,q)_{L_2(\Omega)} + (g,z)_{L_2(\Omega)} \quad \text{for all } (q,z) \in H_0^1(\Omega)\times H_0^1(\Omega). \tag{2.4}$$

# 3 Discrete Problem

In this section, we discretize the saddle point problem (1.8) by a DG method [2, 3, 18]. Let $\mathcal{T}_h$ be a shape regular simplicial triangulation of $\Omega$. The diameter of $T \in \mathcal{T}_h$ is denoted by $h_T$, and $h = \max_{T\in\mathcal{T}_h} h_T$ is the mesh diameter. Let $\mathcal{E}_h = \mathcal{E}_h^b \cup \mathcal{E}_h^i$, where $\mathcal{E}_h^i$ (resp. $\mathcal{E}_h^b$) represents the set of interior edges (resp. boundary edges).

We further decompose the boundary edges $\mathcal{E}_h^b$ into the inflow part $\mathcal{E}_h^{b,-}$ and the outflow part $\mathcal{E}_h^{b,+}$, which are defined as follows:

$$\mathcal{E}_h^{b,-} = \{e \in \mathcal{E}_h^b : e \subset \{x \in \partial\Omega : \boldsymbol{\zeta}(x)\cdot\mathbf{n}(x) < 0\}\}, \quad \mathcal{E}_h^{b,+} = \mathcal{E}_h^b \setminus \mathcal{E}_h^{b,-}.$$

For an edge $e \in \mathcal{E}_h^i$, let $h_e$ be the length of $e$. For each edge, we associate a fixed unit normal $\mathbf{n}$. We denote by $T^+$ the element for which $\mathbf{n}$ is the outward normal, and $T^-$ the element for which $-\mathbf{n}$ is the outward normal. We define the discontinuous finite element space $V_h$ as

$$V_h = \{v \in L_2(\Omega) : v|_T \in \mathbb{P}_1(T) \text{ for all } T \in \mathcal{T}_h\}.$$

For $v \in V_h$ on an edge $e$, we define $v^+ = v|_{T^+}$ and $v^- = v|_{T^-}$. We define the jump and average for $v \in V_h$ on an edge $e$ as follows:

$$[v] = v^+ - v^-, \quad \{v\} = \frac{v^+ + v^-}{2}.$$

For $e \in \mathcal{E}_h^b$ with $e \in \partial T$, we let $[v] = \{v\} = v|_T$. We also denote

$$(w, v)_e := \int_e wv\,ds \quad\text{and}\quad (w, v)_T := \int_T wv\,dx.$$

## 3.1 Discontinuous Galerkin Methods

The DG methods for (2.1) is to find $(\tilde{p}_h, \tilde{y}_h) \in V_h \times V_h$ such that

$$\mathcal{B}_h((\tilde{p}_h, \tilde{y}_h), (q, z)) = -\beta^{\frac{1}{4}}(y_d, q)_{L_2(\Omega)} \quad \text{for all } (q, z) \in V_h \times V_h, \tag{3.1}$$

where

$$\mathcal{B}_h((p, y), (q, z)) = \beta^{\frac{1}{2}} a_h(q, p) - (y, q)_{L_2(\Omega)} - (p, z)_{L_2(\Omega)} - \beta^{\frac{1}{2}} a_h(y, z). \tag{3.2}$$

The bilinear form $a_h(\cdot,\cdot)$ is defined by

$$a_h(u, v) = a_h^{\mathrm{sip}}(u, v) + a_h^{\mathrm{ar}}(u, v) \quad \text{for all } u, v \in V_h, \tag{3.3}$$

where

$$a_h^{\mathrm{sip}}(u, v) = \sum_{T\in\mathcal{T}_h}(\nabla u, \nabla v)_T - \sum_{e\in\mathcal{E}_h}(\{\mathbf{n}\cdot\nabla u\}, [v])_e - \sum_{e\in\mathcal{E}_h}(\{\mathbf{n}\cdot\nabla v\}, [u])_e + \sigma\sum_{e\in\mathcal{E}_h} h_e^{-1}([u], [v])_e \tag{3.4}$$

is the bilinear form of the SIP method with sufficiently large penalty parameter $\sigma$ and the unstabilized DG scheme (cf. [18, 24]) for the advection-reaction term is defined as

$$a_h^{\mathrm{ar}}(u, v) = \sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla u + \gamma u, v)_T - \sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}(\mathbf{n}\cdot\boldsymbol{\zeta}[u], \{v\})_e. \tag{3.5}$$

**Remark 3.1.** We do not consider convection-dominated case in this paper. Therefore, the bilinear form $a_h^{\mathrm{ar}}(\cdot,\cdot)$ does not contain any upwind stabilization terms. However, if we consider convection-dominated case, the well-known upwind schemes [4, 18, 34] could be utilized.

It is also necessary to consider the general problem (2.3) and the dual problem (2.4). The DG method for (2.3) is to find $(p_h, y_h) \in V_h \times V_h$ such that

$$\mathcal{B}_h((p_h, y_h), (q, z)) = (f, q)_{L_2(\Omega)} + (g, z)_{L_2(\Omega)} \quad \text{for all } (q, z) \in V_h \times V_h. \tag{3.6}$$

Similarly, the DG method for (2.4) is to find $(p_h, y_h) \in V_h \times V_h$ such that

$$\mathcal{B}_h((q,z),(p_h,y_h)) = (f,q)_{L_2(\Omega)} + (g,z)_{L_2(\Omega)} \quad \text{for all } (q,z) \in V_h \times V_h. \tag{3.7}$$

Let the norm $\|\cdot\|_{H^1_\beta(\Omega;\mathcal{T}_h)}$ be defined as

$$\|p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} = \beta^{\frac{1}{2}}\|p\|^2_{1,h} + \|p\|^2_{L_2(\Omega)}, \tag{3.8}$$

where

$$\|p\|^2_{1,h} = \sum_{T\in\mathcal{T}_h}\|\nabla p\|^2_{L_2(T)} + \sum_{e\in\mathcal{E}_h}\frac{1}{h_e}\|[p]\|^2_{L_2(e)} + \sum_{e\in\mathcal{E}_h} h_e\|\{\mathbf{n}\cdot\nabla p\}\|^2_{L_2(e)}.$$

Let $V = H^2(\Omega)\cap H^1_0(\Omega)$. For the bilinear form $a_h(\cdot,\cdot)$, we have

$$a_h(w,v) \le C\|w\|_{1,h}\|v\|_{1,h} \quad \text{for all } w,v \in V + V_h, \tag{3.9}$$

$$a_h(v,v) \ge C\|v\|^2_{1,h} \quad \text{for all } w,v \in V_h, \tag{3.10}$$

for sufficiently large $\sigma$. A proof is provided in Appendix A. Note that the constants in (3.9)–(3.10) might depend on $\boldsymbol{\zeta}$ and $\gamma$.

It follows from (3.2), (3.8), (3.9) and the Cauchy–Schwarz inequality that

$$\mathcal{B}_h((p,y),(q,z)) \le (\|p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)})^{\frac{1}{2}}(\|q\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|z\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)})^{\frac{1}{2}} \tag{3.11}$$

for any $(p,y),(q,z) \in (V+V_h)\times(V+V_h)$.

We also have, by (3.2), (3.10) and a direct calculation,

$$\begin{aligned}\mathcal{B}_h((p,y),(p-y,-y-p)) &= \beta^{\frac{1}{2}}a_h(p,p) + (p,p)_{L_2(\Omega)} + \beta^{\frac{1}{2}}a_h(y,y) + (y,y)_{L_2(\Omega)}\\ &\gtrsim \|p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)}\end{aligned} \tag{3.12}$$

and

$$\|p-y\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|-y-p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} = 2(\|p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)}) \tag{3.13}$$

by the parallelogram law. It follows from (3.11)–(3.13) that

$$\|p_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} \approx \sup_{(q,z)\in V_h\times V_h}\frac{\mathcal{B}_h((p_h,y_h),(q,z))}{\|q\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|z\|_{H^1_\beta(\Omega;\mathcal{T}_h)}} \quad \text{for all } (p_h,y_h) \in V_h\times V_h. \tag{3.14}$$

Similarly, we have

$$\|p_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} \approx \sup_{(q,z)\in V_h\times V_h}\frac{\mathcal{B}_h((q,z),(p_h,y_h))}{\|q\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|z\|_{H^1_\beta(\Omega;\mathcal{T}_h)}} \quad \text{for all } (p_h,y_h) \in V_h\times V_h. \tag{3.15}$$

It follows immediately from (3.14) and (3.15) that the discrete problems (3.6) and (3.7) are uniquely solvable. We also need the following lemma, which follows from a standard inverse estimate (cf. [15]) and trace inequalities (cf. [21, Proposition 3.1]),

**Lemma 3.2.** *Assuming $\mathcal{T}_h$ is a quasi-uniform triangulation of $\Omega$, we have the following inverse estimate:*

$$\|v\|_{1,h} \lesssim h^{-1}\|v\|_{L_2(\Omega)} \quad \textit{for all } v \in V_h. \tag{3.16}$$

## 3.2 Interpolation Operator $\Pi_h$

We use the usual continuous nodal interpolant (which belongs to $V_h$) [3, 15, 37] such that the following estimates hold.

