MULTIGRID METHODS FOR OPTIMALITY SYSTEMS

ALFIO BORZÌ*

Abstract. Multigrid methods for the solution of optimal control optimality systems are discussed. These problems arise in the optimal control of systems governed by partial differential equations of elliptic, parabolic, and hyperbolic type.

Key words. Optimal control problems, elliptic pde, parabolic pde, hyperbolic pde, eigenvalue problems, finite differences, accuracy estimates, multigrid methods, convergence theory.

AMS subject classifications. 35J55, 35K57, 35L45, 35L50, 35L65, 49J20, 49K20, 65F15, 65M06, 65M12, 65M55, 65M99, 65N06, 65N12, 65N25, 65N30, 65N55, 65P20, 65P40, 68T99, 76B15, 76N15, 92C35.

1. Introduction. The ultimate aim of modelling and simulating application problems is to achieve better understanding of real world systems eventually with the purpose of being able to influence these systems in a desired way. This purpose has motivated the formulation of optimal control problems. These consist of a dynamical or equilibrium system, a description of the control mechanism, and a criterion defining the cost functional, that models the purpose of the control and describes the cost of its action. An optimal control problem is then formulated as the minimization of the cost functional where the state of the system is characterized by the modelling equations and the action of the control. This is a constrained minimization problem. The necessary conditions for such a minimum result in a set of coupled equations called the optimality system.

We are concerned with the numerical solution of optimality systems corresponding to optimal control problems governed by partial differential equations. Regarding optimal control theory we follow the approach given in [79, 80, 84] and the references given there. Other references relevant for our work are, e.g., [1, 4, 9, 10, 49, 81]. The increasing interest and applications of optimal control methods give account of the large number of recent mathematical contributions to this field. For a partial overview of methods and application areas see [42, 46, 66, 67].

The ever growing computational capabilities allow to realize optimal control strategies in current practical applications having increasing complexity. The aim is to solve large-scale optimization problems in an accurate and computationally efficient way. Most contributions to the development and analysis of discretization schemes and solvers for optimal control problems are rather recent; see, e.g., [65, 66, 67, 73] and references given there. In this paper we present our contribution to this field of scientific computing.

Our contribution to the numerical solution of optimal control problems is twofold. Primarily, we are concerned with the fast solution of discretized optimality systems by means of multigrid methods. The multigrid approach to optimal control problems is, apart from a few contributions, rather recent; see, e.g., [63, 64]. Secondarily, we consider accurate approximation of optimal control solutions by finite difference schemes. In contrast to finite element approximations to optimality systems that are rather well investigated - see, e.g., [84] and the references given there - much less rigorous analysis is available on the finite difference discretization of optimal control problems.

We pursue the one-shot multigrid strategy as proposed in [2, 3, 99]. Related approaches can be found in [47, 51, 95] within the successive quadratic programming method, and in [51, 55, 58, 59, 74] concerning the reformulation of the optimality conditions as fixed-point equations and their solution by multigrid methods.

A one-shot multigrid algorithm means solving the optimality system for the state, the adjoint, and the control variables in parallel in the multigrid process. This is in contrast to solving sequentially the state and the adjoint equations and then updating the control variables along with the gradients provided by the optimality condition. Notice that the one-shot approach of [2, 3] applied to elliptic problems uses the gradient scheme just described for designing the smoothing iteration. This approach may result in a lack of robustness

*Institut für Mathematik, Karl-Franzens-Universität Graz, Heinrichstr. 36, A-8010 Graz, Austria (alfio.borzi@uni-graz.at).

of the multigrid solution process with respect to the value of the weight of the cost of the control, here denoted by ν. Moreover, the 'gradient' approach requires that the uncontrolled state equation be solvable, and thus it cannot be applied to solve singular optimal control problems [80].

In [18, 22] we develop a smoothing scheme that is robust with respect to changes of ν and results in multigrid schemes that successfully apply to singular optimal control problems. Specifically, in these references the optimal control of the solid fuel ignition model was considered that is characterized by an exponential nonlinearity such that the uncontrolled model may not admit solutions. Accuracy of the finite difference approximation of this model is discussed in [18].

The results presented in [18, 22] motivate the multigrid convergence analysis presented in [21]. In this reference, sharp multigrid convergence factor estimates are obtained by means of local Fourier analysis and a multigrid convergence theory in the framework of [32] is provided which guarantees mesh-independent convergence of the multigrid process under weak regularity assumptions on the solutions of the optimality system. In particular, in [21] it is demonstrated that the convergence behavior of our multigrid scheme does not depend on the value of ν. Comparing the results in [18] and [21] it is shown that the presence of nonlinearities in the model does not remarkably influence the multigrid convergence factor.

Based on the finite difference theoretical framework of [61] we present in [21] optimal-order accuracy estimates under minimum regularity assumptions on the solution of the optimality system.

While [18, 21, 22] deal with geometric multigrid schemes, in [14, 15] we extend our approach to algebraic multigrid (AMG) methods. The advantage of the algebraic multigrid formulation is its general applicability to linear optimality systems defined on complex geometries. By taking advantage of the relationship between state and adjoint equations we construct an algebraic multigrid scheme which is able to solve optimality systems with a computational complexity close to that typical of AMG for scalar problems. In applications, optimality systems where the state equation is given by an anisotropic elliptic operator with strongly discontinuous coefficients [14] and by a convection diffusion operator with recirculating convection [15] are considered. Also within this setting numerical experiments demonstrate robustness with respect to the weight of the cost of the control. Within the pure algebraic formulation in [14] conditions are given that guarantee convergence of the AMG iteration.

In [20] a new smoothing iteration for control-constrained optimality systems is presented. It allows to construct robust multigrid schemes that apply also in case $\nu = 0$, thus allowing the investigation of bang-bang control problems. In particular, using the multigrid scheme in [20] it is possible to show the phenomenon of 'chattering control' [9] for elliptic systems which appears to be a problem which received little attention. Also in [20] extensions of our multigrid schemes to solve boundary optimal control problems and optimal control problems where the control acts only in a sub-domain of the computational domain are presented. Using results in [82] we obtain in [20] error estimates for the case of constrained optimal control problems.

The results described above concern multigrid methods for elliptic optimality systems. In the following, we discuss our contribution [11, 12, 19] to the development of multigrid schemes for parabolic optimality systems. These systems are characterized by a set of parabolic partial differential equations with opposite orientation. The starting point for our development is represented by space-time parabolic multigrid methods [57, 103] and the approach presented in [56]. The use of parabolic multigrid methods is suggested by the need of a setting that allows a robust implementation of the time coupling between state and adjoint variables. The coupling is then realized within our smoothing scheme proposed in [11, 19]. This iterative scheme has been successfully applied in combination with different coarsening strategies to solve singular parabolic optimal control problems. In fact, in contrast to [58, 59] our smoothing scheme implements the coupling between state and adjoint variables at each time step, thus avoiding to solve the state equation without effective control.

Results of numerical experiments in [11, 19] demonstrate the ability of our space-time multigrid scheme to solve parabolic optimality systems. We obtain optimal convergence

factors and robustness with respect to changes of the value of the weight of the cost of the control. These facts are confirmed by the results of the multigrid convergence analysis presented in [12].

Multigrid methods for constrained parabolic optimal control problems are discussed in [11, 12]. In these references the presence of constraints is reformulated as a nonlinear term relating the control to the adjoint variable. Within this setting, it is shown in [12] that the W-cycle multigrid scheme is superior to the V-cycle scheme. Optimal error estimates are presented in [12] in case of unconstrained parabolic optimal control problems.

In [16, 17] we present an optimal control formulation of the optical flow problem. Optical flow is the field of apparent velocities of objects in a sequence of images. We apply multigrid schemes to solve the resulting optimality conditions that consist of a system of two elliptic equations for the two optical flow components. These are the flow components for the hyperbolic optical flow equation and for the adjoint equation. The hyperbolic-elliptic optimality system is solved iteratively in a segregation loop. Well posedness of the proposed iteration is analyzed in [16]. Results presented in [16, 17] considering images of noisy and fast-moving objects demonstrate the advantage of our formulation.

In the following we report results of our work mentioned above and comment on our ongoing research projects on multigrid methods and optimal control problems.

The next three sections provide an introduction to the optimal control problems we consider, their approximation, and a description of geometric and algebraic multigrid methods. In Section 5 we report results on the accuracy of finite difference discretization and on the convergence of multigrid methods applied to elliptic optimality systems using local Fourier analysis and general multigrid theory. In Section 6 efficiency and robustness of multigrid methods solving singular optimal control problems is discussed. In Section 7 the multigrid solution of constrained optimal control problems also in the limit case of bang-bang solutions is discussed. In Section 8 we describe the development and analysis of algebraic multigrid methods for optimality systems. In Section 9 multigrid methods for parabolic optimal control problems are described. In Section 10 we discuss an optimal control formulation of the optical flow problem and its solution by multigrid and time-marching methods. In the final section we give account of our ongoing research related to the work presented in this paper and draw conclusions.

2. Optimal control problems. The formulation of optimal control of systems governed by partial differential equations requires the following terms: 1) The definition of a control function u that represents the driving influence of the environment on the system. 2) The partial differential equations modelling the controlled system, represented by the state function $y(u)$. 3) The cost functional which models the purpose of the control on the system. We now discuss these terms that enter in the definition of optimal control problems. We consider steady and evolutionary problems thus the optimal control problems are defined on space or space-time domains.

With u we denote the control function belonging to the closed and convex set of admissible controls $U_{ad} \subset U$, where U is a real Hilbert space with inner product and norm denoted by $(\cdot,\cdot)$ and $|\cdot|$, respectively.

The state of the system as a function of the control is denoted by $y(u) \in Y$, where Y is a real Hilbert space with inner product and norm denoted by $((\cdot,\cdot))$ and $||\cdot||$. It is given by the solution of a partial differential equation which is formally expressed as $e(y,u) = 0$ where $e : Y \times U \to Y^*$. It is required that the solution of this equation with given u defines a continuous affine mapping $u \to y(u)$. Let us denote its first derivative at u in the direction δu by $y'(u,\delta u)$. It is characterized as the solution to $e(y',\delta u) = 0$. The second derivative of $u \to y(u)$ is zero.

The cost functional is formally given by

$$J(\cdot,\cdot) : Y \times U \to \mathbf{R}.$$

It is assumed that $J(y,u)$ is twice Frechet-differentiable and that the second Frechet derivative J'' is locally Lipschitz-continuous. Using the mapping $u \to y(u)$ we can define $\hat{J}(u) = J(y(u),u)$.

The optimal control problem can be formulated as follows: Find $u^* \in U_{ad}$ such that

$$\hat{J}(u^*) = \inf_{u \in U_{ad}} \hat{J}(u).$$

We are interested in cost functionals of tracking type given by

$$J(y,u) = \frac{1}{2}||y-z||^2 + \frac{\nu}{2}|u|^2, \tag{2.1}$$

where $z \in Z$ is the given objective function, Z being a real Hilbert space. In the following we let $Y \subseteq Z$ and when no confusion may occur we denote with $((\cdot,\cdot))$ and $||\cdot||$ the inner product and norm of Z as well. Here $\nu > 0$ is the weight of the cost of the control. (The case $\nu = 0$ is discussed in Section 7.)

We find for the second derivative of $u \to \hat{J}(u)$

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

and thus $u \to \hat{J}(u)$ is uniformly convex. This implies existence of a unique solution $u^* \in U_{ad}$ to the optimal control problem which can be characterized by the following optimality condition

$$\hat{J}'(u^*, v-u^*) = ((y^*-z, y'(u^*, v-u^*))) + \nu(u^*, v-u^*) \geq 0, \text{ for all } v \in U_{ad},$$

where $y^* = y(u^*)$.

Introduce $p^* \in Y$ as the unique solution to $e_y^*(y^*,u^*)p^* = -(y^*-z)$ where $e_y^* : Y \to Y^*$. Then via $e(y',\delta u) = 0$, $\delta u = v - u^*$, we have

$$\hat{J}'(u^*, v-u^*) = -((p^*, v-u^*)) + \nu(u^*, v-u^*) \geq 0 \text{ for all } v \in U_{ad}. \tag{2.2}$$

In case $U_{ad} = U$ this condition becomes $\hat{J}'(u)(u^*; v-u^*) = 0$.

Summarizing, the solution of the optimal control problem is characterized by the following optimality system

$$\begin{aligned} e(y^*,u^*) &= 0, \\ e_y^*(y^*,u^*)p^* &= -(y^*-z), \\ -((p^*, v-u^*)) + \nu(u^*, v-u^*) &\geq 0, \end{aligned} \tag{2.3}$$

for all $v \in U_{ad}$.

Now we give three examples of some of the optimal control problems and of the corresponding optimality systems which are discussed in detail in the sections that follow.

Example I: An elliptic optimal control problem.

$$\begin{cases} \min_{u \in L^2(\Omega)} J(y,u), \\ \qquad -\Delta y = u + g \quad \text{in } \Omega, \\ \qquad\quad y = 0 \quad \text{on } \partial\Omega, \end{cases} \tag{2.4}$$

subject to $u \in U_{ad} = L^2(\Omega)$, where $J(y,u)$ is given by

$$J(y,u) = \frac{1}{2}||y-z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\Omega)}. \tag{2.5}$$

Here and in the following we assume that $\Omega \subset \mathbf{R}^2$ is convex or $\partial\Omega$ is $C^{1,1}$ smooth and g, $z \in L^2(\Omega)$.

The corresponding optimality system is given by

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

From the optimality system one concludes the following regularity property

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

Example II: A parabolic optimal control problem.

$$\text{(2.7)} \qquad \begin{cases} \min_{u \in L^2(Q)} J(y,u), & & & \\ \quad -\partial_t y + \Delta y &=& u & \text{in } Q = \Omega \times (0,T), \\ \quad y &=& y_0 & \text{in } \Omega \text{ at } t = 0, \\ \quad y &=& 0 & \text{on } \Sigma = \partial\Omega \times (0,T), \end{cases}$$

where we take $y_0(\mathbf{x}) \in H_0^1(\Omega)$. We consider a cost functional of tracking type given by

$$\text{(2.8)} \qquad J(y,u) = \frac{1}{2}||y-z||^2_{L^2(Q)} + \frac{\nu}{2}||u||^2_{L^2(Q)}.$$

where $z \in L^2(Q)$.

The optimality system related to this problem is given by

$$\text{(2.9)} \qquad \begin{array}{rcll} -\partial_t y + \Delta y &=& u & \text{in } Q, \\ y &=& 0 & \text{on } \Sigma, \\ \partial_t p + \Delta p &=& -(y-z) & \text{in } Q, \\ p &=& 0 & \text{on } \Sigma, \\ \nu u - p &=& 0 & \text{in } Q, \end{array}$$

with initial condition $y(\mathbf{x},0) = y_0(\mathbf{x})$ and terminal condition $p(\mathbf{x},T) = 0$.

The following regularity property results

LEMMA 2.2. *If* $z \in L^2(Q)$ *then* y^*, u^*, $p^* \in H^{2,1}(Q)$ *where* $H^{2,1}(Q) = L^2(0,T;H^2(\Omega) \cap H_0^1(\Omega)) \cap H^1(0,T;L^2(\Omega))$.

Example III: A hyperbolic optimal control problem.

$$\text{(2.10)} \qquad \begin{cases} \min_{\vec{w} \in V} J(y,\vec{w}), & & & \\ \quad y_t + \vec{w} \cdot \nabla y &=& 0 & \text{in } Q = \Omega \times (0,T), \\ \quad y &=& Y_0 & \text{in } \Omega \text{ at } t = 0, \end{cases}$$

where $\vec{w} = (u,v) \in V$, V is a class of admissible flow fields and J is the cost functional

$$\text{(2.11)} \qquad \begin{aligned} J(y,\vec{w}) = \frac{1}{2}\int_\Omega \sum_{k=1}^{N} |y(x_1,x_2,t_k) - Y_k|^2 d\Omega \\ + \frac{\alpha}{2}\int_Q |\frac{\partial \vec{w}}{\partial t}|^2 dq + \frac{\beta}{2}\int_Q (|\nabla u|^2 + |\nabla v|^2) dq + \frac{\gamma}{2}\int_Q |\nabla \cdot \vec{w}|^2 dq. \end{aligned}$$

Here, $\{Y_k\}_{k=0,N}$ is a sequence of given 'frames' defined at increasing times $t_k \in [0,T]$ with $t_N = T$. The coefficients α, β, and γ are predefined positive weights.

The corresponding optimality system results in

$$\text{(2.12)} \qquad \begin{aligned} y_t + \vec{w} \cdot \nabla y = 0, \text{ with } y(\cdot,0) = Y_0, \\ p_t + \nabla \cdot (\vec{w} p) = \sum_{k=1}^{N-1} \left[\delta(t-t_k)(y(\cdot,t_k) - Y_k)\right], \text{ with } p(\cdot,T) = -(y(\cdot,T) - Y_N), \\ \alpha \frac{\partial^2 u}{\partial t^2} + \beta \nabla \cdot [(|\nabla u|^2 + |\nabla v|^2)\nabla u] + \gamma \frac{\partial}{\partial x_1}(\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_1}, \\ \alpha \frac{\partial^2 v}{\partial t^2} + \beta \nabla \cdot [(|\nabla u|^2 + |\nabla v|^2)\nabla v] + \gamma \frac{\partial}{\partial x_2}(\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_2}, \end{aligned}$$

where, e.g., $\vec{w} \in V$ is required to satisfy prescribed boundary conditions on the spatial boundary and at the temporal boundaries of Q.

We have that

LEMMA 2.3. *Suppose that* $Y_k \in H^1_{per}(\Omega) \cap W^{1,q}(\Omega)$, *for some* $q \in [2,\infty]$ *and all* $k = 0,\ldots,N$ *then* y^*, $p^* \in C(0,T;L^q(\Omega)) \cap L^\infty(0,T;H^1_{per}(\Omega) \cap W^{1,q}(\Omega))$ *and* $\vec{w}^* \in L^\infty(0,T;Lip(Q)) \times L^\infty(0,T;Lip(Q))$.

3. Approximation of optimal control problems. For the approximation of optimal control problems we use a finite difference framework and follow the notation and terminology of [61].

Consider a sequence of grids $\{\Omega_h\}_{h>0}$ defined by

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

To avoid certain technicalities we assume that Ω is a square and that the values of h are chosen such that the boundaries of Ω coincide with grid lines. The negative Laplacian with homogeneous Dirichlet boundary conditions is approximated by the common five-point stencil 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-norm 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. For the definition of H_h^2 we refer to [61], as well. We have the inverse property $|v_h|_2 \le ch^{-1}|v_h|_1$ and the Poincaré-Friedrichs inequality for finite differences (see, e.g., [97]): There exists a constant c_* such that for any grid function $v_h \in V_h$

$$|v_h|_0^2 \le\ c_* \sum_{i=1}^{2} |\partial_i^- v_h|_0^2.$$

Denote with V_h the vector space of nodal functions v_h defined on Ω_h which have prespecified values on the boundary. The system of nodal functions (v_h, w_h) is denoted by $\mathcal{V}_h = V_h \times V_h$.

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 [61]. The following property can be proved

$$|\tilde{R}_h v - R_h v|_0 \le c\, h^2\, |v|_{H^2(\Omega)} \text{ for all } v \in H^2(\Omega). \tag{3.1}$$

Here and below, c denotes a positive constant which does not depend on the discretization parameters.

In case of time-dependent optimal control problems we also need to define the time step size $\delta t = T/N_t$ and the space-time mesh is given by

$$Q_{h,\delta t} = \{(\mathbf{x}, t_m) : \mathbf{x} \in \Omega_h,\, t_m = (m-1)\delta t,\, 1 \le m \le N_t + 1\}.$$

On this grid, y_h^m denotes a grid function at time level m. The action of the backward and forward time difference operators on this function are denoted by

$$\partial_t^+ y_h^m = \frac{y_h^m - y_h^{m-1}}{\delta t} \quad \text{and} \quad \partial_t^- y_h^m = \frac{y_h^{m+1} - y_h^m}{\delta t},$$

respectively.

For grid functions defined on $Q_{h,\delta t}$ we use the discrete $L^2(Q)$ scalar product with norm $||v_{h,\delta t}||_0 = (v_{h,\delta t}, v_{h,\delta t})_{L_{h,\delta t}^2(Q_{h,\delta t})}^{1/2}$.

For convenience, it is assumed that there exist positive constants $c_1 \leq c_2$ such that $c_1 h^2 \leq \delta t \leq c_2 h^2$. Hence h can be considered as the only discretization parameter. Therefore, in the following, the subscript δt is omitted.

On the cylinder Q_h define the family of piecewise constant functions on the intervals $[t_m, t_{m+1})$ as follows

$$V_h = \{v_h \,|\, v_h(t) = v_h(t_m) \text{ for } t \in [t_m, t_{m+1}),\, v_h(t_m) \in L_h^2(\Omega_h)\}.$$

The space-time extension of the operators $\tilde{R}_h$ and R_h are denoted by

$$\tilde{R}_{h,\,Q} : L^2(Q) \to V_h \text{ and } R_{h,\,Q} : H^{2,1}(Q) \to V_h.$$

Based on the present finite difference framework the discretization of the optimal control problems given in the examples above is as follows.

Example I. The discretization of the optimal control problem (2.4) is specified by

$$\left\{ \begin{array}{c} \min \frac{1}{2}|y_h - \tilde{R}_h z|_0^2 + \frac{\nu}{2}|u_h|_0^2, \\ -\Delta_h y_h = u_h + \tilde{R}_h g. \end{array} \right. \tag{3.2}$$

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 (dropping the $*$ superscript)

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

Example II. The discrete optimal control problem approximating (2.7) is given by

$$\left\{ \begin{array}{c} \min \frac{1}{2}||y_h - \tilde{R}_{h,\,Q} z||_0^2 + \frac{\nu}{2}||u_h||_0^2, \\ -\partial_t^+ y_h^m + \Delta_h y_h^m = u_h^m. \end{array} \right. \tag{3.4}$$

The optimality system related to (3.4) follows

$$\begin{aligned} -\partial_t^+ y_h^m + \Delta_h y_h^m &= u_h^m, \\ \partial_t^- p_h^m + \Delta_h p_h^m &= -(y_h^m - \tilde{R}_{h,Q} z^m), \\ \nu u_h^m - p_h^m &= 0. \end{aligned} \tag{3.5}$$

Example III. In this case appropriate finite difference methods for hyperbolic problems are in order. To solve the optimality system (2.12) explicit second-order TVD schemes for both the state equation and the adjoint equation are used. For example, in the one-dimensional case, the TVD scheme for the variable y is written as

$$\begin{aligned} \frac{y_{i,\kappa+1} - y_{i,\kappa}}{\tau} = &- \left\{ \frac{1}{h}[1 + \frac{1}{2}\chi(r_{i-1/2}^+) - \frac{1}{2}\frac{\chi(r_{i-3/2}^+)}{r_{i-3/2}^+}](u)_{i-1/2}^+ (y_{i,\kappa} - y_{i-1,\kappa}) \right. \\ &\left. + \frac{1}{h}[1 + \frac{1}{2}\chi(r_{i+1/2}^-) - \frac{1}{2}\frac{\chi(r_{i+3/2}^-)}{r_{i+3/2}^-}](u)_{i+1/2}^- (y_{i+1,\kappa} - y_{i,\kappa}), \right. \end{aligned} \tag{3.6}$$

where $(u)^+ = \max(0, u)$ and $(u)^- = \min(0, u)$. The limiter function χ is defined as function of the flux difference ratios $r_{i\pm 1/2}^{\mp}$ and $r_{i\pm 3/2}^{\mp}$ given by

$$r_{i-1/2}^+ = \frac{(u)_{i+1/2}^+ (y_{i+1} - y_i)}{(u)_{i-1/2}^+ (y_i - y_{i-1})}, \quad r_{i+1/2}^- = \frac{(u)_{i-1/2}^- (y_i - y_{i-1})}{(u)_{i+1/2}^- (y_{i+1} - y_i)},$$

$$r_{i-3/2}^+ = \frac{(u)_{i-1/2}^+ (y_i - y_{i-1})}{(u)_{i-3/2}^+ (y_{i-1} - y_{i-2})}, \quad r_{i+3/2}^- = \frac{(u)_{i+1/2}^- (y_{i+1} - y_i)}{(u)_{i+3/2}^- (y_{i+2} - y_{i+1})}.$$

We use the Superbee limiter given by $\chi(r) = \max[0, \min(2r, 1), min(r, 2)]$.

The elliptic equations in (2.12) are discretized as follows. Consider the elliptic equation for u

$$\alpha\frac{u^h_{i,j,\kappa+1} - 2u^h_{i,j,\kappa} + u^h_{i,j,\kappa-1}}{\tau^2} + \beta\{\nabla^h \cdot [(|\nabla^h u^h|^2 + |\nabla^h v^h|^2)\nabla^h u^h]\}_{i,j,\kappa}$$
$$+\gamma\frac{u^h_{i+1,j,\kappa} - 2u^h_{i,j,\kappa} + u^h_{i-1,j,\kappa}}{h^2} = [p\frac{\partial I}{\partial x}]_{i,j,\kappa} - \gamma\frac{v^h_{i+1,j+1,\kappa} - v^h_{i+1,j-1,\kappa} - v^h_{i-1,j+1,\kappa} + v^h_{i-1,j-1,\kappa}}{4h^2},$$

and similarly for the v component.

4. Multigrid methods. Solvers of discretized partial differential equations based on the multigrid strategy [28, 35, 60, 100] typically show optimal computational complexity, i.e. the number of computer operations required scales linearly with the number of unknowns. This is also true in case of multigrid methods applied to optimal control problems [2, 18, 19, 21, 22, 11, 12, 14, 15, 47, 55, 95]. Furthermore, it was demonstrated that multigrid methods are robust with respect to changes of the weight of the cost of the control.

