ACCURACY AND CONVERGENCE PROPERTIES OF THE FINITE DIFFERENCE MULTIGRID SOLUTION OF AN OPTIMAL CONTROL OPTIMALITY SYSTEM[*]

ALFIO BORZÌ[†], KARL KUNISCH[†], AND DO Y. KWAK[‡]

Abstract. The finite difference multigrid solution of an optimal control problem associated with an elliptic equation is considered. Stability of the finite difference optimality system and optimal-order error estimates in the discrete L^2 norm and in the discrete H^1 norm under minimum smoothness requirements on the exact solution are proved. Sharp convergence factor estimates of the two grid method for the optimality system are obtained by means of local Fourier analysis. A multigrid convergence theory is provided which guarantees convergence of the multigrid process towards weak solutions of the optimality system.

Key words. optimal control problem, Poisson equation, finite differences, accuracy estimate, convergence theory, multigrid method

AMS subject classifications. 49K20, 65N06, 65N12, 65N55

PII. S0363012901393432

1. Introduction. Optimal control problems involving partial differential equations [17, 18] are nowadays receiving much attention because of their importance in the industrial design process. Especially, the need for accurate and efficient solution methods for these problems has become an important issue.

We consider a finite difference framework and multigrid methods for the case of distributed optimal control of an elliptic problem and provide for this case optimal estimates for the accuracy of the solution and for the convergence factor of the multigrid process. The present work is characterized by the fact that we extend known analytic tools for scalar elliptic problems to the case of a (nonsymmetric) system of elliptic partial differential equations, called an optimality system.

In our finite difference analysis, based on results stated in [14, 20], we prove stability of the finite difference optimality system and prove optimal-order error estimates in the discrete L^2 norm and in the discrete H^1 norm under minimum smoothness requirements on the analytic solution.

It is known that multigrid methods [5, 13, 21] solve elliptic problems with optimal computational order, i.e., the number of computer operations required scales linearly with respect to the number of unknowns. This fact has been demonstrated in the case of multigrid applied to a singular optimal control problem associated with a nonlinear elliptic equation [2]. In particular, results in [2] show that the convergence properties of the multigrid method do not deteriorate as the weight of the cost of the control tends to zero, demonstrating the robustness of this method.

We prove convergence of the multigrid method applied to the optimality system within two analytic frameworks which have complementary features. We use two grid

[†]Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstr. 36, A-8010 Graz, Austria (alfio.borzi@uni-graz.at, karl.kunisch@uni-graz.at).

[‡]Department of Mathematics, Korea Advanced Institute of Science and Technology, Taejon, Korea 305-701 (dykwak@math.kaist.ac.kr).

local Fourier analysis [21, 10] with simplifying assumptions on the boundary conditions to obtain sharp convergence estimates of the multigrid method. These convergence estimates agree very well with results of numerical experiments and appear to be independent of the mesh size and of the value of the control parameter.

While the extension of two grid local Fourier analysis to systems of partial differential equations may be considered straightforward, the extension of the multigrid theory provided in [6, 8, 9, 15] to the case of optimality systems requires additional analysis, which is presented in this paper. The resulting multigrid theory does not require special assumptions on the boundary, it applies to polygonal domains, and guarantees convergence of the multigrid method to weak solutions of the optimality system.

In the following section we introduce and analyze our model problem. The finite difference discretization of this model problem and the corresponding stability and accuracy analysis are considered in section 3. In section 4 we describe the multigrid method and define and analyze its components. Two grid local Fourier analysis is presented in section 5. In section 6, a general convergence theory for multigrid applied to the optimality system is provided.

2. Optimal control problem. We consider the optimal control problem

$$(2.1) \qquad \min J(y,u) = \frac{1}{2}||y - z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\Omega)},$$

subject to $u \in L^2(\Omega)$ and

$$(2.2) \qquad -\Delta y = u + g \text{ in } \Omega,$$
$$y = 0 \text{ on } \partial\Omega,$$

where $\Omega = (0,1) \times (0,1)$, $g \in L^2(\Omega)$, $z \in L^2(\Omega)$ is the objective function, and $\nu > 0$ is the weight of the cost of the control. Existence of a unique solution to (2.1) and its characterization are well known. Let us, for the sake of completeness, give a short derivation and denote by $\hat{J}(u) = J(y(u), u)$, where $y(u)$ denotes the solution of (2.2) as a function of u. Recall that $\Delta : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$ is a homeomorphism. Here we use the fact that Ω is convex. The mapping $u \to y(u)$ from $L^2(\Omega)$ to $H_0^1(\Omega) \cap H^2(\Omega)$ is affine and continuous. Let us denote its first derivative at u in the direction δu by $y'(u, \delta u)$. It is characterized as the solution to

$$(2.3) \qquad -\Delta y'(u, \delta u) = \delta u \text{ in } \Omega,$$
$$y'(u, \delta u) = 0 \text{ on } \partial\Omega.$$

The second derivative of $u \to y(u)$ is zero. Hence we find for the second derivative of $u \to \hat{J}(u)$

$$\hat{J}''(u)(\delta u, \delta u) = ||y'(u, \delta u)||^2_{L^2(\Omega)} + \nu||\delta u||^2_{L^2(\Omega)},$$

and thus $u \to \hat{J}(u)$ is uniformly convex. This implies existence of a unique solution u^* to (2.1). Moreover, the solution is characterized by $\hat{J}'(u)(u^*; \delta u) = 0$ for all δu and consequently

$$\hat{J}'(u^*, \delta u) = (y^* - z, y'(u^*, \delta u))_{L^2(\Omega)} + \nu(u^*, \delta u)_{L^2(\Omega)} = 0 \qquad \text{for all } \delta u \in L^2(\Omega),$$

where $y^* = y(u^*)$. Introduce $\lambda^* \in H_0^1(\Omega) \cap H^2(\Omega)$ as the unique solution to

$$(2.4) \qquad -\Delta\lambda^* = -(y^* - z) \text{ in } \Omega,$$
$$\lambda^* = 0 \text{ on } \partial\Omega.$$

Then by (2.3) and (2.4) we have

$$(2.5) \qquad \hat{J}'(u^*, \delta u) = -(\lambda^*, \delta u)_{L^2(\Omega)} + \nu(u^*, \delta u)_{L^2(\Omega)} = 0 \qquad \text{for all } \delta u \in L^2(\Omega),$$

which constitutes the necessary and sufficient optimality condition for (2.1). In (2.5), λ^* is defined via (2.2) and (2.4).

For later reference let us summarize (2.2), (2.4), and (2.5):

$$(2.6) \quad \begin{aligned} -\Delta y &= \frac{1}{\nu}\lambda + g \text{ in } \Omega, \\ y &= 0 \text{ on } \partial\Omega, \\ -\Delta\lambda &= -(y - z) \text{ in } \Omega, \\ \lambda &= 0 \text{ on } \partial\Omega, \\ \nu u - \lambda &= 0 \text{ in } \Omega. \end{aligned}$$

Here, for convenience, we dropped the $*$-notation. System (2.6) is referred to as the optimality system for (2.1). From the optimality system one concludes the following regularity property.

COROLLARY 2.1. *If* $z, g \in L^2(\Omega)$, *then* $(y^*, u^*, \lambda^*) \in (H_0^1(\Omega) \cap H^2(\Omega))^3$.

In the following section we address finite difference approximations to (2.1) and (2.6).

3. Finite difference approximation of the optimality system. While finite element approximations to (2.1) are rather well investigated, see [18] and the references given there, much less rigorous analysis is available for finite difference methods. Thus, before addressing multigrid methods in the remainder of the paper, we investigate convergence of finite difference approximations to (2.1). We consider a sequence of grids $\{\Omega_h\}_{h>0}$ defined by

$$\Omega_h = \{\mathbf{x} \in \mathbf{R}^2 : x_i = s_i\,h, \quad s_i \in \mathbb{Z}\} \cap \Omega.$$

Here and below we follow the notation and terminology of [14], especially section 9. To avoid certain technicalities we assume also in this section that Ω is a square and that the values of h are chosen such that the boundaries of Ω coincide with grid lines. The case of general convex domains is addressed in Remark 1 below. The negative Laplacian with homogeneous Dirichlet boundary conditions is approximated by the common five-point stencil as in [14, section 4] and denoted by $-\Delta_h$.

For grid functions v_h and w_h defined on Ω_h we introduce the discrete L^2-scalar product

$$(v_h, w_h)_{L_h^2} = h^2 \sum_{\mathbf{x} \in \Omega_h} v_h(\mathbf{x})\, w_h(\mathbf{x}),$$

with associated norm $|v_h|_0 = (v_h, v_h)_{L_h^2}^{1/2}$. We require as well the discrete H^1-product given by

$$|v_h|_1 = \left(|v_h|_0^2 + \sum_{i=1}^2 |\partial_i^- v_h|_0^2 \right)^{1/2},$$

where ∂_i^- denotes the backward difference quotient in the x_i direction and v_h is extended by 0 on grid points outside of Ω. The spaces L_h^2 and H_h^1 consist of the sets

of grid functions v_h endowed with $|v_h|_0$, respectively, $|v_h|_1$, as norm. Further denote with M_h the vector space of nodal functions v_h defined on Ω_h which are zero on the boundary. The system of nodal functions (v_h, w_h) is denoted by $\mathcal{M}_h = M_h \times M_h$.

We need the following lemma.

LEMMA 3.1 (Poincaré–Friedrichs inequality for finite differences). *For any grid function $v_h \in M_h$, there exists a constant c_*, independent of v_h and h, such that*

$$(3.1) \qquad |v_h|_0^2 \leq c_* \sum_{i=1}^{2} |\partial_i^- v_h|_0^2,$$

where $c_* = \frac{1}{4}$.

Proof. For the proof see [20]. ∎

Functions in $L^2(\Omega)$ and $H^2(\Omega)$ are approximated by grid functions defined through their mean values with respect to elementary cells $[x_1 - \frac{h}{2}, x_1 + \frac{h}{2}] \times [x_2 - \frac{h}{2}, x_2 + \frac{h}{2}]$. This gives rise to the restriction operators $\tilde{R}_h : L^2(\Omega) \to L_h^2$ and $R_h : H_0^1(\Omega) \cap H^2(\Omega) \to L_h^2$ defined in [14, p. 232]. For the definition of H_h^2 we refer to [14], as well. Further, we define $\tilde{R}_h^2 : L^2(\Omega) \times L^2(\Omega) \to L_h^2 \times L_h^2$ by $\tilde{R}_h^2 = (\tilde{R}_h, \tilde{R}_h)$ and analogously $R_h^2 = (R_h, R_h)$.