**Lemma 3.3.** *We have*

$$\|z-\Pi_h z\|_{L_2(\Omega)} + h|z-\Pi_h z|_{H^1(\Omega)} \lesssim h^2|z|_{H^2(\Omega)} \quad \textit{for all } z \in H^2(\Omega)\cap H^1_0(\Omega).$$

**Remark 3.4.** We also have (cf. [15])

$$\|z-\Pi_h z\|_{1,h} \lesssim h|z|_{H^2(\Omega)} \quad \text{for all } z \in H^2(\Omega)\cap H^1_0(\Omega).$$

### 3.3 Error Estimates

It is well known [3, 18, 37] that the DG method (3.1) is consistent. Hence, we have the following Galerkin orthogonality:

$$\mathcal{B}_h((p-p_h, y-y_h),(q,z)) = 0 \quad \text{for all } (q,z) \in V_h \times V_h. \tag{3.17}$$

**Lemma 3.5.** *Let the functions* $(p,y)$ *(resp.* $(p_h,y_h)$*) be the solutions of* (2.3) *or* (2.4) *(resp.* (3.6) *or* (3.7)*), we have*

$$\|p-p_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-y_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} \le C(\beta^{\frac12}h^{-2}+1)^{\frac12}\beta^{-\frac12}h^2(\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}), \tag{3.18}$$

$$\|p-p_h\|_{L_2(\Omega)} + \|y-y_h\|_{L_2(\Omega)} \le C(\beta^{\frac12}h^{-2}+1)\beta^{-1}h^4(\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)}). \tag{3.19}$$

*Proof.* We only establish the estimates involving (2.3) and (3.6). Using estimate (3.14) and relation (3.17), we have

$$\|p-p_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-y_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} \lesssim \inf_{(q,z)\in V_h\times V_h}(\|p-q\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-z\|_{H^1_\beta(\Omega;\mathcal{T}_h)}).$$

By Lemma 3.3 and Lemma 2.4, we have

$$\beta^{\frac12}\|p-\Pi_h p\|^2_{1,h} \le C\beta^{\frac12}h^2\|p\|^2_{H^2(\Omega)} \le \frac{C}{\beta^{\frac12}}h^2(\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)})^2,$$

$$\|p-\Pi_h p\|^2_{L_2(\Omega)} \le Ch^4\|p\|^2_{H^2(\Omega)} \le \frac{C}{\beta}h^4(\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)})^2.$$

Thus we obtain

$$\|p-\Pi_h p\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} \le C(\beta^{-\frac12}h^2 + \beta^{-1}h^4)(\|f\|_{L_2(\Omega)} + \|g\|_{L_2(\Omega)})^2.$$

It is similar to estimate $\|y-\Pi_h y\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)}$. Therefore, we have estimate (3.18). Estimate (3.19) is established by a duality argument. Let $(\xi,\theta)$ satisfy

$$\begin{aligned} \beta^{\frac12}(-\Delta\xi + \boldsymbol{\zeta}\cdot\nabla\xi + \gamma\xi) - \theta &= p - p_h, \quad \xi = 0 \quad \text{on } \partial\Omega,\\ -\xi - \beta^{\frac12}(-\Delta\theta - \boldsymbol{\zeta}\cdot\nabla\theta + (\gamma - \nabla\cdot\boldsymbol{\zeta})\theta) &= y - y_h, \quad \theta = 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{3.20}$$

The weak form of (3.20) is to find $(\xi,\theta) \in H^1_0(\Omega)\times H^1_0(\Omega)$ such that

$$\mathcal{B}((q,z),(\xi,\theta)) = (q, p-p_h)_{L_2(\Omega)} + (z, y-y_h)_{L_2(\Omega)} \quad \text{for all } (q,z)\in H^1_0(\Omega)\times H^1_0(\Omega). \tag{3.21}$$

We can show that (cf. [33])

$$\|p-p_h\|^2_{L_2(\Omega)} + \|y-y_h\|^2_{L_2(\Omega)} = \mathcal{B}_h((p-p_h, y-y_h),(\xi,\theta)). \tag{3.22}$$

A proof is provided in Appendix B.

Then it follows from the Cauchy–Schwarz inequality, (3.17), (3.22) and Lemma 2.4 (apply to (3.21)) that

$$\begin{aligned} \|p-p_h\|^2_{L_2(\Omega)} + \|y-y_h\|^2_{L_2(\Omega)} &= \mathcal{B}_h((p-p_h, y-y_h),(\xi,\theta))\\ &= \mathcal{B}_h((p-p_h, y-y_h),(\xi-\Pi_h\xi, \theta-\Pi_h\theta))\\ &\lesssim (\|p-p_h\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-y_h\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)})^{\frac12}\\ &\qquad \times (\|\xi-\Pi_h\xi\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|\theta-\Pi_h\theta\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)})^{\frac12}\\ &\lesssim (\beta^{-\frac12}h^2 + \beta^{-1}h^4)^{\frac12}(\|p-p_h\|^2_{L_2(\Omega)} + \|y-y_h\|^2_{L_2(\Omega)})^{\frac12}\\ &\qquad \times (\|p-p_h\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-y_h\|^2_{H^1_\beta(\Omega;\mathcal{T}_h)})^{\frac12}. \end{aligned} \tag{3.23}$$

Estimate (3.19) then follows from (3.23) and (3.18). □

**Remark 3.6.** Note that the performance of the DG method (3.1) in the norms $\|\cdot\|_{1,h}$ and $\|\cdot\|_{L_2(\Omega)}$ will deteriorate when $\beta$ goes to 0 due to Lemma 3.5. Therefore, a very fine mesh should be used when $\beta$ is small, in which case it is necessary to have an efficient iterative solver.

# 4 Multigrid Methods

In this section, we introduce the multigrid methods for (2.3). A crucial block-diagonal preconditioner $\mathfrak{C}_k$ is introduced. The general idea is to convert the saddle point problem into an SPD problem. Then the well-established multigrid theories for SPD system [7, 16, 29] can be utilized. We omit some proofs since they are identical to those of [13].

## 4.1 Set-Up

Let the triangulation $\mathcal{T}_1, \mathcal{T}_2, \dots$ be generated from the triangulation $\mathcal{T}_0$ through uniform subdivisions such that $h_k \approx \frac{1}{2} h_{k-1}$ and let $V_k$ be the DG space associated with $\mathcal{T}_k$. Our goal is to design multigrid methods for the problem of finding $(p_k, y_k) \in V_k \times V_k$ such that

$$\mathcal{B}_k((p_k, y_k), (q, z)) = F(q) + G(z) \quad \text{for all } (q, z) \in V_k \times V_k, \tag{4.1}$$

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

$$\mathcal{B}_k((q, z), (p_k, y_k)) = F(q) + G(z) \quad \text{for all } (q, z) \in V_k \times V_k. \tag{4.2}$$

Here $\mathcal{B}_k$ represents the bilinear form $\mathcal{B}_h$ on $V_k \times V_k$.

Let $(\cdot, \cdot)_k$ be a mesh-dependent inner product on $V_k$,

$$(v, w)_k = h_k^n \sum_{i=1}^{n_k} v(p_i) w(p_i) \quad \text{for all } v, w \in V_k,$$

where $h_k = \max_{T \in \mathcal{T}_k} \operatorname{diam} T$ and $\{p_i\}_{i=1}^{n_k}$ are the nodes in $\mathcal{T}_k$.

**Remark 4.1.** It can be shown that $(v, v)_k \approx \|v\|_{L_2(\Omega)}^2$ for $v \in V_k$ (cf. [15]).

Then the mesh-dependent inner product $[\cdot, \cdot]_k$ on $V_k \times V_k$ is defined by

$$[(p, y), (q, z)]_k = (p, q)_k + (y, z)_k.$$

It is easy to see that

$$[(p, y), (p, y)]_k \approx \|p\|_{L_2(\Omega)}^2 + \|y\|_{L_2(\Omega)}^2 \quad \text{for all } (p, y) \in V_k \times V_k. \tag{4.3}$$

The coarse-to-fine operator $I_{k-1}^k \colon V_{k-1} \times V_{k-1} \to V_k \times V_k$ is the natural injection and the fine-to-coarse operator $I_k^{k-1} \colon V_k \times V_k \to V_{k-1} \times V_{k-1}$ is the transpose of $I_{k-1}^k$ with respect to the mesh-dependent inner product, namely,

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

Let the system operator $\mathfrak{B}_k \colon V_k \times V_k \to V_k \times V_k$ be defined by

$$[\mathfrak{B}_k(p, y), (q, z)]_k = \mathcal{B}_k((p, y), (q, z)) \quad \text{for all } (p, y), (q, z) \in V_k \times V_k.$$

Then the $k$-th level problem (4.1) is equivalent to

$$\mathfrak{B}_k(p, y) = (f, g), \tag{4.4}$$

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

$$[(f, g), (q, z)]_k = F(q) + G(z) \quad \text{for all } (q, z) \in V_k \times V_k,$$

and the dual problem (4.2) becomes

$$\mathfrak{B}_k^t(p, y) = (f, g). \tag{4.5}$$

Here, for all $(p, y), (q, z) \in V_k \times V_k$, we have

$$[\mathfrak{B}_k^t(p, y), (q, z)]_k = [(p, y), \mathfrak{B}_k(q, z)]_k = \mathcal{B}_k((q, z), (p, y)).$$

## 4.2 A Block-Diagonal Preconditioner

Let $L_k\colon V_k \to V_k$ be an operator that is SPD with respect to $(\cdot,\cdot)_k$ and satisfies

$$(L_k v, v)_k \approx \beta^{\frac{1}{2}}\|v\|_{1,h}^2 + \|v\|_{L_2(\Omega)}^2 \quad \text{for all } v \in V_k.$$

Then the operator $\mathfrak{C}_k\colon V_k \times V_k \to V_k \times V_k$ given by $\mathfrak{C}_k(p,y) = (L_k p, L_k y)$ is SPD with respect to $[\cdot,\cdot]_k$ and we have

$$[\mathfrak{C}_k(p,y),(p,y)]_k \approx \|p\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 + \|y\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 \quad \text{for all } (p,y) \in V_k \times V_k.$$

Here the hidden constant is independent of $k$ and $\beta$.