Before discussing multigrid methods for optimality systems in detail in the forthcoming sections, we give here an introduction of the multigrid framework used and introduce the notation.

Let us index operators and variables defined on a grid with mesh size $h = h_k = 1/2^k$, $k = 1, \ldots, L$, with the index k, the level number. Here L denotes the finest level. We introduce vector notation: we let $\mathbf{w} = (u, v)$ and $|\mathbf{w}|_0 = |(u, v)|_0$, etc..

Consider the following discrete problem

$$\mathcal{A}_k\mathbf{w}_k = \mathbf{f}_k. \tag{4.1}$$

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

$$\mathbf{r}_k = \mathbf{f}_k - \mathcal{A}_k\mathbf{w}_k. \tag{4.2}$$

Initially we can assume that $\mathbf{w}_k$ is affected by errors having a large spectrum of frequencies. 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 $\mathbf{w}^{(m_1)}_k = \mathcal{S}^{m_1}_k(\mathbf{w}_k, \mathbf{f}_k)$, where $\mathcal{S}^{m_1}_k$ is a linear smoothing operator applied m_1 times. For example $\mathcal{S}_k$ can be the pointwise (collective) Gauss-Seidel iteration. One sweep of this iteration is written in the form: $\mathbf{w}^{(m_1)}_k = \mathbf{w}^{(m_1-1)}_k + \mathcal{R}_k\,(\mathbf{f}_k - \mathcal{A}_k\mathbf{w}^{(m_1-1)}_k)$ 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}. That is,

$$\mathcal{A}_{k-1}\mathbf{w}_{k-1} = \mathcal{I}^{k-1}_k\mathbf{r}_k, \tag{4.3}$$

where $\mathbf{w}_{k-1}$ represents on the coarse grid Ω_{k-1} the error $\mathbf{w}^e_k$ on the next finer grid. The operator $\mathcal{I}^{k-1}_k : \mathcal{V}_k \to \mathcal{V}_{k-1}$ restricts the residual computed at level k to the grid with level $k-1$.

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

$$\mathbf{w}^{new}_k = \mathbf{w}_k + \mathcal{I}^k_{k-1}\,\mathbf{w}_{k-1}, \tag{4.4}$$

where $\mathcal{I}^k_{k-1} : \mathcal{V}_{k-1} \to \mathcal{V}_k$ is an interpolation operator. Here $\mathbf{w}_k$ 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 are well damped, then

the solution at level $\mathbf{w}_{k-1}$ should provide enough resolution for the error of $\mathbf{w}_k$ through $\mathcal{I}_{k-1}^k \mathbf{w}_{k-1}$.

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 (e.g., zero) and applies the smoothing iteration m_1 times. The residual is then computed and transferred to the next coarser grid while $\mathbf{w}_k$ obtained by smoothing is left unchanged. On the coarse grid with index $k-1$ again the smoothing process is applied. This procedure is repeated until the coarsest grid is reached.

On the coarsest grid, one solves the problem using a direct method or by iteration and the result is used to improve $\mathbf{w}_k$ via (4.4). The coarse grid correction is then followed by m_2 post-smoothing steps at level k before the CGC procedure followed by post-smoothing is repeated for the next (if any) finer level. This entire process represents one multigrid cycle. It can be interpreted as a linear iteration of the type $\mathbf{w}_L^{\ell+1} = \mathbf{w}_L^{\ell} + \mathcal{M}_L(\mathbf{f}_L - \mathcal{A}_L \mathbf{w}_L^{\ell})$ where $\mathcal{M}_k$ is the multigrid iteration operator. The multigrid algorithm expressed in terms of $\mathcal{M}_k$ is given as follows.

Multigrid Algorithm

Set $\mathcal{M}_1 = \mathcal{A}_1^{-1}$. For $k \geq 2$ define $\mathcal{M}_k : \mathcal{V}_k \to \mathcal{V}_k$ in terms of $\mathcal{M}_{k-1}$ as follows. Let $\mathbf{g} \in \mathcal{V}_k$ and $\mathbf{q}^0 = 0$.

1. Set $\mathbf{w}^0 = 0$.
2. Define $\mathbf{w}^l$ for $l = 1, \ldots, m_1$ by
$$\mathbf{w}^l = \mathbf{w}^{l-1} + \mathcal{R}_k(\mathbf{g} - \mathcal{A}_k \mathbf{w}^{l-1}).$$
3. Set $\mathbf{w}^{m_1+1} = \mathbf{w}^{m_1} + \mathcal{I}_{k-1}^k \mathbf{q}^m$ where $\mathbf{q}^i$ for $i = 1, \ldots, m$ is defined by
$$\mathbf{q}^i = \mathbf{q}^{i-1} + \mathcal{M}_{k-1}[\mathcal{I}_k^{k-1}(\mathbf{g} - \mathcal{A}_k \mathbf{w}^{m_1}) - \mathcal{A}_{k-1}\mathbf{q}^{i-1}].$$
4. Set $\mathcal{M}_k \mathbf{g} = \mathbf{w}^{m_1+m_2+1}$ where $\mathbf{w}^{\ell}$ for $\ell = m_1+2, \cdots, m_1+m_2+1$ is given by Step 2.

Notice that we can perform m two-grid iterations at each working level. For $m = 1$ we have a $V(m_1, m_2)$-cycle and for $m = 2$ we have a $W(m_1, m_2)$-cycle; m is called the cycle index [100]. In the following, N is the number of V- or W-cycles that are applied to solve the problem at hand.

The algorithm above uses the equivalence of equation (4.1) with the residual equation. This holds true for $\mathcal{A}$ being a linear operator. In case $\mathcal{A}(\cdot)$ is a nonlinear operator, the multigrid algorithm described above must be modified. The most used nonlinear multigrid method is the *full approximation scheme* (FAS) [34]. It consists of the same steps as the multigrid algorithm previously described but instead of solving for the error function $\mathbf{w}_k^e$ it applies to the variable $\mathbf{w}_{k-1} = \hat{\mathcal{I}}_k^{k-1} \mathbf{w}_k + \mathbf{w}_k^e$ as the coarse-grid unknown function.

To describe this method consider the discrete problem

$$\mathcal{A}_k(\mathbf{w}_k) = \mathbf{f}_k. \tag{4.5}$$

As in the linear case we need to define a smoothing procedure also denoted by $\mathcal{S}_k$. This corresponds, for example, to a Newton-Gauss-Seidel scheme. The coarse grid correction is defined in the following way. First, a coarse grid problem is constructed on the next coarser grid with index $k-1$. We have

$$\mathcal{A}_{k-1}(\mathbf{w}_{k-1}) = \mathcal{I}_k^{k-1} \mathbf{f}_k + \tau_k^{k-1}, \tag{4.6}$$

where τ_k^{k-1} is the fine-to-coarse defect correction defined by

$$\tau_k^{k-1} = \mathcal{A}_{k-1}(\hat{\mathcal{I}}_k^{k-1} \mathbf{w}_k) - \mathcal{I}_k^{k-1} \mathcal{A}_k(\mathbf{w}_k). \tag{4.7}$$

Notice that $\hat{\mathcal{I}}_k^{k-1}$ is a restriction operator which is not necessarily equal to $\mathcal{I}_k^{k-1}$. We usually choose $\hat{\mathcal{I}}_k^{k-1}$ be straight injection. Once the coarse grid problem is solved, the coarse grid correction follows

$$\mathbf{w}_k^{new} = \mathbf{w}_k + \mathcal{I}_{k-1}^k (\mathbf{w}_{k-1} - \hat{\mathcal{I}}_k^{k-1} \mathbf{w}_k). \tag{4.8}$$

Assuming that the smoothing iteration has reduced the high frequency components of the error on the finer grid, then the difference $\mathbf{w}_{k-1} - \hat{\mathcal{I}}_k^{k-1}\mathbf{w}_k$ provides a good approximation for the smooth error of $\mathbf{w}_k$.

The FAS algorithm can be summarized as follows.

FAS Algorithm

Set $\mathcal{M}_1 \approx \mathcal{A}_1^{-1}$ (e.g., iterating with $\mathcal{S}_1$). For $k \geq 2$ define $\mathcal{M}_k : \mathcal{V}_k \to \mathcal{V}_k$ in terms of $\mathcal{M}_{k-1}$ as follows. Let $\mathbf{g} \in \mathcal{V}_k$ and $\mathbf{q}^0 = 0$.

1. Set $\mathbf{w}^0 = \tilde{\mathbf{w}}$ (starting approximation).
2. Define $\mathbf{w}^l$ for $l = 1, \ldots, m_1$ by
$$\mathbf{w}^l = \mathbf{w}^{l-1} + \mathcal{R}_k(\mathbf{g} - \mathcal{A}_k\,\mathbf{w}^{l-1}).$$
3. Set $\mathbf{w}^{m_1+1} = \mathbf{w}^{m_1} + \mathcal{I}_{k-1}^k(\mathbf{q}^m - \hat{\mathcal{I}}_k^{k-1}\mathbf{w}^{m_1})$ where $\mathbf{q}^i$ for $i = 1, \ldots, m$ is defined by
$$\mathbf{q}^i = \mathbf{q}^{i-1} + \mathcal{M}_{k-1}[\mathcal{I}_k^{k-1}(\mathbf{g} - \mathcal{A}_k\,\mathbf{w}^{m_1}) + \mathcal{A}_{k-1}(\hat{\mathcal{I}}_k^{k-1}\mathbf{w}^{m_1}) - \mathcal{A}_{k-1}\mathbf{q}^{i-1}].$$
4. Set $\mathcal{M}_k\mathbf{g} = \mathbf{w}^{m_1+m_2+1}$ where $\mathbf{w}^\ell$ for $\ell = m_1 + 2, \cdots, m_1 + m_2 + 1$ is given by Step 2.

One can choose a starting grid with a level number $K < L$ which is coarser than the finest grid where the solution is desired. In this case one applies N multigrid cycles on level K and then the solution is interpolated on the next finer grid. The interpolation provides a first approximation for the multigrid solution process on this finer level and so on until the finest grid is reached. The combination of the nested iteration technique with the multigrid scheme is called the full multigrid (FMG) scheme.

Because of the action of the smoothing iteration and since we deal with elliptic-type operators, linear or bilinear interpolation operators can be used to transfer error functions from coarser to finer grids and to transfer residuals from finer to coarser grids. Many choices of interpolation and restriction operators are possible. For example for two-dimensional problems, the bilinear interpolation operator $I_{k-1}^k : V_{k-1} \to V_k$ is given in stencil form by

$$I_{k-1}^k = \frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{4.9}$$

This choice is consistent with the assumption of *bilinear* finite elements on each square partition of the discretization.

To transfer residuals one can use the full-weighting restriction operator $I_k^{k-1} : V_k \to V_{k-1}$ given in stencil form by

$$I_k^{k-1} = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{4.10}$$

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

We complete this section describing the algebraic multigrid (AMG) approach to the solution of $A_k u_k = f_k$, representing a discretized partial differential equation. Algebraic multigrid solvers [26, 37, 38, 41, 54, 92, 96, 104] have been developed to resolve the difficulty of designing geometric multigrid schemes for problems which are highly non-regular (i.e. discontinuous coefficients, singularities, etc.) and in case complex computational domains are considered. AMG schemes resemble the geometric multigrid process utilizing only information contained in the algebraic system to be solved. For a review see [96].

We now outline the main concepts of the algebraic multigrid approach. The algebraic multigrid algorithm constructs a hierarchy of problems indexed by $k = 1, \ldots, L$ where L denotes the coarsest level. Notice that in this case the level index k is reversed: The finest problem corresponds to $k = 1$, coarser problems correspond to larger k. We denote by N_k the total number of (variables) points at level k.

In the development of AMG, smooth and rough components of the error are defined in an algebraic sense. Roughly speaking, $\mathbf{e}$ is smooth for the given iteration operator $\mathcal{S}$ if $|\mathcal{S}\mathbf{e}| \approx |\mathbf{e}|$, in the energy norm. The construction of the prolongation operator I_{k+1}^k is based on the definition of smooth and rough error components and it is such to have the smooth error vectors in its range.

Using the notion of algebraic smoothness it is assumed to approximate the smooth error component $\mathbf{e}$ at point i as a linear combination of its neighboring error components $\mathbf{e}_j$. That is, the following *direct interpolation* formula is used [96]

$$e_i = \sum_{j \in P_i} w_{ij}\, e_j, \quad i \in F, \tag{4.11}$$

where $P_i \subset \tilde{N}_i \cap C$, $\tilde{N}_i = \{j \neq i : a_{ij} \neq 0\}$, and $w_{ij} = -\alpha_i a_{ij}/a_{ii}$. The w_{ij}'s measure the coupling between variables and α_i is introduced to take into account that in general $P_i \neq \tilde{N}_i$. Here, we have partitioned the whole set of unknown variables (also called points) in two subsets, C and F. The subset C has been selected as the subset of linearly independent components for the prolongation operator. It represents a coarse level of variables. The remaining variables belonging to the complement F of C are assumed to be expressed by means of (4.11).

With the sets C and F given, one can define the AMG components. The restriction operator is given by the transpose of the prolongation operator, multiplied by a normalization coefficient as follows $I_k^{k+1} = c\left(I_{k+1}^k\right)^T$. In our implementation we choose $c = \frac{N_{k+1}}{N_k}$. With this choice the restriction operator maps a constant to an almost constant vector with the same mean value.

The coarse matrix of coefficients A_{k+1} is defined by the Galerkin formula

$$A_{k+1} = I_k^{k+1}\, A_k\, I_{k+1}^k.$$

With these components, the solving phase of AMG can be implemented according to the classical multigrid schemes like the V-cycle and the W-cycle.

Notice that the main step in the implementation of AMG schemes is the construction of the sets P_i. The set P_i should contain those indices j for which the absolute value of w_{ij} is larger, or at least some of them. To this end the notion of strong coupling between variables is introduced [96]. A point i is said to be *strongly negatively coupled* (or *strongly n-coupled*) to another point j if

$$-a_{ij} > \varepsilon_{str} \max_{a_{ik}<0} |a_{ik}| \quad with\ 0 < \varepsilon_{str} < 1. \tag{4.12}$$

The set of points to which i is strongly n-coupled is denoted by S_i. Clearly the variables in S_i should be in P_i to have an effective interpolation formula. Consequently a point which has a large amount of points strongly n-coupled to it should be put in C. This amount is the number of elements in the set $S_i^T = \{j : i \in S_j\}$. Once the point i has been put in C, the points in S_i^T which are not already in C are put in F.

For the purpose of further reducing memory complexity, *aggressive coarsening* is used. This corresponds to a generalization of the notion of connectivity between points: the point i has a (p, l) long-range strong n-connection with the point j if there are at least p paths of length l, each given by a sequence of points $i_0, i_1, \ldots, i_l$ with $i_0 = i$ and $i_l = j$ such that $i_{q+1} \in S_{i_q}$ for $q = 0, 1, \ldots, l-1$ [96]. Aggressive coarsening is made by substituting the set $S_i^{p,l} = \{j \in \Omega : i \text{ strongly n} - \text{connected to j w.r.t. (p,l)}\}$ to S_i in the algorithm previously described. Using $S_i^{2,2}$ or $S_i^{1,2}$ the computational burden for the C/F splitting is high. In [14, 15, 25] we propose a modification of the splitting algorithm [96] which keeps the computational effort to a minimum. Specifically, our aggressive coarsening strategy differs from that in [96] where standard coarsening followed by a further coarsening for C variables is used.

Corresponding to the long-range coarsening strategy, multi-pass interpolation is performed. With our splitting algorithm, a three-pass interpolation is enough, whereas using the approach in [96] requires a four-pass interpolation.

5. Analysis of the multigrid solution of an elliptic optimality system. In this section, the main results of the work presented in [21] are reported. In this paper, finite difference discretization of the optimal control problem (2.4) and its multigrid solution are investigated. Results on stability of the finite difference optimality system and optimal-order error estimates in appropriate norms are discussed next.

Consider the following discrete optimality system

$$\left\{ \begin{array}{rcl} -\nu\,\Delta_h y_h - p_h &=& \nu\,\tilde{R}_h g, \\ -\Delta_h p_h + y_h &=& \tilde{R}_h z. \end{array} \right. \tag{5.1}$$

Which is obtained from (3.3) by eliminating u_h using the control equation.

To investigate the convergence of the solution of (5.1) to the solution of (2.6) as $h \to 0^+$, we introduce the family of operators

$$\mathcal{A}_h = \left(\begin{array}{cc} -\nu\,\Delta_h & -I_h \\ I_h & -\Delta_h \end{array} \right), \tag{5.2}$$

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. The cases $\mathcal{A}_h : H_h^1 \times H_h^1 \to H_h^{-1} \times H_h^{-1}$ and $\mathcal{A}_h : H_h^2 \times H_h^2 \to L_h^2 \times L_h^2$ were considered.

For $\mathcal{A}_h : H_h^1 \times H_h^1 \to H_h^{-1} \times H_h^{-1}$ it is proved that the family $\{\mathcal{A}_h\}_{h>0}$ is H_h^1-regular, that is, A_h is invertible and there exists a constant C_1 independent of h such that

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

Analogously, it is proved that $\mathcal{A}_h : H_h^2 \times H_h^2 \to L_h^2 \times L_h^2$ is H_h^2-regular.

The next step in order to get accuracy estimates is to prove consistency of $\mathcal{A}_h$ with its infinite dimensional analog given by

$$\mathcal{A} = \left(\begin{array}{cc} -\nu\,\Delta & -I \\ I & -\Delta \end{array} \right), \tag{5.3}$$

where Δ is understood with homogeneous Dirichlet boundary conditions. This operator is well defined from $H_0^1(\Omega) \times H_0^1(\Omega)$ to $H^{-1}(\Omega) \times H^{-1}(\Omega)$ as well as from $(H^2(\Omega) \cap H_0^1(\Omega)) \times (H^2(\Omega) \cap H_0^1(\Omega))$ to $L^2(\Omega) \times L^2(\Omega)$. We have the following consistency result.

LEMMA 5.1. *There exists a constant C_K independent of h such that*

$$||\mathcal{A}_h R_h^2 - \tilde{R}_h^2 \mathcal{A}||_{\mathcal{L}((H^2 \cap H_0^1)^2, (H_h^{-1} \times H_h^{-1}))} \leq C_K\, h.$$

Using these preparatory results the following two theorems are stated.

THEOREM 5.2. *There exists a constant K_1, depending on Ω, g, z, and independent of h such that*

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

Assuming that the boundaries of Ω coincide with grid lines we have

THEOREM 5.3. *There exists a constant K_2, depending on Ω, g, z, and independent of h such that*

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

We now discuss the derivation of sharp convergence factor estimates of the two grid method solving the optimality system. These estimates are obtained by means of local Fourier analysis [36]. Then, a multigrid convergence theory in the framework of [32] is provided that guarantees convergence of the multigrid process towards weak solutions of the optimality system.

TABLE 5.1
Convergence factors.

	LFA	Experim.
(m_1, m_2)	$\eta(TG_k^{k-1})$	$V(m_1, m_2)$
(1,1)	0.25	0.30
(2,1)	0.12	0.12
(2,2)	0.08	0.08
(3,2)	0.06	0.06
(3,3)	0.05	0.05

Consider (5.1) expressed as

$$\mathcal{A}_k \mathbf{w}_k = \mathbf{f}_k \text{ on } \Omega_k, \tag{5.4}$$

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

As smoother we choose the collective Gauss-Seidel scheme (CGS). It can be written as $\mathcal{S}_h = I_h - \mathcal{R}_h \mathcal{A}_h$ where $\mathcal{R}_h$ is the iteration matrix. To analyze this iteration by local Fourier analysis, consider the Fourier space spanned by the functions

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

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

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

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

$$\mu = \mu(\mathcal{S}_h) = \sup\{|\rho(\widehat{\mathcal{S}_h}(\boldsymbol{\theta}))| : \boldsymbol{\theta} \text{ high frequency }\}, \tag{5.5}$$

where $\widehat{\mathcal{S}_h}(\boldsymbol{\theta})$ is the Fourier symbol [100] of the CGS scheme and ρ denotes the spectral radius.

By inspection in the range of high frequencies for $h \in [0.01, 0.25]$ and ν ranging in the interval $[10^{-6}, 1]$, we find the upper bound $\mu(\mathcal{S}_h) \leq 0.5$.

Local Fourier analysis is also applied to investigate 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

$$TG_k^{k-1} = \mathcal{S}_k^{m_2} \left[\mathcal{I}_k - \mathcal{I}_{k-1}^k \left(\mathcal{A}_{k-1}\right)^{-1} \mathcal{I}_k^{k-1} \mathcal{A}_k\right] \mathcal{S}_k^{m_1}. \tag{5.6}$$

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

The convergence factor is given by

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

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

In Table 5.1 the values of $\eta(TG_k^{k-1})$ obtained with the two grid analysis are reported. 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 $[10^{-6}, 1]$.

In the framework of local Fourier analysis it is possible to obtain sharp convergence estimates which are 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 possible to prove optimal convergence of the multigrid process under weak regularity assumptions. The general multigrid convergence theory presented in [21] and discussed here is developed in two steps. First, the multigrid method applied to the uncoupled differential system is considered. Then, the non differential coupling term 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.

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

$$\begin{aligned} -\Delta y &= f \text{ in } \Omega, \\ y &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{5.7}$$

The matrix form of this problem is

$$\hat{A}_k y_k = f_k. \tag{5.8}$$

Let $\hat{P}_{k-1} : V_k \to V_{k-1}$ (resp. $I_k^{k-1} : V_k \to V_{k-1}$) 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 $u \in V_k$ and $v \in V_{k-1}$. Let $\hat{R}_k : V_k \to V_k$ be the iteration operator of the smoothing process. Then the V-cycle multigrid algorithm to solve (5.8) in recursive form is given as follows:

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

1. Set $y^0 = 0$.
2. Define y^l for $l = 1, \ldots, m_1$ 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
$$q = \hat{M}_{k-1} I_k^{k-1}(g - \hat{A}_k\, y^{m_1}).$$
4. Set $\hat{M}_k g = y^{m_1+m_2+1}$ where y^ℓ for $\ell = m_1 + 2, \cdots, m_1 + m_2 + 1$ 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 V_k. Then $\hat{S}_k\, y = y - y^1$. Now for $y \in V_k$, $k = 2, \cdots, L$, we have

$$\begin{aligned} (I_k - \hat{M}_k\hat{A}_k)\, y &= y - y^1 - I_{k-1}^k q \\ &= \hat{S}_k\, y - I_{k-1}^k \hat{M}_{k-1}\hat{A}_{k-1}\hat{P}_{k-1}\hat{S}_k\, y \\ &= [I_k - I_{k-1}^k \hat{M}_{k-1}\hat{A}_{k-1}\hat{P}_{k-1}]\, \hat{S}_k\, y \\ &= [(I_k - I_{k-1}^k\hat{P}_{k-1}) + I_{k-1}^k(I_{k-1} - \hat{M}_{k-1}\hat{A}_{k-1})\, \hat{P}_{k-1}]\hat{S}_k\, y. \end{aligned} \tag{5.9}$$

The convergence results of the multigrid method are expressed in terms of the error operators $\hat{E}_k := I_k - \hat{M}_k\hat{A}_k$ and $\hat{E} := \hat{E}_L$.

We have the following result from [32].

THEOREM 5.4. *Let $\hat{R}_k$ be a suitable iteration operator. Then there exists a positive constant $\hat{\delta} < 1$ such that*

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

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

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

$$\begin{array}{rcll} -\nu\Delta y &=& \nu g & \text{in } \Omega, \\ y &=& 0 & \text{on } \partial\Omega, \\ -\Delta p &=& z & \text{in } \Omega, \\ p &=& 0 & \text{on } \partial\Omega. \end{array} \tag{5.10}$$

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

$$\hat{\mathcal{A}}_k = \begin{pmatrix} \nu\,\hat{A}_k & 0 \\ 0 & \hat{A}_k \end{pmatrix}, \tag{5.11}$$

and analogously $\hat{\mathcal{M}}_k$, $\hat{\mathcal{E}}_k$ etc. as the system counterparts of $\hat{M}_k$, $\hat{E}_k$ etc., then the multigrid algorithm has exactly the same form as (5.9) with $\hat{\mathcal{M}}_k$, $\hat{\mathcal{A}}_k$, etc. replacing $\hat{M}_k$, $\hat{A}_k$, etc.. As a consequence we have

THEOREM 5.5. *Under the assumption of Theorem 5.4, there exists a positive constant* $\hat{\delta} < 1$ *such that*

$$(\hat{\mathcal{A}}_L\hat{\mathcal{E}}_L(y,p), \hat{\mathcal{E}}_L(y,p))_L \le \hat{\delta}^2(\hat{\mathcal{A}}_L(y,p),(y,p))_L \quad \textit{for all } (y,p)\in\mathcal{V}_L \tag{5.12}$$

where $\hat{\delta}$ *has the same form as in Theorem 5.4.*

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

$$|(d_k(u,v),(y,p))| \le C_0\,|(u,v)|_0\,|(y,p)|_0, \tag{5.13}$$

for some constant C_0. Now, the multigrid algorithm corresponding to this nonsymmetric problem has exactly the same recursive form as (5.9) with $\mathcal{M}_k$, $\mathcal{A}_k$, etc. replacing $\hat{M}_k$, $\hat{A}_k$, etc. and thus,

$$\mathcal{E}_k = \mathcal{I}_k - \mathcal{M}_k\mathcal{A}_k = [\mathcal{I}_k - \mathcal{I}_{k-1}^k\mathcal{P}_{k-1} + \mathcal{I}_{k-1}^k(\mathcal{I}_{k-1} - \mathcal{M}_{k-1}\mathcal{A}_{k-1})\,\mathcal{P}_{k-1}]\,\mathcal{S}_k, \tag{5.14}$$

where $\mathcal{I}_k$ is the identity operator on $\mathcal{V}_k$.