The discrete optimal control problems are specified next:

$$(3.2) \qquad \begin{cases} \min \frac{1}{2}|y_h - \tilde{R}_h z|_0^2 + \frac{\nu}{2}|u_h|_0^2, \\ -\Delta_h y_h = u_h + \tilde{R}_h g, \qquad u_h \in L_h^2. \end{cases}$$

Let u_h^* denote the unique solution to (3.2) and set $y_h^* = y_h(u_h^*)$. The optimality system related to (3.2) is found to be

$$(3.3) \qquad \begin{aligned} -\Delta_h y_h^* &= u_h^* + \tilde{R}_h g, \\ -\Delta_h \lambda_h^* &= -(y_h^* - \tilde{R}_h z), \\ \nu u_h^* - \lambda_h^* &= 0. \end{aligned}$$

We can eliminate u_h^* from this system and obtain, dropping the superscript $*$,

$$(3.4) \qquad \begin{cases} -\nu \Delta_h y_h - \lambda_h &= \nu \tilde{R}_h g, \\ -\Delta_h \lambda_h + y_h &= \tilde{R}_h z. \end{cases}$$

To investigate the convergence of the solution of (3.4) to the solution of (2.6) as $h \to 0^+$, we introduce the family of operators

$$(3.5) \qquad \mathcal{A}_h = \begin{pmatrix} -\nu \Delta_h & -I_h \\ I_h & -\Delta_h \end{pmatrix},$$

where I_h is the identity operator on grid functions v_h. The operators $\mathcal{A}_h$ are defined between product spaces of grid functions. For us the cases $\mathcal{A}_h : H_h^1 \times H_h^1 \to H_h^{-1} \times H_h^{-1}$ and $\mathcal{A}_h : H_h^2 \times H_h^2 \to L_h^2 \times L_h^2$ are important. Here H_h^{-1} denotes the dual space of H_h^1 with L_h^2 as pivot space.

The family $\{\mathcal{A}_h\}_{h>0}$ is called H_h^1-regular if A_h is invertible and there exists a constant C_1 independent of h such that

$$\|\mathcal{A}_h^{-1}\|_{\mathcal{L}(H_h^{-1} \times H_h^{-1}, H_h^1 \times H_h^1)} \leq C_1,$$

and analogously it is called H_h^2-regular if

$$\|\mathcal{A}_h^{-1}\|_{\mathcal{L}(L_h^2 \times L_h^2, H_h^2 \times H_h^2)} \leq C_2,$$

for C_2 independent of h.

LEMMA 3.2. *The family of operators $\{\mathcal{A}_h\}_{h>0}$, with h such that the boundaries of Ω are grid lines, is H^1_h-regular.*

Proof. Let $(v_h, w_h) \in \mathcal{M}_h$ be a pair of grid functions. Then

$$(\mathcal{A}_h(v_h, w_h), (v_h, w_h))_{L^2_h \times L^2_h} = \nu(-\Delta_h v_h, v_h)_{L^2_h} + (-\Delta_h w_h, w_h)_{L^2_h}$$

$$(3.6) \qquad\qquad \geq \min(\nu, 1)\, C \sum_{i=1}^{2}(|\partial_i^- v_h|_0^2 + |\partial_i^- w_h|_0^2),$$

where C is independent of h and arises from the coercivity estimate for $-\Delta_h$, i.e.,

$$(3.7) \qquad (-\Delta_h v_h, v_h)_{L^2_h} \geq C \sum_{i=1}^{2} |\partial_i^- v_h|_0^2 \qquad \text{for all } v_h;$$

see, e.g., [14, p. 231]. Using Poincaré inequality in (3.6) results in

$$(\mathcal{A}_h(v_h, w_h), (v_h, w_h))_{L^2_h \times L^2_h} \geq C_1^{-2} |(v_h, w_h)|_{H^1_h \times H^1_h}^2 \qquad \text{for all } (v_h, w_h) \in L^2_h \times L^2_h,$$

with $C_1^{-2} = \min(\nu, 1)\, C\, c_0$. Due to the Lax–Milgram lemma $\mathcal{A}_h$ is invertible. Moreover,

$$||\mathcal{A}_h^{-1}||_{\mathcal{L}(H^{-1}_h \times H^{-1}_h, H^1_h \times H^1_h)} \leq C_1 \qquad \text{for all } h. \qquad \square$$

The infinite dimensional analogue of $\mathcal{A}_h$ is the operator

$$(3.8) \qquad\qquad \mathcal{A} = \begin{pmatrix} -\nu\,\Delta & -I \\ I & -\Delta \end{pmatrix},$$

where Δ is understood with homogeneous Dirichlet boundary conditions. It is well defined from $H^1_0(\Omega) \times H^1_0(\Omega)$ to $H^{-1}(\Omega) \times H^{-1}(\Omega)$ as well as from $(H^2(\Omega) \cap H^1_0(\Omega)) \times (H^2(\Omega) \cap H^1_0(\Omega))$ to $L^2(\Omega) \times L^2(\Omega)$. We have the following consistency result.

LEMMA 3.3. *There exists a constant C_K independent of h such that*

$$||\mathcal{A}_h R_h^2 - \tilde{R}_h^2 \mathcal{A}||_{\mathcal{L}((H^2 \cap H^1_0)^2, (H^{-1}_h \times H^{-1}_h))} \leq C_K\, h.$$

Proof. Let $(v, w) \in (H^2(\Omega) \cap H^1_0(\Omega))^2$ and note that, due to the consistency property of $-\Delta_h$ as discretization of $-\Delta$, we have

$$|\mathcal{A}_h R_h^2(v, w) - \tilde{R}_h^2 \mathcal{A}(v, w)|_{H^{-1}_h \times H^{-1}_h}^2$$

$$\leq \nu\,|(-\Delta_h)R_h v - \tilde{R}_h(-\Delta)v|_{H^{-1}_h}^2 + |(-\Delta_h)R_h w - \tilde{R}_h(-\Delta)w|_{H^{-1}_h}^2$$

$$+ |R_h v - \tilde{R}_h v|_{H^{-1}_h}^2 + |R_h w - \tilde{R}_h w|_{H^{-1}_h}^2$$

$$\leq C_K^2\, h^2\, |(v, w)|_{H^2(\Omega)^2};$$

see [14, p. 232]. $\square$

THEOREM 3.4. *There exists a constant K_1, depending on Ω, g, z, and independent of h, such that*

$$|y_h^* - R_h y^*|_1 + |u_h^* - R_h u^*|_1 + |\lambda_h^* - R_h \lambda^*|_1 \leq K_1\, h.$$

Proof. From (2.6) and (3.4) we have

$$(3.9) \qquad (y_h^*, \lambda_h^*) - R_h^2 (y^*, \lambda^*) = \mathcal{A}_h^{-1}(\tilde{R}_h^2 \mathcal{A} - \mathcal{A}_h R_h^2)(y^*, \lambda^*).$$

Lemmas 3.2 and 3.3 imply the existence of $\bar{K}_1$ such that

$$|y_h^* - R_h y^*|_1 + |\lambda_h^* - R_h \lambda^*|_1 \leq \bar{K}_1 h.$$

Using $\nu u^* = \lambda^*$ and its discrete analogue, we have the following claim. $\qquad \square$

Remark 1. In the case of a general convex domain attention must be paid to the discretization of $-\Delta$ along the boundary. The literature offers several options. For the Shortley–Weller discretization, as described in [14, p. 78], $-\Delta_h$ is H_h^1-regular and consistent with $-\Delta$ from $H^2(\Omega)$ to H_h^{-1}. Using these facts the generalization of Theorem 3.4 to convex domains is straightforward.

In the following result the assumption that the boundaries of Ω coincide with grid lines is used.

THEOREM 3.5. *There exists a constant K_2, depending on Ω, g, z, and independent of h, such that*

$$|y_h^* - R_h y^*|_0 + |u_h^* - R_h u^*|_0 + |\lambda_h^* - R_h \lambda^*|_0 \leq K_2 \, h^2.$$

Proof. We start by showing that $\mathcal{A}_h^T$ is H_h^2-regular. For this purpose it suffices to show the existence of a constant C_2 independent of h such that for all $(f_h, g_h) \in L_h^2 \times L_h^2$

$$(3.10) \qquad |(v_h, w_h)|_{H_h^2 \times H_h^2} \leq C_2 \, |(f_h, g_h)|_{L_h^2 \times L_h^2},$$

where $\mathcal{A}_h^T(v_h, w_h) = (f_h, g_h)$. Proceeding as in Lemma 3.2 one shows that $\mathcal{A}_h^T$ is H_h^1-regular. In particular, there exists $\tilde{C}_2 \geq 1$, independent of h, such that

$$(3.11) \qquad |(v_h, w_h)|_{H_h^1 \times H_h^1} \leq \tilde{C}_2 \, |(f_h, g_h)|_{L_h^2 \times L_h^2}.$$

Since $-\Delta_h$ is H_h^2-regular [14, p. 242], $\tilde{C}_2$ can also be chosen such that

$$(3.12) \qquad ||(-\Delta_h)^{-1}||_{\mathcal{L}(L_h^2, H_h^2)} \leq \tilde{C}_2.$$

Note that v_h satisfies $\nu \, v_h = (-\Delta_h)^{-1}(f_h - w_h)$. Hence by (3.11) and (3.12)

$$(3.13) \qquad |v_h|_{H_h^2} \leq \frac{2}{\nu} \, \tilde{C}_2^2 \, |(f_h, g_h)|_{L_h^2 \times L_h^2}.$$

Similarly $w_h = (-\Delta_h)^{-1}(g_h + v_h)$ and hence

$$(3.14) \qquad |w_h|_{H_h^2} \leq 2 \, \tilde{C}_2^2 \, |(f_h, g_h)|_{L_h^2 \times L_h^2}.$$

Combining (3.13) and (3.14) we have (3.10). From (3.10) it follows by duality that

$$(3.15) \qquad ||\mathcal{A}_h^{-1}||_{\mathcal{L}(H_h^{-2} \times H_h^{-2}, L_h^2 \times L_h^2)} \leq C_2.$$

Turning to consistency, due to the assumption that the boundary of Ω coincides with grid lines, we have

$$(3.16) \qquad ||(-\Delta_h)R_h - \tilde{R}_h(-\Delta)||_{\mathcal{L}(H^2, H_h^{-2})} \leq K \, h^2,$$

$$(3.17) \qquad ||R_h - \tilde{R}_h||_{\mathcal{L}(H^2, L_h^2)} \leq K \, h^2,$$

for a constant K independent of h. Estimate (3.16) is given in [14, p. 239] and (3.17) follows from a direct computation. From (3.9) and (3.16) we have

$$(3.18) \qquad |(y_h^*, \lambda_h^*) - R_h^2(y^*, \lambda^*)|_{L_h^2 \times L_h^2} \leq C_2 |(\tilde{R}_h^2 \mathcal{A} - \mathcal{A}_h R_h^2)(y^*, \lambda^*)|_{H_h^{-2} \times H_h^{-2}}.$$

Using (3.16) and (3.17), we proceed as in the proof of Lemma 3.3 to obtain the desired result.   □

The approximation results stated in Theorem 3.5 are demonstrated in numerical experiments with global mesh refinement (multigrid); see [2].