**Remark 4.2.** In practice, we use $\mathfrak{C}_k^{-1}$ as a block preconditioner. The operation $L_k^{-1}\phi$ can be computed approximately by solving the following boundary value problem:

$$\begin{aligned} -\beta^{\frac{1}{2}}\Delta u + u &= \phi && \text{in } \Omega,\\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{4.6}$$

using a SIP discretization. This can be constructed by multigrid [16, 28].

**Lemma 4.3.** *We have*

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k \approx \|p\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 + \|y\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 \quad \textit{for all } (p,y) \in V_k \times V_k, \tag{4.7}$$

$$[\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t(p,y),(p,y)]_k \approx \|p\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 + \|y\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 \quad \textit{for all } (p,y) \in V_k \times V_k. \tag{4.8}$$

**Lemma 4.4.** *There exists positive constants $C_{\min}$ and $C_{\max}$, independent of $k$ and $\beta$, such that*

$$\lambda_{\min}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k) \geq C_{\min}, \qquad \lambda_{\min}(\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t) \geq C_{\min}, \tag{4.9}$$

$$\lambda_{\max}(\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k) \leq C_{\max}(\beta^{\frac{1}{2}}h_k^{-2}+1), \quad \lambda_{\max}(\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t) \leq C_{\max}(\beta^{\frac{1}{2}}h_k^{-2}+1). \tag{4.10}$$

*Proof.* We only prove the estimates involving $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$; other estimates are similar. It follows from (4.7) and (4.3) that

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k \geq \|p\|_{L_2(\Omega)}^2 + \|y\|_{L_2(\Omega)}^2 \approx [(p,y),(p,y)]_k.$$

Estimate (4.9) then is trivial by the Rayleigh quotient formula. We also have, by (4.7), (3.16) and (4.3), that

$$[\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k(p,y),(p,y)]_k \lesssim (\beta^{\frac{1}{2}}h_k^{-2}+1)[(p,y),(p,y)]_k.$$

Estimate (4.10) is immediate by the Rayleigh quotient formula. □

**Remark 4.5.** Lemma 4.4 implies that the operators $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$, $\mathfrak{B}_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t$ are well-conditioned when $\beta^{\frac{1}{2}}h_k^{-2} \leq 1$.

## 4.3 *W*-Cycle Algorithm

Let the output of the $W$-cycle algorithm for (4.4) 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 $\mathcal{B}_0^{-1}(f,g)$. For $k \geq 1$, we compute $MG_W(k,(f,g),(p_0,y_0),m_1,m_2)$ in three steps.

- *Pre-Smoothing*. The approximate solutions $(p_1,y_1),\dots,(p_{m_1},y_{m_1})$ are computed recursively by

$$(p_j,y_j) = (p_{j-1},y_{j-1}) + \lambda_k\mathfrak{C}_k^{-1}\mathfrak{B}_k^t((f,g)-\mathfrak{B}_k(p_{j-1},y_{j-1})) \tag{4.11}$$

for $1 \leq j \leq m_1$. The choice of the damping factor $\lambda_k$ is determined by the following criteria:

$$\lambda_k = \frac{C}{\beta^{\frac{1}{2}}h_k^{-2}+1} \quad \text{when } \beta^{\frac{1}{2}}h_k^{-2} \geq 1, \tag{4.12a}$$

$$\lambda_k = \frac{2}{\lambda_{\min}+\lambda_{\max}} \quad \text{when } \beta^{\frac{1}{2}}h_k^{-2} < 1, \tag{4.12b}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ are the smallest and largest eigenvalues of $\mathfrak{B}_k^t\mathfrak{C}_k^{-1}\mathfrak{B}_k$ respectively.

- *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 compute $(p'_1, y'_1), (p'_2, y'_2) \in V_{k-1} \times V_{k-1}$ by

$$(p'_1, y'_1) = \mathrm{MG}_W(k-1, (f', g'), (0,0), m_1, m_2),$$
$$(p'_2, y'_2) = \mathrm{MG}_W(k-1, (f', g'), (p'_1, y'_1), m_1, m_2).$$

  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.* The approximate solutions $(p_{m_1+2}, y_{m_1+2}), \dots, (p_{m_1+m_2+1}, y_{m_1+m_2+1})$ are computed recursively by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{B}_k^t \mathfrak{C}_k^{-1}((f, g) - \mathfrak{B}_k(p_{j-1}, y_{j-1})) \tag{4.13}$$

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

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

**Remark 4.6.** The choice of (4.12a) is motivated by Lemma 4.4 such that $\lambda_{\max}(\lambda_k \mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k) \le 1$. The choice of (4.12b) is motivated by the optimal choice of Richardson iteration [38] and the well-conditioning of $\mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k$ (cf. Remark 4.5).

**Remark 4.7.** Note that the post-smoothing step is the Richardson iteration of the SPD system

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

which is equivalent to (4.4).

### 4.4 *V*-Cycle Algorithm

Let the output of the $V$-cycle algorithm for (4.4) with initial guess $(p_0, y_0)$ and $m_1$ (resp. $m_2$) pre-smoothing (resp. post-smoothing) steps be denoted by $\mathrm{MG}_V(k, (f, g), (p_0, y_0), m_1, m_2)$.

The computation of $\mathrm{MG}_V(k, (f, g), (p_0, y_0), m_1, m_2)$ differs from the computation of the $W$-cycle algorithm only in the coarse grid correction step, where we compute

$$(p'_1, y'_1) = \mathrm{MG}_V(k-1, (f', g'), (0,0), m_1, m_2)$$

and take $(p_{m_1+1}, y_{m_1+1})$ to be $(p_{m_1}, y_{m_1}) + I_{k-1}^k(p'_1, y'_1)$.

### 4.5 Multigrid Algorithms for (4.5)

We define $W$-cycle and $V$-cycle algorithms for (4.5) by simply interchanging $\mathfrak{B}_k^t$ and $\mathfrak{B}_k$ in Sections 4.3 and 4.4. The pre-smoothing step is given by

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

and the post-smoothing step is given by

$$(p_j, y_j) = (p_{j-1}, y_{j-1}) + \lambda_k \mathfrak{B}_k \mathfrak{C}_k^{-1}((f, g) - \mathfrak{B}_k^t(p_{j-1}, y_{j-1})). \tag{4.14}$$

## 5 Smoothing and Approximation Properties

In this section, we establish the smoothing property and the approximation property of the $W$-cycle algorithm. These results can then be used to establish the convergence of $W$-cycle algorithm as in [13]. We omit some proofs since they are identical to those of [13].

## 5.1 A Scale of Mesh-Dependent Norms

For $0 \le s \le 1$, we define

$$\begin{aligned}
|\!|\!|(p,y)|\!|\!|_{s,k} &= [(\mathfrak{B}_k^t \mathfrak{C}_k^{-1}\mathfrak{B}_k)^s(p,y),(p,y)]_k^{\frac{1}{2}} \quad \text{for all } (p,y) \in V_k \times V_k,\\
|\!|\!|(p,y)|\!|\!|_{\tilde{s},k} &= [(\mathfrak{B}_k \mathfrak{C}_k^{-1}\mathfrak{B}_k^t)^s(p,y),(p,y)]_k^{\frac{1}{2}} \quad \text{for all } (p,y) \in V_k \times V_k.
\end{aligned}$$

Note that

$$|\!|\!|(p,y)|\!|\!|_{0,k}^2 \approx \|p\|_{L_2(\Omega)}^2 + \|y\|_{L_2(\Omega)}^2 \approx (|\!|\!|(p,y)|\!|\!|_{\tilde{0},k})^2 \quad \text{for all } (p,y) \in V_k \times V_k,$$

by (4.3), and

$$|\!|\!|(p,y)|\!|\!|_{1,k}^2 \approx \|p\|_{H_\beta^1(\Omega;\mathfrak{T}_h)}^2 + \|y\|_{H_\beta^1(\Omega;\mathfrak{T}_h)}^2 \approx (|\!|\!|(p,y)|\!|\!|_{\tilde{1},k})^2 \quad \text{for all } (p,y) \in V_k \times V_k,$$

by (4.7) and (4.8).