Next, we recall some preparatory lemmata; see [21, 44].

LEMMA 5.6. *There exists some constant* $C_{\mathcal{S}}$ *independent of* k *such that*

$$|(\hat{\mathcal{A}}_k(\mathcal{S}_k - \hat{\mathcal{S}}_k)\mathbf{w},\mathbf{v})_k| \le C_{\mathcal{S}}\,h_k\,|\mathbf{w}|_1\,|\mathbf{v}|_1, \tag{5.15}$$

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

LEMMA 5.7. *The following inequalities hold*

$$|(\hat{\mathcal{A}}_{k-1}(\hat{\mathcal{P}}_{k-1} - \mathcal{P}_{k-1})\mathbf{w},\mathbf{v})_{k-1}| \le C_{\mathcal{P}}\,h_{k-1}\,|\mathbf{w}|_1|\mathbf{v}|_1, \ \textit{for } \mathbf{w}\in\mathcal{V}_k,\ \mathbf{v}\in\mathcal{V}_{k-1} \tag{5.16}$$

and

$$|(\hat{\mathcal{A}}_k(\mathcal{I}_k - \mathcal{I}_{k-1}^k\mathcal{P}_{k-1})\mathbf{w},\mathbf{v})_k| \le C_I\,h_k\,|\mathbf{w}|_1\,|\mathbf{v}|_1, \ \textit{for } \mathbf{w}\in\mathcal{V}_k,\ \mathbf{v}\in\mathcal{V}_k \tag{5.17}$$

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

In the present framework the main result is given by the following theorem.

THEOREM 5.8. *There exist positive constants h_0 and $\delta < 1$ such that for all $h_1 < h_0$ we have*

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

where $\delta = \hat{\delta} + Ch_1$ and $\hat{\delta}$ is as in Theorem 5.5.

The requirement for h_1 to be sufficiently small was not found to be relevant in our numerical experience (using collective Gauss-Seidel). However, the estimate of Theorem 5.8 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.

6. Optimal control of the steady state solid fuel ignition model. In this section we present results given in [18, 22] for the case of singular optimal control problems discretized by finite differences and solved by multigrid techniques. Two optimal control strategies for solid fuel ignition phenomena are considered and compared. The corresponding optimality systems are solved to second-order accuracy by a multigrid method whose convergence properties are independent of the values of the weights in the cost functionals and of the number of grid points.

The steady state solid fuel ignition model which can give rise to blow-up phenomena is given by

$$\begin{aligned} \Delta y + \delta \exp(y) &= u \text{ in } \Omega, \\ y &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{6.1}$$

Due to the non-monotonic sign in front of the exponential term, this nonlinear indefinite problem may admit multiple solutions or no solution at all depending on Ω and on the value of the positive coefficient δ. Actually, the absence of coerciveness represents a major difficulty in the analysis and in the solution of this problem.

A necessary condition for the existence of at least one solution is expressed by the following theorem [53].

THEOREM 6.1. *Let λ_0 be the smallest eigenvalue of $-\Delta$ under homogeneous Dirichlet boundary conditions. If $\lambda_0 < \delta e$, then the set of solutions is empty.*

Thus, in order to have at least one solution we must have $\lambda_0 \geq \delta e$. If Ω is a unit square, then there exists a critical value $\delta^* > 0$ such that

- If $\delta > \delta^*$ there is no solution.
- If $\delta \in]0, \delta^*]$ ($\delta \in]0, \delta^*[$) then there exists at least one and at most two solution which belong to $W^{2,q}(\Omega)$, $q \geq 1$.
- If $\delta = \delta^*$ there exists one solution $y^* \in H_0^1(\Omega) \cap W^{2,q}(\Omega)$, $q \geq 1$, and δ^* is referred to as the turning point.

For the case of $\Omega = [0,1] \times [0,1]$, the turning point has been estimated, based on continuation techniques and obtained $\delta^* \approx 6.80$; see, e.g., [50].

Sufficient conditions for existence of solutions to (6.1) where u is a continuous function on $\overline{\Omega}$ can be obtained by means of the upper and lower solutions method [53, 87, 94]. We have the following result [18].

THEOREM 6.2. *Let $\tilde{y}$ and $\hat{y}$ be ordered upper and lower solutions of (6.1). Then $\{\overline{y}^{(k)}\}$ converges monotonically from above to a solution $\bar{y}$, and $\{\underline{y}^{(k)}\}$ converges monotonically from below to a solution $\underline{y}$, and both solutions of (6.1) belong to $C^{2+\alpha}(\overline{\Omega})$, $\alpha > 0$. Moreover, $\underline{y} \leq \bar{y}$.*

Control on (6.1) can be exerted by adding or subtracting thermal energy to the fuel through a source term. The optimal control problem is formulated as follows

$$\begin{cases} \min_{u \in L^2(\Omega)} J(y(u), u), \\ \Delta y + \delta \exp(y) = u \quad \text{in } \Omega, \\ y = 0 \quad \text{on } \partial\Omega, \end{cases} \tag{6.2}$$

Due to the lack of coercivity of the state equation it is at first not obvious how to properly choose the cost functional $J(y,u)$. In previous work two formalisms were considered:

$$J_1(y,u) = \frac{1}{2}||y-z||^2_{L^2(\Omega)} + \frac{\beta}{2}||e^y - e^z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\Omega)} \tag{6.3}$$

(see [72]), or as,

$$J_2(y,u) = \frac{1}{2}||y-z||^2_{L^2(\Omega)} + \frac{1}{2}||\nabla y - \nabla z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\Omega)} \tag{6.4}$$

(see [70]). In (6.3) and (6.4) the next to last terms guarantee the existence of solutions to (6.2). In these cost functionals ν is the weight of the cost of the control, β is a positive scaling factor, and $z \in H^2(\Omega)$ is the desired state. In both cases radial unboundedness of J in (6.2) is guaranteed and existence of a solution $(u^*, y^*) = (u^*, y(u^*))$ to the optimal control problem (6.2) with $J = J_1$ or $J = J_2$ can be established.

For $J = J_1$ the solution of the optimal control problem is characterized by the following optimality system

$$(OPC1) \qquad \begin{aligned} \Delta y + \delta \exp(y) - u &= 0, \\ \Delta p + \delta \exp(y)p + (y-z) + \beta e^y(e^y - e^z) &= 0, \\ \nu u - p &= 0. \end{aligned}$$

Similarly, the optimality system corresponding to $J = J_2$ is found to be

$$(OPC2) \qquad \begin{aligned} \Delta y + \delta \exp(y) - u &= 0, \\ \Delta p + \delta \exp(y)p + (y-z) - \Delta(y-z) &= 0, \\ \nu u - p &= 0. \end{aligned}$$

We now discuss the discretization of (6.1) on Ω_h. We have

$$\begin{aligned} \Delta_h y_h + \delta \exp(y_h) &= u_h \text{ in } \Omega_h, \\ y_h &= 0 \text{ on } \partial\Omega_h. \end{aligned} \tag{6.5}$$

Necessary and sufficient conditions for existence of solutions to the steady state solid fuel ignition model can be extended to the present discrete case. A sufficient condition for existence of solutions to (6.5) can be obtained by adaptation of the upper and lower solutions method; see [88].

Second-order accuracy of the solution to (6.5) is stated in the following lemma [18]. It uses results given in [27] for a linear indefinite problem.

LEMMA 6.3. *Let $y \in C^4(\overline{\Omega})$ be a bounded solution to the steady state solid fuel ignition model (6.1) and let y_h satisfy (6.5), with $\lim_{h\to 0}||y - y_h|| = 0$, where $||\cdot||$ denotes the maximum norm on Ω_h. Then for sufficiently small mesh sizes h we have*

$$||y - y_h|| = O(h^2).$$

The discrete system corresponding to (OPC1) is denoted by

$$\begin{aligned} \Delta^h y_h + \delta \exp(y_h) - u_h &= 0, \\ \Delta^h p_h + \delta \exp(y_h)p_h + (y_h - z_h) + \beta e^{y_h}(e^{y_h} - e^{z_h}) &= 0, \\ \nu u_h - p_h &= 0. \end{aligned} \tag{6.6}$$

For (OPC2) we have

$$\begin{aligned} \Delta^h y_h + \delta \exp(y_h) - u_h &= 0, \\ \Delta^h p_h + \delta \exp(y_h)p_h + (y_h - z_h) - \Delta(y_h - z_h) &= 0, \\ \nu u_h - p_h &= 0. \end{aligned} \tag{6.7}$$

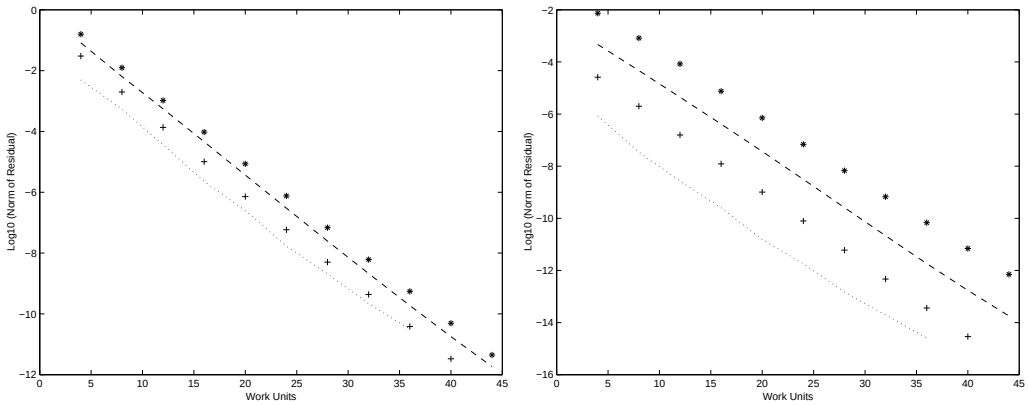

FIG. 6.1. *Convergence history of the GSN-FAS method for solving the* **state equation** *(left) and the* **co-state equation** *(right) of (OPC1) for different* ν. *We have* $\nu = 10^{-3}$ *with '*';* $\nu = 10^{-5}$ *with '–',* $\nu = 10^{-7}$ *with '+',* $\nu = 10^{-9}$ *with '. . .'*

Using the result of Lemma 6.3 and results of [27] it is possible to prove second-order accuracy of the solution of the optimality system.

To solve (6.6) and (6.7) using the FAS Algorithm two smoothing schemes were considered [18]. Following [2, 3] the first smoothing iteration is given by the following steps: One step of an iterative solver applied to the state equation, then one step of an iterative solver applied to the co-state equation and finally update of the control function by means of $u = p/\nu$. As iterative solver for each equation we take one Gauss-Seidel-Picard (GSP) iteration.

Our second choice is the Gauss-Seidel-Newton (GSN) iteration. It is a local Newton step applied at each grid point to the set of variables $\phi = (y, p, u)$. It is defined by

$$\phi_{ij}^{(new)} = \phi_{ij}^{(old)} - [\mathbf{G}'(\phi_{ij})]^{-1}\mathbf{G}(\phi_{ij}). \tag{6.8}$$

The vector equation $\mathbf{G}(y_h, p_h, u_h) = \mathbf{0}$ represents (6.6) or (6.7) at any grid point and $\mathbf{G}'(y_h, p_h, u_h)$ denotes the Jacobian of $\mathbf{G}$.

Notice that by this approach no global linearization of the optimal control problem is involved. The FAS method is applied directly to the optimal control system.

We give examples demonstrating the efficiency of the multigrid method to solve singular optimal control problems. We report the values of the tracking error $||y_h - z_h||$ (maximum norm) as well as of the control costs $|u_h|_0 = |p_h|_0/\nu$. The numerical accuracy of the solution can be expressed by $||E_H|| = \max_{\Omega_H} |y_H - \hat{I}_h^H y_h|$ and the norm of the error of the adjoint variable p is defined by $||V_H|| = \max_{\Omega_H} |p_H - \hat{I}_h^H p_h|$. Recall that the critical value for δ is $\delta^* = 6.80$. Consider the desired state given by $z(x_1, x_2) = \frac{1}{\pi^2}\sin(\pi x_1)\sin(\pi x_2)$ and set $\delta = \beta = 6.0$, and $\nu = 5 \cdot 10^{-3}$.

With this setting, the FAS algorithm with GSP-smoothing show typical multigrid convergence behavior. However, for smaller values of ν the performance of the GSP-FAS method worsen. The GSP method is not robust with respect to this parameter.

A significant improvement in robustness of the FAS algorithm is obtained using the GSN iteration for smoothing. The resulting full multigrid version of the FAS scheme computes the solution of (OPC1) and of (OPC2) in a few work units and is robust with respect to value of ν. The convergence history of the GSN-FAS method solving (OPC1) is illustrated in in Figure 6.1. Observe that the convergence factor is almost independent of ν. Similar results are obtained in solving the (OPC2) problem; see Figure 6.2.

Results obtained in [18, 22] confirm the expected optimal control behavior: As ν decreases, $|u_h|_0$ increases and $||y_h - z_h||$ decreases, both for (OPC1) and (OPC2); See Tables 6.1 and 6.2. In Table 6.2 we report results obtained using the FAS based FMG scheme. Notice that the value of $||y_h - z_h||$ is almost independent of the discretization level, showing that the numerical solution quickly attains its optimal value. Comparing the tracking error obtained with the two formalisms we note that, with the present choice of z, the J_2-based formulation allows better tracking than the J_1-based formulation.

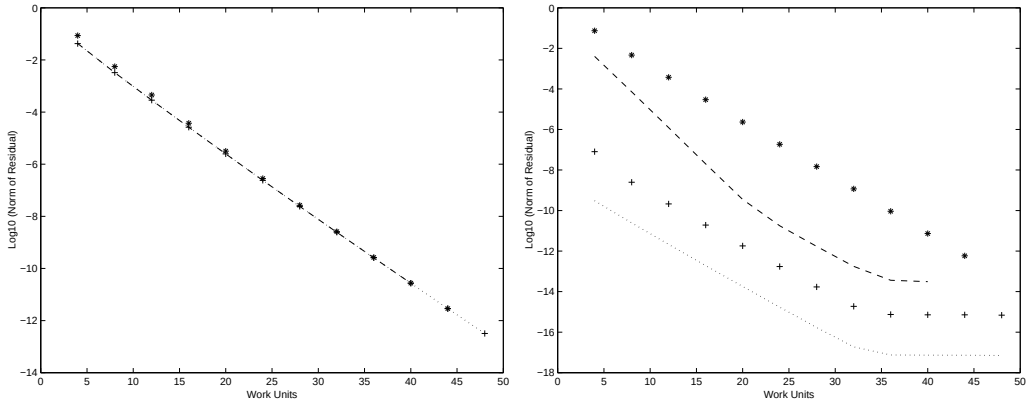

FIG. 6.2. *Convergence history of the GSN-FAS method for solving the* **state equation** *(left) and the* **co-state equation** *(right) of (OPC2) for different ν. We have $\nu = 10^{-3}$ with '*'; $\nu = 10^{-5}$ with '–', $\nu = 10^{-7}$ with '+', $\nu = 10^{-9}$ with '. . .'*

TABLE 6.1
Tracking ability of the GSN-FAS method; $z(x_1, x_2) = \sin(\pi x_1)\sin(\pi x_2)/\pi^2$.

	OPC1	OPC2
ν	$\|\|y_h - z_h\|\|$	$\|\|y_h - z_h\|\|$
10^{-1}	3.4 (-1)	2.5 (-1)
10^{-3}	3.0 (-2)	5.6 (-3)
10^{-5}	3.4 (-3)	6.7 (-5)

TABLE 6.2
Tracking and accuracy results with (OPC1) and (OPC2) obtained by the GSN-FAS based FMG scheme. Use $\beta = \delta = 6.8$, and $\nu = 10^{-5}$.

	OPC1				OPC2			
	3-FMG				3-FMG			
k	$\|\|y_h - z_h\|\|$	$\|\|E_H\|\|$	$\|\|V_H\|\|$	$\|u_h\|_0$	$\|\|y_h - z_h\|\|$	$\|\|E_H\|\|$	$\|\|V_H\|\|$	$\|u_h\|_0$
3	0.26(-2)	0.41(-3)	0.35(-5)	5.72	0.64(-4)	0.27(-4)	0.33(-6)	5.82
4	0.30(-2)	0.15(-3)	0.12(-5)	5.81	0.65(-4)	0.11(-5)	0.11(-5)	6.05
5	0.33(-2)	0.45(-4)	0.30(-6)	5.83	0.66(-4)	0.29(-5)	0.28(-5)	6.15
6	0.34(-2)	0.11(-4)	0.77(-7)	5.84	0.67(-4)	0.35(-5)	0.35(-5)	6.20
7	0.34(-2)	–	–	5.84	0.67(-4)	–	–	6.21
	5-FMG				5-FMG			
k	$\|\|y_h - z_h\|\|$	$\|\|E_H\|\|$	$\|\|V_H\|\|$	$\|\|u_h\|\|_0$	$\|\|y_h - z_h\|\|$	$\|\|E_H\|\|$	$\|\|V_H\|\|$	$\|\|u_h\|\|_0$
3	0.26(-2)	0.42(-3)	0.35(-5)	5.72	0.64(-4)	0.33(-6)	0.33(-6)	5.82
4	0.30(-2)	0.15(-3)	0.12(-5)	5.81	0.65(-4)	0.11(-5)	0.11(-5)	6.05
5	0.33(-2)	0.45(-4)	0.30(-6)	5.83	0.66(-4)	0.29(-5)	0.28(-5)	6.15
6	0.33(-2)	0.11(-4)	0.77(-7)	5.84	0.67(-4)	0.35(-5)	0.35(-5)	6.20
7	0.34(-2)	–	–	5.84	0.67(-4)	–	–	6.21

In Table 6.3 we depict the convergence factors for the state and the co-state equations of (OPC1) and (OPC2), with various choices for δ, $\beta = \delta$, and $\nu = 10^{-5}$. We observe that for a wide range of δ-values the convergence factors are almost independent of the number of variables. Notice that the controlled system is considered for values of δ significantly larger than the critical value δ^* beyond which the uncontrolled system does not admit solutions. A slight deterioration is revealed as the coefficient of the nonlinearity becomes very large.

When solving (OPC1) a value of the scaling factor β must be chosen. Only a weak influence of this factor on the convergence property of the multigrid iteration is observed. Tracking improves as β is increased. Finally, in [18] examples are given where the desired function z is not attainable by any control. Convergence behavior similar to that observed above is observed also in these cases. Results depicted in Figure 6.3 show that the (OPC1) approach is more suitable for tracking objective functions that are not compatible with the boundary conditions. The (OPC2) approach is more suitable when smooth objective functions are considered that satisfy the boundary conditions.

TABLE 6.3
Convergence factors of the GSN-multigrid method on different grids and various δ.

OPC1				
	L=4		L=7	
δ	ρ_y	ρ_p	ρ_y	ρ_p
1	0.09	0.10	0.09	0.09
10	0.07	0.09	0.08	0.09
50	0.15	0.16	0.12	0.12
100	0.19	0.19	0.13	0.12
OPC2				
	L=4		L=7	
δ	ρ_y	ρ_p	ρ_y	ρ_p
1	0.09	0.09	0.10	0.11
10	0.09	0.09	0.10	0.05
50	0.09	0.09	0.10	0.05
100	0.07	0.07	0.09	0.05

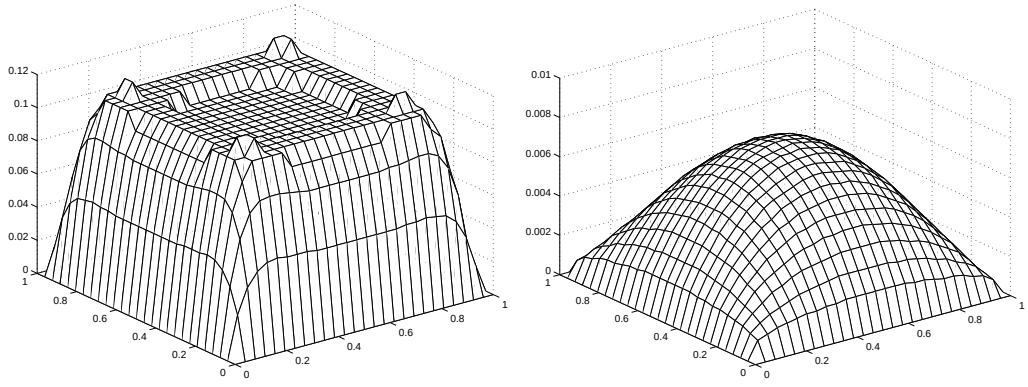

FIG. 6.3. *State solutions of (OPC1) and (OPC2) with $z = 1/\pi^2$ (not attainable).*

7. Multigrid schemes for elliptic constrained optimal control problems. By reformulating the presence of constraints as a nonlinear equation relating the control to the adjoint variable, it is possible to solve constrained optimal control problems using nonlinear multigrid methods. This approach was considered in [11]. The resulting multigrid scheme shows typical multigrid efficiency for sufficiently large values of the weight of the cost of the control. For small values of the weight, convergence of the multigrid iteration may deteriorate, showing a lack of robustness of this approach.

In [20] we propose a different technique where the constraints are enforced at each grid point in the smoothing procedure. This procedure appears to be robust with respect to changes of the value of the weight and, in particular, it allows the choice $\nu = 0$. This fact makes our multigrid algorithm a useful tool to investigate bang-bang type control phenomena for elliptic problems.

To describe our multigrid approach consider the following optimal control problem

$$\left\{ \begin{array}{rcll} \min_{u \in U_{ad}} J(y,u), & & & \\ -\Delta y & = & Bu + g & \text{in } \Omega, \\ y & = & 0 & \text{on } \partial\Omega, \end{array} \right. \tag{7.1}$$

where Ω is a open bounded set in $\mathbf{R}^2$, with boundary $\partial\Omega$ and $u \in U_{ad} \subset L^2(\Omega)$. We assume that Ω is convex or that $\partial\Omega$ is $C^{1,1}$ smooth. The cost functional J is of the tracking type and is given by

$$J(y,u) = \frac{1}{2}||y - z||^2_{L^2(\Omega)} + \frac{\nu}{2}||Bu||^2_{L^2(\Omega)}, \tag{7.2}$$

where $\nu \geq 0$. The set of admissible controls is the closed convex subset of $L^2(\Omega)$ given by

$$U_{ad} = \{u \in L^2(\omega) \,|\, \underline{u}(\mathbf{x}) \leq u(\mathbf{x}) \leq \overline{u}(\mathbf{x}) \text{ a.e in } \omega \subset \Omega\}, \tag{7.3}$$

where $\underline{u}$ and $\overline{u}$ are elements of $L^\infty(\Omega)$ and ω is a subset of Ω. The extension operator $B : L^2(\omega) \to L^2(\Omega)$ is defined as follows

$$Bu = \begin{cases} u & \text{in } \omega, \\ 0 & \text{in } \Omega \setminus \omega. \end{cases}$$

Existence of a unique solution to (7.1) and its characterization are well known [20, 79, 82]. We have that the solution to (7.1) is characterized by the following optimality system

$$\begin{aligned} -\Delta y &= Bu + g && \text{in } \Omega, \\ y &= 0 && \text{on } \partial\Omega, \\ -\Delta p &= -(y - z) && \text{in } \Omega, \\ p &= 0 && \text{on } \partial\Omega, \\ (\nu u - B^* p,\, v - u) &\geq 0 && \text{for all } v \in U_{ad}, \end{aligned} \tag{7.4}$$

where B^* denotes the adjoint of B.

Notice that the last equation in (7.4) giving the optimality condition is equivalent to (see [79, 82])

$$u = \max\{\underline{u}, \min\{\overline{u}, \frac{1}{\nu} p(u)\}\} \text{ in } \omega, \quad \text{if } \nu > 0. \tag{7.5}$$

We obtain the following regularity result.

LEMMA 7.1. If $\nu > 0$ and $\underline{u}$, $\overline{u} \in H^1(\Omega)$ then $u \in H^1(\Omega)$.

This regularity result is used [20] in the analysis of the finite difference approximation of the solution to the optimality system.

The unique solution u to (7.1) with $\nu = 0$ satisfies

$$\begin{aligned} -\Delta y &= Bu + g && \text{in } \Omega, \\ y &= 0 && \text{on } \partial\Omega, \\ -\Delta p &= -(y - z) && \text{in } \Omega, \\ p &= 0 && \text{on } \partial\Omega, \\ p &= \min\{0, p + u - \underline{u}\} + \max\{0, p + u - \overline{u}\} && \text{in } \Omega. \end{aligned} \tag{7.6}$$

We are interested in solutions for which the inactive set $\mathcal{I} = \{\mathbf{x} : \underline{u} < u < \overline{u}\}$ is small. For this case, sufficient conditions for the construction of test examples are given in [20].

We propose to call controls for which $\mathcal{I}$ contains no interior points almost bang-bang. Sufficient conditions which guarantee $meas(\mathcal{I}) = 0$, i.e. that the control is bang-bang, appear to be an interesting open problem.

Now consider the discrete optimal control problem

$$\begin{cases} \min \frac{1}{2}|y_h - \tilde{R}_h z|_0^2 + \frac{\nu}{2}|u_h|_0^2, \\ -\Delta_h y_h = B_h u_h + \tilde{R}_h g, \end{cases} \tag{7.7}$$

where $u_h \in U_{adh} = U_{ad} \cap L_h^2$.