4. The multigrid method. Multigrid methods have been extensively used to solve discretized partial differential equations; see, e.g., [21] and the references given there. This fact has motivated intensive research towards the determination of convergence properties of multigrid schemes; see [5, 6, 9, 10, 11, 14, 15]. Multigrid methods have also been used to solve optimal control problems. Most of these contributions except for [12] are rather recent, e.g., [1, 2, 3, 4, 19]. Concerning the convergence theory of multigrid applied to systems of partial differential equations and, in particular, to optimality systems, the theory is far from being complete.

The purpose of the present work is to analyze a multigrid algorithm that solves the optimality system (2.6) with typical multigrid efficiency. We briefly describe the multigrid framework to keep this paper self-contained. Let us index the operators and variables defined on the grid with mesh size $h = h_k = 1/2^k$, $k = 1, \ldots, L$, with the index k, and for simplicity of presentation let us introduce vector notation: we let $\mathbf{w} = (u, v)$ and $|\mathbf{w}|_0 = |(u, v)|_0$, etc.

Consider the discrete problem (3.4) expressed as

$$(4.1) \qquad \mathcal{A}_k \boldsymbol{\phi}_k = \mathbf{f}_k \text{ on } \Omega_k,$$

where $\boldsymbol{\phi}_k = (y_k, \lambda_k)$ and $\mathbf{f}_k = (g_k, z_k)$ are defined on the mesh Ω_{h_k}.

For the purpose of multigrid methods it is important to utilize the fact that the solution of (4.1) is equivalent to solve $\mathcal{A}_k \boldsymbol{\phi}_k^e = \mathbf{r}_k$, where $\boldsymbol{\phi}_k^e = \bar{\boldsymbol{\phi}}_k - \boldsymbol{\phi}_k$ is the error grid function between the solution $\bar{\boldsymbol{\phi}}_k$ to (4.1) and its current approximation $\boldsymbol{\phi}_k$, and $\mathbf{r}_k$ is the residual defined by

$$(4.2) \qquad \mathbf{r}_k = \mathbf{f}_k - \mathcal{A}_k \boldsymbol{\phi}_k.$$

In fact, the multigrid strategy is to solve for all frequency components of the error using multiple grids.

On the grid of level k, a smoothing procedure is applied in order to solve for the high-frequency components of the error. This is an iterative scheme denoted by $\boldsymbol{\phi}_k^{(m)} = (\mathcal{S}_k)^m(\boldsymbol{\phi}_k, \mathbf{f}_k)$, where $(\mathcal{S}_k)^m$ is a linear smoothing operator applied m times. One sweep of this iteration is written in the form $\boldsymbol{\phi}_k^{(m)} = \boldsymbol{\phi}_k^{(m-1)} + \mathcal{R}_k(\mathbf{f}_k - \mathcal{A}_k \boldsymbol{\phi}_k^{(m-1)})$, where the operator $\mathcal{R}_k$ applies to the residual.

To correct for the smooth components of the error, a coarse grid correction (CGC) is defined. For this purpose a coarse grid problem for the error function is constructed on the grid with mesh size h_{k-1}:

$$(4.3) \qquad \mathcal{A}_{k-1} \boldsymbol{\phi}_{k-1} = \mathcal{I}_k^{k-1} \mathbf{r}_k,$$

where $\boldsymbol{\phi}_{k-1}$ aims to represent, on the coarse grid Ω_{k-1}, the error $\boldsymbol{\phi}_k^e$ on the next finer grid. Because of Dirichlet boundary conditions, we have $\boldsymbol{\phi}_{k-1} = 0$ at the boundary.

The operator $\mathcal{I}_k^{k-1} : \mathcal{M}_k \to \mathcal{M}_{k-1}$ restricts the residual computed at level k to the grid with level $k-1$.

Once the coarse grid problem is solved, the CGC follows:

$$(4.4) \qquad \phi_k^{new} = \phi_k + \mathcal{I}_{k-1}^k \, \phi_{k-1},$$

where $\mathcal{I}_{k-1}^k : \mathcal{M}_{k-1} \to \mathcal{M}_k$ is an interpolation operator. Here ϕ_k represents the current approximation at level k as it was obtained by the smoothing process and before coarsening. If the high-frequency components of the error on the finer grid k were well damped, then the solution at level ϕ_{k-1} should provide enough resolution for the error of ϕ_k through $\mathcal{I}_{k-1}^k \phi_{k-1}$.

The idea of transferring to a coarser grid can be applied along the set of nested meshes. One starts at level k with a given initial approximation (zero) and applies the smoothing iteration m_1 times. The residual is then computed and transferred to the next coarser grid while ϕ_k obtained by smoothing is left unchanged. On the coarse grid with index $k-1$ the smoothing process is again applied. This procedure is repeated until the coarsest grid is reached.

On the coarsest grid, one solves the problem exactly and the result is used to improve ϕ_k via (4.4). The CGC is then followed by m_2 postsmoothing steps at level k before the CGC procedure followed by postsmoothing is repeated for the next (if any) finer level. This entire process represents one multigrid $V(m_1, m_2)$-cycle.

A compact description of the multigrid method is given in section 6.

In the following sections we specify and analyze the multigrid components introduced here.

4.1. Smoothing iterations. Numerical experience [2] has shown that in order to obtain a multigrid algorithm which is robust with respect to changes of ν, care must be taken in the choice of the smoother. For example, when using the Picard–Gauss–Seidel iteration [1, 2], difficulties arise when the value of the weight of the cost of the control is smaller than h^2, which may easily occur when coarse grids are used. On the other hand, the collective Gauss–Seidel (CGS) scheme appears to be a reasonable choice [2]. Notice that this iterative method belongs to the class of Vanka smoothers [22].

To analyze the CGS scheme, let us introduce some notation:

$$(4.5) \quad \mathcal{A}_h^+ = \left[\begin{array}{cc} \nu \Sigma_h^+ & 0 \\ 0 & \Sigma_h^+ \end{array} \right], \quad \mathcal{A}_h^- = \left[\begin{array}{cc} \nu \Sigma_h^- & 0 \\ 0 & \Sigma_h^- \end{array} \right], \quad \mathcal{D}_h = \left[\begin{array}{cc} \nu \frac{4}{h^2} I_h & -I_h \\ I_h & \frac{4}{h^2} I_h \end{array} \right],$$

where I_h is the identity operator on Ω_h and the operators Σ_h^+ and Σ_h^- are given in stencil form by

$$(4.6) \qquad \Sigma_h^+ = \frac{1}{h^2} \left[\begin{array}{ccc} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{array} \right], \qquad \Sigma_h^- = \frac{1}{h^2} \left[\begin{array}{ccc} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{array} \right].$$

Thus a sweep of the forward CGS scheme and of a backward CGS scheme are expressed by

$$(4.7) \qquad (\mathcal{D}_h - \mathcal{A}_h^+) \, \phi^{(1)} - \mathcal{A}_h^- \, \phi^{(0)} = \mathbf{f} \quad \text{and} \quad (\mathcal{D}_h - \mathcal{A}_h^-) \, \phi^{(2)} - \mathcal{A}_h^+ \, \phi^{(1)} = \mathbf{f},$$

respectively. In the symmetric version of the CGS smoother, the forward CGS step is followed by a backward CGS step. The resulting iteration can be written in the

linear form

$$\phi^{(2)} = \phi^{(0)} + \mathcal{R}_h \left[f_h - \mathcal{A}_h \phi^{(0)}\right], \qquad \text{where } \mathcal{R}_h = (\mathcal{D}_h - \mathcal{A}_h^-)^{-1} \mathcal{D}_h (\mathcal{D}_h - \mathcal{A}_h^+)^{-1},$$

which gives the smoothing operator $\mathcal{S}_h = I_h - \mathcal{R}_h \mathcal{A}_h$.

We analyze this iteration by local Fourier analysis [10, 21]. Consider the Fourier space spanned by the functions

$$\phi(\boldsymbol{\theta}, \mathbf{x}) = \mathbf{a} \, e^{i\theta_1 x/h} \, e^{i\theta_2 y/h}, \qquad \boldsymbol{\theta} = (\theta_1, \theta_2),$$

where $\mathbf{a} = (1,1)^T$. One defines

$$\phi \text{ low-frequency component} \iff \boldsymbol{\theta} \in [-\frac{\pi}{2}, \frac{\pi}{2})^2,$$

$$\phi \text{ high-frequency component} \iff \boldsymbol{\theta} \in [-\pi, \pi)^2 \setminus [-\frac{\pi}{2}, \frac{\pi}{2})^2.$$

In the Fourier space consider the symbols of the discrete operators $(\mathcal{D}_h - \mathcal{A}_h^+)$ and $\mathcal{A}_h^-$ for the forward CGS iteration. We have

$$(4.8) \qquad \overline{(\mathcal{D}_h - \mathcal{A}_h^+)}(\boldsymbol{\theta}) = -\frac{1}{h^2} \left[\begin{array}{cc} \nu \left(e^{-i\theta_1} + e^{-i\theta_2} - 4\right) & h^2 \\ -h^2 & \left(e^{-i\theta_1} + e^{-i\theta_2} - 4\right) \end{array} \right],$$

$$(4.9) \qquad \overline{(\mathcal{A}_h^-)}(\boldsymbol{\theta}) = -\frac{1}{h^2} \left[\begin{array}{cc} \nu \left(e^{i\theta_1} + e^{i\theta_2}\right) & 0 \\ 0 & \left(e^{i\theta_1} + e^{i\theta_2}\right) \end{array} \right].$$

Thus, the symbol of the forward CGS scheme is given by

$$\overline{\mathcal{S}_h^+}(\boldsymbol{\theta}) = (\overline{(\mathcal{D}_h - \mathcal{A}_h^+)}(\boldsymbol{\theta}))^{-1} \, \overline{(\mathcal{A}_h^-)}(\boldsymbol{\theta}).$$

In this framework, the smoothing factor of the forward CGS scheme for the optimality system is defined by

$$(4.10) \qquad \mu = \mu(\mathcal{S}_h^+) = \sup\{|\rho(\overline{\mathcal{S}_h^+}(\boldsymbol{\theta}))| : \boldsymbol{\theta} \text{ high frequency }\},$$

where ρ denotes the spectral radius. In the same way one defines the smoothing factor for the backward CGS step: $\mathcal{S}_h^- = (\mathcal{D}_h - \mathcal{A}_h^-)^{-1} (\mathcal{A}_h^+)$. The symmetric CGS scheme is then given by $\mathcal{S}_h^s = \mathcal{S}_h^- \mathcal{S}_h^+$. Since the symbols associated with the CGS iterative schemes considered here are 2×2 operators with entries being functions of (θ_1, θ_2), it is possible, by any symbolical package, to obtain the eigenvalues of the symbols. Thus we have the following.