## 5.2 Post-Smoothing Properties

The error propagation operator for one post-smoothing step defined by (4.13) is given by

$$R_k = \mathrm{Id}_k - \lambda_k \mathfrak{B}_k^t \mathfrak{C}_k^{-1} \mathfrak{B}_k, \tag{5.1}$$

where $\mathrm{Id}_k$ is the identity operator on $V_k \times V_k$. We also need the error propagation operator for one post-smoothing step defined by (4.14), which is

$$\tilde{R}_k = \mathrm{Id}_k - \lambda_k \mathfrak{B}_k \mathfrak{C}_k^{-1} \mathfrak{B}_k^t. \tag{5.2}$$

**Lemma 5.1** (Smoothing Properties). *In the case where $\beta^{\frac{1}{2}} h_k^{-1} < 1$, we have*

$$\begin{aligned}
|\!|\!|R_k(p,y)|\!|\!|_{1,k} &\le \tau |\!|\!|(p,y)|\!|\!|_{1,k} \quad \textit{for all } (p,y) \in V_k \times V_k,\\
|\!|\!|\tilde{R}_k(p,y)|\!|\!|_{\tilde{1},k} &\le \tau |\!|\!|(p,y)|\!|\!|_{\tilde{1},k} \quad \textit{for all } (p,y) \in V_k \times V_k,
\end{aligned}$$

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

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

$$\begin{aligned}
|\!|\!|R_k^m(p,y)|\!|\!|_{1,k} &\le C(\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{s}{2}} m^{-\frac{s}{2}} |\!|\!|(p,y)|\!|\!|_{1-s,k} \quad \textit{for all } (p,y) \in V_k \times V_k,\\
|\!|\!|\tilde{R}_k^m(p,y)|\!|\!|_{\tilde{1},k} &\le C(\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{s}{2}} m^{-\frac{s}{2}} |\!|\!|(p,y)|\!|\!|_{\widetilde{1-s},k} \quad \textit{for all } (p,y) \in V_k \times V_k,
\end{aligned}$$

*where the constant $C$ is independent of $k$ and $\beta$.*

## 5.3 Approximation Properties

The operators $P_k^{k-1}\colon V_k \times V_k \to V_{k-1} \times V_{k-1}$ and $\tilde{P}_k^{k-1}\colon V_k \times V_k \to V_{k-1} \times V_{k-1}$ are defined as follows. For all $(p,y) \in V_k \times V_k$, $(q,z) \in V_{k-1} \times V_{k-1}$,

$$\mathcal{B}_{k-1}(P_k^{k-1}(p,y),(q,z)) = \mathcal{B}_k((p,y), I_{k-1}^k(q,z)) = \mathcal{B}_k((p,y),(q,z)), \tag{5.3}$$

$$\mathcal{B}_{k-1}((q,z), \tilde{P}_k^{k-1}(p,y)) = \mathcal{B}_k(I_{k-1}^k(q,z),(p,y)) = \mathcal{B}_k((q,z),(p,y)). \tag{5.4}$$

**Lemma 5.2.** *We have the following properties:*

$$(I_{k-1}^k P_k^{k-1})^2 = I_{k-1}^k P_k^{k-1}, \quad (\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})^2 = \mathrm{Id}_k - I_{k-1}^k P_k^{k-1}, \tag{5.5}$$

$$(I_{k-1}^k \tilde{P}_k^{k-1})^2 = I_{k-1}^k \tilde{P}_k^{k-1}, \quad (\mathrm{Id}_k - I_{k-1}^k \tilde{P}_k^{k-1})^2 = \mathrm{Id}_k - I_{k-1}^k \tilde{P}_k^{k-1}. \tag{5.6}$$

*Proof.* We only prove the identities in (5.5), the ones in (5.6) are similar. Notice that, for $(p, y) \in V_{k-1} \times V_{k-1}$, we have

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

for all $(q, z) \in V_{k-1} \times V_{k-1}$. This implies $P_k^{k-1} I_{k-1}^k = \mathrm{Id}_{k-1}$. Then equalities (5.5) are immediate by a direct calculation. □

**Lemma 5.3** (Approximation Properties). *We have, for all $(p, y) \in V_k \times V_k$ and $k \geq 1$,*

$$|||(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})(p, y)|||_{0,k} \lesssim (\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 |||(p, y)|||_{1,k}, \tag{5.7}$$

$$|||(\mathrm{Id}_k - I_{k-1}^k \tilde{P}_k^{k-1})(p, y)|||_{\tilde{0},k} \lesssim (\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 |||(p, y)|||_{\tilde{1},k}. \tag{5.8}$$

*Here the hidden constant is independent of k and β.*

*Proof.* We will only prove (5.7); the argument for (5.8) is similar. Let $(p, y) \in V_k \times V_k$ be arbitrary and

$$(\zeta, \mu) = (\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})(p, y).$$

By (4.3), it suffices to show that

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

Estimate (5.7) is established through a duality argument (cf. [16]). Let $(\xi, \theta) \in H_0^1(\Omega) \times H_0^1(\Omega)$ satisfy

$$\mathcal{B}((q, z), (\xi, \theta)) = (\zeta, q)_{L_2(\Omega)} + (\mu, z)_{L_2(\Omega)} \quad \text{for all } (q, z) \in H_0^1(\Omega) \times H_0^1(\Omega).$$

Moreover, we define $(\xi_k, \theta_k) \in V_k \times V_k$ and $(\xi_{k-1}, \theta_{k-1}) \in V_{k-1} \times V_{k-1}$ by

$$\mathcal{B}_k((q, z), (\xi_k, \theta_k) = (\zeta, q)_{L_2(\Omega)} + (\mu, z)_{L_2(\Omega)} \quad \text{for all } (q, z) \in V_k \times V_k, \tag{5.9}$$

$$\mathcal{B}_{k-1}((q, z), (\xi_{k-1}, \theta_{k-1})) = (\zeta, q)_{L_2(\Omega)} + (\mu, z)_{L_2(\Omega)} \quad \text{for all } (q, z) \in V_{k-1} \times V_{k-1}. \tag{5.10}$$

Note that (5.9) and (5.10) imply

$$\mathcal{B}_k((q, z), (\xi_k, \theta_k)) = \mathcal{B}_{k-1}((q, z), (\xi_{k-1}, \theta_{k-1})) \quad \text{for all } (q, z) \in V_{k-1} \times V_{k-1}. \tag{5.11}$$

It follows from (5.11) and (5.4) that

$$(\xi_{k-1}, \theta_{k-1}) = \tilde{P}_k^{k-1}(\xi_k, \theta_k). \tag{5.12}$$

We have the following by (3.18) and $h_k \approx \frac{1}{2} h_{k-1}$:

$$\|\xi - \xi_k\|_{H_\beta^1(\Omega;\mathcal{T}_h)} + \|\theta - \theta_k\|_{H_\beta^1(\Omega;\mathcal{T}_h)} \leq C(\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 (\|\zeta\|_{L_2(\Omega)} + \|\mu\|_{L_2(\Omega)}), \tag{5.13}$$

$$\|\xi - \xi_{k-1}\|_{H_\beta^1(\Omega;\mathcal{T}_h)} + \|\theta - \theta_{k-1}\|_{H_\beta^1(\Omega;\mathcal{T}_h)} \leq C(\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 (\|\zeta\|_{L_2(\Omega)} + \|\mu\|_{L_2(\Omega)}). \tag{5.14}$$

Therefore, by (3.14), (5.3), (5.9), (5.12), (5.13) and (5.14), we have

$$\begin{aligned}
\|\zeta\|_{L_2(\Omega)}^2 + \|\mu\|_{L_2(\Omega)}^2 &= \mathcal{B}_k((\zeta, \mu), (\xi_k, \theta_k)) \\
&= \mathcal{B}_k((\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})(p, y), (\xi_k, \theta_k)) \\
&= \mathcal{B}_k((p, y), (\xi_k, \theta_k)) - \mathcal{B}_k(I_{k-1}^k P_k^{k-1}(p, y), (\xi_k, \theta_k)) \\
&= \mathcal{B}_k((p, y), (\xi_k, \theta_k)) - \mathcal{B}_{k-1}(P_k^{k-1}(p, y), \tilde{P}_k^{k-1}(\xi_k, \theta_k)) \\
&= \mathcal{B}_k((p, y), (\xi_k, \theta_k)) - \mathcal{B}_{k-1}(P_k^{k-1}(p, y), (\xi_{k-1}, \theta_{k-1})) \\
&= \mathcal{B}_k((p, y), (\xi_k, \theta_k)) - \mathcal{B}_k((p, y), I_{k-1}^k(\xi_{k-1}, \theta_{k-1})) \\
&= \mathcal{B}_k((p, y), (\xi_k, \theta_k) - I_{k-1}^k(\xi_{k-1}, \theta_{k-1})) \\
&\lesssim (\|\xi_k - \xi_{k-1}\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 + \|\theta_k - \theta_{k-1}\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2)^{\frac{1}{2}} (\|p\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2 + \|y\|_{H_\beta^1(\Omega;\mathcal{T}_h)}^2)^{\frac{1}{2}} \\
&\lesssim (\beta^{\frac{1}{2}} h_k^{-2} + 1)^{\frac{1}{2}} \beta^{-\frac{1}{2}} h_k^2 (\|\zeta\|_{L_2(\Omega)} + \|\mu\|_{L_2(\Omega)}) |||(p, y)|||_{1,k},
\end{aligned}$$

which implies (5.7). □

# 6 Convergence Analysis

Let $E_k\colon V_k \times V_k \to V_k \times V_k$ be the error propagation operator for the $k$-th level $W$-cycle algorithm for (4.4). The following recursive relationship is well known (cf. [15, 29]):

$$E_k = R_k^{m_2}(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1} + I_{k-1}^k E_{k-1}^2 P_k^{k-1}) S_k^{m_1}, \tag{6.1}$$

where $R_k$ is defined in (5.1) and

$$S_k = \mathrm{Id}_k - \lambda_k \mathfrak{C}_k^{-1} \mathfrak{B}_k^t \mathfrak{B}_k$$

measures the effect of one pre-smoothing step (4.11).

**Remark 6.1.** We have the following adjoint relation:

$$\mathcal{B}_k(S_k(p,y),(q,z)) = \mathcal{B}_k((p,y),\tilde{R}_k(q,z)) \quad \text{for all } (p,y),(q,z) \in V_k \times V_k, \tag{6.2}$$

where $\tilde{R}_k$ is defined in (5.2). Relation (6.2) is the reason why we consider the multigrid algorithms for (4.4) and (4.5) simultaneously.