Let u_h^* denote the unique solution to (7.7) and set $y_h^* = y_h(u_h^*)$. The optimality system related to (7.7) is found to be

$$\begin{aligned} -\Delta_h y_h^* &= B_h u_h^* + \tilde{R}_h g, \\ -\Delta_h p_h^* &= -(y_h^* - \tilde{R}_h z), \\ (\nu u_h^* - B_h^* p_h^*) \cdot (v_h - u_h^*) &\geq 0 \text{ for all } v_h \in U_{adh}. \end{aligned} \tag{7.8}$$

To investigate the accuracy of the solution to (7.8) we use the approach of [82] extended to the finite difference framework introduced in Section 3. We require $\nu > 0$ and $\underline{u}$, $\overline{u} \in C^{0,1}(\Omega)$. The following accuracy estimates are obtained in [20]. We have

$$|u_h^* - R_h u^*|_0 \leq c\, h, \quad |y_h^* - R_h y^*|_0 \leq c\, h, \text{ and } |p_h^* - R_h p^*|_0 \leq c\, h.$$

These estimates are sub-optimal in the sense that $H^2(\Omega)$-regularity of y^* and p^* would suggest $\mathcal{O}(h^2)$ convergence estimates. Such results are impeded by the lack of the estimate

TABLE 7.1
Accuracy results; $\nu = 10^{-8}$.

mesh	$\|y - y_h\|_0$	$\|p - p_h\|_0$	$\|u - u_h\|_0$
129×129	$1.63\,10^{-7}$	$1.60\,10^{-3}$	$1.54\,10^{-3}$
257×257	$4.20\,10^{-8}$	$4.03\,10^{-4}$	$5.26\,10^{-4}$
513×513	$1.11\,10^{-8}$	$1.00\,10^{-4}$	$1.80\,10^{-4}$
1025×1025	$2.93\,10^{-9}$	$2.51\,10^{-5}$	$6.26\,10^{-5}$

$|y_h^* - \bar{y}_h|_0 \leq c\,h\,|y_h^* - \bar{y}_h|_1$. For finite element approximations this is a consequence of the Aubin-Nitsche duality argument. If the estimate were to hold, we would obtain h^2. Moreover, numerical experiments show that the estimate for the control function is pessimistic. In fact, we observe

$$|u_h^* - R_h u^*|_0 \leq c\,h^{3/2},$$

and $\mathcal{O}(h^2)$ convergence for the state and the adjoint variables. In a recent publication [90] these convergence rates could be verified for finite element approximations.

To validate these accuracy results consider the following exact solution to (7.4) with

$$g(x_1, x_2) = -u + 2\pi^2 \sin(\pi x_1)\sin(\pi x_2), \tag{7.9}$$

$$z(x_1, x_2) = -\Delta p + y, \tag{7.10}$$

where

$$y(x_1, x_2) = \sin(\pi x_1)\sin(\pi x_2), \tag{7.11}$$

$$p(x_1, x_2) = \sin(8\pi x_1)\sin(8\pi x_2), \tag{7.12}$$

$$u(x_1, x_2) = \max\{-1,\, \min\{1,\, p/\nu\}\}. \tag{7.13}$$

Note that the control is active for any $\nu < 1$.

Results of experiments with this test case are reported in Table 7.1. We observe second-order convergence of the state and adjoint variables. In fact, the solution errors reduce approximately as a factor of four by halving the mesh size. On the other hand, the error for the control scales as $h^{3/2}$.

These results have been obtained using our multigrid scheme. For the formulation of this algorithm notice that in order to represent (7.8) on all grids we need to have a multigrid full approximation storage representation of the problem. That is, a representation where the solution and not the error is computed on all grids. In fact, in order to impose the constraints the variables u_h and p_h must be available at all levels.

Let (7.8) with $h = h_k$ represents the fine grid problem to be solved, then the corresponding coarse grid problem on the grid with mesh size $H = h_{k-1}$ is given by

$$-\Delta_H y_H - B_H u_H = I_h^H g_h + \tau(y)_h^H, \tag{7.14}$$

$$-\Delta_H p_H + y_H = I_h^H z_h + \tau(p)_h^H, \tag{7.15}$$

$$(\nu u_H - B_H^* p_H) \cdot (v_H - u_H) \geq 0 \text{ for all } v_H \in U_{adH},$$

where $\tau(y)_h^H$ and $\tau(y)_h^H$ are fine-to-coarse defect corrections defined by

$$\tau(y)_h^H = -\Delta_H \hat{I}_h^H y_h - B_H \hat{I}_h^H u_h - I_h^H(-\Delta_h y_h - B_h u_h), \tag{7.16}$$

$$\tau(p)_h^H = -\Delta_H \hat{I}_h^H p_h + \hat{I}_h^H y_h - I_h^H(-\Delta_h p_h + y_h), \tag{7.17}$$

with $\hat{I}_h^H : L_h^2 \to L_H^2$ a restriction operator not necessarily equal to I_h^H. We choose $\hat{I}_h^H$ to be straight injection. Once the coarse grid problem is solved, the coarse grid correction follows

$$y_h^{new} = y_h + I_H^h(y_H - \hat{I}_h^H y_h), \tag{7.18}$$

$$p_h^{new} = p_h + I_H^h(p_H - \hat{I}_h^H p_h), \tag{7.19}$$

where $I_H^h : L_H^2 \to L_h^2$ represents an interpolation operator.

The coarse grid correction step just described is implemented in the $W(m_1, m_2)$-cycle process. In fact, our numerical experience [12] shows that in case of constrained control problems the use of W-cycles results in a robust multigrid iteration.

In the W-cycle we apply m_1 pre-smoothing and m_2 post-smoothing iterations. In the present context, the smoothing iteration must reduce the high-frequency components of the error and must preserve the inequality constraint. In case of unconstrained optimal control problems of the type considered here, the Fourier analysis presented in [21] and discussed in Section 5 proves that collective Gauss-Seidel iteration has the smoothing property required in the multigrid method. In order to present our modification of this iteration that takes account of the presence of constraints consider (7.8) at $\mathbf{x} \in \Omega_h$, where $\mathbf{x} = (ih, jh)$ and i, j index the grid points lexicographically. We have

$$-(y_{i-1\,j} + y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1}) + 4y_{ij} - h^2\, B_h u_{ij} = h^2 g_{ij} + h^2\, f_{ij}^{(y)}, \tag{7.20}$$

$$-(p_{i-1\,j} + p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1}) + 4p_{ij} + h^2\, y_{ij} = h^2\, z_h + h^2\, f_{ij}^{(p)}, \tag{7.21}$$

$$(\nu u_{ij} - B_h^*\, p_{ij}) \cdot (v_{ij} - u_{ij}) \geq 0 \text{ for all } v_h \in U_{adh}, \tag{7.22}$$

where $f^{(y)}$ and $f^{(p)}$ have been introduced to take into account the presence of defect corrections in (7.14) and (7.15).

A Gauss-Seidel step at $\mathbf{x}$ consists in updating the values y_{ij} and p_{ij} such that the resulting residuals of the two equations at that point are zero. The neighboring variables are considered constant during this process. Therefore, define the two constants

$$C_y = (y_{i-1\,j} + y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1}) + h^2 g_{ij} + h^2\, f_{ij}^{(y)},$$

and

$$C_p = (p_{i-1\,j} + p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1}) + h^2\, f_{ij}^{(p)}.$$

Replacing these two constants in (7.20) and (7.21), we obtain y_{ij} and p_{ij} as functions of u_{ij} as follows

$$y_{ij} = (C_y + h^2\, B_h u_{ij})/4, \tag{7.23}$$

and

$$p_{ij} = (4\, C_p - h^2\, C_y + 4\, h^2\, z_{ij} - h^4\, B_h u_{ij})/16. \tag{7.24}$$

Now to obtain the u_{ij} update, replace the expression for p_{ij} in the inequality constraint and define the auxiliary variable

$$\tilde{u}_{ij} = \frac{1}{16\,\nu + h^4} B_h^*\, (4\, C_p - h^2\, C_y + 4\, h^2\, z_{ij}). \tag{7.25}$$

Then, the new value for u_{ij} resulting from our Gauss-Seidel step is given by

$$u_{ij} = \begin{cases} \overline{u}_{ij} & \text{if } \tilde{u}_{ij} > \overline{u}_{ij} \\ \tilde{u}_{ij} & \text{if } \underline{u}_{ij} \leq \tilde{u}_{ij} \leq \overline{u}_{ij} \\ \underline{u}_{ij} & \text{if } \tilde{u}_{ij} < \underline{u}_{ij} \end{cases} \tag{7.26}$$

for all $\mathbf{x} = (ih, jh) \in \omega_h$, $u_{ij} = 0$ otherwise. With the new value of u_{ij} given, new values for y_{ij} and p_{ij} are obtained. This completes the Gauss-Seidel step.

The collective Gauss-Seidel step defined by (7.23), (7.24), (7.25), and (7.26) satisfies the inequality constraint; see [20]. Further, in case $\nu = 0$ the Gauss-Seidel iteration defined above satisfies (7.6). Because of (7.26) we can consider the present iteration belongs to the class of projected Gauss-Seidel schemes [39].

In the following we present numerical results with constrained optimal control problems obtained using our multigrid scheme. The multigrid setting is as follows. For the coarsest grid we have $h_1 = 1/8$ and we use up to eleven levels. For $L = 11$ we have a 8193×8193 mesh. In all experiments we use $m_1 = m_2 = 2$ smoothing steps and the FMG-W-cycle version of

TABLE 7.2
Case $\omega = \Omega$.

$\nu = 10^{-4}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.04, 0.04	$1.11\,10^{-1}$	$3.1\,10^{-10}$, $1.2\,10^{-13}$
257×257	0.03, 0.04	$1.11\,10^{-1}$	$6.8\,10^{-10}$, $7.1\,10^{-14}$
513×513	0.03, 0.04	$1.11\,10^{-1}$	$4.9\,10^{-10}$, $1.5\,10^{-13}$
1025×1025	0.03, 0.03	$1.11\,10^{-2}$	$3.2\,10^{-10}$, $7.2\,10^{-13}$
$\nu = 10^{-6}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.56, 0.56	$5.30\,10^{-2}$	$1.3\,10^{-6}$, $2.2\,10^{-10}$
257×257	0.52, 0.51	$5.30\,10^{-2}$	$1.5\,10^{-7}$, $1.3\,10^{-11}$
513×513	0.03, 0.03	$5.30\,10^{-2}$	$3.5\,10^{-10}$, $5.3\,10^{-14}$
1025×1025	0.03, 0.03	$5.30\,10^{-2}$	$2.2\,10^{-10}$, $2.2\,10^{-13}$
$\nu = 10^{-8}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.63, 0.63	$5.28\,10^{-2}$	$1.6\,10^{-3}$, $8.3\,10^{-8}$
257×257	0.54, 0.54	$5.28\,10^{-2}$	$2.4\,10^{-6}$, $7.4\,10^{-11}$
513×513	0.64, 0.60	$5.28\,10^{-2}$	$2.5\,10^{-7}$, $3.7\,10^{-12}$
1025×1025	0.68, 0.66	$5.28\,10^{-2}$	$2.7\,10^{-7}$, $2.1\,10^{-12}$
2049×2049	0.74, 0.71	$5.28\,10^{-2}$	$7.8\,10^{-7}$, $3.5\,10^{-12}$
4097×4097	0.76, 0.70	$5.28\,10^{-2}$	$7.4\,10^{-8}$, $2.9\,10^{-12}$

the algorithm with initial level $K = 3$. Recall that with this setting the typical multigrid convergence factor for the unconstrained optimal problem is $\rho \approx 0.08$ independently of ν and of the mesh size; see Section 5.

Consider the following objective function

$$z(x_1, x_2) = \sin(2\pi x_1)\sin(\pi x_2).$$

We make the choice $\omega = \Omega$ and the constraints are given by $\underline{u} = -30$ and $\overline{u} = 30$. For this case, the constraints are active in large portions of the domain for all three choices of $\nu = \{10^{-4}, 10^{-6}, 10^{-8}\}$ considered here.

From the results of numerical experiments reported in Table 7.2 we observe that for $\nu = 10^{-4}$ the multigrid convergence behavior is similar to that observed in the unconstrained case. Reducing the value of ν results in steeper gradients of the adjoint and control variables, particularly close to the boundary where p and u are required to be zero. Furthermore, decreasing ν results in an increasingly more complex switching structure of the control between upper and lower bounds; see Figure 7.1. The results for $\nu = 10^{-6}$ in Table 7.2 suggest that once the mesh size is sufficiently fine to resolve completely the switching structure the typical multigrid convergence rate is obtained. They further indicate that the multigrid convergence factor depends only weakly on the mesh size provided the problem is sufficiently well resolved on the mesh.

The ability of the multigrid scheme in solving constrained control problems with very small value of ν allows to investigate the occurrence of almost bang-bang control for the present class of problems. In particular, with the choice of z given above we can observe fast switching of the control function in the x_2 direction as depicted in Figure 7.1. In this figure we give plots of the control function for $x_1 = 3/4$ and $x_2 \in [0, 1]$ for the following choices of $\nu \in \{10^{-8}, 10^{-10}, 10^{-12}, 0\}$. We can see that as the value of ν is reduced the number of switching points increases.

The solution obtained for $\nu = 0$ is of interest: By further refining the mesh size additional switching points can be seen as shown in Figure 7.2.

Now consider the same setting and another desired state given by

$$z_1(x_1, x_2) = \sin(4\pi x_1)\sin(2\pi x_2).$$

The difference between this objective function and the previous one is that the gradient of z_1 is larger close to the boundary. For the choice $\nu = 0$ the constraints are everywhere active, i.e. differently from the previous case with desired state z the control is bang-bang. Moreover no fast switching of the control occurs. In Figure 7.3 the optimal control and the

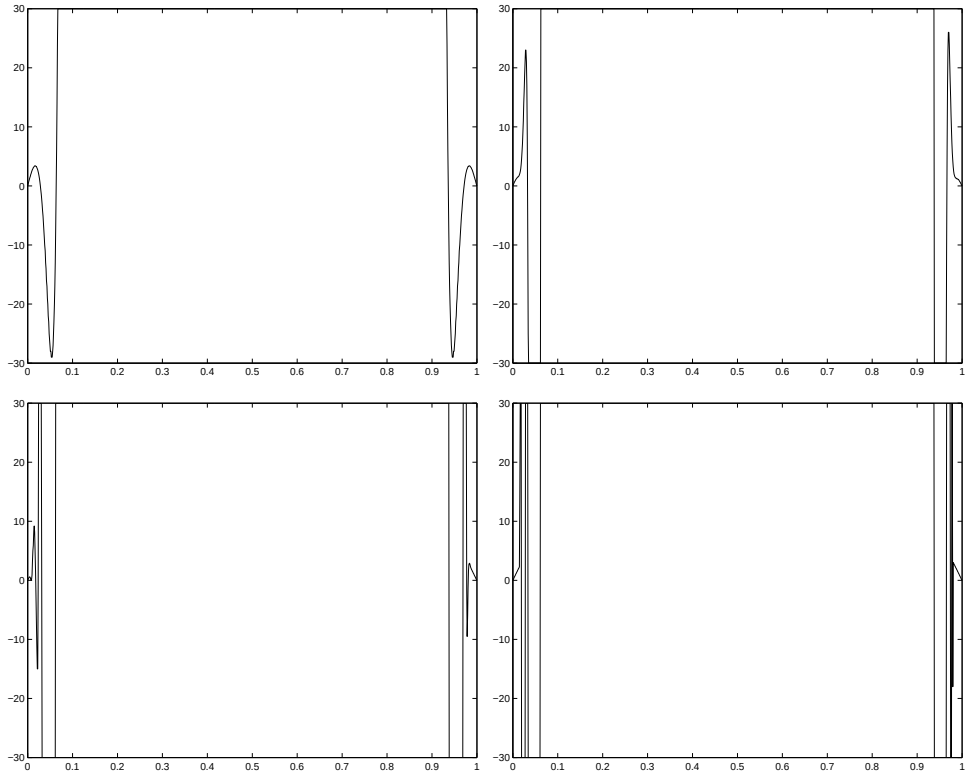

FIG. 7.1. *The control function for* $x_1 = 3/4$ *and* $x_2 \in [0,1]$ *obtained with* $\nu = 10^{-8}$ *(top left),* $\nu = 10^{-10}$ *(top right),* $\nu = 10^{-12}$ *(bottom left), and* $\nu = 0$ *(bottom right);* 2049×2049 *mesh.*

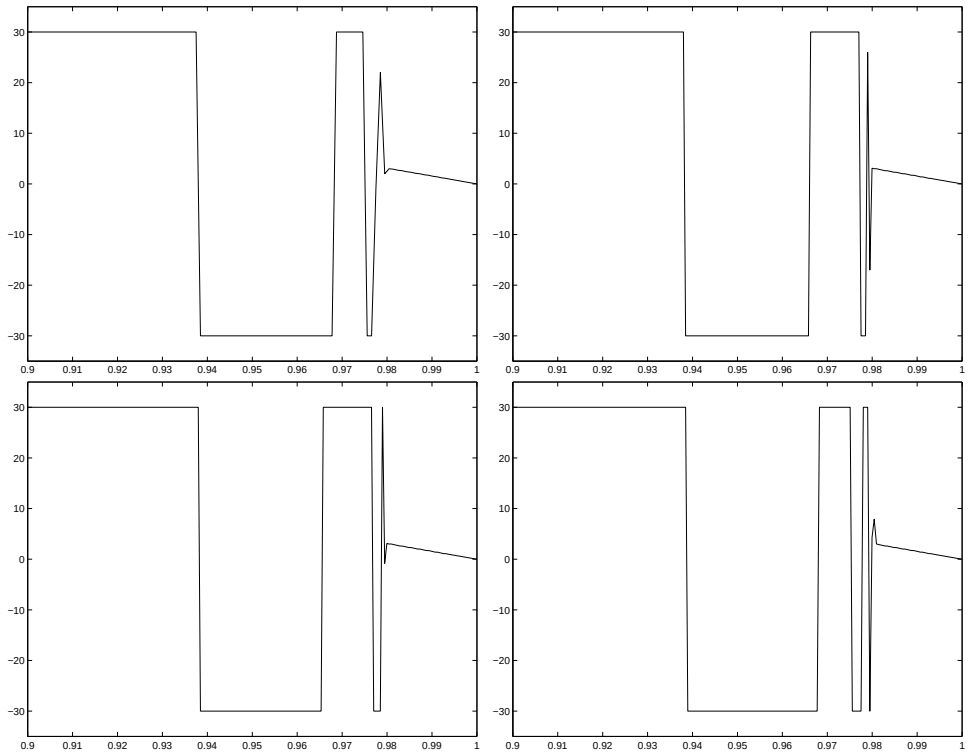

FIG. 7.2. *Switching of the control function for* $x_1 = 3/4$ *and* $x_2 \in [0.9,1]$ *(notice the scaling) obtained with* $\nu = 0$ *on increasingly finer meshes:* 1025×1025 *(top left),* 2049×2049 *(top right),* 4097×4097 *(bottom left), and* 8193×8193 *(bottom right).*

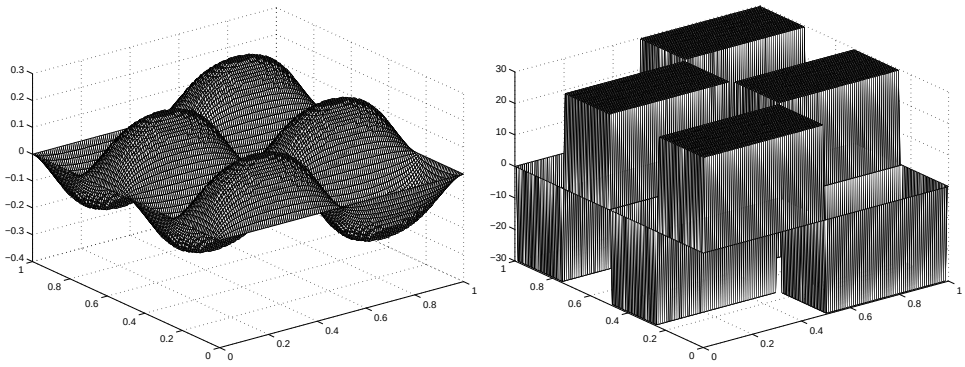

FIG. 7.3. *Numerical solutions with z_1 and $\nu = 0$. The state (left) and the control (right);* 257×257 *mesh.*

corresponding state for $\nu = 0$ are depicted. The numerical results in Table 7.3 document the convergence factors.

TABLE 7.3
Results of experiments with z_1 and $\nu = 0$.

mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
513×513	0.12, 0.13	$3.70\,10^{-1}$	$2.9\,10^{-8}$, $1.3\,10^{-13}$
1025×1025	0.12, 0.13	$3.70\,10^{-1}$	$2.5\,10^{-8}$, $4.2\,10^{-13}$
2049×2049	0.12, 0.16	$3.70\,10^{-1}$	$1.9\,10^{-8}$, $1.6\,10^{-12}$

Next consider the case where the control acts on ω which is given by

$$\omega = \{\mathbf{x} \in \Omega : (x_1 - 1/2)^2 + (x_2 - 1/2)^2 < \sqrt{(7/160)}\} \subset \Omega,$$

and z is the desired state.

In the unconstrained case typical multigrid convergence behavior is obtained. In the constrained case the obtained convergence factors are similar to those reported for the previous case with $\omega = \Omega$ and are reported in Table 7.4. Notice that the convergence factors are almost mesh independent.

TABLE 7.4
Case $\omega \subset \Omega$.

$\nu = 10^{-4}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.05, 0.05	$4.24\,10^{-1}$	$4.4\,10^{-10}$, $5.0\,10^{-11}$
257×257	0.05, 0.05	$4.23\,10^{-1}$	$3.2\,10^{-10}$, $1.4\,10^{-13}$
513×513	0.05, 0.05	$4.23\,10^{-1}$	$2.3\,10^{-10}$, $5.6\,10^{-13}$
1025×1025	0.05, 0.05	$4.23\,10^{-1}$	$2.7\,10^{-10}$, $2.6\,10^{-12}$
$\nu = 10^{-6}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.61, 0.61	$4.23\,10^{-1}$	$6.9\,10^{-6}$, $1.1\,10^{-9}$
257×257	0.67, 0.67	$4.22\,10^{-1}$	$3.0\,10^{-6}$, $2.7\,10^{-10}$
513×513	0.71, 0.70	$4.22\,10^{-1}$	$1.1\,10^{-6}$, $5.7\,10^{-11}$
1025×1025	0.72, 0.70	$4.22\,10^{-1}$	$3.7\,10^{-7}$, $1.0\,10^{-11}$
$\nu = 10^{-8}$			
mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
129×129	0.57, 0.57	$4.23\,10^{-1}$	$7.8\,10^{-5}$, $1.3\,10^{-8}$
257×257	0.58, 0.57	$4.22\,10^{-1}$	$1.2\,10^{-5}$, $6.9\,10^{-10}$
513×513	0.64, 0.63	$4.22\,10^{-1}$	$5.5\,10^{-6}$, $1.2\,10^{-10}$
1025×1025	0.70, 0.68	$4.22\,10^{-1}$	$2.5\,10^{-6}$, $2.7\,10^{-11}$

For comparison we report results obtained with the choice $\nu = 0$ and z_1 as objective function. Results of numerical experiments for this case are reported in Table 7.5. In Figure 7.4, bang-bang control in ω can be seen.

TABLE 7.5
Results of experiments with z_1, $\omega \subset \Omega$, and $\nu = 0$.

mesh	ρ_y, ρ_p	$\|y - z\|_0$	$\|r_y\|_0$, $\|r_p\|_0$
513×513	0.27, 0.23	$4.87\,10^{-1}$	$6.6\,10^{-8}$, $2.4\,10^{-13}$
1025×1025	0.22, 0.21	$4.87\,10^{-1}$	$1.6\,10^{-7}$, $6.3\,10^{-13}$
2049×2049	0.20, 0.21	$4.87\,10^{-1}$	$9.0\,10^{-9}$, $2.5\,10^{-12}$

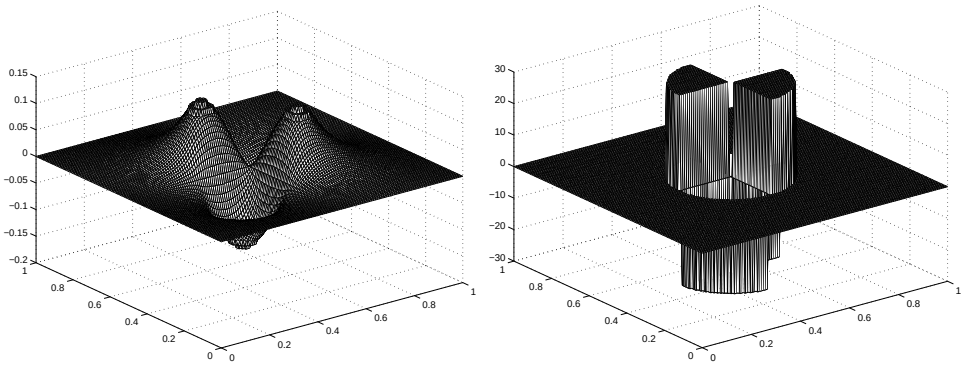

FIG. 7.4. *Numerical solutions with z_1 and $\omega \subset \Omega$ and $\nu = 0$. The state (left) and the control (right); 257×257 mesh.*

We now turn to the following boundary optimal control problem with constraints. Minimize

$$J(y,u) = \frac{1}{2}||y-z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\partial\Omega)}, \tag{7.27}$$

subject to $u \in U_{ad} \subset L^2(\partial\Omega)$ and

$$\begin{array}{rcll} -\Delta y + y & = & g & \text{in } \Omega, \\ \frac{\partial y}{\partial n} & = & u & \text{on } \partial\Omega, \end{array} \tag{7.28}$$

where Ω is a open bounded set of $\mathbf{R}^2$, $g \in L^2(\Omega)$, $z \in L^2(\Omega)$ is the objective function, and $\nu \geq 0$. The set of admissible controls is given in this case by

$$U_{ad} = \{u \in L^2(\partial\Omega) \,|\, \underline{u}(\mathbf{x}) \leq u(\mathbf{x}) \leq \overline{u}(\mathbf{x}) \text{ a.e in } \partial\Omega\}, \tag{7.29}$$

where $\underline{u}$ and $\overline{u}$ are functions of $L^\infty(\partial\Omega)$.