Remark 2. By inspection in the range of high frequencies for $h \in [0.01, 0.25]$ and ν ranging in the interval $[10^{-6}, 1]$, the following upper bounds for the smoothing factor are found:

$$\mu(\mathcal{S}_h^+) \le 0.5, \quad \mu(\mathcal{S}_h^-) \le 0.5, \quad \text{and } \mu(\mathcal{S}_h^s) \le 0.25.$$

Therefore, we can conclude that the forward CGS, the backward CGS, and the symmetric CGS are all good smoothers for the purpose of the multigrid scheme.

4.2. Intergrid transfer operators. Among two grids $\overline{\Omega}_k$ and $\overline{\Omega}_{k-1}$, corresponding to mesh sizes h_k and h_{k-1}, we define a prolongation operator, $I_{k-1}^k : M_{k-1} \to M_k$, given in stencil form by

$$(4.11) \qquad I_{k-1}^k = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}.$$

This choice is consistent with the assumption of *bilinear* finite elements on each square partition of the discretization. That is, on each square partition $[x_i, x_{i+1}] \times [y_j, y_{j+1}]$ of $\overline{\Omega}_{k-1}$, the piecewise bilinear function which interpolates U at the nodes is given by

$$\tilde{u}(\tilde{x}, \tilde{y}) = (1 - \tilde{x})(1 - \tilde{y})\, u_{ij} + \tilde{x}(1 - \tilde{y})\, u_{i+1\,j} + \tilde{y}(1 - \tilde{x})\, u_{ij+1} + \tilde{x}\tilde{y}\, u_{i+1\,j+1}.$$

Here, $0 \leq \tilde{x}, \tilde{y} \leq 1$ are local coordinates such that $x = x_i + \tilde{x}\, h_{k-1}$ and $y = y_j + \tilde{y}\, h_{k-1}$. Thus the prolongation of u on a grid point of $\overline{\Omega}_k$ is the value of $\tilde{u}$ corresponding to that grid point.

Next, we define the full-weighting restriction operator, $I_k^{k-1} : M_k \to M_{k-1}$, given in stencil form by

$$(4.12) \qquad I_k^{k-1} = \frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix},$$

with the inner product

$$(4.13) \qquad (v, w)_k = \sum_{i=2}^{N_k} \sum_{j=2}^{N_k} h_k^2\, v_{kij}\, w_{kij},$$

where $N_k = 2^k$. We have that the restriction operator is the adjoint of the prolongation operator [13], in the sense that

$$(I_k^{k-1} v_k, w_{k-1})_{k-1} = (v_k, I_{k-1}^k w_{k-1})_k \quad \text{for all } v_k \in M_k,\ w_{k-1} \in M_{k-1}.$$

The action of I_{k-1}^k (resp., I_k^{k-1}) on pairs of grid functions is denoted by $\mathcal{I}_{k-1}^k$ (resp., $\mathcal{I}_k^{k-1}$).

In order to extend the multigrid convergence theory formulated in [9, 15, 6] to the present multigrid method for optimality systems, we need the following lemma [9].

LEMMA 4.1. *Let us introduce the bilinear form $a_k(u, v) = (-\Delta_k u, v)_k$, $u,\ v \in M_k$. The prolongation operator* (4.11) *satisfies the following conditions:*

$$(4.14) \quad a_k(I_{k-1}^k u_{k-1}, I_{k-1}^k u_{k-1}) \leq a_{k-1}(u_{k-1}, u_{k-1}) \qquad \text{for all } u_{k-1} \in M_{k-1},$$

$$(4.15) \quad (I_{k-1}^k u_{k-1}, I_{k-1}^k u_{k-1})_k \leq (u_{k-1}, u_{k-1})_{k-1} \qquad \text{for all } u_{k-1} \in M_{k-1}.$$

In particular, the result of Lemma 4.1 applied to the operator (3.5) results in the following:

$$(4.16) \qquad (\mathcal{A}_k \mathcal{I}_{k-1}^k \mathbf{w}_{k-1}, \mathcal{I}_{k-1}^k \mathbf{w}_{k-1})_k \leq (\mathcal{A}_{k-1} \mathbf{w}_{k-1}, \mathbf{w}_{k-1})_{k-1}$$

for all $\mathbf{w}_{k-1} = (u_{k-1}, v_{k-1}) \in \mathcal{M}_{k-1}$.

5. Two grid local Fourier analysis. In this section we perform local Fourier analysis [10, 21] of the two grid solution process for the optimal control optimality system. That is, we apply the local Fourier analysis to the two grid operator given by

$$(5.1) \qquad TG_k^{k-1} = \mathcal{S}_k^{m_2} \left[\mathcal{I}_k - \mathcal{I}_{k-1}^k \left(\mathcal{A}_{k-1} \right)^{-1} \mathcal{I}_k^{k-1} \mathcal{A}_k \right] \mathcal{S}_k^{m_1}.$$

Here, the coarse grid operator is $CG_k^{k-1} = \left[\mathcal{I}_k - \mathcal{I}_{k-1}^k \left(\mathcal{A}_{k-1} \right)^{-1} \mathcal{I}_k^{k-1} \mathcal{A}_k \right]$.

The local Fourier analysis considers infinite grids, $G_k = \{ (ih_k, jh_k), i, j \in \mathbb{Z} \}$, and therefore the influence of boundary conditions is not taken into account. Nevertheless, experience shows that local Fourier analysis provides predictions of multigrid convergence which are very sharp. This analysis is based on the quadruples of Fourier components

$$\phi_k(\boldsymbol{\theta}, \mathbf{x}) = e^{i\theta_1 x / h_k} e^{i\theta_2 y / h_k}$$

that coincide on G_{k-1}. For any low frequency $\boldsymbol{\theta} = (\theta_1, \theta_2) \in [-\pi/2, \pi/2)^2$, we consider

$$
\begin{aligned}
\boldsymbol{\theta}^{(0,0)} &:= (\theta_1, \theta_2), & \boldsymbol{\theta}^{(1,1)} &:= (\overline{\theta_1}, \overline{\theta_2}), \\
\boldsymbol{\theta}^{(1,0)} &:= (\overline{\theta_1}, \theta_2), & \boldsymbol{\theta}^{(0,1)} &:= (\theta_1, \overline{\theta_2}),
\end{aligned}
$$

where

$$
\overline{\theta_i} = \begin{cases} \theta_i + \pi & \text{if } \theta_i < 0, \\ \theta_i - \pi & \text{if } \theta_i \geq 0. \end{cases}
$$

We have $\phi(\boldsymbol{\theta}^{(0,0)}, \cdot) = \phi(\boldsymbol{\theta}^{(1,1)}, \cdot) = \phi(\boldsymbol{\theta}^{(1,0)}, \cdot) = \phi(\boldsymbol{\theta}^{(0,1)}, \cdot)$ for $\boldsymbol{\theta}^{(0,0)} \in [-\pi/2, \pi/2)^2$ and $(x, y) \in G_{k-1}$. Denote with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ and consider $\boldsymbol{\alpha} \in \{ (0,0), (1,1), (1,0), (0,1) \}$; then on G_{k-1} we have $\phi_k(\boldsymbol{\theta}^{\boldsymbol{\alpha}}, \mathbf{x}) = \phi_{k-1}(2\boldsymbol{\theta}^{(0,0)}, \mathbf{x})$. The four components $\phi_k(\boldsymbol{\theta}^{\boldsymbol{\alpha}}, \cdot)$ are called harmonics. For a given $\boldsymbol{\theta} = \boldsymbol{\theta}^{(0,0)} \in [-\pi/2, \pi/2)^2$, the four dimensional space of harmonics is defined by

$$E_k^\theta = span[\phi_k(\boldsymbol{\theta}^{\boldsymbol{\alpha}}, \cdot) : \boldsymbol{\alpha} \in \{ (0,0), (1,1), (1,0), (0,1) \}].$$

For each θ, the spaces $E_k^\theta \times E_k^\theta$ are invariant under the action of TG_k^{k-1}; see [21]. We now study the action of TG_k^{k-1} on an arbitrary couple $(\psi_y, \psi_\lambda) \in E_k^\theta \times E_k^\theta$, where

$$\psi_y = \sum_{\boldsymbol{\alpha}} A^{\boldsymbol{\alpha}} \phi_k(\boldsymbol{\theta}^{\boldsymbol{\alpha}}, \cdot) \quad \text{and} \quad \psi_\lambda = \sum_{\boldsymbol{\alpha}} B^{\boldsymbol{\alpha}} \phi_k(\boldsymbol{\theta}^{\boldsymbol{\alpha}}, \cdot).$$

We analyze how the vector of coefficients $(A^{(0,0)}, \dots, B^{(0,0)}, \dots)$ is transformed if the two grid iteration (5.1) is applied to (ψ_y, ψ_λ). We use the following theorem, which is an extension of Theorem 4.4.1 of [21] to our system of equations.

THEOREM 5.1. *Under the assumption that all multigrid components in* (5.1) *are linear and that* $(\mathcal{A}_{k-1})^{-1}$ *exists, the coarse grid operator* CG_k^{k-1} *is represented on* E_k^θ *by the* 8×8 *matrix* $\widehat{CG}_k^{k-1}(\boldsymbol{\theta})$,

$$\widehat{CG}_k^{k-1}(\boldsymbol{\theta}) = [\hat{\mathcal{I}}_k - \hat{\mathcal{I}}_{k-1}^k(\boldsymbol{\theta}) \, (\widehat{\mathcal{A}}_{k-1}(2\boldsymbol{\theta}))^{-1} \, \hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta}) \, \widehat{\mathcal{A}}_k(\boldsymbol{\theta})],$$

for each $\boldsymbol{\theta} \in [-\pi/2, \pi/2)^2$. *Here,* $\hat{\mathcal{I}}_k$ *and* $\widehat{\mathcal{A}}_k(\boldsymbol{\theta})$ *are* 8×8 *matrices,* $\hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta})$ *is a* 2×8 *matrix,* $\hat{\mathcal{I}}_{k-1}^k(\boldsymbol{\theta})$ *is a* 8×2 *matrix, and* $\widehat{\mathcal{A}}_{k-1}(2\boldsymbol{\theta})$ *is a* 2×2 *matrix.*

If the spaces $E_k^\theta \times E_k^\theta$ are invariant under the smoothing operator $\mathcal{S}_k$, i.e., (the 8×8 matrix) $\hat{\mathcal{S}}_k(\boldsymbol{\theta}) : E_k^\theta \times E_k^\theta \to E_k^\theta \times E_k^\theta$ for all $\boldsymbol{\theta} \in [-\pi/2, \pi/2)^2$, we also have a representation of TG_k^{k-1} on $E_k^\theta \times E_k^\theta$ by a 8×8 matrix given by

$$\widehat{TG}_k^{k-1}(\boldsymbol{\theta}) = \hat{\mathcal{S}}_k(\boldsymbol{\theta})^{m_2} \, \widehat{CG}_k^{k-1}(\boldsymbol{\theta}) \, \hat{\mathcal{S}}_k(\boldsymbol{\theta})^{m_1}.$$

We now give the symbols of the operators above in explicit form.