**Lemma 6.2.** *We have*

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

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

*Proof.* For all $(p,y) \in V_k \times V_k$, it follows from (3.14) that

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

This implies $\|(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1}) S_k^m\| \lesssim \|\tilde{R}_k^m (\mathrm{Id}_k - I_{k-1}^k \tilde{P}_k^{k-1})\|$. The other direction of the estimate is similar. □

## 6.1 Convergence of the Two-Grid Algorithm

In the two-grid algorithm, the coarse grid residual equation is solved exactly ($E_{k-1} = 0$ in (6.1)). We therefore obtain the error propagation operator of the two-grid algorithm $R_k^{m_2}(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1}) S_k^{m_1}$ with $m_1$ pre-smoothing steps and $m_2$ post-smoothing steps.

We have the following lemma for the convergence of the two-grid algorithm.

**Lemma 6.3.** *In the case of $\beta^{-\frac{1}{2}} h_k^2 < 1$, we have*

$$\|R_k^{m_2}(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1}) S_k^{m_1}\| \lesssim \tau^{m_1+m_2}.$$

*In the case of $\beta^{-\frac{1}{2}} h_k^2 \geq 1$, we have*

$$\|R_k^{m_2}(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1}) S_k^{m_1}\| \lesssim [\max(1,m_2)\max(1,m_1)]^{-\frac{1}{2}}.$$

*Proof.* The proof can be found in [13, Section 5.1]. The key ingredients are (5.6), Lemma 5.1, Lemma 5.3 and Lemma 6.2. □

## 6.2 Convergence of the *W*-Cycle Algorithm

It is well known that the convergence of the two-grid algorithm implies the convergence of the $W$-cycle algorithm by a standard perturbation argument (cf. [15, 19, 29]). A delicate modification [13] of the standard argument leads to the following theorem.

**Theorem 6.4.** *There exists a positive integer $m_*$, independent of $k$ and $\beta$, such that*

$$\|E_k\| \le C_\sharp \tau^{m_1+m_2} \qquad \text{for all } 1 \le k \le k_*, \tag{6.3}$$

$$\|E_k\| \le C_\flat [\max(1,m_2)\max(1,m_1)]^{-\frac{1}{2}} + 4^{1-2^{k-k_*}}(C_\sharp \tau^{m_1+m_2}) \qquad \text{for all } k \ge k_* + 1. \tag{6.4}$$

*provided* $[\max(1,m_2)\max(1,m_1)] \ge m_*$. *Here $C_\sharp$ and $C_\flat$ are constants independent of $k$ and $\beta$, and the integer $k_*$ is the largest positive integer such that $\beta^{\frac{1}{2}} h_k^{-2} < 1$.*

**Remark 6.5.** The interpretations and implications of Theorem 6.4 are as follows.
(1) The $W$-cycle algorithm for the $k$-th level problem (4.4) is a contraction in the energy norm $|||\cdot|||_{1,k}$ if the number of smoothing steps is large enough. The contraction number is bounded away from 1 uniformly in $k$ and $\beta$. Therefore, the $W$-cycle algorithm is robust with respect to $k$ and $\beta$.
(2) At coarse levels (where $\beta^{\frac{1}{2}} h_k^{-2} < 1$), estimate (6.3) indicates that the contraction numbers decrease exponentially with respect to the number of smoothing steps. Estimate (6.4) implies that the contraction numbers will be dominated by the term $[\max(1,m_2)\max(1,m_1)]^{-\frac{1}{2}}$ at finer levels (where $\beta^{\frac{1}{2}} h_k^{-2} \ge 1$) eventually.

# 7 Numerical Results

In this section, we report the numerical results of the symmetric $W$-cycle and $V$-cycle algorithms ($m_1 = m_2 = m$). The preconditioner $\mathfrak{C}_k^{-1}$ is computed using a $V(4,4)$ multigrid solve for (4.6) based on a SIP discretization [16]. The eigenvalues $\lambda_{\max}$ and $\lambda_{\min}$ in (4.12b) are estimated using power iterations. We employed the MATLAB/C++ toolbox FELICITY [43] in our computation.

**Example 7.1** (Unit Square). In this example, we take $\Omega = (0,1)^2$ and $\sigma = 6$ in (3.4). For simplicity, we take $\boldsymbol{\zeta} = [1,0]^t$ and $\gamma = 0$ in (3.5). See Figure 1 for the initial triangulation $\mathcal{T}_0$ and the uniform refinements $\mathcal{T}_1$ and $\mathcal{T}_2$.

We report the contraction numbers of the $W$-cycle algorithm in Tables 1–3 for $\beta = 10^{-2}$, $\beta = 10^{-4}$ and $\beta = 10^{-6}$. We observe that the contraction numbers of the symmetric $W$-cycle algorithm decay exponentially at coarse levels and then approach the standard $O(m^{-1})$ behavior at finer levels for all choices of $\beta$. Notice that our $W$-cycle algorithm is clearly robust with respect to $\beta$ and the performance agrees with Remark 6.5.

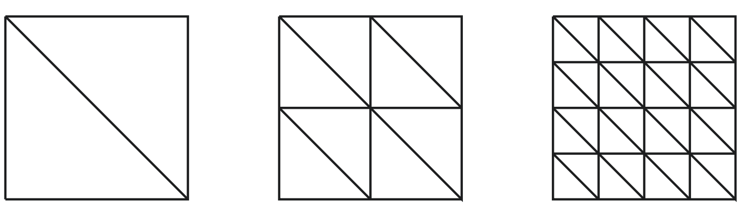

**Figure 1:** Triangulations $\mathcal{T}_0$, $\mathcal{T}_1$ and $\mathcal{T}_2$ for the unit square in Example 7.1.

| | *m* | | | | | | |
|---|---|---|---|---|---|---|---|
| ***k*** | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 8.17e−01 | 6.87e−01 | 5.08e−01 | 2.94e−01 | 1.02e−01 | 1.25e−02 | 1.89e−04 |
| 2 | 8.31e−01 | 7.02e−01 | 5.31e−01 | 3.29e−01 | 1.43e−01 | 4.19e−02 | 8.24e−03 |
| 3 | 8.96e−01 | 8.08e−01 | 6.74e−01 | 4.84e−01 | 2.92e−01 | 1.31e−01 | 4.48e−02 |
| 4 | 8.64e−01 | 7.55e−01 | 5.93e−01 | 4.05e−01 | 2.17e−01 | 1.01e−01 | 4.58e−02 |
| 5 | 8.49e−01 | 7.36e−01 | 5.63e−01 | 3.71e−01 | 1.95e−01 | 9.95e−02 | 4.59e−02 |
| 6 | 8.46e−01 | 7.35e−01 | 5.55e−01 | 3.61e−01 | 1.90e−01 | 9.50e−02 | 4.68e−02 |
| 7 | 8.45e−01 | 7.34e−01 | 5.52e−01 | 3.57e−01 | 1.90e−01 | 9.54e−02 | 4.70e−02 |

**Table 1:** The contraction numbers of the $k$-th level ($k = 1, \ldots, 7$) symmetric $W$-cycle algorithm for Example 7.1 with $\beta = 10^{-2}$ and $m = 2^0, \ldots, 2^6$.

| | $m$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $k$ | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 5.85e−01 | 3.74e−01 | 1.48e−01 | 2.42e−02 | 6.54e−04 | 4.81e−07 | 4.97e−14 |
| 2 | 8.08e−01 | 6.71e−01 | 4.65e−01 | 2.28e−01 | 5.52e−02 | 3.34e−03 | 1.13e−05 |
| 3 | 8.38e−01 | 7.26e−01 | 5.36e−01 | 3.31e−01 | 1.54e−01 | 3.87e−02 | 3.66e−03 |
| 4 | 9.36e−01 | 8.78e−01 | 7.82e−01 | 6.32e−01 | 4.38e−01 | 2.49e−01 | 1.01e−01 |
| 5 | 8.85e−01 | 7.92e−01 | 6.54e−01 | 4.85e−01 | 2.94e−01 | 1.47e−01 | 6.19e−02 |
| 6 | 8.95e−01 | 7.46e−01 | 5.77e−01 | 3.87e−01 | 2.09e−01 | 1.06e−01 | 5.49e−02 |
| 7 | 8.48e−01 | 7.37e−01 | 5.58e−01 | 3.63e−01 | 1.95e−01 | 9.75e−02 | 4.89e−02 |

**Table 2:** The contraction numbers of the $k$-th level ($k = 1, \ldots, 7$) symmetric $W$-cycle algorithm for Example 7.1 with $\beta = 10^{-4}$ and $m = 2^0, \ldots, 2^6$.

| | $m$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $k$ | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 4.18e−01 | 1.75e−01 | 3.02e−02 | 9.45e−04 | 8.71e−07 | 1.07e−13 | 1.86e−16 |
| 2 | 4.35e−01 | 1.90e−01 | 3.60e−02 | 1.24e−03 | 1.53e−06 | 4.60e−13 | 2.30e−16 |
| 3 | 7.06e−01 | 5.10e−01 | 2.84e−01 | 8.65e−02 | 7.88e−03 | 7.14e−05 | 5.31e−07 |
| 4 | 8.33e−01 | 7.10e−01 | 5.22e−01 | 2.78e−01 | 9.24e−02 | 1.13e−02 | 1.76e−04 |
| 5 | 8.43e−01 | 7.30e−01 | 5.44e−01 | 3.50e−01 | 1.79e−01 | 5.68e−02 | 1.13e−02 |
| 6 | 9.14e−01 | 8.41e−01 | 7.24e−01 | 5.45e−01 | 3.50e−01 | 1.78e−01 | 8.27e−02 |
| 7 | 8.70e−01 | 7.67e−01 | 6.14e−01 | 4.38e−01 | 2.52e−01 | 1.24e−01 | 5.87e−02 |