For the existence of a unique solution to (7.27)-(7.29) we refer to [79]. The solution is characterized by the following optimality system

$$\begin{array}{rcll} -\Delta y + y & = & g & \text{in } \Omega, \\ \frac{\partial y}{\partial n} & = & u & \text{on } \partial\Omega, \\ -\Delta p + p & = & -(y-z) & \text{in } \Omega, \\ \frac{\partial p}{\partial n} & = & 0 & \text{on } \partial\Omega, \\ (\nu u - p, v - u) & \geq & 0 & \text{for all } v \in U_{ad}. \end{array} \tag{7.30}$$

After discretization the optimal control problem becomes

$$\left\{ \begin{array}{c} \min \frac{1}{2}|y_h - \tilde{R}_h z|^2_0 + \frac{\nu}{2}|u_h|^2_0, \\ -\Delta_h y_h = \tilde{R}_h g, \\ \partial^n_h y_h = u_h. \end{array} \right. \tag{7.31}$$

Here, ∂^n_h denotes the second-order centered difference quotient with orientation normal to the boundary. The optimality system related to (7.31) is found to be

$$-\Delta_h y_h + y_h = g_h,$$

$$
\begin{aligned}
\partial_h^n y_h &= u_h \\
-\Delta_h p_h + p_h &= -(y_h - z_h), \\
\partial_h^n p_h &= 0 \\
(\nu u_h - p_h)\cdot(v_h - u_h) &\geq 0 \text{ for all } v_h \in U_{adh},
\end{aligned} \tag{7.32}
$$

where $g_h = \tilde{R}_h g$ and $z_h = \tilde{R}_h z$.

Notice that to solve (7.32) we need to realize the control on the boundary. For this purpose we eliminate the Neumann boundary conditions including them in the stencil of the differential operator considered at the boundary. We discuss this approach explicitly for one lateral boundary of a rectangular domain Ω_h.

Let $\mathbf{x} = (ih,\, jh)$ be a boundary grid point on the side $x = 0$. We have

$$
\begin{aligned}
-(y_{i-1\,j} + y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1}) + (4+h^2)y_{ij} &= h^2 g_{ij} + h^2\, f_{ij}^{(y)}, \\
y_{i-1\,j} - y_{i+1\,j} &= 2h u_{ij}, \\
-(p_{i-1\,j} + p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1}) + (4+h^2)p_{ij} + h^2\, y_{ij} &= h^2\, z_h + h^2\, f_{ij}^{(p)}, \\
p_{i-1\,j} - p_{i+1\,j} &= 0.
\end{aligned}
$$

Summing up the minus Laplacian stencil with the normal derivative the (ghost) variables outside of Ω are eliminated. We have

$$
\begin{aligned}
-(2y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1}) + (4+h^2)y_{ij} - 2h u_{ij} &= h^2 g_{ij} + h^2\, f_{ij}^{(y)}, \\
-(2p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1}) + (4+h^2)p_{ij} + h^2\, y_{ij} &= h^2\, z_h + h^2\, f_{ij}^{(p)}.
\end{aligned}
$$

The equations obtained in this way have the same structure as (7.20) and (7.21) and can be solved by our multigrid scheme.

The application of the collective Gauss-Seidel iteration follows along the same lines as described above. In the interior of the computational domain the collective Gauss-Seidel iteration reduces in this case to the single Gauss-Seidel iteration for the state equation while the residuals of the state equation and of the adjoint equation both enter in the relaxation of the adjoint variable.

For restricting the residuals at the boundary we use the full-weighting restriction operator [100]. This choice is necessary to guarantee the right scaling for the coarse-grid problem formulation; see the discussion in [100]. Clearly, on the boundary the restriction operator is mirrored.

To numerically validate the present algorithm for solving boundary optimal control problems, consider the desired state given by

$$
z(x_1, x_2) = (x_1^2 - x_2^2)\sin(\pi x_1)\sin(\pi x_2),
$$

and $g = 0$. We choose constraints given by $\underline{u} = -1$ and $\overline{u} = 1$ which are active in part of the boundary for $\nu \leq 10^{-6}$. The multigrid setting is the same as above. Results for this case are reported in Table 7.6.

TABLE 7.6
Results of experiments, boundary control problems; 1025×1025 *mesh.*

ν	ρ_y, ρ_p	$\lvert y - z\rvert_0$	$\lvert r_y\rvert_0$, $\lvert r_p\rvert_0$
10^{-6}	0.05, 0.05	$8.09\,10^{-2}$	$1.7\,10^{-10}$, $2.9\,10^{-13}$
10^{-8}	0.14, 0.12	$8.09\,10^{-2}$	$3.7\,10^{-8}$, $2.9\,10^{-13}$
10^{-10}	0.28, 0.28	$8.09\,10^{-2}$	$4.7\,10^{-5}$, $9.9\,10^{-11}$
0	0.25, 0.26	$8.09\,10^{-2}$	$3.5\,10^{-5}$, $4.8\,10^{-11}$

8. Algebraic multigrid methods for optimality systems. In this section we report our contribution to the development of algebraic multigrid schemes for solving elliptic differential systems [14, 15] with a focus on optimality systems arising from optimal control of convection-diffusion problems.

Present algebraic multigrid methods have been designed with a focus on scalar problems which limits the applicability of AMG to differential systems; see, e.g., [86]. In particular, a straightforward application of AMG schemes to solve optimality systems results in more copies of equal operators which results in unnecessary high memory complexity. Furthermore, the standard approach results in coarsening operators that not only depend on the 'differential' part of the problem but also on zero-order coupling terms. This must be avoided when solving optimality systems where the coefficient modulating the action of the control on the system scales as $1/\nu$, where ν may be taken as small as possible.

We are interested in the AMG solution of optimality systems of the following form

$$\begin{aligned} Ay &= p/\nu, \\ A^*p &= -(y - z). \end{aligned}$$

At the discrete level we have that in general the discrete operators A_h and A_h^* do not coincide. This is the case when, e.g., A is not self-adjoint. Nevertheless, under suitable conditions it may occur that $A_h^* = A_h^T$, where T means transpose. Assuming that the optimal control problem is discretized and then the optimality system is derived, we have that $A_h^* = A_h^T$ holds always true. In an optimal control context this approach is considered valid.

Therefore, under this assumption, to represent the optimality system one needs only store the matrix A_h and the mass matrix needed to represent the zero-order coupling terms. In the storage of the mass matrix we may assume a sparsity pattern equal to (or a subset of) that of the matrix of coefficients. Thus we do not need to add new pointers to the matrix elements. By this approach we obtain AMG algorithms for optimality systems having memory complexity close to that of algebraic multigrid methods for scalar problems.

In our approach the intergrid transfer operators are constructed based only on A_h. The advantage of this approach is that the resulting AMG setup effort for solving optimality systems is comparable to the algebraic multigrid setup work needed in scalar cases. This is a remarkable fact since AMG setup may cost as much as, e.g., six V-cycles of the AMG solver. Notice that, in our approach all algebraic multigrid components need be stored only once. Additional storage is required only to store an additional array corresponding the adjoint variables.

Results of numerical experiments demonstrate that the present approach results in AMG solvers that are robust with respect to changes of ν.

For the purpose of our discussion assume $\Omega \subset \mathbf{R}^3$ and consider the following elliptic problem

$$-\sum_{k=1}^{3} \frac{\partial}{\partial x_k} \left(d_k \frac{\partial y}{\partial x_k} \right) + \sum_{k=1}^{3} c_k \frac{\partial y}{\partial x_k} = u + g \quad \text{in } \Omega, \tag{8.1}$$

$$\alpha \frac{\partial y}{\partial \tau} + \beta\, y = \gamma \quad \text{on } \partial\Omega, \tag{8.2}$$

where, for $\mathbf{x} \in \Omega$, the functions $d_k = d_k(\mathbf{x})$ and $c_k = c_k(\mathbf{x})$, represent the diffusion coefficients and the convection coefficients, respectively. The right-hand side is given by u, $g \in L^2(\Omega)$. Equation (8.2) is defined on $\partial\Omega$, where α, β, $\gamma \in \mathbf{R}$ and τ is the outward normal to the boundary. It describes general Robin boundary conditions.

The formulation of optimal control problems requires to find u such that the following cost functional of tracking type is minimized

$$J(y, u) = \frac{1}{2}||y - z||^2_{L^2(\Omega)} + \frac{\nu}{2}||u||^2_{L^2(\Omega)}, \tag{8.3}$$

under the constrain that y is solution of (8.1) and (8.2). To minimize (8.3) with y governed by (8.1)-(8.2) one derives the necessary optimality conditions which result in the following optimality system

$$-\sum_{k=1}^{3} \frac{\partial}{\partial x_k} \left(d_k \frac{\partial y}{\partial x_k} \right) + \sum_{k=1}^{3} c_k \frac{\partial y}{\partial x_k} - \frac{1}{\nu}\, p = g \quad \text{in } \Omega, \tag{8.4}$$

$$\alpha \frac{\partial y}{\partial \tau} + \beta\, y = \gamma \quad \text{on } \partial\Omega, \tag{8.5}$$

$$-\sum_{k=1}^{3} \frac{\partial}{\partial x_k} \left(d_k \frac{\partial p}{\partial x_k} \right) - \sum_{k=1}^{3} c_k \frac{\partial p}{\partial x_k} + y = z \quad \text{in } \Omega, \tag{8.6}$$

$$\alpha \frac{\partial p}{\partial \tau} + \beta\, p = 0 \quad \text{on } \partial\Omega, \tag{8.7}$$

where we have substituted the control equation $\nu\, u - p = 0$. For simplicity of presentation we assumed that $\sum_{k=1}^{d} \partial c_k / \partial x_k = 0$. The differential operator in (8.6) is defined to be the adjoint of the differential operator in (8.4). The optimality system (8.4), (8.5), (8.6), and (8.7) characterizes the optimal control solution.

To describe our algebraic multigrid approach to solve the optimality system (8.4)-(8.7) denote with A the differential part of (8.4) and with A^* denote the differential part of (8.6). Let us represent the discretized version of the optimality system for the state and adjoint variable $\mathbf{w} = (y, p)$ as follows

$$\hat{\mathcal{A}}_k \mathbf{w}_k + \mathcal{B}_k \mathbf{w}_k = \mathcal{F}_k, \tag{8.8}$$

where

$$\hat{\mathcal{A}}_k = \begin{pmatrix} A_k & 0 \\ 0 & A_k^* \end{pmatrix}, \quad \mathcal{B}_k = \begin{pmatrix} 0 & -\frac{1}{\nu} B_k \\ B_k & 0 \end{pmatrix}, \text{ and } \mathcal{F}_k = \begin{pmatrix} g_k \\ z_k \end{pmatrix}, \tag{8.9}$$

and where the boundary conditions enter in the definition of $\hat{\mathcal{A}}_k$, $\mathcal{B}_k$, and of $\mathcal{F}_k$. The algebraic multigrid algorithm constructs a hierarchy of coarser problems denoted by (8.8) with $k = 2, \ldots, L$, where L indexes the coarsest level. In our approach this setup phase is based on the operator A_k. A detailed description of our AMG setup phase is given in [14]. Based on the entries of this matrix the sets of coarse and fine points are defined and the restriction and prolongation operators are constructed. The coarse matrix of coefficients $\hat{\mathcal{A}}_{k+1}$ and the mass matrix $\mathcal{B}_{k+1}$ are given by the Galerkin formula

$$\hat{\mathcal{A}}_{k+1} = \mathcal{I}_k^{k+1}\, \hat{\mathcal{A}}_k\, \mathcal{I}_{k+1}^k \quad \text{and} \quad \mathcal{B}_{k+1} = \mathcal{I}_k^{k+1}\, \mathcal{B}_k\, \mathcal{I}_{k+1}^k.$$

The AMG components are denoted in the following way. A smoothing procedure to solve for the rough components of the error at level k is denoted by $\mathcal{S}_k$. The operator $\mathcal{I}_k^{k+1}$ restricts the residual computed at level k to the level $k+1$. It represents the action of I_k^{k+1} on the residuals of the state and of the adjoint equations. I_k^{k+1} is the AMG restriction operator based on A_k. The prolongation operator is denoted by $\mathcal{I}_{k+1}^k$, it represents the action of I_{k+1}^k on the variables y and p. We sketch one V-cycle of the AMG solution process:

ALGORITHM 8.1. *AMG - $V(m_1, m_2)$-cycle to solve $\hat{\mathcal{A}}_k \mathbf{w}_k + \mathcal{B}_k \mathbf{w}_k = \mathcal{F}_k$.*
AMG SETUP (A_k,B_k)
AMG($\hat{\mathcal{A}}_k$,$\mathcal{B}_k$,$\mathbf{w}_k$,$\mathcal{F}_k$,ε)
begin
  if $\hat{\mathcal{A}}_k$ is the coarsest matrix then
    solve $\hat{\mathcal{A}}_k \mathbf{w}_k + \mathcal{B}_k \mathbf{w}_k = \mathcal{F}_k$ directly
  else
    apply $\mathcal{S}_k$, m_1 times, on $\hat{\mathcal{A}}_k \mathbf{w}_k + \mathcal{B}_k \mathbf{w}_k = \mathcal{F}_k$
    $\mathbf{r}_k = \mathcal{F}_k - \hat{\mathcal{A}}_k \mathbf{w}_k - \mathcal{B}_k \mathbf{w}_k$
    if k is the finest level and $\|\mathbf{r}_k\| < \varepsilon$ exit
    $\mathcal{F}_{k+1} = \mathcal{I}_k^{k+1} \mathbf{r}_k$, $\mathbf{w}_{k+1} = 0$
    AMG($\hat{\mathcal{A}}_{k+1}$,$\mathcal{B}_{k+1}$,$\mathbf{w}_{k+1}$,$\mathcal{F}_{k+1}$, ε)
    $\mathbf{w}_k = \mathbf{w}_k + \mathcal{I}_{k+1}^k \mathbf{w}_{k+1}$
    apply $\mathcal{S}_k$, m_2 times, on $\hat{\mathcal{A}}_k \mathbf{w}_k + \mathcal{B}_k \mathbf{w}_k = \mathcal{F}_k$
  endif
end

Numerical experiments were performed based on finite differences and finite elements matrices [14]. For the former we used the SPARSKIT software [93]. These matrices result

TABLE 8.1
Number of variables at various levels k of the coarsening process. Standard coarsening.

	$n = 50$	$n = 80$	$n = 100$
k	N_i	N_i	N_i
1	115200	486720	960400
2	57600	243360	480200
3	28800	121641	240051
4	11913	50452	99600
5	3450	14587	29159
6	830	3464	6787
7	169	658	1298
8	36	118	224
9	-	22	30

from the second-order centered finite-difference discretization of convection-diffusion equations in a cubic domain $\Omega = (0,1)^3$ discretized by a uniform grid with $n_1 = n_2 = n_3 = n$ grid points in each direction. Therefore a total of $N = n_1 n_2 n_3$ of state variables and of adjoint variables define the set of unknowns. In case of Dirichlet boundary conditions the corresponding boundary entries are eliminated from the algebraic system; we denote with N_i the effective number of grid points.

The number of iterations (No. iter.) reported are those needed by AMG to reach a value of the sum of the discrete L^2-norm $|\cdot|_0$ of the residuals which is less than the tolerance, $tol = 10^{-10}$, times the sum of the discrete L^2-norm of the right-hand sides. We denote with c_i and c_r the total storage complexity corresponding to the integer and real variables stored in the AMG code, respectively. Storage complexity is defined as the ratio of the total memory needed by the solver to the memory required to define the problem at the finest level (the input).

A challenging optimal control problem is formulated by (8.4) - (8.7) with boundary conditions given by

$$\begin{array}{ll} \text{planes } x_1 = 0,\, x_1 = 1: & \alpha = 0, \quad \beta = 1, \quad \gamma = 0; \\ \text{planes } x_2 = 0,\, x_2 = 1: & \alpha = 1, \quad \beta = 0, \quad \gamma = 0; \\ \text{planes } x_3 = 0,\, x_3 = 1: & \alpha = 0, \quad \beta = 1, \quad \gamma = 0. \end{array} \tag{8.10}$$

Further, we assume no convection, $c_k = 0$, $k = 1, 2, 3$, and the following discontinuous anisotropic diffusion coefficients

$$d_1(\mathbf{x}) = \begin{cases} 1 & x_2 \geq x_1, \\ 10^2 & x_2 < x_1, \end{cases} \quad d_2(\mathbf{x}) = \begin{cases} 10^2 & x_2 \geq x_1, \\ 1 & x_2 < x_1, \end{cases} \quad d_3(\mathbf{x}) = \begin{cases} 10^1 & x_2 + x_3 \geq 1, \\ 10^{-1} & x_2 + x_3 < 1. \end{cases}$$

As desired state we take

$$z(x_1, x_2, x_3) = \sin(2\pi x_1)\, \cos(2\pi x_2)\, \sin(2\pi x_3).$$

In Table 8.1 the coarsening history relative to this case is reported. Standard coarsening is approximately halving the number of variables at each coarsening step and this reduction factor is almost independent of the number of initial points.

When aggressive coarsening is performed, it is used only to pass from the finest to the next coarser level. This step reduces the number of variables by a factor of approximately eight.

Results regarding the computational complexity of our algorithm are reported in Table 8.2. These results are obtained with one pre-smoothing step and one post-smoothing step. In case of standard coarsening, the convergence factor is similar to that obtained by multigrid for simple scalar problems. In case of aggressive coarsening larger values of ρ are obtained. This is the disadvantage of using aggressive coarsening technique; compare with [96]. On the other hand, the use of aggressive coarsening results in better storage complexity as can be seen comparing the complexity factors.

Results on tracking ability and robustness of our algorithm are reported in Table 8.3. Notice that the AMG convergence factor does not deteriorate by reducing the value of ν.

TABLE 8.2
Convergence properties depending on N ($\nu = 10^{-6}$).

	Standard coarsening				Aggressive coarsening			
N_i	ρ	*No.iter.*	c_i/c_r	$\|y-z\|_0$	ρ	*No.iter.*	c_i/c_r	$\|y-z\|_0$
115200	0.16	17	3.25/3.31	3.23(-1)	0.61	63	2.21/2.19	3.23(-1)
486720	0.28	22	3.33/3.39	3.23(-1)	0.66	73	2.25/2.23	3.23(-1)
960400	0.30	25	3.35/3.42	3.22(-1)	0.67	80	2.26/2.25	3.22(-1)

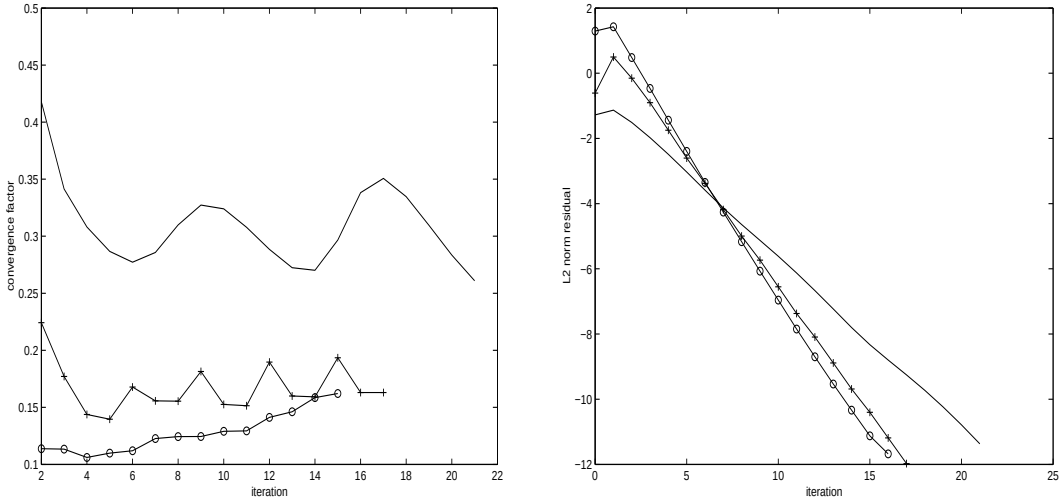

FIG. 8.1. *Computational performance of AMG for three values of the weight; $n = 50$. Standard coarsening. Left: convergence factor; Right: L^2-norm of residual. Solid line, $\nu = 10^{-4}$; +-Solid line $\nu = 10^{-6}$; o-Solid line $\nu = 10^{-8}$.*

In Figure 8.1 the convergence history of AMG for various choices of ν is depicted. Notice that initially the rate of reduction of the residuals may be different for different cycles. This fact may result in different total numbers of iterations also in the case of (almost) equal asymptotic convergence factors.

TABLE 8.3
Tracking properties depending on ν; $n = 50$.

ν	$\|y-z\|_0$	ρ	*No.iter.*
10^{-4}	3.56(-1)	0.30	21
10^{-6}	3.23(-1)	0.16	17
10^{-8}	4.57(-2)	0.13	16

Next consider the optimal control problem governed by a convection diffusion equation given by (8.4) - (8.7) with $d_k(\mathbf{x}) = 1$, $k = 1, 2, 3$, and with convection giving recirculating flow as follows

$$\begin{aligned} c_1 &= -\sin \pi x_1 \cos \pi x_2, \\ c_2 &= \sin \pi x_2 \cos \pi x_1, \\ c_3 &= 0. \end{aligned} \tag{8.11}$$

The boundary conditions are given by

$$\begin{aligned} &\text{planes } x_1 = 0,\ x_1 = 1: \quad \alpha = 1, \quad \beta = 0, \quad \gamma = 0; \\ &\text{planes } x_2 = 0,\ x_2 = 1: \quad \alpha = 1, \quad \beta = 0, \quad \gamma = 0; \\ &\text{planes } x_3 = 0,\ x_3 = 1: \quad \alpha = 0, \quad \beta = 1, \quad \gamma = 0. \end{aligned} \tag{8.12}$$

For these experiments the desired state is given by

$$z(x_1, x_2, x_3) = \sin(3\pi x_1) \cos(3\pi x_2) \sin(\pi x_3).$$

The number of coarse points obtained for the present test case are reported in Table 8.4. The coarsening behavior observed in this case is similar to that seen in the discontinuous diffusion case.

TABLE 8.4
Number of variables at various levels of the coarsening process.

	Standard coarsening		Aggressive coarsening	
k	N_i	N_i	N_i	N_i
1	120000	499200	120000	499200
2	60000	249600	14198	58067
3	10000	41608	6259	25454
4	1255	5225	991	3985
5	272	1210	161	639
6	41	193	31	98
7	-	34	-	18

In Table 8.5 we report the convergence behavior of AMG solving the present optimal control problem. We use two pre-smoothing and two post-smoothing steps. Notice a weak dependency of the convergence factor ρ on the size of the problem. Also in Table 8.5 we compare the performance of AMG when using standard coarsening and aggressive coarsening. The clear advantage of aggressive coarsening is smaller values of the complexity factors compared with those obtained by standard coarsening.

TABLE 8.5
Convergence properties depending on N_i ($\nu = 10^{-6}$).

	Standard coarsening				Aggressive coarsening			
N_i	ρ	*No.iter.*	c_i/c_r	$\lvert y-z\rvert_0$	ρ	*No.iter.*	c_i/c_r	$\lvert y-z\rvert_0$
120000	0.14	15	2.65/2.68	$2.79\,10^{-2}$	0.40	33	1.70/1.69	$2.79\,10^{-2}$
499200	0.19	21	2.69/2.73	$2.61\,10^{-2}$	0.51	42	1.72/1.72	$2.60\,10^{-2}$
712800	0.23	25	2.70/2.74	$2.59\,10^{-2}$	0.60	47	1.72/1.72	$2.60\,10^{-2}$

In Table 8.6 tracking errors and convergence factors depending on ν are reported. In all experiments we observed AMG convergence that does not deteriorates as ν tends to be small.

We conclude this section discussing a theoretical framework that makes possible to extend convergence results for AMG for scalar problems to the present AMG approach [14]. Convergence theories available [32, 33, 37, 92, 104] focus on symmetric positive definite scalar problems. Our approach is based on the theory developed in [21, 29].

Denote with V_k the space of vector functions u_k defined on level k. We introduce the discrete L^2-scalar product

$$(u,v)_k = \frac{1}{N_k} < u,\, v >_k, \quad u,\, v \in V_k, \tag{8.13}$$

where $< \cdot,\cdot >_k$ denotes the Euclidean scalar product for the N_k-dimensional vector space V_k. In the following we also denote with $(\cdot,\cdot)_k$ the L^2-scalar product for the system of point functions $\mathbf{w}_k = (u_k, v_k) \in \mathcal{V}_k$ where $\mathcal{V}_k = V_k \times V_k$. Furthermore, denote with $|\cdot|_{0,k} = (\cdot,\cdot)_k^{1/2}$ the discrete L^2-norm and $|\cdot|_{1,k} = (\hat{\mathcal{A}}_k\cdot,\cdot)_k^{1/2}$, where $\hat{\mathcal{A}}_k$ is defined as in (8.9) by multiplying the first equation by ν.