The coarse grid operator $\mathcal{A}_{k-1}$ is

$$\widehat{\mathcal{A}}_{k-1}(2\boldsymbol{\theta}) = \left[\begin{array}{cc} \nu \frac{4-2(\cos(2\theta_1)+\cos(2\theta_2))}{h_{k-1}^2} & -1 \\ 1 & \frac{4-2(\cos(2\theta_1)+\cos(2\theta_2))}{h_{k-1}^2} \end{array} \right].$$

The fine grid operator is $\mathcal{A}_k$. The symbol $\widehat{\mathcal{A}}_k(\boldsymbol{\theta})$ is given by

$$\left[\begin{array}{cccccccc} \nu\, l(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & \nu\, l(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & \nu\, l(\boldsymbol{\theta}^{(1,0)}) & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \nu\, l(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & l(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & l(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & l(\boldsymbol{\theta}^{(1,0)}) & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & l(\boldsymbol{\theta}^{(0,1)}) \end{array} \right],$$

where

$$l(\boldsymbol{\theta}^{\boldsymbol{\alpha}}) = \frac{4 - 2(\cos(\theta_1^{\alpha_1}) + \cos(\theta_2^{\alpha_2}))}{h_k^2}.$$

The restriction operator is $\mathcal{I}_k^{k-1}$. The symbol $\hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta})$ is given by

$$\left[\begin{array}{cccccccc} I(\boldsymbol{\theta}^{(0,0)}) & I(\boldsymbol{\theta}^{(1,1)}) & I(\boldsymbol{\theta}^{(1,0)}) & I(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I(\boldsymbol{\theta}^{(0,0)}) & I(\boldsymbol{\theta}^{(1,1)}) & I(\boldsymbol{\theta}^{(1,0)}) & I(\boldsymbol{\theta}^{(0,1)}) \end{array} \right],$$

where

$$I(\boldsymbol{\theta}^{\boldsymbol{\alpha}}) = I_k^{k-1}(\boldsymbol{\theta}^{\boldsymbol{\alpha}}) = \frac{1}{4}(1 + \cos(\theta_1^{\alpha_1}))(1 + \cos(\theta_2^{\alpha_2})).$$

For the prolongation operator we have $\hat{\mathcal{I}}_{k-1}^k(\boldsymbol{\theta}) = \hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta})^T$.

For the smoothing iteration $\mathcal{S}_k$ consider the forward CGS scheme as described in section 4.1. On $E_k^\theta \times E_k^\theta$ it is given by $(\overline{(\mathcal{D}_h - \mathcal{A}_h^+)(\boldsymbol{\theta})})^{-1} \overline{(\mathcal{A}_h^-)}(\boldsymbol{\theta})$, where $(\overline{(\mathcal{D}_h - \mathcal{A}_h^+)(\boldsymbol{\theta})})$ is as follows:

$$\frac{1}{h^2} \left[\begin{array}{cccccccc} \nu\, s_+(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 & -h^2 & 0 & 0 & 0 \\ 0 & \nu\, s_+(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 & 0 & -h^2 & 0 & 0 \\ 0 & 0 & \nu\, s_+(\boldsymbol{\theta}^{(1,0)}) & 0 & 0 & 0 & -h^2 & 0 \\ 0 & 0 & 0 & \nu\, s_+(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & -h^2 \\ h^2 & 0 & 0 & 0 & s_+(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 \\ 0 & h^2 & 0 & 0 & 0 & s_+(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 \\ 0 & 0 & h^2 & 0 & 0 & 0 & s_+(\boldsymbol{\theta}^{(1,0)}) & 0 \\ 0 & 0 & 0 & h^2 & 0 & 0 & 0 & s_+(\boldsymbol{\theta}^{(0,1)}) \end{array} \right],$$

and the operator $\overline{(\mathcal{A}_h^-)}(\boldsymbol{\theta})$ is given by

$$\frac{1}{h^2} \left[\begin{array}{cccccccc} \nu\, s_-(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \nu\, s_-(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \nu\, s_-(\boldsymbol{\theta}^{(1,0)}) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \nu\, s_-(\boldsymbol{\theta}^{(0,1)}) & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & s_-(\boldsymbol{\theta}^{(0,0)}) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s_-(\boldsymbol{\theta}^{(1,1)}) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & s_-(\boldsymbol{\theta}^{(1,0)}) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & s_-(\boldsymbol{\theta}^{(0,1)}) \end{array} \right],$$

(m_1, m_2)	Local Fourier analysis		Experim.
	$\mu_{loc}^{m_1+m_2}$	$\eta(TG_k^{k-1})$	$V(m_1, m_2)$
(1,1)	0.25	0.25	0.30
(2,1)	0.125	0.12	0.12
(2,2)	0.06	0.08	0.08
(3,2)	0.03	0.06	0.06
(3,3)	0.01	0.05	0.05

where

$$s_+(\boldsymbol{\theta^\alpha}) = 4 - e^{-i\theta_1^{\alpha_1}} - e^{-i\theta_2^{\alpha_2}} \quad \text{and} \quad s_-(\boldsymbol{\theta^\alpha}) = -e^{i\theta_1^{\alpha_1}} - e^{i\theta_2^{\alpha_2}}.$$

Based on the representation on TG_k^{k-1} by a 8×8 matrix $\widehat{TG}_k^{k-1}(\boldsymbol{\theta})$, we can calculate the *convergence factor*:

$$\eta(TG_k^{k-1}) = \sup\{\rho(\widehat{TG}_k^{k-1}(\boldsymbol{\theta})) : \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2\}.$$

Here, $\rho(\widehat{TG}_k^{k-1}(\boldsymbol{\theta}))$ is the spectral radius of $\widehat{TG}_k^{k-1}(\boldsymbol{\theta})$.

Under the invariance property advocated by Theorem 5.1, to measure the smoothing property of the iteration one can assume an *ideal coarse grid correction* which annihilates the low-frequency error components and leaves the high-frequency error components unchanged. That is, one defines the projection operator Q_k^{k-1} on E_k^θ by

$$Q_k^{k-1} \phi(\boldsymbol{\theta}, \cdot) = \begin{cases} 0 \text{ if } \boldsymbol{\theta} = \boldsymbol{\theta}^{(0,0)} \in [-\pi/2, \pi/2)^2, \\ \phi(\boldsymbol{\theta}, \cdot) \text{ if } \boldsymbol{\theta} \in \{\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(1,0)}, \boldsymbol{\theta}^{(0,1)}\}. \end{cases}$$

On the space $E_k^\theta \times E_k^\theta$ we then have

$$(5.2) \qquad \widehat{Q}_k^{k-1}(\boldsymbol{\theta}) = \begin{bmatrix} Q_k^{k-1} & 0 \\ 0 & Q_k^{k-1} \end{bmatrix} \qquad \text{for } \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2.$$

In this framework the smoothing property of $\mathcal{S}_k$ is defined as follows:

$$(5.3) \qquad \mu_{loc} = \mu(\mathcal{S}_k, m) = \sup\left\{ \sqrt[m]{|\rho(\widehat{Q}_k^{k-1} \hat{\mathcal{S}}_k(\boldsymbol{\theta})^m)|} : \boldsymbol{\theta} \in [-\pi/2, \pi/2)^2 \right\}.$$

Notice that, assuming an ideal CGC takes place, the convergence factor of the two grid scheme is given by $\mu_{loc}^{m_1+m_2}$.

We complete this section by reporting in Table 5.1 the values of $\eta(TG_k^{k-1})$ and those of $\mu(\mathcal{S}_k, m)$ obtained with the two grid analysis described above. Here the forward Gauss–Seidel smoother is used. For comparison, the observed value of convergence factor defined as the "asymptotic" value of the ratio between the discrete L^2 norms of residuals resulting from two successive multigrid cycles on the finest mesh is reported. Notice that the values reported in Table 5.1 are typical of the standard Poisson model problem. These values have been obtained considering the mesh size value h ranging in the interval $[0.01, 0.25]$ corresponding to the interval of mesh sizes used in the multigrid code. The value of the weight ν has been taken in the interval $[10^{-6}, 1]$.

6. General multigrid convergence theory. In this section, we prove multigrid convergence for the optimal control problem in a more general functional setting. We use the framework in [6, 9, 11] adapted to the nonsymmetric system above. This framework applies directly to elliptic problems with prescribed boundary conditions on bounded polygonal domains.

For the purpose of our analysis, we briefly describe multigrid convergence theory for scalar Poisson equation discretized by the finite difference method on a unit square. Consider

(6.1)
$$\begin{aligned} -\Delta y &= f \text{ in } \Omega, \\ y &= 0 \text{ on } \partial\Omega. \end{aligned}$$

The matrix form of this problem is

(6.2)
$$\hat{A}_k y_k = f_k.$$

Let $\hat{P}_{k-1} : M_k \to M_{k-1}$ (resp., $I_k^{k-1} : M_k \to M_{k-1}$) be the $\hat{A}_k$ (resp., L_k^2) projections defined by

$$(\hat{A}_{k-1}\hat{P}_{k-1}u, v)_{k-1} = (\hat{A}_k u, I_{k-1}^k v)_k \text{ (resp., } (I_k^{k-1}u, v)_{k-1} = (u, I_{k-1}^k v)_k)$$

for all $u \in M_k$ and $v \in M_{k-1}$. Let $\hat{R}_k : M_k \to M_k$ be an iteration operator. Then the V-cycle multigrid algorithm to solve (6.2) in recursive form is given as follows.

MULTIGRID ALGORITHM $V(m_1, m_2)$.
Set $\hat{B}_1 = \hat{A}_1^{-1}$. For $k \geq 2$ define $\hat{B}_k : M_k \to M_k$ in terms of $\hat{B}_{k-1}$ as follows. Let $g \in M_k$.

1. Set $y^0 = 0$.
2. Define y^l for $l = 1, \dots, m_1$ by

$$y^l = y^{l-1} + \hat{R}_k(g - \hat{A}_k\, y^{l-1}).$$

3. Set $y^{m_1+1} = y^{m_1} + I_{k-1}^k q$, where

$$q = \hat{B}_{k-1}I_k^{k-1}(g - \hat{A}_k\, y^{m_1}).$$

4. Set $\hat{B}_k g = y^{m_1+m_2+1}$, where y^ℓ for $\ell = m_1 + 2, \dots, m_1 + m_2 + 1$ is given by step 2 ($\hat{R}_k^t$ instead of $\hat{R}_k$).