**Table 3:** The contraction numbers of the $k$-th level ($k = 1, \ldots, 7$) symmetric $W$-cycle algorithm for Example 7.1 with $\beta = 10^{-6}$ and $m = 2^0, \ldots, 2^6$.

| | $k$ | | | | | | |
|---|---|---|---|---|---|---|---|
| $m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | | | | $\beta = 10^{-2}$ | | | |
| $2^2$ | 5.08e−01 | 5.55e−01 | 6.79e−01 | 6.54e−01 | 6.50e−01 | 6.23e−01 | 6.15e−01 |
| $2^3$ | 2.94e−01 | 3.45e−01 | 4.97e−01 | 4.68e−01 | 4.83e−01 | 4.54e−01 | 4.41e−01 |
| $2^4$ | 1.02e−01 | 1.52e−01 | 2.92e−01 | 2.87e−01 | 2.88e−01 | 2.80e−01 | 2.52e−01 |
| | | | | $\beta = 10^{-4}$ | | | |
| $2^2$ | 1.49e−01 | 4.65e−01 | 5.40e−01 | 7.84e−01 | 7.31e−01 | 7.07e−01 | 7.05e−01 |
| $2^3$ | 2.42e−02 | 2.28e−01 | 3.30e−01 | 6.33e−01 | 5.82e−01 | 5.74e−01 | 5.66e−01 |
| $2^4$ | 6.53e−04 | 5.60e−02 | 1.53e−01 | 4.38e−01 | 3.82e−01 | 3.75e−01 | 3.77e−01 |
| | | | | $\beta = 10^{-6}$ | | | |
| $2^2$ | 3.08e−02 | 3.60e−02 | 2.72e−01 | 5.18e−01 | 5.54e−01 | 7.23e−01 | 6.94e−01 |
| $2^3$ | 9.47e−04 | 1.26e−03 | 8.61e−02 | 2.83e−01 | 3.55e−01 | 5.47e−01 | 5.31e−01 |
| $2^4$ | 7.11e−07 | 1.70e−06 | 8.02e−03 | 9.24e−02 | 1.80e−01 | 3.53e−01 | 3.33e−01 |

**Table 4:** The contraction numbers of the $k$-th level ($k = 1, \ldots, 7$) symmetric $V$-cycle algorithm for Example 7.1 with $\beta = 10^{-2}, 10^{-4}, 10^{-6}$ and $m = 2^2, 2^3, 2^4$.

We have also tested the symmetric $V$-cycle algorithm for $k$-th level problem (4.4) and briefly report the results in Table 4. We observe that our $V$-cycle algorithm is also a contraction with slightly more numbers of smoothing steps ($m = 4$) and the contraction numbers are robust with respect to $k$ and $\beta$.

**Example 7.2** (L-Shaped Domain). We also test our multigrid methods on nonconvex domains. In this example, we take $\Omega = (0,1)^2 \setminus (0.5,1)^2$ and $\sigma = 6$ in (3.4). We also take $\boldsymbol{\zeta} = [1,0]^t$ and $\gamma = 0$ in (3.5). See Figure 2 for the initial triangulation $\mathcal{T}_0$ and the uniform refinement $\mathcal{T}_1$.

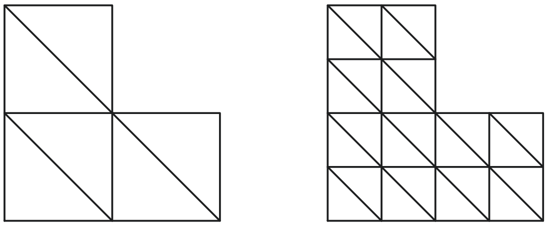

**Figure 2:** Triangulations $\mathcal{T}_0$ and $\mathcal{T}_1$ for the L-shaped domain in Example 7.2.

| | ***m*** | | | | | | |
|---|---|---|---|---|---|---|---|
| ***k*** | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 8.34e−01 | 7.18e−01 | 5.31e−01 | 3.25e−01 | 1.30e−01 | 2.83e−02 | 1.62e−03 |
| 2 | 8.95e−01 | 8.09e−01 | 6.74e−01 | 4.84e−01 | 2.84e−01 | 1.22e−01 | 3.10e−02 |
| 3 | 8.64e−01 | 7.55e−01 | 5.96e−01 | 4.02e−01 | 2.17e−01 | 9.76e−02 | 3.76e−02 |
| 4 | 8.50e−01 | 7.35e−01 | 5.65e−01 | 3.72e−01 | 1.95e−01 | 9.48e−02 | 4.58e−02 |
| 5 | 8.46e−01 | 7.33e−01 | 5.56e−01 | 3.61e−01 | 1.91e−01 | 9.53e−02 | 4.70e−02 |
| 6 | 8.46e−01 | 7.34e−01 | 5.53e−01 | 3.59e−01 | 1.90e−01 | 9.52e−02 | 4.62e−02 |
| 7 | 8.45e−01 | 7.34e−01 | 5.52e−01 | 3.55e−01 | 1.91e−01 | 9.50e−02 | 4.71e−02 |

**Table 5:** The contraction numbers of the $k$-th level ($k = 1, \dots, 7$) symmetric $W$-cycle algorithm for Example 7.2 with $\beta = 10^{-2}$ and $m = 2^0, \dots, 2^6$.

| | ***m*** | | | | | | |
|---|---|---|---|---|---|---|---|
| ***k*** | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 8.02e−01 | 6.63e−01 | 4.67e−01 | 2.26e−01 | 5.76e−02 | 3.57e−03 | 1.55e−05 |
| 2 | 8.37e−01 | 7.20e−01 | 5.33e−01 | 3.26e−01 | 1.52e−01 | 3.67e−02 | 2.71e−03 |
| 3 | 9.36e−01 | 8.78e−01 | 7.82e−01 | 6.29e−01 | 4.38e−01 | 2.50e−01 | 9.60e−02 |
| 4 | 8.85e−01 | 7.90e−01 | 6.52e−01 | 4.96e−01 | 2.91e−01 | 1.44e−01 | 6.13e−02 |
| 5 | 8.56e−01 | 7.47e−01 | 5.77e−01 | 3.87e−01 | 2.10e−01 | 1.06e−01 | 5.59e−02 |
| 6 | 8.48e−01 | 7.37e−01 | 5.58e−01 | 3.64e−01 | 1.95e−01 | 9.73e−02 | 4.83e−02 |
| 7 | 8.46e−01 | 7.35e−01 | 5.53e−01 | 3.56e−01 | 1.92e−01 | 9.57e−02 | 4.73e−02 |

**Table 6:** The contraction numbers of the $k$-th level ($k = 1, \dots, 7$) symmetric $W$-cycle algorithm for Example 7.2 with $\beta = 10^{-4}$ and $m = 2^0, \dots, 2^6$.

| | ***m*** | | | | | | |
|---|---|---|---|---|---|---|---|
| ***k*** | $2^0$ | $2^1$ | $2^2$ | $2^3$ | $2^4$ | $2^5$ | $2^6$ |
| 1 | 4.26e−01 | 1.88e−01 | 3.61e−02 | 1.23e−03 | 1.18e−06 | 2.27e−13 | 2.26e−16 |
| 2 | 7.07e−01 | 5.24e−01 | 2.84e−01 | 8.75e−02 | 8.40e−03 | 8.03e−05 | 3.31e−07 |
| 3 | 8.30e−01 | 7.04e−01 | 5.15e−01 | 2.81e−01 | 9.01e−02 | 1.06e−02 | 1.53e−04 |
| 4 | 8.42e−01 | 7.33e−01 | 5.43e−01 | 3.49e−01 | 1.78e−01 | 5.61e−02 | 1.28e−02 |
| 5 | 9.14e−01 | 8.41e−01 | 7.23e−01 | 5.45e−01 | 3.52e−01 | 1.78e−01 | 8.26e−02 |
| 6 | 8.70e−01 | 7.67e−01 | 6.14e−01 | 4.37e−01 | 2.50e−01 | 1.24e−01 | 5.85e−02 |
| 7 | 8.52e−01 | 7.42e−01 | 5.66e−01 | 3.73e−01 | 2.03e−01 | 1.03e−01 | 5.28e−02 |

**Table 7:** The contraction numbers of the $k$-th level ($k = 1, \dots, 7$) symmetric $W$-cycle algorithm for Example 7.2 with $\beta = 10^{-6}$ and $m = 2^0, \dots, 2^6$.

We report the contraction numbers of the $W$-cycle algorithm in Tables 5–7 for $\beta = 10^{-2}$, $\beta = 10^{-4}$ and $\beta = 10^{-6}$. We observe that the contraction numbers of the symmetric $W$-cycle algorithm decay exponentially at coarse levels and then approach the standard $O(m^{-\frac{2}{3}})$ behavior for L-shaped domains at finer levels for all choices of $\beta$. Notice that our $W$-cycle algorithm is clearly robust with respect to $\beta$ and the performance agrees with Remark 6.5.

# 8 Concluding Remarks

We proposed and analyzed multigrid methods for an elliptic optimal control problem based on DG discretizations. We proved that, for a sufficiently large number of smoothing steps, the $W$-cycle algorithm is uniformly convergent with respect to mesh refinements and a regularizing parameter. The numerical results coincide with the theoretical findings. Our multigrid analysis can be also be extended to other DG methods. In fact, our multigrid analysis should work for any DG methods with convergence results in $H^1_\beta(\Omega;\mathcal{T}_h)$ norm as that of Lemma 3.5.