In order to prove convergence of multigrid, the following two conditions are usually required (see [32] and the discussion given there)

(A.1) The spectrum of $\hat{\mathcal{S}}_k\hat{\mathcal{S}}_k^*$ is in the interval $[0,1)$;
(A.2) $(\hat{\mathcal{A}}_k\,\mathcal{I}_{k+1}^k\mathbf{w}_{k+1}, \mathcal{I}_{k+1}^k\mathbf{w}_{k+1})_k \le (\hat{\mathcal{A}}_{k+1}\mathbf{w}_{k+1}, \mathbf{w}_{k+1})_{k+1}$ for all $\mathbf{w}_{k+1} \in \mathcal{V}_{k+1}$;

where $\hat{\mathcal{S}}_k^*$ is the adjoint of $\hat{\mathcal{S}}_k$ with respect to $(\hat{\mathcal{A}}_k\cdot,\cdot)_k$. Condition (A.1) is satisfied by the collective Gauss-Seidel iteration under the assumption that $\hat{\mathcal{A}}_k$ is positive definite and its diagonal blocks are positive definite; see [32, 62, 85]. Notice that the operator $\mathcal{A}_k = \hat{\mathcal{A}}_k + \mathcal{B}_k$ corresponding to the complete system satisfies these two conditions. Therefore condition (A.1) is satisfied by the collective Gauss-Seidel and (A.2) holds with equality due to the AMG variational formulation, $\mathcal{A}_{k+1} = \mathcal{I}_k^{k+1}\,\mathcal{A}_k\,\mathcal{I}_{k+1}^k$.

Conditions (A.1) and (A.2) are essential to ensure convergence of algebraic multigrid methods as stated in [32] (Theorem 1).

TABLE 8.6
Tracking properties depending on ν; $N_i = 120000$, standard coarsening.

ν	$\|y - z\|_0$	ρ	*No.iter.*
10^{-4}	$2.61\,10^{-1}$	0.08	12
10^{-6}	$2.79\,10^{-2}$	0.14	15
10^{-8}	$3.89\,10^{-3}$	0.03	8

THEOREM 8.2. *Assume that (A.1) and (A.2) hold. Then for some $\hat{\delta} < 1$ we have*

$$(\hat{\mathcal{A}}_1\hat{\mathcal{E}}_1\mathbf{u}, \hat{\mathcal{E}}_1\mathbf{u})_1 \leq \hat{\delta}^2\,(\hat{\mathcal{A}}_1\mathbf{u}, \mathbf{u})_1 \quad \textit{for all } \mathbf{u} \in \mathcal{V}_1,$$

where $\hat{\mathcal{E}}$ is the multigrid error operator.

This estimate does not guarantee AMG convergence independent of the number of unknowns. In order to have convergence independent of the size of the problem, additional assumptions on $\hat{\mathcal{A}}_k$ and on the construction of the intergrid transfer operators are needed.

In particular, $\mathcal{I}_{k+1}^k$ should satisfy the following weak approximation property: For any $\mathbf{u} \in \mathcal{V}_k$ there exists $\mathbf{v} \in \mathcal{V}_{k+1}$ such that

$$|\mathbf{u} - \mathcal{I}_{k+1}^k\mathbf{v}|_{0,k}^2 \leq \frac{\tilde{C}}{\rho(\hat{\mathcal{A}}_k)}|\mathbf{u}|_1^2.$$

(See [33, 37, 83, 104].)

We assume that $\hat{\delta}$ as in Theorem 5.4 is independent of the numbers of variables and formulate two additional conditions that provide convergence of our AMG algorithm:

LEMMA 8.3. *There exist two positive constants $\beta_{\mathcal{S}}$ and $C_{\mathcal{S}}$ independent of k such that*

$$\textit{(C.1)} \qquad |(\hat{\mathcal{A}}_k(\mathcal{S}_k - \hat{\mathcal{S}}_k)\mathbf{w}, \mathbf{v})_k| \leq C_{\mathcal{S}}\,\tfrac{1}{N_L^{\beta_{\mathcal{S}}}}\,|\mathbf{w}|_{1,k}\,|\mathbf{v}|_{1,k},$$

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

LEMMA 8.4. *There exist two positive constants $\beta_{\mathcal{P}}$ and $C_{\mathcal{P}}$ independent of k such that*

$$\textit{(C.2)} \qquad |(\hat{\mathcal{A}}_{k+1}(\hat{\mathcal{P}}_{k+1} - \mathcal{P}_{k+1})\mathbf{w}, \mathbf{v})_{k+1}| \leq C_{\mathcal{P}}\,\tfrac{1}{N_L^{\beta_{\mathcal{P}}}}\,|\mathbf{w}|_{1,k}|\mathbf{v}|_{1,k+1},$$

for all $\mathbf{w} \in \mathcal{V}_k$, $\mathbf{v} \in \mathcal{V}_{k+1}$.

The proof of the next theorem results from techniques given in [14, 21].

THEOREM 8.5. Assume that (A.1), (A.2), (C.1), and (C.2) hold. Then there exist positive constants N_0 and $\delta < 1$ such that for all $N_L > N_0$ we have

$$(\hat{\mathcal{A}}_1\mathcal{E}_1\mathbf{w}, \mathcal{E}_1\mathbf{w})_1 \leq \delta^2(\hat{\mathcal{A}}_1\mathbf{w}, \mathbf{w})_1, \quad \text{for all } \mathbf{w} \in \mathcal{V}_1,$$

where $\delta = \hat{\delta} + C\frac{1}{N_L^{\beta}}$, $\hat{\delta}$ is as in Theorem 5.4, and $\beta = \min\{\beta_{\mathcal{S}}, \beta_{\mathcal{P}}\}$.

In the context of geometric multigrid theory, Conditions (C.1) and (C.2) are proved in, e.g., [19, 32, 29] for various discretization schemes. While we were not able to prove the validity of these two conditions in general within the pure algebraic setting, we can give examples where (C.1) and (C.2) hold. In fact, we can argue that there exists a constant C such that $\bar{\lambda}_k^{-1} \leq C\,N_L^{-\beta}$ and therefore (C.1) holds in case of Richardson smoothing. Condition (C.2) is satisfied if we require that the norm of the difference between the $\hat{\mathcal{A}}_k$ projection and the discrete L^2 projection can be bounded by $C\,N_L^{-\beta}$.

The estimate of Theorem 5.8 states that for sufficiently large number of unknowns we have $\delta \approx \hat{\delta}$, that is, the convergence factor of the algebraic multigrid method applied to the optimality system is close to the convergence factor of the algebraic multigrid scheme applied to the corresponding scalar problem in agreement with our numerical experience.

9. Multigrid methods for parabolic optimal control problems. In this section we describe our development and analysis of space-time multigrid schemes for the solution of parabolic optimal control problems [11, 12, 19].

We consider parabolic multigrid methods that solve distributed parabolic optimal control systems in the whole space-time cylinder. The advantage of our approach, in contrast

to the sequential one, is the ability to implement time coupling in the optimality system consisting of parabolic partial differential equations with opposite orientation. The present space-time multigrid strategy results in fast solvers whose convergence factors are mesh independent and do not deteriorate as the weight of the cost of the control tends to be small.

The disadvantage of this approach is the need to store the dependent variables for all time steps. This fact may limit the choice of the size of the time interval. Nevertheless, the present approach may be used in the framework of receding horizon techniques [69] to solve optimal control problems in unbounded time intervals.

Consider the following optimal control problem

$$\left\{\begin{array}{rcll} \min_{u\in L^2(Q)} J(y,u), & & & \\ -\partial_t y+\Delta y &=& u & \text{in } Q=\Omega\times(0,T),\\ y(\mathbf{x},0) &=& y_0(\mathbf{x}) & \text{in } \Omega \text{ at } t=0,\\ y(\mathbf{x},t) &=& 0 & \text{on } \Sigma=\partial\Omega\times(0,T), \end{array}\right. \tag{9.1}$$

where we take $y_0(\mathbf{x})\in H_0^1(\Omega)$. We consider cost functionals of the tracking type given by

$$J(y,u)=\frac{1}{2}||y-z||^2_{L^2(Q)}+\frac{\nu}{2}||u||^2_{L^2(Q)}, \tag{9.2}$$

where $z\in L^2(Q)$ denotes the desired state. Then there exists a unique solution $(y^*,u^*)=(y^*(u^*),u^*)$ to the optimal control problem above; see [79]. Corresponding to our setting we have $y^*(u^*)\in H^{2,1}(Q)$ where $H^{2,1}(Q)=L^2(0,T;H^2(\Omega)\cap H_0^1(\Omega))\cap H^1(0,T;L^2(\Omega))$.

The solution to (9.1) is characterized by the following optimality system

$$-\partial_t y+\Delta y=u, \tag{9.3}$$

$$\partial_t p+\Delta p+(y-z)=0, \tag{9.4}$$

$$\nu u-p=0, \tag{9.5}$$

with initial condition $y(\mathbf{x},0)=y_0(\mathbf{x})$ for the state equation (evolving forward in time) and terminal condition $p(\mathbf{x},T)=0$ for the adjoint equation (evolving backward in time). Here, for convenience we dropped the $*$-notation. From (9.4) and (9.5) we have p, $u\in H^{2,1}(Q)$. The a priori knowledge of the regularity of solution is essential for the numerical analysis results which follow.

We consider the numerical solution of the optimality system (9.3)-(9.5) in the framework of finite differences and backward Euler scheme as discussed in Section 3.

The discrete optimal control problem is given by

$$\left\{\begin{array}{c} \min \frac{1}{2}||y_h-\tilde{R}_{h,Q}z||_0^2+\frac{\nu}{2}||u_h||_0^2,\\ -\partial_t^+ y_h+\Delta_h y_h=u_h. \end{array}\right. \tag{9.6}$$

Let $u_h\in L_h^2(Q_h)$ denote the unique solution to (9.6) and set $y_h=y_h(u_h)$. The optimality system related to (9.6) is found to be

$$\begin{aligned} -\partial_t^+ y_h+\Delta_h y_h&=u_h,\\ \partial_t^- p_h+\Delta_h p_h&=-(y_h-\tilde{R}_{h,Q}z),\\ \nu u_h-p_h&=0. \end{aligned} \tag{9.7}$$

Using the control equation we eliminate u_h from this system. We obtain

$$\begin{aligned} -[1+4\gamma]\,y_{i\,j\,m}+\gamma\,[y_{i+1\,j\,m}+y_{i-1\,j\,m}+y_{i\,j+1\,m}+y_{i\,j-1\,m}]+y_{i\,j\,m-1}&\\ =\frac{\delta t}{\nu}\,p_{i\,j\,m},\quad 2\le m\le N_t+1,& \end{aligned} \tag{9.8}$$

$$\begin{aligned} -[1+4\gamma]\,p_{i\,j\,m}+\gamma\,[p_{i+1\,j\,m}+p_{i-1\,j\,m}+p_{i\,j+1\,m}+p_{i\,j-1\,m}]+p_{i\,j\,m+1}&\\ +\delta t\,(y_{i\,j\,m}-\tilde{z}_{i\,j\,m})=0,\quad 1\le m\le N_t,& \end{aligned} \tag{9.9}$$

where $\gamma = \frac{\delta t}{h^2}$, i, j index the internal grid points, m is the time-step index, and $\tilde{z} = \tilde{R}_{h,Q} z$. The implementation of the boundary conditions on Σ, of the initial condition at $t = 0$, and of the terminal condition at $t = T$ should be clear.

For convenience, it is assumed that there exist positive constants $c_1 \leq c_2$ such that $c_1 h^2 \leq \delta t \leq c_2 h^2$. Hence h can be considered as the only discretization parameter. Therefore, in the following, the parameter δt is omitted.

Recall condition (3.1) given in Section 3. It implies

$$||\tilde{R}_{h,Q}\, v - R_{h,Q}\, v||_0 \leq c\, h^2\, |v|_{H^{2,1}(Q)}. \tag{9.10}$$

Using Lemma 1.1 and the approach of Theorem 1.2 of [82], and (9.10) we have

$$||u_h^* - R_{h,Q}\, u^*||_0 \leq c\, h^2. \tag{9.11}$$

Next, assume that y be solution of the state equation with any given $u \in L^2(Q)$ and let y_h be its finite difference approximation where $u_h = \tilde{R}_{h,Q}\, u$. Then the following estimate holds

$$||y_h - R_{h,Q}\, y||_0 \leq c\, h^2\, |y|_{H^{2,1}(Q)},$$

and in a similar way one has that $||p_h - R_{h,Q}\, p||_0 \leq c\, h^2 |p|_{H^{2,1}(Q)}$. Using these estimates and (9.11), we obtain the following estimates for the approximation of the state variable and of the adjoint variable

$$||y_h^* - R_{h,Q}\, y^*||_0 \leq c\, h^2 \text{ and } ||p_h^* - R_{h,Q}\, p^*||_0 \leq c\, h^2. \tag{9.12}$$

To validate the accuracy estimates (9.11) and (9.12), consider the following exact solution

$$y(\mathbf{x}, t) = t^2\, (1 - t)^2\, \sin(\pi x_1) \sin(\pi x_2),$$

$$p(\mathbf{x}, t) = 2\, \nu\, (1 - t) t (\pi^2 t^2 - (\pi^2 - 2) t - 1) \sin(\pi x_1) \sin(\pi x_2),$$

for the optimal control problem with objective function given by

$$z(\mathbf{x}, t) = ((t - 1)^2 t^2 + 2\nu\, (2\pi^4 t^4 - 4\pi^4 t^3 + 2(\pi^4 - 3) t^2 + 6t - 1)) \sin(\pi x_1) \sin(\pi x_2),$$

in $\Omega = (0, 1) \times (0, 1)$ and $T = 1$.

In Table 9.1 results of numerical experiments with this choice of $z(\mathbf{x}, t)$ and using $y_0(\mathbf{x}) = y(\mathbf{x}, 0)$ are reported. We observe second-order convergence. In fact, the solutions errors reduce approximately as a factor of four by halving the space mesh size.

TABLE 9.1
Accuracy results; $\nu = 10^{-4}$, $\delta t = 32\, h^2$.

$(\mathbf{x}, t)$-mesh	$\|\|y - y_h\|\|_0$	$\|\|p - p_h\|\|_0$
$32 \times 32 \times 32$	$2.63\, 10^{-5}$	$5.64\, 10^{-7}$
$64 \times 64 \times 128$	$7.05\, 10^{-6}$	$1.47\, 10^{-7}$
$128 \times 128 \times 512$	$1.78\, 10^{-6}$	$3.73\, 10^{-8}$

To solve (9.7) we consider two multigrid schemes corresponding to two different coarsening strategies. These methods solve (9.8) and (9.9) for all time levels simultaneously. We use a set of grids with space mesh size $h = h_k = h_1/2^{k-1}$, $k = 1, \ldots, L$, that is, standard coarsening in the space directions. In the time direction we set $\delta t = \delta t_k = \delta t_1 / s^{k-1}$, $s \in \{1, 2\}$. If $s = 1$ we have semicoarsening in space; the case $s = 2$ is referred to as standard time coarsening. The choice of different coarsening strategies can be motivated by memory needs. Clearly, larger values of the coarsening factor s results in less memory requirements.

The mesh of level k is denoted by Q_k. Any operator and variable defined on Q_k is indexed by k. Denote by $\mathcal{V}_k = V_k \times \tilde{V}_k$ the space of the system of nodal functions (y_k, p_k).

The algebraic problem given by (9.8) and (9.9) at level L with given initial, terminal, and boundary conditions is represented by the following equation

$$\mathcal{A}_L(\mathbf{w}_L) = \mathbf{f}_L, \tag{9.13}$$

where $\mathbf{w}_L$ is the pair $(y_L, p_L) \in \mathcal{V}_L$.

To solve (9.13) we use the FAS algorithm described in Section 4. To transfer the residuals we choose the half-weighted restriction operator in space with no averaging in the time direction. The prolongation is defined by bilinear interpolation in space. If $s = 1$ no interpolation in time is needed, whereas if $s = 2$ then bilinear interpolation also in time is applied.

The choice of the smoothing operator $\mathcal{S}_k$ is a delicate one. We need guarantee a robust coupling between state and control variables and take care of the fact that the state variable evolves forward in time while the adjoint variable evolves backwards. Robust coupling is essential in case of optimal control of nonlinear parabolic problems with possible finite-time blow-up [19, 22, 11]; see later in this section. In the following we describe a smoothing iteration which guarantees robust coupling.

Consider a collective Gauss-Seidel iteration which is applied at each grid point to the set of variables $\mathbf{w}_k = (y_k, p_k)$. For this purpose denote with $\mathcal{E}(\mathbf{w}_{i\,j\,m}) = [\mathbf{f} - \mathcal{A}(\mathbf{w})]_{i\,j\,m} = 0$, the two algebraic equations (9.8) and (9.9) for the two variables $y_{i\,j\,m}$ and $p_{i\,j\,m}$ at the grid point $i\,j\,m$. Further denote with $\mathcal{E}'$ the Jacobian of $\mathcal{E}$ with respect to these two variables. A sweep of the collective Gauss-Seidel scheme is given by

$$\mathbf{w}^{(1)}_{i\,j\,m} = \mathbf{w}^{(0)}_{i\,j\,m} - [\mathcal{E}'(\mathbf{w}^{(0)}_{i\,j\,m})]^{-1}\mathcal{E}(\mathbf{w}^{(0)}_{i\,j\,m}). \tag{9.14}$$

This iteration was discussed in Section 6 (see [18]) and successfully used to solve optimality systems relative to (steady-state) explosive phenomena. In case of time-dependent phenomena, iteration (9.14) will eventually diverge because the information of the opposite orientation of the state equation and of the adjoint equation is not taken into account. To add this information we use (9.14) to update the state component y marching in the forward direction and to update the adjoint variable p using (9.14) but marching backwards in time. In this way a robust iteration is obtained given by the following

Time-Splitted Collective Gauss-Seidel Iteration (TS-CGS)

1. Set $\mathbf{w}^0 = \tilde{\mathbf{w}}$.
2. For $m = 2, \ldots, N_t$ do
3. For ij in lexicographic order do

$$y^{(1)}_{i\,j\,m} = y^{(0)}_{i\,j\,m} - [\mathcal{E}'(\mathbf{w}_{i\,j\,m})]^{-1}\mathcal{E}(\mathbf{w}_{i\,j\,m})|_y,$$

$$p^{(1)}_{i\,j\,N_t-m+2} = p^{(0)}_{i\,j\,N_t-m+2} - [\mathcal{E}'(\mathbf{w}_{i\,j\,N_t-m})]^{-1}\mathcal{E}(\mathbf{w}_{i\,j\,N_t-m+2})|_p,$$

4. end.

Here, $[\mathcal{E}'(\mathbf{w}_{i\,j\,m'})]^{-1}\mathcal{E}(\mathbf{w}_{i\,j\,m'})$ is a two-component column vector corresponding to the variables y and p. A more explicit form of Step 3. follows

$$y^{(1)}_{i\,j\,m} = y^{(0)}_{i\,j\,m} - \frac{(1+4\gamma)\,r_y(\mathbf{w}) - \frac{\delta t}{\nu}\,r_p(\mathbf{w})}{(1+4\gamma)^2 + \frac{\delta t^2}{\nu}}|^{(0)}_{i\,j\,m},$$

$$p^{(1)}_{i\,j\,N_t-m+2} = p^{(0)}_{i\,j\,N_t-m+2} - \frac{(1+4\gamma)\,r_p(\mathbf{w}) + \delta t\,r_y(\mathbf{w})}{(1+4\gamma)^2 + \frac{\delta t^2}{\nu}}|^{(0)}_{i\,j\,N_t-m+2},$$

where $r_y(\mathbf{w})$ denotes the residual of (9.8) and $r_p(\mathbf{w})$ denotes the residual of (9.9) prior update. Obvious modifications are required to define time-splitted Red-Black collective Gauss-Seidel scheme or time-splitted collective Jacobi scheme.

The multigrid methods discussed in this section are designed to solve parabolic optimal control problems where the time discretization is by backward Euler scheme. In case of

Crank-Nicolson discretization our multigrid approach can be successfully applied only for a small range of values of $\gamma \approx 1$. For $\gamma >> 1$ our multigrid schemes possibly diverge. This fact is in agreement with results in [101] where it is shown that for large values of γ space-time multigrid solvers of parabolic problems are not robust in solving for Crank-Nicolson discretization. Following [101] three-level backward Euler discretization could be used for large γ when second-order time accuracy is required.

We now investigate the convergence properties of the proposed multigrid methods. We use the FAS $V(1,1)$-cycle with the time-splitted collective Gauss-Seidel scheme as the smoothing iteration. The discretization parameters are chosen such that $\gamma = \delta t/h^2 >> 1$, which is the most common situation where implicit time discretization is chosen.

Consider $\Omega = (0,1) \times (0,1)$, $T = 1$, and the following objective function

$$z(\mathbf{x}, t) = (x_1 - x_1^2)\,(x_2 - x_2^2)\,\cos(4\pi\, t).$$

We take $y_0(\mathbf{x}) = z(\mathbf{x}, 0)$.

The results reported in the following tables are obtained with $N = 10$ FAS $V(1,1)$ cycles. In Table 9.2 we report results of numerical experiments with multigrid using semi-coarsening strategy. Three different grids with increasing refinement in space directions are considered. Similar results are obtained also on meshes with $N_t = \{128, 256, 512\}$. The observed convergence factors demonstrate usual multigrid convergence speeds and appear to be almost independent of the value of γ and weakly dependent on the value of ν. This fact shows robustness of the multigrid scheme with TS-CGS smoothing. As the value of ν increases then larger values of $||y - z||_0$ are obtained.

TABLE 9.2
Results of experiments with semicoarsening.

$\nu = 10^{-4}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.146	$1.55\,10^{-3}$	$4.5\,10^{-10}$, $7.6\,10^{-12}$
$64 \times 64 \times 64$	64	0.164	$1.55\,10^{-3}$	$9.1\,10^{-10}$, $1.0\,10^{-11}$
$128 \times 128 \times 64$	256	0.159	$1.55\,10^{-3}$	$1.1\,10^{-9}$, $8.1\,10^{-12}$
$\nu = 10^{-6}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.147	$4.03\,10^{-5}$	$1.4\,10^{-10}$, $1.9\,10^{-13}$
$64 \times 64 \times 64$	64	0.140	$4.23\,10^{-5}$	$2.6\,10^{-10}$, $2.1\,10^{-13}$
$128 \times 128 \times 64$	256	0.165	$4.27\,10^{-5}$	$3.3\,10^{-10}$, $5.8\,10^{-13}$
$\nu = 10^{-8}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.008	$9.09\,10^{-7}$	$4.7\,10^{-15}$, $1.1\,10^{-18}$
$64 \times 64 \times 64$	64	0.06	$1.73\,10^{-6}$	$9.1\,10^{-12}$, $7.6\,10^{-16}$
$128 \times 128 \times 64$	256	0.134	$2.06\,10^{-6}$	$9.1\,10^{-11}$, $8.1\,10^{-15}$

In case of multigrid coarsening in space and in time we observe slow convergence; see Table 9.3. Nevertheless, for sufficiently small values of ν typical multigrid convergence factors are obtained. This fact is confirmed by the Fourier analysis reported later in this section.

The setting $s = 2$ provides acceptable multigrid convergence rates also in case where T and correspondingly γ tend to be small. Small time intervals are of interest when considering control of transient phenomena. For example, consider the case $T = 0.01$ and a highly oscillating objective function given by

$$\tilde{z}(\mathbf{x}, t) = (x_1 - x_1^2)\,(x_2 - x_2^2)\,\sin(100\,\pi t).$$

In this case the choices $s = 2$ gives good results even for moderate values of ν as can be seen in Table 9.4.

Further numerical experiments demonstrate that the multigrid convergence behavior as observed in this section appears to be insensitive to the particular choice of the objective function, which may not be attainable.

In the framework of Fourier mode analysis [36, 68, 100, 101] it is possible to analyze the convergence properties of the twogrid version of our parabolic optimal control solver.

TABLE 9.3
Results of experiments with standard coarsening.

$\nu = 10^{-4}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.929	$1.60\,10^{-3}$	$1.9\,10^{-2}$, $2.6\,10^{-4}$
$64 \times 64 \times 64$	64	0.976	$1.69\,10^{-3}$	$7.3\,10^{-2}$, $5.5\,10^{-4}$
$128 \times 128 \times 64$	256	0.999	$2.20\,10^{-3}$	$1.4\,10^{-1}$, $1.2\,10^{-3}$
$\nu = 10^{-6}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.905	$4.07\,10^{-5}$	$3.1\,10^{-3}$, $1.9\,10^{-6}$
$64 \times 64 \times 64$	64	0.140	$4.23\,10^{-5}$	$2.6\,10^{-10}$, $2.1\,10^{-13}$
$128 \times 128 \times 64$	256	0.165	$4.27\,10^{-5}$	$3.3\,10^{-10}$, $5.8\,10^{-13}$
$\nu = 10^{-8}$				
$(\mathbf{x}, t)$-mesh	γ	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
$32 \times 32 \times 64$	16	0.008	$9.09\,10^{-7}$	$4.7\,10^{-15}$, $1.1\,10^{-18}$
$64 \times 64 \times 64$	64	0.06	$1.73\,10^{-6}$	$9.1\,10^{-12}$, $7.6\,10^{-16}$
$128 \times 128 \times 64$	256	0.134	$2.06\,10^{-6}$	$9.1\,10^{-11}$, $8.1\,10^{-15}$

TABLE 9.4
Results of experiments with standard time coarsening for the case $T = 0.01$ and $\tilde{z}$ as objective function. The finest grid is $N_x \times N_y \times N_t = 64 \times 64 \times 128$.

ν	ρ	$\|\|y - z\|\|_0$	$\|\|r_y(\mathbf{w})\|\|_0$, $\|\|r_p(\mathbf{w})\|\|_0$
10^{-3}	0.154	$2.29\,10^{-3}$	$1.1\,10^{-8}$, $3.2\,10^{-10}$
10^{-5}	0.117	$9.63\,10^{-4}$	$2.2\,10^{-9}$, $6.0\,10^{-12}$
10^{-7}	0.110	$4.61\,10^{-5}$	$1.4\,10^{-9}$, $3.7\,10^{-13}$

We investigate the dependence of the twogrid convergence factor on the weight ν and on the ratio $\gamma = \delta t/h^2$. The Fourier analysis provides convergence results that closely predict the convergence factors observed experimentally. For ease of presentation the one space dimensional case is considered.