For the purpose of analysis, we take $m_1 = 1$ and $m_2 = 0$.
From the definition of $\hat{P}_{k-1}$, we see that

$$I_k^{k-1}\hat{A}_k = \hat{A}_{k-1}\hat{P}_{k-1}.$$

Let $\hat{S}_k = I_k - \hat{R}_k\hat{A}_k$ for $k > 1$, where I_k denotes the identity on M_k. Then $\hat{S}_k y = y - y^1$. Now for $y \in M_k$, $k = 2, \dots, L$, we have

(6.3)
$$\begin{aligned} (I_k - \hat{B}_k\hat{A}_k)\, y &= y - y^1 - I_{k-1}^k q \\ &= \hat{S}_k\, y - I_{k-1}^k \hat{B}_{k-1}\hat{A}_{k-1}\hat{P}_{k-1}\hat{S}_k\, y \\ &= [I_k - I_{k-1}^k \hat{B}_{k-1}\hat{A}_{k-1}\hat{P}_{k-1}]\, \hat{S}_k\, y \\ &= [(I_k - I_{k-1}^k\hat{P}_{k-1}) + I_{k-1}^k(I_{k-1} - \hat{B}_{k-1}\hat{A}_{k-1})\, \hat{P}_{k-1}]\hat{S}_k\, y. \end{aligned}$$

The convergence results of the multigrid method are expressed in terms of the error operators $\hat{E}_k := I_k - \hat{B}_k\hat{A}_k$ and $\hat{E} := \hat{E}_L$. In the following, let C denote a

generic constant independent of k that can have different values in different places, unless otherwise stated.

In order to prove convergence of the multigrid algorithm, the following two conditions are required. There exists a constant $\bar{C}_{\hat{R}}$ independent of y and k such that

$$(6.4) \qquad \frac{|y|_0^2}{\mu(\hat{A}_k)} \leq \bar{C}_{\hat{R}}(\bar{\bar{R}}y, y) \quad \text{for all } y \in M_k,$$

where $\mu(\hat{A}_k)$ denotes the maximum eigenvalue of $\hat{A}_k$, $\bar{\bar{R}} = (I_k - \hat{S}_k^* \hat{S}_k) \hat{A}_k^{-1}$, $\hat{S}_k^* = I - \hat{R}_k^t \hat{A}_k$, and $*$ denotes adjoint with respect to the inner product $(\hat{A}_k \cdot, \cdot)$. Next, for $k > 1$ define $\hat{T}_k = \hat{R}_k \hat{A}_k$. We assume that there exists a constant θ, $0 < \theta < 2$, independent of y such that

$$(6.5) \qquad (\hat{A}_k \hat{T}_k y, \hat{T}_k y)_k \leq \theta(\hat{A}_k \hat{T}_k y, y)_k \quad \text{for all } y \in M_k.$$

In this paper, we are dealing with multigrid for finite difference method applied to the Poisson equation on rectangular domains. In this case the stiffness matrix is exactly the same as that arising from the finite element case. Hence we have the following result from [9].

THEOREM 6.1. *Let $\hat{R}_k$ satisfy (6.4) and (6.5) for $k > 1$. Then there exists a positive constant $\hat{\delta} < 1$ such that*

$$(\hat{A}_L \hat{E}_L y, \hat{E}_L y)_L \leq \hat{\delta}^2 (\hat{A}_L y, y)_L \quad \text{for all } y \in M_L,$$

where $\hat{\delta} = CL/(CL + 1)$.

Remark 3. The dependence of $\hat{\delta}$ on L can be removed by a perturbation analysis given in [16].

Remark 4. For the multigrid algorithm $V(m, 0)$ one obtains $\hat{\delta} = CL/(CL + m)$; see [9]. The constant C depends linearly on $\bar{C}_{\hat{R}}$; see [7, 8, 9] to find estimates of these constants. As discussed in [7], the $\hat{\delta}$ estimate in Theorem 6.1 is pessimistic, in the sense that the observed $\hat{\delta}$ is smaller than the theoretical one and their difference becomes larger for larger values of m.

To prove convergence of multigrid for the optimal control optimality system, we first consider the decoupled symmetric system:

$$(6.6) \qquad \begin{aligned} -\nu \Delta y &= \nu g \text{ in } \Omega, \\ y &= 0 \text{ on } \partial\Omega, \\ -\Delta \lambda &= z \text{ in } \Omega, \\ \lambda &= 0 \text{ on } \partial\Omega. \end{aligned}$$

This system is exactly two copies of Poisson equation, hence the multigrid convergence theory for this system inherits the properties of the scalar case. In fact, if we define

$$(6.7) \qquad \hat{\mathcal{A}}_k = \begin{pmatrix} \nu \, \hat{A}_k & 0 \\ 0 & \hat{A}_k \end{pmatrix},$$

and analogously $\hat{\mathcal{B}}_k$, $\hat{\mathcal{E}}_k$, etc., as the system counterparts of $\hat{B}_k$, $\hat{E}_k$, etc., then the multigrid algorithm has exactly the same form as (6.3) with $\hat{\mathcal{B}}_k$, $\hat{\mathcal{A}}_k$, etc., replacing $\hat{B}_k$, $\hat{A}_k$, etc. As a consequence we have the following theorem.

THEOREM 6.2. *Under the assumption of Theorem* 6.1, *there exists a positive constant* $\hat{\delta} < 1$ *such that*

$$(6.8) \qquad (\hat{\mathcal{A}}_L \hat{\mathcal{E}}_L(y, \lambda), \hat{\mathcal{E}}_L(y, \lambda))_L \leq \hat{\delta}^2 (\hat{\mathcal{A}}_L(y, \lambda), (y, \lambda))_L \quad \text{for all } (y, \lambda) \in \mathcal{M}_L,$$

where $\hat{\delta}$ *has the same form as in Theorem* 6.1.

To analyze the optimality system we let

$$\mathcal{A}_k = \hat{\mathcal{A}}_k + d_k,$$

where

$$d_k = \begin{pmatrix} 0 & -I_k \\ I_k & 0 \end{pmatrix}.$$

We note that

$$(6.9) \qquad |(d_k(u, v), (y, \lambda))| \leq C \, |(u, v)|_0 \, |(y, \lambda)|_0,$$

for some constant C. Now, the multigrid algorithm corresponding to this nonsymmetric problem has exactly the same recursive form as (6.3) with $\mathcal{B}_k$, $\mathcal{A}_k$, etc., replacing $\hat{B}_k$, $\hat{A}_k$, etc., and thus,

$$(6.10) \qquad \mathcal{E}_k = \mathcal{I}_k - \mathcal{B}_k \mathcal{A}_k = [\mathcal{I}_k - \mathcal{I}_{k-1}^k \mathcal{P}_{k-1} + \mathcal{I}_{k-1}^k (\mathcal{I}_{k-1} - \mathcal{B}_{k-1} \mathcal{A}_{k-1}) \mathcal{P}_{k-1}] \mathcal{S}_k,$$

where $\mathcal{I}_k$ is the identity operator on $\mathcal{M}_k$. We need a subspace decomposition of $\mathcal{M}_k$. Let

$$(6.11) \qquad \mathcal{M}_k = \sum_{i=1}^{\ell} \mathcal{M}_k^i,$$

where ℓ is the number of grid points of the discrete domain and $\mathcal{M}_k^i$ is a two dimensional subspace of $\mathcal{M}_k$ consisting of nodal functions with zero nodal values except at the grid point i. Denote the decomposition of $\mathcal{A}_k$ (resp., $\hat{\mathcal{A}}_k$) with respect to the subspace $\mathcal{M}_k^i$ by $\mathcal{A}_k^i : \mathcal{M}_k^i \to \mathcal{M}_k^i$ (resp., $\hat{\mathcal{A}}_k^i$), satisfying

$$(\mathcal{A}_k^i \mathbf{w}, \boldsymbol{\chi})_k = (\mathcal{A}_k \mathbf{w}, \boldsymbol{\chi})_k \quad \text{for all } \boldsymbol{\chi} \in \mathcal{M}_k^i, \, \mathbf{w} \in \mathcal{M}_k^i.$$

Define $\mathcal{P}_k^i : \mathcal{M}_k \to \mathcal{M}_k^i$ (resp., $\hat{\mathcal{P}}_k^i$) by

$$(6.12) \qquad (\mathcal{A}_k \mathcal{P}_k^i \mathbf{w}, \boldsymbol{\chi})_k = (\mathcal{A}_k \mathbf{w}, \boldsymbol{\chi})_k \quad \text{for all } \boldsymbol{\chi} \in \mathcal{M}_k^i, \, \mathbf{w} \in \mathcal{M}_k.$$

We use the notation $(\mathbf{w}, \boldsymbol{\chi})_{0,i} = (\mathbf{w}, \boldsymbol{\chi})_0$ and $(\mathbf{w}, \boldsymbol{\chi})_{1,i} = (\mathbf{w}, \boldsymbol{\chi})_1$ for $\boldsymbol{\chi} \in \mathcal{M}_k^i$. In the case of a CGS smoother, we obtain a product representation (see [8]) of $\mathcal{S}_k$: For this purpose we set $\mathbf{w}^0 = 0$, for $i = 1, \dots, \ell$,

$$(6.13) \qquad \mathbf{w}^i = \mathbf{w}^{i-1} + (\mathcal{A}_k^i)^{-1} \mathcal{Q}_k^i \, (\mathbf{f}_k - \mathcal{A}_k \mathbf{w}^{i-1}),$$

and $\mathcal{R}_k \mathbf{f}_k = \mathbf{w}^\ell$. From the identity $\mathcal{A}_k^i \mathcal{P}_k^i = \mathcal{Q}_k^i \mathcal{A}_k$ on $\mathcal{M}_k^i$ it follows that $\mathcal{S}_k = \mathcal{I}_k - \mathcal{R}_k \mathcal{A}_k = \prod_{i=1}^{\ell} (\mathcal{I}_k - \mathcal{P}_k^i)$. Here, the operator $\mathcal{Q}_k^i : \mathcal{M}_k \to \mathcal{M}_k^i$ represents the orthogonal projection onto $\mathcal{M}_k^i$ with respect to $(\cdot, \cdot)_k$. Theorem 3.2 of [8] applies here to prove that (6.13) satisfies (6.4) and (6.5).