The analysis of our multigrid methods can also be extended to higher order polynomials assuming higher regularity of the solutions to the optimal control problems. We briefly discuss the strategy. Note that one has the following interpolation estimate [37]:

$$\|z-\Pi_h z\|_{L_2(\Omega)} + h|z-\Pi_h z|_{H^1(\Omega)} \lesssim h^{\min(l+1,s)}|z|_{H^s(\Omega)}, \tag{8.1}$$

where $l$ is the order of the polynomials. From (8.1), one can easily obtain the error estimates for the DG methods as follows:

$$\|p-p_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} + \|y-y_h\|_{H^1_\beta(\Omega;\mathcal{T}_h)} \lesssim (\beta^{\frac{1}{2}}h^{-2}+1)^{\frac{1}{2}}\beta^{-\frac{1}{2}}h^{\min(l+1,s)}.$$

Note that the inverse estimate (3.16) remains the same for higher order polynomials [37], and hence the smoothing property (cf. Lemma 5.1) remains unchanged, which is

$$|||R_k^m(p,y)|||_{1,k} \le C(\beta^{\frac{1}{2}}h_k^{-2}+1)^{\frac{1}{2}}m^{-\frac{1}{2}}|||(p,y)|||_{0,k}.$$

The approximation property can be proved using the same idea as in Lemma 5.3, and it becomes

$$|||(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})(p,y)|||_{0,k} \lesssim (\beta^{\frac{1}{2}}h_k^{-2}+1)^{\frac{1}{2}}\beta^{-\frac{1}{2}}h_k^{\min(l+1,s)}|||(p,y)|||_{1,k}.$$

The convergence of the multigrid can be obtained by combining the smoothing property and the approximation property. In fact, we have

$$\begin{aligned}\|R_k^m(\mathrm{Id}_k - I_{k-1}^k P_k^{k-1})\| &\le C(\beta^{\frac{1}{2}}h_k^{-2}+1)^{\frac{1}{2}}m^{-\frac{1}{2}}(\beta^{\frac{1}{2}}h_k^{-2}+1)^{\frac{1}{2}}\beta^{-\frac{1}{2}}h_k^{\min(l+1,s)}\\ &= C(1+\beta^{-\frac{1}{2}}h_k^2)h_k^{\min(l+1,s)-2}m^{-\frac{1}{2}}\\ &\le C(1+\beta^{-\frac{1}{2}}h_k^2)m^{-\frac{1}{2}}.\end{aligned}$$

We use the fact that $\min(l+1,s) \ge 2$ in the last inequality. Therefore, the assumptions on $\beta$ and $h_k$ in Sections 4–6 are still valid in the case of higher order polynomials. The theoretical results then follows as that of Section 6.

A more interesting problem is to consider an advection-dominated state equation. DG methods are promising for advection-dominated problem due to the natural built-in upwind stabilization and the weak treatment of the boundary conditions. Related work can be found, for example, in [34]. However, the challenge for extending our result is to design proper preconditioner so that the multigrid methods are robust for the advection-dominated case. This is under investigation in our ongoing projects.

# A Proofs of (3.9) and (3.10)

For $T \in \mathcal{T}_h$ and $v \in H^{1+s}(\Omega)$, where $s \in (\frac{1}{2},1]$, the following trace inequalities with scaling is standard (cf. [25, Lemma 7.2] and [21, Proposition 3.1]):

$$\|v\|_{L_2(\partial T)} \le C(h_T^{-\frac{1}{2}}\|v\|_{L_2(T)} + h_T^{s-\frac{1}{2}}|v|_{H^s(T)}). \tag{A.1}$$

The following discrete Poincaŕe inequality for DG functions [4, 8, 20] is valid for all $v \in V_h$:

$$\|v\|^2_{L_2(\Omega)} \le C\Big(\sum_{T\in\Omega}\|\nabla v\|^2_{L_2(T)} + \sum_{e\in\partial T}\frac{1}{h_e}\|[v]\|^2_{L_2(e)}\Big). \tag{A.2}$$

*Proof.* It is well known that [3, 15, 37]

$$
\begin{aligned}
a_h^{\mathrm{sip}}(w,v) &\le C\|w\|_{1,h}\|v\|_{1,h} && \text{for all } w,v\in V+V_h,\\
a_h^{\mathrm{sip}}(v,v) &\ge C\|v\|_{1,h}^2 && \text{for all } v\in V_h.
\end{aligned}
$$

For the advection-reaction term, we have, for all $w, v \in V + V_h$,

$$
\begin{aligned}
a_h^{\mathrm{ar}}(w,v) &= \sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla w+\gamma w,v)_T-\sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}(\mathbf{n}\cdot\boldsymbol{\zeta}[w],\{v\})_e\\
&\lesssim\Big(\sum_{T\in\mathcal{T}_h}\|\nabla w\|_{L_2(T)}^2\Big)^{\frac12}\|v\|_{L_2(\Omega)}+\|w\|_{L_2(\Omega)}\|v\|_{L_2(\Omega)}\\
&\qquad+\Big(\sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}\frac{\sigma}{h_e}\|[w]\|_{L_2(e)}^2\Big)^{\frac12}\Big(\sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}\frac{h_e}{\sigma}\|\{v\}\|_{L_2(e)}^2\Big)^{\frac12}\\
&\lesssim |\!|\!|w|\!|\!|_h|\!|\!|v|\!|\!|_h,
\end{aligned}
$$

where we use $\boldsymbol{\zeta}\in[W^{1,\infty}(\Omega)]^2$, $\gamma\in L_\infty(\Omega)$, (A.1) and (A.2). Furthermore, upon integration by parts, we have, for all $v \in V_h$,

$$
\begin{aligned}
a_h^{\mathrm{ar}}(v,v) &= \sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla v+\gamma v,v)_T-\sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}(\mathbf{n}\cdot\boldsymbol{\zeta}[v],\{v\})_e\\
&=\sum_{T\in\mathcal{T}_h}\Big(\Big(\gamma-\frac12\nabla\cdot\boldsymbol{\zeta}\Big)v,v\Big)_T+\sum_{T\in\mathcal{T}_h}\int_{\partial T}\frac12(\boldsymbol{\zeta}\cdot\mathbf{n})v^2\,ds-\sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}\int_e\boldsymbol{\zeta}\cdot\mathbf{n}[v]\{v\}\,ds\\
&=\sum_{T\in\mathcal{T}_h}\Big(\Big(\gamma-\frac12\nabla\cdot\boldsymbol{\zeta}\Big)v,v\Big)_T+\int_{\partial\Omega}\frac12|\boldsymbol{\zeta}\cdot\mathbf{n}|v^2\,ds.
\end{aligned}
$$

By assumption (1.4), we immediately have $a_h^{\mathrm{ar}}(v,v)\ge 0$. This finishes the proof. □

# B A Proof of (3.22)

*Proof.* It follows from (3.20) that

$$
\begin{aligned}
\|p-p_h\|_{L_2(\Omega)}^2+\|y-y_h\|_{L_2(\Omega)}^2 &= (\beta^{\frac12}(-\Delta\xi+\boldsymbol{\zeta}\cdot\nabla\xi+\gamma\xi)-\theta,p-p_h)_{L_2(\Omega)}\\
&\quad+(-\xi+\beta^{\frac12}(\Delta\theta+\boldsymbol{\zeta}\cdot\nabla\theta-(\gamma-\nabla\cdot\boldsymbol{\zeta})\theta),y-y_h)_{L_2(\Omega)}\\
&=\beta^{\frac12}(-\Delta\xi,p-p_h)_{L_2(\Omega)}+\beta^{\frac12}\sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla\xi+\gamma\xi,p-p_h)_T\\
&\quad-(\theta,p-p_h)_{L_2(\Omega)}-(\xi,y-y_h)_{L_2(\Omega)}\\
&\quad+\beta^{\frac12}(\Delta\theta,y-y_h)_{L_2(\Omega)}+\beta^{\frac12}\sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla\theta-(\gamma-\nabla\cdot\boldsymbol{\zeta})\theta,y-y_h)_T.
\end{aligned}
\tag{B.1}
$$

By the consistency of the SIP method (cf. [3, 37]), we have

$$
(-\Delta\xi,p-p_h)=a_h^{\mathrm{sip}}(\xi,p-p_h)\quad\text{and}\quad(\Delta\theta,y-y_h)=-a_h^{\mathrm{sip}}(y-y_h,\theta).
\tag{B.2}
$$

For the last term in (B.1), it follows from integration by parts that

$$
\begin{aligned}
\sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla\theta-(\gamma-\nabla\cdot\boldsymbol{\zeta})\theta,y-y_h)_T &= \sum_{T\in\mathcal{T}_h}(-\boldsymbol{\zeta}\cdot\nabla(y-y_h),\theta)_T-(\gamma(y-y_h),\theta)_T\\
&\quad+\sum_{T\in\mathcal{T}_h}\int_{\partial T}(\boldsymbol{\zeta}\cdot\mathbf{n})(y-y_h)\theta\,ds.
\end{aligned}
\tag{B.3}
$$

The last term in (B.3) can be rewritten as the following [3, 24]:

$$
\begin{aligned}
\sum_{T\in\mathcal{T}_h}\int_{\partial T}(\boldsymbol{\zeta}\cdot\mathbf{n})(y-y_h)\theta\,ds &= \sum_{e\in\mathcal{E}_h^i}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}[(y-y_h)\theta]\,ds + \sum_{e\in\mathcal{E}_h^b}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}(y-y_h)\theta\,ds\\
&= \sum_{e\in\mathcal{E}_h^i}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}[y-y_h]\{\theta\}\,ds + \sum_{e\in\mathcal{E}_h^i}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}\{y-y_h\}[\theta]\,ds\\
&\quad + \sum_{e\in\mathcal{E}_h^b}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}(y-y_h)\theta\,ds.
\end{aligned}
\tag{B.4}
$$