We use Fourier mode analysis assuming infinite grids; see [12] for all the details. On the fine grid consider the Fourier components $\phi(\boldsymbol{j}, \boldsymbol{\theta}) = e^{i\boldsymbol{j}\cdot\boldsymbol{\theta}}$ where i is the imaginary unit, $\boldsymbol{j} = (j_x, j_t) \in \mathbb{Z} \times \mathbb{Z}$, $\boldsymbol{\theta} = (\theta_x, \theta_t) \in [-\pi, \pi)^2$, and $\boldsymbol{j} \cdot \boldsymbol{\theta} = j_x\theta_x + j_t\theta_t$.

First, consider the case of semicoarsening in space. The frequency domain is spanned by the following two sets of frequencies

$$\boldsymbol{\theta}^{(0,0)} := (\theta_x, \theta_t) \text{ and } \boldsymbol{\theta}^{(1,0)} := (\overline{\theta}_x, \theta_t),$$

where $(\theta_x, \theta_t) \in ([-\pi/2, \pi/2) \times [-\pi, \pi))$ and $\overline{\theta}_x = \theta_x - sign(\theta_x)\pi$. The components $\phi(\cdot, \boldsymbol{\theta}^{\boldsymbol{\alpha}})$ are called harmonics. The first harmonics $\phi(\cdot, \boldsymbol{\theta}^{(0,0)})$ represents low frequencies components in space. The second harmonics $\phi(\cdot, \boldsymbol{\theta}^{(1,0)})$ contains the high frequencies components in space direction. Both have all frequencies components in time direction. Using semicoarsening, we have that $\phi(\boldsymbol{j}, \boldsymbol{\theta}^{(0,0)}) = \phi(\boldsymbol{j}, \boldsymbol{\theta}^{(1,0)})$ on the coarse grid.

With this setting the twogrid operator $\boldsymbol{TG}_k^{k-1}$ results in a 4×4 matrix given by

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

where $\widehat{\boldsymbol{CG}}_k^{k-1}$ is the Fourier symbol of the coarse grid correction.

Consider to apply the TS-CGS step first to all state variables leaving the adjoint variables unchanged and then vice versa. Under this assumption substituting (**??**) into (9.8) and (9.9) we obtain

$$\widehat{\mathcal{S}}_k(\boldsymbol{\theta}) = diag\{\sigma(\boldsymbol{\theta}^{(0,0)}), \sigma(\boldsymbol{\theta}^{(1,0)}), \sigma(\boldsymbol{\theta}^{(0,0)}), \sigma(\boldsymbol{\theta}^{(1,0)})\},$$

where

$$\sigma(\boldsymbol{\theta}^{(p,q)}) = \frac{\nu\gamma(2\gamma+1)e^{i\theta_x^p}}{\delta t^2 + \nu[(2\gamma+1)^2 - \gamma(2\gamma+1)e^{-i\theta_x^p} - (2\gamma+1)e^{-i\theta_t^q}]}.$$

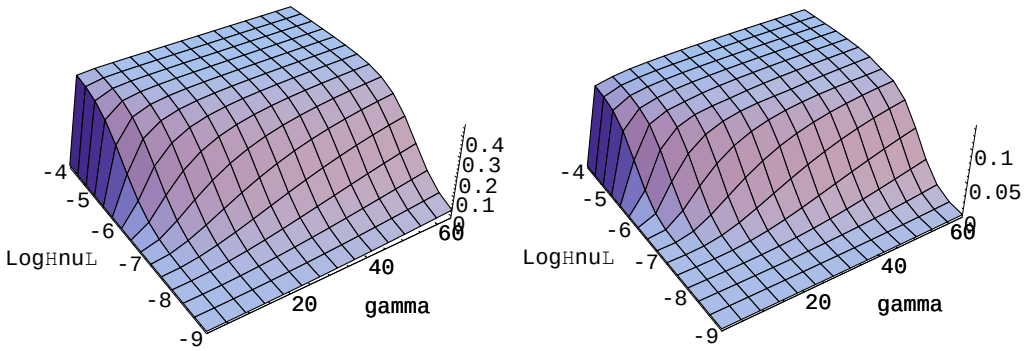

FIG. 9.1. *Semicoarsening: (left) smoothing factor as a function of ν and γ; (right) twogrid convergence factor as a function of ν and γ ($m_1 = m_2 = 1$).*

In space, we consider a full-weighting restriction operator whose symbol is given by

$$\hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta}) = \frac{1}{2}\begin{bmatrix} (1+\cos(\theta_x)) & (1-\cos(\theta_x)) & 0 & 0 \\ 0 & 0 & (1+\cos(\theta_x)) & (1-\cos(\theta_x)) \end{bmatrix}.$$

For the linear prolongation operator we have $\hat{\mathcal{I}}_{k-1}^{k}(\boldsymbol{\theta}) = \hat{\mathcal{I}}_k^{k-1}(\boldsymbol{\theta})^T$. The symbol of the fine grid operator is

$$\widehat{\mathcal{A}}_k(\boldsymbol{\theta}) = \begin{bmatrix} a_y(\boldsymbol{\theta}^{(0,0)}) & 0 & -\delta t/\nu & 0 \\ 0 & a_y(\boldsymbol{\theta}^{(1,0)}) & 0 & -\delta t/\nu \\ \delta t & 0 & a_p(\boldsymbol{\theta}^{(0,0)}) & 0 \\ 0 & \delta t & 0 & a_p(\boldsymbol{\theta}^{(1,0)}) \end{bmatrix},$$

where

$$a_y(\boldsymbol{\theta}^{(p,q)}) = 2\gamma\cos(\theta_x^p) - e^{-i\theta_t^q} - 2\gamma - 1 \text{ and } a_p(\boldsymbol{\theta}^{(p,q)}) = 2\gamma\cos(\theta_x^p) - e^{i\theta_t^q} - 2\gamma - 1.$$

The symbol of the coarse grid operator follows

$$\widehat{\mathcal{A}}_{k-1}(\boldsymbol{\theta}) = \begin{bmatrix} \gamma\cos(2\theta_x)/2 - e^{-i\theta_t} - \gamma/2 - 1 & -\delta t/\nu \\ \delta t & \gamma\cos(2\theta_x)/2 - e^{i\theta_t} - \gamma/2 - 1 \end{bmatrix}.$$

Notice that on the coarser grid δt remains unchanged while $\gamma \to \gamma/4$ by coarsening.

Based on the representation on $\boldsymbol{TG}_k^{k-1}$ by a 4×4 matrix $\widehat{\boldsymbol{TG}}_k^{k-1}(\boldsymbol{\theta})$ we can calculate the convergence factor given by

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

By Fourier mode analysis the problem of determining the convergence factor of a twogrid scheme is reduced to that of determining the spectral radius of a 4×4 matrix. This task may be performed using any symbolic package (we use Mathematica). In Figure 9.1 we report the smoothing factor and the convergence factor depending on the value of ν and on the value of γ for $\delta t = 1/64$. Similar figures are obtained with different choices of time-step size. We observe that the values of convergence factors predicted by Fourier mode analysis are very close to those obtained experimentally and reported in Table 9.2.

In case of standard coarsening the analysis performed in [12] is similar to that performed in Section 5 for the case of two-dimensional elliptic problems, considering the time coordinate as the second space coordinate.

The values of the smoothing factor and of the convergence factor as functions of ν and γ are reported in Figure 9.2 (for $\delta t = 1/64$). Comparing with the semicoarsening case the smoothing factor and correspondingly the convergence factor have deteriorated.

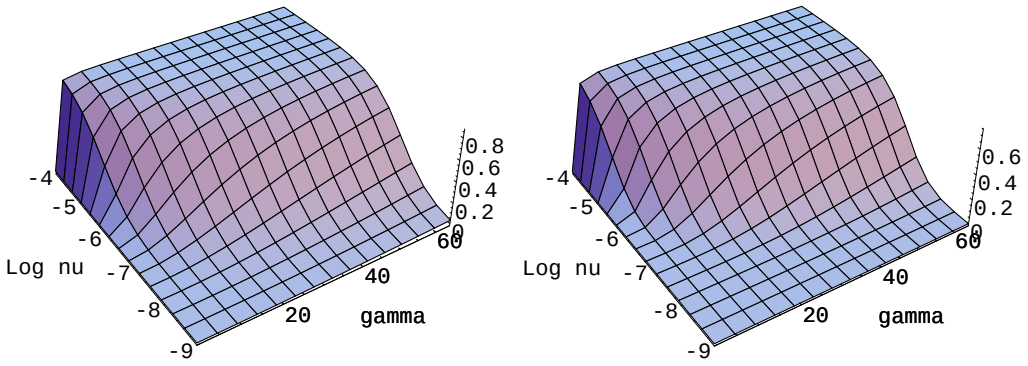

FIG. 9.2. *Standard coarsening: (left) smoothing factor as a function of ν and γ; (right) twogrid convergence factor as a function of ν and γ ($m_1 = m_2 = 1$).*

Nevertheless, as γ (correspondingly δt) and ν are sufficiently small the convergence factor improves.

As demonstrated in [19, 22, 11], the convergence performance and behavior of the space-time multigrid algorithms discussed in this section remain similar also when solving singular optimal control problems.

We conclude this section discussing the multigrid solution of constrained optimal control problems. In [11] two multigrid approaches were investigated. The first one considers the presence of constraints as a generic nonlinearity within the FAS multigrid approach. The second one uses the primal-dual strategy [8] and defines a primal-dual-multigrid algorithm. In this section we report further results regarding the former 'direct' multigrid approach.

Consider the optimal control problem (9.1) with the control u belonging to the following admissible set

$$u \in \{v \in L^2(Q) \,|\, c_1 \le v(\mathbf{x}) \le c_2 \text{ a.e. in } Q\}. \tag{9.15}$$

Existence of a unique solution can be proved [79] and is characterized by the following optimality system

$$-\partial_t y + \Delta y = u, \tag{9.16}$$

$$\partial_t p + \Delta p + (y - z) = 0, \tag{9.17}$$

$$u - \sup(c_1, \inf(c_2, \frac{p}{\nu})) = 0; \tag{9.18}$$

see [79, 82].

The discretization scheme (9.8) and (9.9) applies also to (9.16), (9.17), and (9.18) with the right-hand side of (9.8) being replaced by $\delta t\, G(p_{i\,j\,m})$ where

$$G(p_{i\,j\,m}) = \max(c_1, \min(c_2, \frac{p_{i\,j\,m}}{\nu})). \tag{9.19}$$

Unfortunately the presence of the term (9.19) prevents us from defining $\mathcal{E}'$ in the obvious way. Specifically, we cannot differentiate (9.18) with respect to p. To overcome this difficulty we set $G'(p) = 1/\nu$ as in the unconstrained case in those part of the domain where the constrains are inactive. Whenever a constraint is active we set $G'(p) = 0$. So the TS-CGS scheme is as follows

$$y^{(1)}_{i\,j\,m} = y^{(0)}_{i\,j\,m} - \frac{(1+4\gamma)\, r_y(\mathbf{w}) - \delta t G'(p)\, r_p(\mathbf{w})}{(1+4\gamma)^2 + \delta t^2 G'(p)}|^{(0)}_{i\,j\,m},$$

$$p^{(1)}_{i\,j\,N_t-m+2} = p^{(0)}_{i\,j\,N_t-m+2} - \frac{(1+4\gamma)\, r_p(\mathbf{w}) + \delta t\, r_y(\mathbf{w})}{(1+4\gamma)^2 + \delta t^2 G'(p)}|^{(0)}_{i\,j\,N_t-m+2},$$

where $m = 2, \ldots, N_t$.

Using this relaxation scheme a lack of robustness of the V-cycle with respect to changes of the value of the weight of the cost of the control may be observed [19]. Specifically, as ν tends to be smaller the convergence factor of the V-cycle version of our multigrid scheme worsen. We now show that much better computational properties are obtained when using the W-cycle version of our algorithm.

Consider the following objective function

$$z(\mathbf{x}, t) = \sin(2\pi t)\sin(\pi x_1)\sin(\pi x_2).$$

For the purpose of comparison we report in Table 9.5 results of experiments using the V-cycle and W-cycle algorithms with semicoarsening. Observe the deterioration of the convergence factor when reducing the value of ν in the V-cycle case. In the W-cycle case much better convergence factors can be observed and the algorithm remains competitive also for moderate small values of ν. The use of W-cycles is also beneficial for standard coarsening multigrid schemes.

TABLE 9.5
Convergence factors with semicoarsening on a $128 \times 128 \times 128$ *mesh* $\gamma = 128$*;* $c_1 = -10$ *and* $c_2 = 10$.

ν	$V(2,2)$-cycle	$W(2,2)$-cycle
10^{-4}	0.137	0.052
10^{-6}	0.761	0.143

10. Optimal control approach to optical flow computation. In [16, 17] we formulate an optimal control problem for the determination of optical flow. An optical flow is the field of apparent velocities in a sequence of images. This is a challenging application of hyperbolic optimal control problems. Multigrid methods apply to solve the elliptic sub-system defining the optimality conditions.

Our new framework differs from preceding approaches [5, 6, 52, 75, 107, 108] in that it does not require differentiation of the data and combines optical flow with image reconstruction. It can be considered as a control-in-the-coefficients hyperbolic problem with a cost functional of tracking type.

From the knowledge of the optical flow, information about the spatial arrangement of objects and the rate of change of this arrangement ought to be obtained. Under suitable assumptions, we can assume that the image brightness of an object point remains constant in the images when the object moves. That is, the total time derivative of the brightness at each point (x_1, x_2) at time t is zero:

$$\frac{\partial y}{\partial t} + u\frac{\partial y}{\partial x_1} + v\frac{\partial y}{\partial x_2} = 0, \tag{10.1}$$

where $y = y(x_1, x_2, t)$ denotes the image brightness at (x_1, x_2) and t, and $\vec{w} = (u, v)$ represents the optical flow vector.

We formulate the optimal control problem for optical flow as follows. Consider a sequence of image frames $\{Y_k\}_{k=0,N}$ sampled at increasing time steps, $t_k \in [0, T]$, $k = 0, 2, \ldots, N$, where $t_0 = 0$ and $t_N = T$. Each frame is assumed to be defined on a rectangle which defines the spatial domain Ω. The space-time box in which the optical flow takes place is $\Omega \times [0, T]$. We define the following optimal control problem:

Find $\vec{w} \in V$ (V is a class of admissible optical flow fields) and $y = y(\vec{w})$ such that

$$\begin{cases} y_t + \vec{w} \cdot \nabla y = 0, & \text{in } Q = \Omega \times (0, T], \\ y(\cdot, 0) = Y_0, \end{cases} \tag{10.2}$$

and minimize the cost functional

$$\begin{aligned} J(y, \vec{w}) = {} & \frac{1}{2}\int_\Omega \sum_{k=1}^{N} |y(x_1, x_2, t_k) - Y_k|^2 d\Omega \\ & + \frac{\alpha}{2}\int_Q \Phi(|\frac{\partial \vec{w}}{\partial t}|^2) dq + \frac{\beta}{2}\int_Q \Psi(|\nabla u|^2 + |\nabla v|^2) dq + \frac{\gamma}{2}\int_Q |\nabla \cdot \vec{w}|^2 dq. \end{aligned} \tag{10.3}$$

Here, α, β, and γ are predefined nonnegative weights. Let us briefly discuss the relevance of the additive terms in J. The first one is the least-squares term requiring that $\vec{w}$ is chosen such that $y(\cdot, t_k, \vec{w})$ approximates Y_k at the sampling times. Assuming $\vec{w}$ is smooth with respect to t, we set $\Phi(s) = s$. For the regularization in the spatial direction we use a Gaussian-type regularization in regions where $\vec{w}$ varies smoothly and bounded variation type regularization across edges and corners, where $\nabla\vec{w}$ is large; see, e.g., [70, 91]. These two criteria lead to the following choice

$$\Psi(s) = \begin{cases} 2\sqrt{s} & \text{for } s \in [0, \delta), \\ s + c_1 & \text{for } s \in [\delta, \delta'], \\ 2\sqrt{s} + c_2 & \text{for } s \in (\delta', \infty); \end{cases} \tag{10.4}$$

see [16] for all details. In [17] the case $\Psi(s) = s$ is considered.

To motivate the last term in (10.3), consider the case where the velocity field on the border of a small subregion is uniform. The points in the interior of the subregion should be assigned the same value too. A way of expressing this property is to penalize by $\int_Q |\nabla\cdot\vec{w}|^2 dq$.

From (10.2)-(10.3) we derive the following optimality system

$$\begin{aligned} y_t + \vec{w}\cdot\nabla y = 0, \text{ with } y(\cdot,0) = Y_0, \\ p_t + \nabla\cdot(\vec{w}p) = \sum_{k=1}^{N-1} \left[\delta(t-t_k)(y(\cdot,t_k) - Y_k)\right], \text{ with } p(\cdot,T) = -(y(\cdot,T) - Y_N), \\ \alpha\frac{\partial^2 u}{\partial t^2} + \beta\nabla\cdot[\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla u] + \gamma\frac{\partial}{\partial x_1}(\nabla\cdot\vec{w}) = p\frac{\partial y}{\partial x_1}, \\ \alpha\frac{\partial^2 v}{\partial t^2} + \beta\nabla\cdot[\Psi'(|\nabla u|^2 + |\nabla v|^2)\nabla v] + \gamma\frac{\partial}{\partial x_2}(\nabla\cdot\vec{w}) = p\frac{\partial y}{\partial x_2}, \end{aligned} \tag{10.5}$$

where δ denotes the Dirac δ-function. The interpretation of the second equation in (10.5) is

$$p_t + \nabla\cdot(\vec{w}p) = 0, \text{ on } t \in (t_{k-1}, t_k), \text{ for } k = 1, \ldots, N, \tag{10.6}$$

$$p(\cdot, t_k^+) - p(\cdot, t_k^-) = y(\cdot, t_k) - Y_k, \text{ for } k = 1, \ldots N-1. \tag{10.7}$$

We refer to the first hyperbolic equation of (10.5) as the *optical flow constraint* (OFC) equation (marching forward in time). The second equation is the *adjoint optical flow* equation (marching backward in time). The last two equations are nonlinear elliptic equations. They are referred to as optimality condition. Some aspects of the well-posedness of (10.5) in the case $\Psi = I$ are analyzed in [16].

Concerning boundary conditions for $\vec{w}$ we restrict the admissible optical flow fields to satisfy prescribed homogeneous Dirichlet boundary conditions on the spatial boundary and natural boundary conditions at the temporal boundaries of Q.

In the numerical implementation of our method we use the Horn & Schunck scheme [52] to obtain a starting approximation. Though this scheme was one of the first methods for determining optical flow, it is still popular and one of the most used methods both in its original form or with various modifications; see, e.g., [5, 6, 75, 107, 108].

The method of Horn and Schunck [52] combines the optical flow constraint (10.1) with a global smoothness term, minimizing:

$$\int_D (y_t + \vec{w}\cdot\nabla y)^2 + \lambda^2(|\nabla u|^2 + |\nabla v|^2) d\mathbf{x}. \tag{10.8}$$

In the original formulation [52] the domain D is the space domain Ω and all quantities in (10.8) are considered defined at intermediate time step $(DT/2)$ between two sampled images, $\{Y_k, Y_{k+1}\}$. The spatio-temporal derivatives (y_{x_1}, y_{x_2}, y_t) are obtained by numerical differentiation of the sampled image data.

A minimum of (10.8) satisfies necessarily the Euler equations:

$$\lambda^2\Delta u - y_{x_1}(y_t + uy_{x_1} + vy_{x_2}) = 0, \tag{10.9}$$

$$\lambda^2\Delta v - y_{x_2}(y_t + uy_{x_1} + vy_{x_2}) = 0, \tag{10.10}$$

where Δ is the Laplace operator and homogeneous Dirichlet boundary conditions are used. To discretize (10.9) - (10.10) a uniform sampling rate DT is assumed and set equal to one. The spatial dimension of Ω is given in units of pixels. The distance between pixels is normalized to be one and the spatial discretization in (10.9) - (10.10) is linked to the availability of pixel information by setting $DX_1 = DX_2 = 1$.

The choice of the regularization parameter λ specifies the degree of smoothness of the solution. Though there is no "best choice" available, we shall take $\lambda = 0.5$ unless specified otherwise. This is the value suggested in [5].

In [6], examples are given where the use of accurate discretization schemes for differentiation provide wrong results unless the velocity of the pattern (i.e. $\vec{w}$) is close to the ratio of the spatial to the temporal sampling, that is, $u \approx DX_1/DT$ and $v \approx DX_2/DT$. This may be related to the fact that in order to solve numerically the optical flow equation (10.1) the following CFL-like condition must be satisfied:

$$\tau \leq \frac{C_{CFL}}{\max(|u|_{max}/h, |v|_{max}/h)}. \tag{10.11}$$

Here τ is the time step size, h the spatial mesh size, and $0 < C_{CFL} \leq 1$ is the CFL number.

The discussion above outlines a limitation of the Horn & Schunck scheme which is not present in our approach, since the time discretization for the numerical realization of (10.5) and the sampling times for the images are independent.

The numerical difficulties in solving (10.5) are due to the presence of two coupled subsystems with different characteristics: as pointed out above the first two equations of (10.5) are hyperbolic, the last two are elliptic. To solve (10.5) we combine an explicit time-marching second-order TVD scheme for the hyperbolic part of the system and a FAS multigrid method for the elliptic part.

Some details of the discretization are given in Section 3. Regarding the implementation of the FAS scheme notice that the discretized control equations have strong anisotropy in the coefficients because usually we have $\tau << h$. For this reason t-line relaxation is chosen.

The coupling between the hyperbolic and the elliptic subsystems is obtained in an outer loop where each solver is called sequentially and their solutions are used to update the data for the next subsystem. This outer loop, which we call *segregation loop*, is repeated a fixed number of times ($I_{loop} = 10$) or until a given convergence criterion is reached.

The method is summarized as follows:

Segregation loop for solving the optimal control problem (10.5).

1. Apply the Horn & Schunck method for a starting approximation to the optical flow.
2. Solve the optical flow constraint equation to obtain y.
3. Solve (backward) the adjoint optical flow constraint equation to obtain p.
4. Update the right-hand sides of the elliptic system.
5. Apply a few cycles of multigrid to solve the control equations.
6. Go to 2 and repeat I_{loop} times.

Note that the segregation loop realizes the following sequence of substitutions:

$$\vec{w}^{old} \rightarrow y(\vec{w}^{old}) \rightarrow p(y(\vec{w}^{old})) \rightarrow \vec{w}^{new}.$$

This iteration is well-posed; see [16].

In the description of numerical experiments with images, we need to define quantities describing accuracy of optical flow determination, measures of tracking ability and corresponding costs.

In the optical flow community an angular measure of error is used to measure optical flow accuracy. One considers the pattern displacement as a space-time direction vector $\vec{w} = (u, v, 1)$ in units of (pixel,pixel,frame). The corresponding three-dimensional direction vector is given by:

$$\hat{w} = \frac{1}{\sqrt{1 + u^2 + v^2}}(u, v, 1)^T,$$

where T means transpose. The angular error between the correct velocity $\vec{w}^c$ (assuming it is known) and an estimate $\vec{w}^e$ is as follows [5, 75]

$$\psi^E = \arccos(\hat{w}^c \cdot \hat{w}^e). \tag{10.12}$$

Notice that this represents the *space-time* orientation error. We denote by $\psi^E_{i,j,\kappa} = \arccos(\hat{w}^c_{i,j,\kappa} \cdot \hat{w}^e_{i,j,\kappa})$ the function (10.12) evaluated at grid point (i, j, κ), $i, j = 1, \ldots, L$, where L is the number of pixel in each direction, and $\kappa = 1, \ldots, K$, where K is number of time sub-intervals of size $\tau = T/K$.

The measure (10.12) of the optical flow error is made global by considering the *mean orientation error*:

$$\tilde{\psi} = \frac{1}{KL^2} \sum_{\kappa=1}^{K} \sum_{i,j=1}^{L} \psi^E_{i,j,\kappa},$$

at the space-time mesh points of the evolving images. On this set of points, we also compute the maximum modulus of the functions u and v denoted by $|u|_{max}$ and $|v|_{max}$ respectively. This will help to establish and compare the accuracy of the optical flow with respect to the maximal velocity, obtained with the optimal control approach and with the Horn & Schunck method.

In the tables we also report the tracking error, denoted by $||y - Y||^2$, and defined by

$$\sum_{\kappa=1}^{K} \sum_{i,j=1}^{L} (y^{t_k}_{i,j} - Y(x_{1i}, x_{2j}, t_k))^2. \tag{10.13}$$

Moreover, with $\sum cost$ we denote the value of the discrete version to the following expression:

$$\frac{\alpha}{2} \int_Q |\frac{\partial \vec{w}}{\partial t}|^2 dq + \frac{\beta}{2} \int_Q \Psi(|\nabla u|^2 + |\nabla v|^2) dq + \frac{\gamma}{2} \int_Q |\nabla \cdot \vec{w}|^2 dq. \tag{10.14}$$

Finally, in order to validate the divergence term in the cost functional, we also report the value of the discrete version of the divergence term.

In the following experiments with synthetic images we consider frames with 64×64 ($L = 64$) pixel and $K = 64$ time sub-intervals of size $\tau = T/64$.

If, for example, $T = 4$ we are given five image frames, we have $\tau = 1/16$ and, since $DT = 1$, every $l_t = K/T = 16$ time steps a new image frame is given. In the multigrid solver five levels are used. The coarsest grid is a $4 \times 4 \times 4$ space-time grid, refined by halving the mesh size.