LEMMA 6.3. *For* $\mathbf{v}, \mathbf{w} \in \mathcal{M}_k$, *we have*

$$(6.14) \qquad |(\hat{\mathcal{A}}_k \hat{\mathcal{P}}_k^i \mathbf{w}, \mathbf{v})_k| \le C\, |\mathbf{w}|_1\, |\mathbf{v}|_1$$

and

$$(6.15) \qquad |(\hat{A}_k(\hat{\mathcal{P}}_k^i - \mathcal{P}_k^i)\mathbf{w}, \mathbf{v})| \le C\, h_k\, |\mathbf{w}|_1\, |\mathbf{v}|_1.$$

Proof. By coercivity, Lemma 3.2, and the Poincaré inequality, it follows that there exists a positive constant α such that

$$\alpha\, |\hat{\mathcal{P}}_k^i \mathbf{w}|_1^2 \le (\hat{\mathcal{A}}_k \hat{\mathcal{P}}_k^i \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{w})_k = (\hat{\mathcal{A}}_k \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{w})_k$$
$$\le C\, |\mathbf{w}|_1\, |\hat{\mathcal{P}}_k^i \mathbf{w}|_1.$$

Hence $\hat{\mathcal{P}}_k^i$ is bounded in the discrete energy norm and (6.14) is obtained by the Cauchy–Schwarz inequality. For (6.15), we have

$$(\hat{\mathcal{A}}_k(\hat{\mathcal{P}}_k^i - \mathcal{P}_k^i)\mathbf{w}, \mathbf{v})_k = (\hat{\mathcal{A}}_k(\hat{\mathcal{P}}_k^i - \mathcal{P}_k^i)\mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k$$
$$= (\hat{\mathcal{A}}_k \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k - (\hat{\mathcal{A}}_k \mathcal{P}_k^i \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k$$
$$= (\hat{\mathcal{A}}_k \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k - (\mathcal{A}_k \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k + (d_k\, \mathcal{P}_k^i \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k$$
$$= -(d_k\, \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k + (d_k\, \mathcal{P}_k^i \mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k$$
$$= -(d_k\, (\mathcal{I}_k - \mathcal{P}_k^i)\mathbf{w}, \hat{\mathcal{P}}_k^i \mathbf{v})_k.$$

Taking the absolute value we get by the Poincaré inequality,

$$|(\hat{\mathcal{A}}_k(\hat{\mathcal{P}}_k^i - \mathcal{P}_k^i)\mathbf{w}, \mathbf{v})_k| \le C\, |(\mathcal{I}_k - \mathcal{P}_k^i)\mathbf{w}|_{0,i}\, |\hat{\mathcal{P}}_k^i \mathbf{v}|_{1,i}$$
$$\le C\, h_k\, |\mathbf{w}|_1\, |\mathbf{v}|_1,$$

where the boundedness of $\hat{\mathcal{P}}_k^i$ is used for the second inequality. ☐

The proof of the following lemma is based on subspace decomposition and proved with the aid of Lemma 6.3 exactly in the same way as that of Theorem 3.1 in [11]. We skip the details.

LEMMA 6.4. *There exists some constant $C_\mathcal{S}$ independent of k such that*

$$(6.16) \qquad |(\hat{\mathcal{A}}_k(\mathcal{S}_k - \hat{\mathcal{S}}_k)\mathbf{w}, \mathbf{v})_k| \le C_\mathcal{S}\, h_k\, |\mathbf{w}|_1\, |\mathbf{v}|_1$$

for all $\mathbf{w}, \mathbf{v} \in \mathcal{M}_k$.

LEMMA 6.5. *The following inequalities hold:*

$$(6.17) \quad |(\hat{\mathcal{A}}_{k-1}(\hat{\mathcal{P}}_{k-1} - \mathcal{P}_{k-1})\mathbf{w}, \mathbf{v})_{k-1}| \le C_\mathcal{P}\, h_{k-1}\, |\mathbf{w}|_1 |\mathbf{v}|_1\ \textit{for}\ \mathbf{w} \in \mathcal{M}_k,\ \mathbf{v} \in \mathcal{M}_{k-1}$$

and

$$(6.18) \qquad |(\hat{\mathcal{A}}_k(\mathcal{I}_k - \mathcal{I}_{k-1}^k \mathcal{P}_{k-1})\mathbf{w}, \mathbf{v})_k| \le C_I\, h_k\, |\mathbf{w}|_1\, |\mathbf{v}|_1\ \textit{for}\ \mathbf{w} \in \mathcal{M}_k,\ \mathbf{v} \in \mathcal{M}_k,$$

where $C_\mathcal{P}$ and C_I are some constants independent of k.

Proof. Let us first prove (6.17) with $\hat{\mathcal{A}}_{k-1}$ replaced by $\mathcal{A}_{k-1}$. We have for $\mathbf{w} \in \mathcal{M}_k$ and $\mathbf{v} \in \mathcal{M}_{k-1}$,

$$
\begin{aligned}
&|(\mathcal{A}_{k-1}\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} - (\mathcal{A}_{k-1}\mathcal{P}_{k-1}\mathbf{w}, \mathbf{v})_{k-1}| \\
&= |(\hat{\mathcal{A}}_{k-1}\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} + (d_{k-1}\,\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} - (\mathcal{A}_k\mathbf{w}, \mathcal{I}_{k-1}^k\mathbf{v})_k| \\
&= |(\hat{\mathcal{A}}_k\mathbf{w}, \mathcal{I}_{k-1}^k\mathbf{v})_k + (d_{k-1}\,\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} - (\mathcal{A}_k\mathbf{w}, \mathcal{I}_{k-1}^k\mathbf{v})_k| \\
&= |(d_{k-1}\,\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} - (d_k\,\mathbf{w}, \mathcal{I}_{k-1}^k\mathbf{v})_k| \\
&= |(d_{k-1}\,\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} - (d_{k-1}\,\mathcal{I}_k^{k-1}\mathbf{w}, \mathbf{v})_{k-1}| \\
&= |(d_{k-1}\,(\hat{\mathcal{P}}_{k-1} - \mathcal{I}_k^{k-1})\mathbf{w}, \mathbf{v})_{k-1}| \\
&\leq C\,|(\hat{\mathcal{P}}_{k-1} - \mathcal{I}_k^{k-1})\mathbf{w}|_0\,|\mathbf{v}|_1,
\end{aligned}
$$

where the last inequality is obtained as follows: Let us denote by $\bar{\hat{\mathcal{P}}}_{k-1}$ the elliptic projection of the linear finite element method and denote by $\bar{\mathcal{I}}_{k-1} : \mathcal{M}_k \to \mathcal{M}_{k-1}$ the fine-to-coarse injection. We have

$$|(\hat{\mathcal{P}}_{k-1} - \mathcal{I}_k^{k-1})\mathbf{w}|_0 \leq |(\hat{\mathcal{P}}_{k-1} - \bar{\hat{\mathcal{P}}}_{k-1})\mathbf{w}|_0 + |(\bar{\hat{\mathcal{P}}}_{k-1} - \bar{\mathcal{I}}_{k-1})\mathbf{w}|_0 + |(\bar{\mathcal{I}}_{k-1} - \mathcal{I}_k^{k-1})\mathbf{w}|_0.$$

Here, inequality (6.7), (6.8), (6.9), and (6.11) of [9], and the approximation property of $\mathcal{I}_k^{k-1}$, are used to obtain the estimate $|(\hat{\mathcal{P}}_{k-1} - \mathcal{I}_k^{k-1})\mathbf{w}|_0 \leq C\,h_{k-1}\,|\mathbf{w}|_1$.

It follows that $\|\hat{\mathcal{P}}_{k-1} - \mathcal{P}_{k-1}\|_{\mathcal{A}_{k-1}} \leq C\,h_{k-1}$, where $\|\cdot\|_{\mathcal{A}_{k-1}}$ denotes the usual operator norm induced by $\mathcal{A}_{k-1}$. Because $(\mathcal{A}_{k-1}\mathbf{v}, \mathbf{v})_{k-1} = (\hat{\mathcal{A}}_{k-1}\mathbf{v}, \mathbf{v})_{k-1}$ from the definition of d_{k-1}, we also have $\|\hat{\mathcal{P}}_{k-1} - \mathcal{P}_{k-1}\|_{\hat{\mathcal{A}}_{k-1}} \leq C\,h_{k-1}$, which is the desired result.

The second assertion (6.18) follows directly from (6.17). $\square$

With these preparations we can show the following theorem.

THEOREM 6.6. *There exist positive constants h_0 and $\delta < 1$ such that for all $h_1 < h_0$ we have*

$$(\hat{\mathcal{A}}_L\mathcal{E}_L\mathbf{w}, \mathcal{E}_L\mathbf{w})_L \leq \delta^2(\hat{\mathcal{A}}_L\mathbf{w}, \mathbf{w})_L \quad \text{for all } \mathbf{w} \in \mathcal{M}_L,$$

where $\delta = \hat{\delta} + Ch_1$ and $\hat{\delta}$ is as in Theorem 6.2.

Proof. Denoting the operator norm $\|\cdot\|_{\hat{\mathcal{A}}_k}$ by $\|\cdot\|$, we show that $\|\mathcal{E}_k - \hat{\mathcal{E}}_k\| \leq c_k h_1$, where c_k is uniformly bounded. The error operator $\mathcal{E}_k$ can be written as

$$\mathcal{E}_k = (\mathcal{I}_k - \mathcal{I}_{k-1}^k\mathcal{B}_{k-1}\mathcal{A}_{k-1}\mathcal{P}_{k-1})\,\mathcal{S}_k,$$

and $\hat{\mathcal{E}}_k$ has similar representation. We compare the error operators and write their difference as

$$
\begin{aligned}
\mathcal{E}_k - \hat{\mathcal{E}}_k = {}&(\mathcal{I}_k - \mathcal{I}_{k-1}^k\mathcal{B}_{k-1}\mathcal{A}_{k-1}\mathcal{P}_{k-1})\,(\mathcal{S}_k - \hat{\mathcal{S}}_k) \\
&- \mathcal{I}_{k-1}^k\mathcal{B}_{k-1}\mathcal{A}_{k-1}(\mathcal{P}_{k-1} - \hat{\mathcal{P}}_{k-1})\,\hat{\mathcal{S}}_k + \mathcal{I}_{k-1}^k(\mathcal{E}_{k-1} - \hat{\mathcal{E}}_{k-1})\,\hat{\mathcal{P}}_{k-1}\,\hat{\mathcal{S}}_k.
\end{aligned}
$$

Thus in terms of the operator norm, we have by (4.16)

$$
\begin{aligned}
(6.19) \qquad \|\mathcal{E}_k - \hat{\mathcal{E}}_k\| \leq {}&\|\mathcal{I}_k - \mathcal{I}_{k-1}^k\mathcal{B}_{k-1}\mathcal{A}_{k-1}\mathcal{P}_{k-1}\|\,\|\mathcal{S}_k - \hat{\mathcal{S}}_k\| \\
&+ \|\mathcal{B}_{k-1}\mathcal{A}_{k-1}\|\,\|\mathcal{P}_{k-1} - \hat{\mathcal{P}}_{k-1}\|\,\|\hat{\mathcal{S}}_k\| \\
&+ \|\mathcal{E}_{k-1} - \hat{\mathcal{E}}_{k-1}\|\,\|\hat{\mathcal{P}}_{k-1}\hat{\mathcal{S}}_k\|.
\end{aligned}
$$