It then follows from $[\theta]=0$ on interior edges and $\theta=0$ on $\partial\Omega$ that

$$
\sum_{T\in\mathcal{T}_h}\int_{\partial T}(\boldsymbol{\zeta}\cdot\mathbf{n})(y-y_h)\theta\,ds = \sum_{e\in\mathcal{E}_h^i\cup\mathcal{E}_h^{b,-}}\int_e \boldsymbol{\zeta}\cdot\mathbf{n}[y-y_h]\{\theta\}\,ds.
\tag{B.5}
$$

According to (B.3)–(B.5), we conclude

$$
\sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla\theta-(\gamma-\nabla\cdot\boldsymbol{\zeta})\theta, y-y_h)_T = -a_h^{\mathrm{ar}}(y-y_h,\theta).
\tag{B.6}
$$

Similarly, we can show

$$
\sum_{T\in\mathcal{T}_h}(\boldsymbol{\zeta}\cdot\nabla\xi+\gamma\xi, p-p_h)_T = a_h^{\mathrm{ar}}(\xi, p-p_h).
\tag{B.7}
$$

Therefore, we obtain the following by (B.2), (B.6), (B.7), (3.3) and (3.2):

$$
\begin{aligned}
\|p-p_h\|_{L_2(\Omega)}^2+\|y-y_h\|_{L_2(\Omega)}^2 &= \beta^{\frac{1}{2}}a_h(\xi,p-p_h)-(\theta,p-p_h)_{L_2(\Omega)}-(\xi,y-y_h)_{L_2(\Omega)}-\beta^{\frac{1}{2}}a_h(y-y_h,\theta)\\
&= \mathcal{B}_h((p-p_h,y-y_h),(\xi,\theta)).
\end{aligned}
\qquad\square
$$

**Acknowledgment:** The author would like to thank Prof. Susanne C. Brenner, Prof. Li-Yeng Sung and Prof. Yi Zhang for the helpful discussion regarding this project.

**Funding:** The revision of this material is based upon work supported by the National Science Foundation under Grant No. DMS-1929284 while the author was in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the “Numerical PDEs: Analysis, Algorithms, and Data Challenges” program.

## References

[1] P. F. Antonietti, M. Sarti and M. Verani, Multigrid algorithms for *hp*-discontinuous Galerkin discretizations of elliptic problems, *SIAM J. Numer. Anal.* **53** (2015), no. 1, 598–618.
[2] D. N. Arnold, An interior penalty finite element method with discontinuous elements, *SIAM J. Numer. Anal.* **19** (1982), no. 4, 742–760.
[3] D. N. Arnold, F. Brezzi, B. Cockburn and L. D. Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, *SIAM J. Numer. Anal.* **39** (2001/02), no. 5, 1749–1779.
[4] B. Ayuso and L. D. Marini, Discontinuous Galerkin methods for advection-diffusion-reaction problems, *SIAM J. Numer. Anal.* **47** (2009), no. 2, 1391–1420.
[5] I. Babuška, The finite element method with Lagrangian multipliers, *Numer. Math.* **20** (1973), no. 3, 179–192.
[6] A. Borzi and V. Schulz, Multigrid methods for PDE optimization, *SIAM Rev.* **51** (2009), no. 2, 361–395.
[7] J. H. Bramble, *Multigrid Methods*, Pitman Res. Notes in Math. Ser. 294, John Wiley & Sons, New York, 1993.
[8] S. C. Brenner, Poincaré–Friedrichs inequalities for piecewise $H^1$ functions, *SIAM J. Numer. Anal.* **41** (2003), no. 1, 306–324.
[9] S. C. Brenner, J. Cui, T. Gudi and L.-Y. Sung, Multigrid algorithms for symmetric discontinuous Galerkin methods on graded meshes, *Numer. Math.* **119** (2011), no. 1, 21–47.
[10] S. C. Brenner, J. Cui and L.-Y. Sung, Multigrid methods for the symmetric interior penalty method on graded meshes, *Numer. Linear Algebra Appl.* **16** (2009), no. 6, 481–501.
[11] S. C. Brenner, H. Li and L.-Y. Sung, Multigrid methods for saddle point problems: Stokes and Lamé systems, *Numer. Math.* **128** (2014), no. 2, 193–216.

[12] S. C. Brenner, H. Li and L.-Y. Sung, Multigrid methods for saddle point problems: Oseen system, *Comput. Math. Appl.* **74** (2017), no. 9, 2056–2067.
[13] S. C. Brenner, S. Liu and L.-Y. Sung, Multigrid methods for saddle point problems: Optimality systems, *J. Comput. Appl. Math.* **372** (2020), Article ID 112733.
[14] S. C. Brenner, D.-S. Oh and L.-Y. Sung, Multigrid methods for saddle point problems: Darcy systems, *Numer. Math.* **138** (2018), no. 2, 437–471.
[15] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts Appl. Math. 15, Springer, New York, 2008.
[16] S. C. Brenner and J. Zhao, Convergence of multigrid algorithms for interior penalty methods, *Appl. Numer. Anal. Comput. Math.* **2** (2005), no. 1, 3–18.
[17] F. Brezzi, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, *Rev. Française Automat. Informat. Rech. Opér. Sér. Rouge* **8** (1974), no. R2, 129–151.
[18] F. Brezzi, L. D. Marini and E. Süli, Discontinuous Galerkin methods for first-order hyperbolic problems, *Math. Models Methods Appl. Sci.* **14** (2004), no. 12, 1893–1903.
[19] W. L. Briggs, V. E. Henson and S. F. McCormick, *A Multigrid Tutorial*, Society for Industrial and Applied Mathematics, Philadelphia, 2000.
[20] Z. Chen and H. Chen, Pointwise error estimates of discontinuous Galerkin methods with penalty for second-order elliptic problems, *SIAM J. Numer. Anal.* **42** (2004), no. 3, 1146–1166.
[21] P. Ciarlet, Jr., Analysis of the Scott–Zhang interpolation in the fractional order Sobolev spaces, *J. Numer. Math.* **21** (2013), no. 3, 173–180.
[22] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Stud. Math. Appl. 4, North-Holland, Amsterdam, 1978.
[23] D. A. Di Pietro and A. Ern, Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier–Stokes equations, *Math. Comp.* **79** (2010), no. 271, 1303–1330.
[24] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Math. Appl. (Berlin) 69, Springer, Heidelberg, 2012.
[25] A. Ern and J.-L. Guermond, Finite element quasi-interpolation and best approximation, *ESAIM Math. Model. Numer. Anal.* **51** (2017), no. 4, 1367–1385.
[26] F. Gaspoz, C. Kreuzer, A. Veeser and W. Wollner, Quasi-best approximation in optimization with PDE constraints, *Inverse Problems* **36** (2020), no. 1, Article ID 014004.
[27] W. Gong, Z. Tan and Z. Zhou, Optimal convergence of finite element approximation to an optimization problem with PDE constraint, *Inverse Problems* **38** (2022), no. 4, Article ID 045004.
[28] J. Gopalakrishnan and G. Kanschat, A multilevel discontinuous Galerkin method, *Numer. Math.* **95** (2003), no. 3, 527–550.
[29] W. Hackbusch, *Multigrid Methods and Applications*, Springer Ser. Comput. Math. 4, Springer, Berlin, 1985.
[30] M. Hinze, A variational discretization concept in control constrained optimization: The linear-quadratic case, *Comput. Optim. Appl.* **30** (2005), no. 1, 45–61.
[31] Q. Hong, J. Kraus, J. Xu and L. Zikatanov, A robust multigrid method for discontinuous Galerkin discretizations of Stokes and linear elasticity equations, *Numer. Math.* **132** (2016), no. 1, 23–49.
[32] G. Kanschat and Y. Mao, Multigrid methods for $H^{\mathrm{div}}$-conforming discontinuous Galerkin methods for the Stokes equations, *J. Numer. Math.* **23** (2015), no. 1, 51–66.
[33] D. Leykekhman, Investigation of commutative properties of discontinuous Galerkin methods in PDE constrained optimal control problems, *J. Sci. Comput.* **53** (2012), no. 3, 483–511.
[34] 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), no. 4, 2012–2038.
[35] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Grundlehren Math. Wiss. 170, Springer, New York, 1971.
[36] J. W. Pearson, M. Stoll and A. J. Wathen, Regularization-robust preconditioners for time-dependent PDE-constrained optimization problems, *SIAM J. Matrix Anal. Appl.* **33** (2012), no. 4, 1126–1152.
[37] B. Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations*, Front. Appl. Math. 35, Society for Industrial and Applied Mathematics, Philadelphia, 2008.
[38] Y. Saad, *Iterative Methods for Sparse Linear Systems*, Society for Industrial and Applied Mathematics, Philadelphia, 2003.
[39] J. Schöberl, R. Simon and W. Zulehner, A robust multigrid method for elliptic optimal control problems, *SIAM J. Numer. Anal.* **49** (2011), no. 4, 1482–1503.
[40] 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** (2013), no. 3, 1853–1874.
[41] S. Ta'asan, "One-shot" methods for optimal control of distributed parameter systems 1: Finite dimensional control, Technical Report IICASE-Report 91-2, NASA Langley Research Center, Hampton, 1991.
[42] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*, Grad. Stud. Math. 112, American Mathematical Society, Providence, 2010.
[43] S. W. Walker, FELICITY: A MATLAB/C++ toolbox for developing finite element methods and simulation modeling, *SIAM J. Sci. Comput.* **40** (2018), no. 2, C234–C257.