Let us make the induction hypothesis: $\|\mathcal{E}_{k-1} - \hat{\mathcal{E}}_{k-1}\| \leq c_{k-1}h_1$, where c_{k-1} is a constant to be defined below. By the triangle inequality and Theorem 6.2,

$$(6.20) \qquad\qquad \|\mathcal{E}_{k-1}\| \leq \hat{\delta} + c_{k-1}h_1$$

and

$$(6.21) \qquad\qquad \|\mathcal{B}_{k-1}\mathcal{A}_{k-1}\| \leq 1 + \hat{\delta} + c_{k-1}h_1.$$

Using the induction hypothesis, (4.14), Lemma 6.4, and Lemma 6.5, we have

$$
\begin{aligned}
&\|\mathcal{I}_k - \mathcal{I}_{k-1}^k \mathcal{B}_{k-1}\mathcal{A}_{k-1}\mathcal{P}_{k-1}\| \\
(6.22) \quad &\leq \|\mathcal{I}_k - \mathcal{I}_{k-1}^k \mathcal{P}_{k-1}\| + \|\mathcal{I}_{k-1} - \mathcal{B}_{k-1}\mathcal{A}_{k-1}\| \, \|\mathcal{P}_{k-1}\| \\
(6.23) \quad &\leq C_I h_{k-1} + \|\mathcal{E}_{k-1}\| \, (1 + C_{\mathcal{P}} h_{k-1}) \\
(6.24) \quad &\leq C_I \, (h_{k-1} + \hat{\delta} + c_{k-1}h_1),
\end{aligned}
$$

where we assumed C_I sufficiently large so that $1 + C_{\mathcal{P}} h_{k-1} \leq C_I$. To prove the second inequality (6.23) we used the fact that $\|\hat{\mathcal{P}}_{k-1}\| \leq 1$ and the chain of inequalities $\|\mathcal{P}_{k-1}\| \leq \|\hat{\mathcal{P}}_{k-1}\| + \|\mathcal{P}_{k-1} - \hat{\mathcal{P}}_{k-1}\| \leq 1 + C_{\mathcal{P}} h_{k-1}$. The stability of $\hat{\mathcal{P}}_{k-1}$ results from Lemma 4.1 and the identity $(\hat{\mathcal{A}}_{k-1}\hat{\mathcal{P}}_{k-1}\mathbf{w}, \mathbf{v})_{k-1} = (\hat{\mathcal{A}}_k\mathbf{w}, \mathcal{I}_{k-1}^k\mathbf{v})_k$.

Collecting (6.19) through (6.21), and using (4.14), Lemma 6.4, Lemma 6.5, and (6.22)–(6.24), we see that

$$
\begin{aligned}
\|\mathcal{E}_k - \hat{\mathcal{E}}_k\| &\leq C_I \, C_{\mathcal{S}} \, (h_{k-1} + \hat{\delta} + c_{k-1}h_1) \, h_k \\
&\quad + C_{\mathcal{P}} \, (1 + \hat{\delta} + c_{k-1}h_1) \, h_{k-1} + c_{k-1}h_1 \\
&\leq \left(\frac{C_I C_{\mathcal{S}}}{2} + C_{\mathcal{P}} \right) h_{k-1} \, (1 + \hat{\delta} + c_{k-1}h_1) + c_{k-1}h_1
\end{aligned}
$$

for all k.

Now let $\hat{C} := \frac{C_I C_{\mathcal{S}}}{2} + C_{\mathcal{P}}$ and define

$$(6.25) \qquad\qquad c_k := c_{k-1} + \hat{C}h_1^{-1}h_{k-1}(1 + \hat{\delta} + c_{k-1}h_1).$$

To see that the sequence c_k is uniformly bounded in k, one notes that $c_j \leq c_k$ for $j \leq k$ and hence

$$
\begin{aligned}
c_k &= c_{k-1} + \hat{C}h_1^{-1}(1 + \hat{\delta} + c_{k-1}h_1)h_{k-1} \\
&= c_1 + \hat{C}h_1^{-1}\sum_{j=2}^{k}(1 + \hat{\delta} + c_{j-1}h_1)h_{j-1} \\
&\leq c_1 + \hat{C}h_1^{-1}\sum_{j=2}^{k}(1 + \hat{\delta} + c_k h_1)h_{j-1} \\
&\leq c_1 + 2\hat{C}(1 + \hat{\delta}) + 2\hat{C}h_1 c_k.
\end{aligned}
$$

Now move the c_k term to the left to get

$$
c_k \leq (c_1 + 2\hat{C}(1 + \hat{\delta}))/(1 - 2\hat{C}h_1),
$$

provided that h_1 is small enough. Therefore, if the coarsest grid is sufficiently fine, we have $\delta = \hat{\delta} + Ch_1 < 1$.   □

We conclude this section with remarks on some of the constants appearing in the proofs. In case of collective CGS iteration, the constant $\bar{C}_{\hat{R}}$ appears to be almost independent from the value of the weight of the cost of the control. Its value is close to that of the Gauss–Seidel scheme applied to the scalar Poisson problem.

The constants in (6.16), (6.17), and (6.18) depend on the features of the optimality system as, for example, nonsymmetry. They account for the induction hypothesis where the coarsest mesh size, h_1, enters in the analysis and results in the estimate $\delta = \hat{\delta} + C h_1$. The requirement for a sufficiently small h_1 has no correspondence to our numerical experience (using CGS). However, the estimate of Theorem 6.6 states that, for sufficiently small h_1, we have $\delta \approx \hat{\delta}$, that is, the convergence factor of the multigrid method applied to the optimality system is close to the convergence factor of the multigrid scheme applied to the scalar Poisson problem. This fact agrees with our numerical experience and the results reported in Table 5.1.

7. Conclusions. We have presented a systematic study of a finite difference multigrid method for a class of optimality systems arising from optimal control of elliptic partial differential equations. In this emerging field of scientific computing there is an increasing interest in the development of accurate and efficient solution methods for optimal control problems. In the first part we have proved optimal-order error estimates in the discrete L^2 norm and in the discrete H^1 norm under minimum regularity requirements on the data. In the second part, two complementary analytical tools for multigrid convergence theory have been discussed. In the framework of local Fourier analysis it is possible to obtain sharp convergence estimates which are very important in the first phase of development of the multigrid components. The other analytical tool presented here is important from the theoretical point of view. It makes it possible to prove optimal convergence of the multigrid process under weak regularity assumptions. The general multigrid convergence theory discussed in this paper is developed in two steps. First, the multigrid method applied to the uncoupled differential system is considered. Then, the nondifferential coupling part characterizing the optimality system is introduced. By analyzing the difference between the operators obtained with and without coupling, we are able to estimate the convergence factor of multigrid for optimality systems based on the estimates available for the uncoupled problem.

REFERENCES

[1] E. ARIAN AND S. TA'ASAN, *Smoothers for optimization problems*, in Seventh Copper Mountain Conference on Multigrid Methods, Vol. CP3339, N. Duane Melson, T.A. Manteuffel, S.F. McCormick, and C.C. Douglas, eds., NASA Conference Publication, NASA, Hampton, VA, 1995, pp. 15–30.

[2] A. BORZÌ AND K. KUNISCH, *The numerical solution of the steady state solid fuel ignition model and its optimal control*, SIAM J. Sci. Comput., 22 (2000), pp. 263–284.

[3] A. BORZÌ AND K. KUNISCH, *A multigrid method for optimal control of time-dependent reaction diffusion processes*, in Fast Solution of Discretized Optimization Problems, Internat. Ser. Numer. Math. 138, K.H. Hoffmann, R. Hoppe, and V. Schulz, eds., Birkhäuser, Basel, 2001.

[4] A. BORZÌ, K. KUNISCH, AND M. VANMAELE, *A multi-grid approach to the optimal control of solid fuel ignition problems*, in Multigrid Methods, VI (Ghent, 1999), E. Dick, K. Riemslagh, and J. Vierendeels, eds., Lect. Notes Comput. Sci. Eng. 14, Springer–Verlag, Berlin, 2000, pp. 59–65.

[5] J.H. BRAMBLE, *Multigrid Methods*, Pitman Res. Notes Math. Ser. 294, John Wiley, New York, 1993.

[6] J.H. BRAMBLE, D. Y. KWAK, AND J.E. PASCIAK, *Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems*, SIAM J. Numer. Anal., 31 (1994), pp. 1746–1763.

[7] J.H. BRAMBLE AND J.E. PASCIAK, *New convergence estimates for multigrid algorithms*, Math. Comp., 49 (1987), pp. 311–329.

[8] J.H. BRAMBLE AND J.E. PASCIAK, *The analysis of smoothers for multigrid algorithms*, Math. Comp., 58 (1992), pp. 467–488.

[9] J.H. BRAMBLE, J.E. PASCIAK, AND J. XU, *The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms*, Math. Comp., 56 (1991), pp. 1–34.

[10] A. BRANDT, *Rigorous quantitative analysis of multigrid,* I: *Constant coefficients two-level cycle with L_2-norm*, SIAM J. Numer. Anal., 31 (1994), pp. 1695–1730.

[11] S.H. CHOU AND D.Y. KWAK, *V-cycle multigrid for a vertex-centered covolume method for elliptic problems*, Numer. Math., 90 (2002), pp. 441–458.

[12] W. HACKBUSCH, *Fast solution of elliptic control problems*, J. Optim. Theory Appl., 31 (1980), pp. 565–581.

[13] W. HACKBUSCH, *Multi-Grid Methods and Applications*, Springer–Verlag, New York, 1985.

[14] W. HACKBUSCH, *Elliptic Differential Equations*, Springer–Verlag, New York, 1992.

[15] D.Y. KWAK, *V-cycle multigrid for cell-centered finite differences*, SIAM J. Sci. Comput., 21 (1999), pp. 552–564.

[16] D.Y. KWAK, *A general multigrid framework for a class of perturbed problems*, in Fluid Flow and Transport in Porous Media: Mathematical and Numerical Treatment, AMS, Providence, RI, 2002, pp. 317–325.

[17] J.L. LIONS, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer–Verlag, Berlin, 1971.

[18] P. NEITTAANMÄKI AND D. TIBA, *Optimal Control of Nonlinear Parabolic Systems*, Marcel Dekker, New York, 1994.

[19] V. SCHULZ AND G. WITTUM, *Multigrid optimization methods for stationary parameter identification problems in groundwater flow*, in Multigrid Methods V, Lect. Notes Comput. Sci. Eng. 3, W. Hackbusch and G. Wittum, eds., Springer–Verlag, Berlin, pp. 276–288.

[20] E. SÜLI, *Convergence of finite volume schemes for Poisson's equation on nonuniform meshes*, SIAM J. Numer. Anal., 28 (1991), pp. 1419–1430.

[21] U. TROTTENBERG, C. OOSTERLEE, AND A. SCHÜLLER, *Multigrid*, Academic Press, London, 2001.

[22] S. VANKA, *Block-implicit multigrid calculation of two-dimensional recirculating flows*, Comp. Methods Appl. Mech. Engrg., 59 (1986), pp. 29–48.