A standard test for optical flow solvers with synthetic images is given by a square moving with velocity (u_c, v_c). At time t the frames are defined by means of the following function:

$$I(x, y, t) = \begin{cases} 1 + \eta' & x_{1_t}^l \le x_1 \le x_{1_t}^r \text{ and } x_{2_t}^l \le x_2 \le x_{2_t}^u, \\ \eta' & \text{otherwise,} \end{cases} \tag{10.15}$$

where $x_{1_t}^l$ and $x_{1_t}^r$ are the x_1-coordinates of the left- and right vertical edges of the square (or rectangle) respectively, and $x_{2_t}^l$ and $x_{2_t}^u$ are the x_2-coordinates of the lower- and upper horizontal edges of the square (or rectangle) respectively. The explicit form in terms of t of these coordinate functions will be specified later. The function η' denotes noise which is realized numerically by adding uniformly distributed random numbers in $[-\eta, \eta]$ at pixel values. The difficulty of this test stems from the fact of severe under-determination: in fact the flow field is only available along the edges of the cube.

In the experiments that follow, we consider the translating square (or rectangle) on a sequence of five frames, so that $N = 4$ and $T = 4$ ($l_t = 16$). Further (x_{10}, x_{20}) represents the pixel $(20, 20)$.

A first test case is given by a moving square with $(u_c, v_c) = (1, 1)$. Let us consider the images of a square given by $x_{1_t}^l = x_{10} + u_c t$, $x_{1_t}^r = x_{10} + u_c t + 20$, $x_{2_t}^l = x_{20} + v_c t$, and $x_{2_t}^u = x_{20} + v_c t + 20$. The noise function is $\eta = 0.3$ (this corresponds to 30% of maximal image brightness). Numerical results are depicted in Table 10.1.

TABLE 10.1
Case 1 : $(u_c, v_c) = (1, 1)$ *with 30% added noise. Dependence on* γ.

Dependence on γ; $\alpha = 5.0$, and $\beta = 0.25$.					
γ	$\|u\|_{max}$, $\|v\|_{max}$	ψ	$\|\|y - Y\|\|^2$	$\sum cost$	$\|\|div(w)\|\|^2$
0	1.18, 1.25	35.7	209.3	20.2	0
0.25	1.12, 1.15	33.5	221.6	12.3	9.6(-1)
0.5	**1.08, 1.08**	**31.6**	**233.0**	**9.5**	**7.5(-1)**
1	1.02, 1.04	28.3	253.2	6.8	5.4(-1)
5	0.91, 0.91	18.3	336.3	3.4	2.1(-1)
H & S	1.62, 1.64	41.66			

TABLE 10.2
Two sample images: $(u_c, v_c) = (1.5, 2)$ *with added 30% noise. Dependence on control weights.*

Dependence on α; $\beta = 0.25$, and $\gamma = 0.1$.					
α	$\|u\|_{max}$, $\|v\|_{max}$	ψ	$\|\|y - Y\|\|^2$	$\sum cost$	$\|\|div(w)\|\|^2$
0.5	1.59, 2.08	44.1	107.4	20.1	8.5(-1)
1	1.78, 2.27	45.7	104.7	45.6	1.6
5	2.08, 2.34	48.4	98.5	145.5	3.9
Dependence on γ; $\alpha = 1.0$, and $\beta = 0.25$.					
γ	$\|u\|_{max}$, $\|v\|_{max}$	ψ	$\|\|y - Y\|\|^2$	$\sum cost$	$\|\|div(w)\|\|^2$
0	1.86, 2.34	46.5	102.4	65.1	0
0.25	1.67, 2.16	44.7	107.6	29.2	1.3
0.5	1.53, 2.03	43.5	111.6	16.5	1.0
1	**1.39, 1.82**	**41.8**	**117.9**	**7.9**	**6.9(-1)**
2	1.31, 1.58	39.8	127.6	4.0	4.3(-1)
H & S	2.18, 2.28	49.9			

The actual choice of "optimal parameters" is a delicate matter due to the lack of an indisputable measure. In the tables the result printed in bold fonts represents a combination of a preferable choice of parameters (α, β, γ). The effect of changes of (α, β, γ) on the result is quite the same as that expected from our experience with inverse problems. Since (u_c, v_c) is independent of time, the effect of α is not so significant , as long as $\alpha > 0$. Increasing (β, γ) from $(0, 0)$ until a threshold is reached improves the result both as far as the graphical representation of (u_c, v_c) is concerned as well as the error $\bar{\psi} = 180\,\tilde{\psi}/\pi$. However, increasing (β, γ) also has the effect that the object to be reconstructed, i.e. (u_c, v_c), is increasingly diffused and hence underestimated with respect to its size. Increasing γ has the effect of filling in information available along the edges of the square into its interior.

Similar accurate results are obtained in case $(u_c, v_c) = (1.5, 2)$. This is a more challenging situation because the optical flow orientation is not invariant under coordinate change.

Another challenging test is provided by a moving and dilating rectangle. Let us consider the image of a rectangle given by $x_{1_t}^l = x_{10} + u_c t$, $x_{1_t}^r = x_{10} + 2u_c t + 20$, $x_{2_t}^l = x_{20}$, and $x_{2_t}^u = x_{20} + 20$. No noise is added. This function represents a square of 20×20 pixels that changes to a rectangle linearly with time along the x_1 direction: $u_c = 1.3$ ($v_c = 0$).

By considering translation only in one direction we can test to which extend energy associated to one flow component is diffused to the other component. In Figure 10.1 snapshots of the optical flow solution for this case (at $t = 2$) are shown. We observe that the v-component computed with Horn & Schunck is far from being zero. On the other hand the optimal control approach seems to have a reduced 'mixing' effect and allows a more accurate velocity estimate also in this case.

Our approach is able to determine optical flow also in the case where only two image frames are given. In this case, $N = 1$, $T = 1$. Control applies through the final observation. To test the ability of our algorithm, we consider the translating square case with $(u_c, v_c) = (1.5, 2)$ and 30% added noise. In Table 10.2 we notice that in case of only two given frames a smaller value of α than in the multiple image frame sequence gives best results.

We conclude this section reporting results on a known benchmark for verification of optical flow solvers: the 'Hamburg Taxi Sequence'; see [5]. It consists of a sequence of frames of a taxi coming from the right in the main road and turning right into a side street in Hamburg (Germany). The first frame of the Taxi Sequence is depicted in Figure 10.2

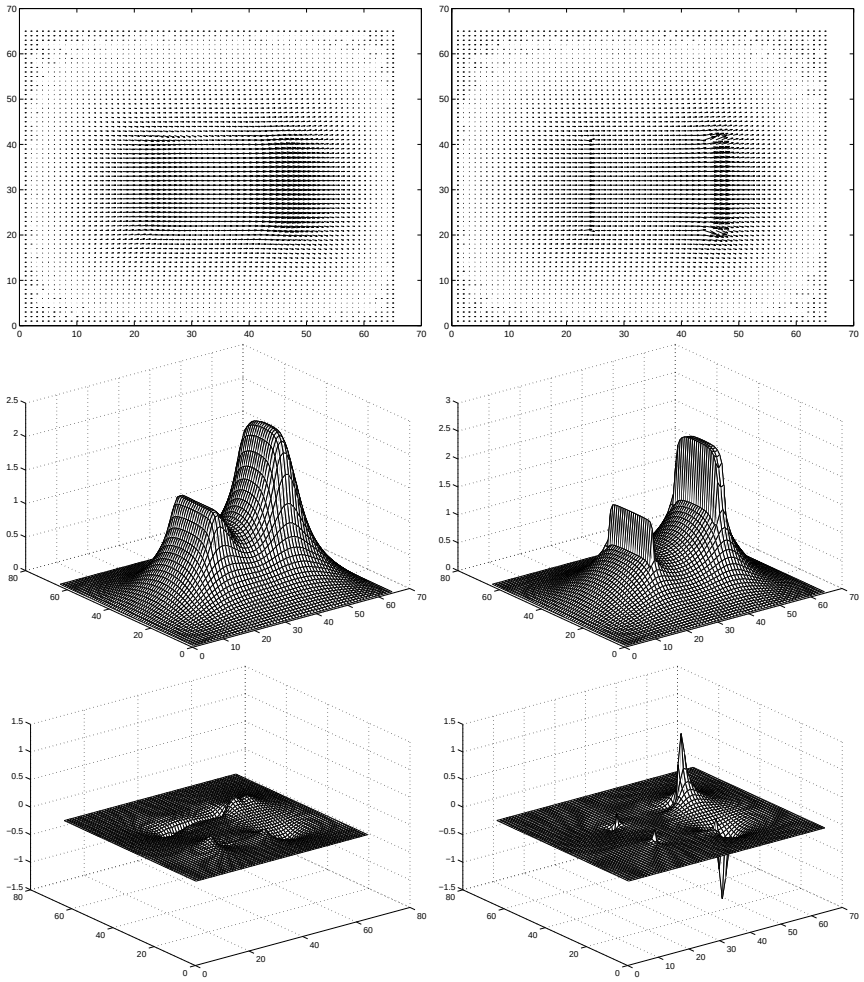

FIG. 10.1. *Optical flow for the moving and dilating rectangle. Optical flow results with optimal control (left) and with Horn & Schunck (right). Optical flow (top); the u component (middle); the v component (bottom).*

(top-left). In the same scene, two dark colored cars are driving in the main road entering from the left and from the right into the scene. One can also see parked cars, houses, and trees. The corresponding (irregular) brightness pattern is shown in Figure 10.2 (top-right).

We consider a sequence of five photos of the moving taxi taken at regular intervals ($T = 4$). Our algorithm is applied with $\alpha = 5.0$, $\beta = 0.25$, $\gamma = 0.5$, and $I_{loop} = 10$. In Figure 10.2 the optical flow computed with the optimal control approach is depicted.

11. Conclusion and outlook. Results of our work on multigrid methods for elliptic, parabolic, and hyperbolic optimality systems were presented.

In case of linear elliptic problems sharp multigrid convergence estimates were obtained using local Fourier analysis. A multigrid theory that guarantees convergence of the multigrid iteration under minimal regularity assumptions was presented. These results show that collective relaxation of state and adjoint variables is essential to obtain typical multigrid convergence, which is robust with respect to the value of the weight of the control in the cost functional. This was proved to be true also in case of multigrid schemes applied to singular (nonlinear) optimal control problems using appropriate local (collective) Newton steps for smoothing. Based on this experience a collective projected iterative scheme was proposed as a smoothing scheme for multigrid applied to constrained optimal control problems. The

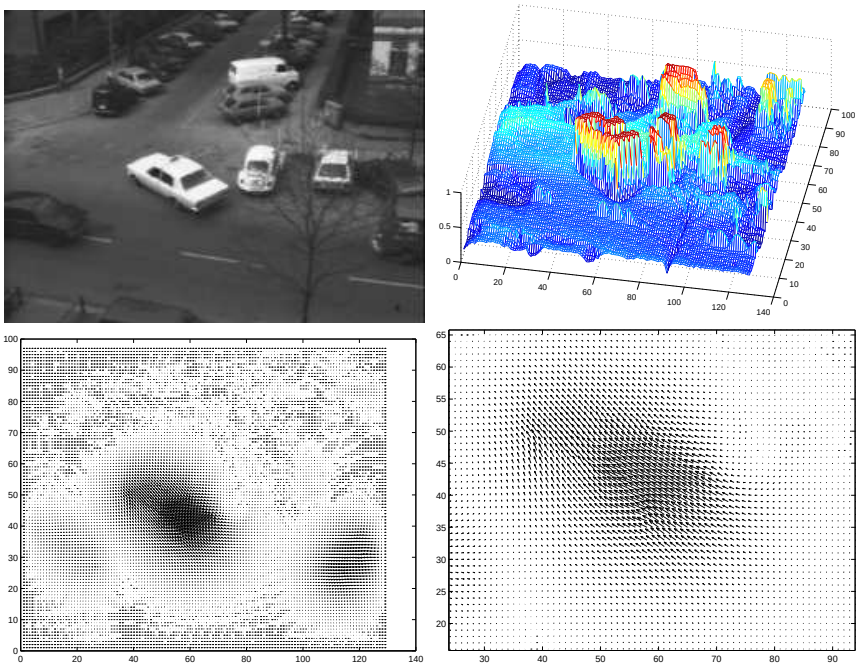

FIG. 10.2. *Top: First frame of the Taxi Sequence and the corresponding brightness distribution. Bottom: Optical flow for the Taxi sequence obtained with the optimal control method. On the right, a close-up of the solution containing the region of the taxi.*

robustness of this scheme allowed the investigation of bang-bang control phenomena. In order to enlarge the range of applicability of our multigrid strategy a new algebraic multigrid approach was formulated possessing optimal computational complexity.

Similar results were obtained in case of parabolic optimal control problems where the development of multigrid methods poses additional requirements. Parabolic optimality systems are characterized by parabolic partial differential equations with opposite orientation. To solve such problems we considered the space-time parabolic multigrid framework and developed a new smoothing scheme that allowed the fast solution of singular parabolic optimal control problems.

The results of the work presented here and in the quoted references demonstrate the ability of multigrid methods in solving optimal control problems. This fact motivates our present research activity on multigrid methods for optimality systems which focuses on the following projects.

An interesting application is eigenvalue optimization; see, e.g., [7]: Consider an elastic membrane Ω with density ρ. Important issues are the determination of the shape of Ω or of the function ρ such that the eigenfrequencies of the vibrating membrane satisfy given conditions. Therefore we need to solve eigenvalue problems with variable (possibly discontinuous) coefficients on general complex domains. For this purpose we present in [25] an efficient algebraic multigrid eigen-solver that satisfies the above requirements and represents our first step towards the development of multigrid methods for eigenvalue optimization.

Another application area of interest for us is optimal control of reaction diffusion processes [11, 19]. Time-dependent nonlinear reaction-diffusion systems are an essential ingredient in the modelling of chemical reaction processes, in the description of time evolution of ecological and biological systems, in the modelling of phase-transitions, and in the study of onset to turbulent behavior; see, e.g., [13, 45] and the references given there.

We consider representative models of reaction-diffusion phenomena provided by lambda omega systems [48]. These describe two-species reaction diffusion dynamics where the reaction kinetics exhibit periodic limit cycle behavior. Therefore, they represent a 'universal model' to investigate travelling waves, spiral waves, and turbulent evolution. In [13] we

investigate the numerical solution of lambda-omega systems by theta-schemes and multigrid methods. Using the analysis presented in this reference and the approach given in [12] we currently investigate the multigrid solution of optimal control problems governed by lambda-omega systems.

Related to our work on multigrid solution of time-dependent optimal control problems is also the investigation presented in [24] concerning optimal control of quantum systems; see, e.g., [43]. These problems require minimizing cost functionals of tracking type where the state of the system is modelled by Schrödinger equations. The control is typically realized by laser pulses which enter in the state equation as a multiplicative function of the state variable. Our purpose is to extend the multigrid approach given in [11, 12, 19] to infinite dimensional quantum models [89].

In [98] the two-dimensional inviscid incompressible Euler system is reformulated in two subsystems having elliptic and hyperbolic character, respectively. The elliptic part of the Euler system consists of the set of Cauchy-Riemann equations. Our contribution to the numerical solution of these equations can be found in [23, 106]. Based on the above results and our experience with elliptic and hyperbolic optimal control problems we intend to develop a new framework for the optimal control of incompressible fluids and its efficient numerical realization by multigrid methods.

In future work we also plan to merge multigrid techniques and globalization strategies for optimal control problems.

12. Acknowledgments. I express my gratitude to all members of the Institut für Mathematik, University of Graz, for the pleasant working atmosphere and for friendship. This work was possible also thanks to continuous stimulating discussions with Karl Kunisch and with my colleagues. It was Kunisch who introduced me to the mathematical discipline of optimal control problems.

REFERENCES

[1] N.U. Ahmed and K.L. Teo, *Optimal control of distributed parameter systems*, Elsevier, New York, 1981.
[2] E. Arian and S. Ta'asan, *Smoothers for optimization problems*, in Seventh Copper Mountain Conference on Multigrid Methods, Vol. CP3339, NASA Conference Publication, NASA, N. Duane Melson, T.A. Manteuffel, S.F. McCormick and C.C. Douglas, eds., Hampton, VA, 1995, pp. 15–30.
[3] E. Arian and S. Ta'asan, *Multigrid one shot methods for optimal control problems: Infinite dimensional control*, ICASE Report No. 94-52, 1994.
[4] H.T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.
[5] J.L. Barron, D.J. Fleet, and S.S. Beauchemin, *Performance of optical flow techniques*, IJCV, 12 (1994), pp. 43–77.
[6] R. Battiti, E. Amaldi, and C. Koch, *Computing optical flow across multiple scales: An adaptive coarse-to-fine strategy*, IJCV, 6 (1991), pp. 133–145.
[7] M.P. Bendsoe, *Optimization of Structural Topology, Shape, and Material*, Springer-Verlag, Berlin, 1995.
[8] M. Bergounioux, K. Ito, and K. Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37(4) (1999), pp. 1176–1194.
[9] L.D. Berkovitz, *Optimal Control Theory*, Springer-Verlag, 1974.
[10] D.P. Bertsekas, *Nonlinear Programming*, Athena Scientific, Belmont, 1995.
[11] A. Borzì, *Fast multigrid methods for parabolic optimal control problems*, Proceedings of the 18^{th} GAMM-Seminar Leipzig, 2002, pp. 1–10.
[12] A. Borzì, *Multigrid methods for parabolic distributed optimal control problems*, to appear in J. Comp. Appl. Math.
[13] A. Borzì, *On the numerical solution of lambda-omega systems: Theta-schemes and multigrid methods*, submitted to Numer. Math.
[14] A. Borzì and G. Borzì, *An algebraic multigrid method for a class of elliptic differential systems*, submitted to SISC.
[15] A. Borzì and G. Borzì, *An efficient algebraic multigrid method for solving optimality systems*, EMG 2002, Hohenwart Forum, 2002, submitted to Comput. Visual. in Science.
[16] A. Borzì, K. Ito, and K. Kunisch, *Optimal control formulation for determining optical flow*, SIAM J. Sci. Comp., 24(3) (2002), pp. 818-847.
[17] A. Borzì, K. Ito, and K. Kunisch, *An optimal control approach to optical flow computation*, Int. J. Numer. Meth. Fluids., 40 (2002), pp. 231-240.
[18] A. Borzì and K. Kunisch, *The numerical solution of the steady state solid fuel ignition model and its optimal control*, SIAM J. Sci. Comp., 22(1) (2000), pp. 263–284.

[19] A. Borzì and K. Kunisch, *A multigrid method for optimal control of time-dependent reaction diffusion processes*, In K.H. Hoffmann, R. Hoppe, and V. Schulz (Eds.), Fast Solution of Discretized Optimization Problems, International Series on Numerical Mathematics, Vol. 138, Birkhäuser, 2001.
[20] A. Borzì and K. Kunisch, *A multigrid scheme for elliptic constrained optimal control problems*, submitted to Comp. Opt. Appl.
[21] A. Borzì, K. Kunisch, and D.Y. Kwak, *Accuracy and convergence properties of the finite difference multigrid solution of an optimal control optimality system*, SIAM J. Control Opt., 41(5) (2003), pp. 1477-1497.
[22] A. Borzì, K. Kunisch, and M. Vanmaele, *A multi-grid approach to the optimal control of solid fuel ignition problems*, In E. Dick, K. Riemslagh and J. Vierendeels (Eds.), Lecture Notes in Computer Science and Engineering, European Multigrid Meeting 1999, Springer-Verlag, 2000.
[23] A. Borzì, K.W. Morton, E. Süli and M. Vanmaele, *Multilevel solution of cell vertex Cauchy-Riemann equations*, SIAM J. Sci. Comp., 18(2) (1997), pp. 441-459.
[24] A. Borzì, G. Stadler, and U. Hohenester, *Optimal quantum control in nanostructures: Theory and application to a generic three-level system*, Phys. Rev. A, 66 (2002), 053811.
[25] G. Borzì and A. Borzì, *Algebraic multigrid methods for solving eigenvalue problems*, EMG 2002, Hohenwart Forum, 2002, submitted to Comput. Visual. in Science.
[26] D. Braess, *Towards algebraic multigrid for elliptic problems of second order*, Computing 55, (1995), pp. 379–393.
[27] J.H. Bramble, *Error estimates for difference methods in forced vibration problems*, J. SIAM Numer. Anal., 3 (1966), pp. 1–12.
[28] J.H. Bramble, *Multigrid Methods*, Pitman Research Notes in Mathematics Series, Essex, 1993.
[29] J.H. Bramble, Do Y. Kwak, and J.E. Pasciak, *Uniform convergence of multigrid V-cycle iterations for indefinite and nonsymmetric problems*, SIAM J. Numer. Anal. 31 (1994), 1746–1763.
[30] J.H. Bramble and J.E. Pasciak, *New convergence estimates for multigrid algorithms*, Math. Comp., 49 (1987), 311–329.
[31] J.H. Bramble and J.E. Pasciak, *The analysis of smoothers for multigrid algorithms*, Math. Comp., 58 (1992), 467–488.
[32] 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), 1–34.
[33] J.H. Bramble, J.E. Pasciak, J. Wang, and J. Xu, *Convergence estimates for multigrid algorithms without regularity assumptions*, Math. Comp., 57 (1992), pp. 23–45.
[34] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.
[35] A. Brandt, *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamic*, GMD-Studien no. 85, St. Augustin, Germany, 1984.
[36] A. Brandt, *Rigorous quantitative analysis of multigrid: I. Constant coefficients two level cycle with L_2 norm*, SIAM J. Numer. Anal., 31 (1994), 1695–1730.
[37] A. Brandt, *Algebraic multigrid theory: The symmetric case*, Proc. Int. Multigrid Conf., Copper Mountain, Colorado, 1983; Appl. Math. Comp., 19 (1986), pp. 23–56.
[38] A. Brandt, *General highly algebraic coarsening*, ETNA, 10 (2000), pp. 1–20.
[39] A. Brandt and C.W. Cryer, *Multi-grid algorithms for the solution of linear complementarity problems arising from free boundary problems*, SIAM J. Sci. Stat. Comp., 4 (1983), pp. 655–684.
[40] A. Brandt, J. Greenwald *Parabolic multigrid revisited*, In W. Hackbusch and U. Trottenberg (Eds.), Multigrid Methods III, Int. Series of Numerical Mathematics, Vol. 98, Birkhäuser, Basel, 1991.
[41] M. Brezina, A. J. Cleary, R. D. Falgout, V. E. Henson, J. E. Jones, T. A. Manteuffel, S. F. McCormick, J. W. Ruge, *Algebraic multigrid based on element interpolation (AMGe)*, SIAM J. Sci. Comput., 22(5) (2000), pp. 1570–1592.
[42] R. Bulirsch, A. Miele, J. Stoer, and K.H. Well (Eds.), *Optimal Control*, International Series on Numerical Mathematics, Vol. 111, Birkhäuser, 1993.
[43] A.G. Butkovskiy and Yu.I. Samoilenko, *Control of Quantum-Mechanical Processes and Systems, Kluwer Acad. Publ., 1990.*
[44] S. H. Chou and Do Y. Kwak, *Multigrid algorithms for a vertex-centered covolume method for elliptic problems*, Numer. Math., 90(3) (2002), pp. 441–458.
[45] M. Cross and P. Hohenberg, *Pattern formation outside of equilibrium*, Rev. Mod. Phys. 65, (1993), pp. 851-1112.
[46] W. Desch, F. Kappel, and K. Kunisch (Eds.), *Control and Estimation of Distributed Parameter Systems*, International Series on Numerical Mathematics, Vol. 143, Birkhäuser, 2003.
[47] Th. Dreyer, B. Maar, V. Schulz, *Multigrid optimization in applications*, J. Comput. Appl. Math., 120 (2000), pp. 67–84.
[48] A. Duffy, K. Britton, and J. Murray, *Spiral wave solutions of practical reaction-diffusion systems*, SIAM J. Appl. Math., 39(1) (1980), pp. 8–13.
[49] H.O. Fattorini, *Infinite dimensional optimization and control theory*, Cambridge Univ. Press, 1999.
[50] R. Glowinski, H. B. Keller, and L. Rheinhart, *Continuation-conjugate gradient methods for the least-squares solution of nonlinear boundary value problems*, SIAM J. Sci. Statist. Comput., 6 (1985), pp. 793–832.
[51] H. Goldberg and F. Tröltzsch, *On a SQP–multigrid technique for nonlinear parabolic boundary control problems*, in W.W. Hager and P.M. Pardalos (Eds.), Optimal Control: Theory, Algorithms, and Applications, Kluwer Academic Publishers B.V. 1998, pp. 154–174.
[52] B.K.P. Horn and B.G. Schunck, *Determining optical flow*, AI, 17 (1981), pp. 185–204.

[53] H. Fujita, *On the nonlinear equations* $\Delta u + \exp(u) = 0$ *and* $v_t = \Delta v + \exp(u)$, Bull. Amer. Math. 75 (1968) pp. 132–135.
[54] J.E. Jones and P.S. Vassilevski, *AMGe Based on Element Agglomeration*, SIAM J. Sci. Comput., 23(1) (2001), pp. 109–133.
[55] W. Hackbusch, *Fast solution of elliptic control problems*, Journal of Optimization Theory and Application, 31 (1980), 565–581.
[56] W. Hackbusch, *A numerical method for solving parabolic equations with opposite orientations*, Computing, 20 (1978), pp. 229–240.
[57] W. Hackbusch, *Parabolic multigrid methods*, In R. Glowinski and J.-L. Lions, Computing Methods in Applied Sciences and Engineering VI, North-Holland, Amsterdam, 1984.
[58] W. Hackbusch, *On the fast solving of parabolic boundary control problems*, SIAM J. Control and Optim., 17 (1979), pp. 231–244.
[59] W. Hackbusch, *Numerical solution of linear and nonlinear parabolic optimal control problems*, Lecture Notes in Control and Information Science 30, Springer Verlag, Berlin, 1981.
[60] W. Hackbusch, *Multi-grid Methods and Applications*, Springer-Verlag, New York, 1985.
[61] W. Hackbusch, *Elliptic Differential Equations*, Springer-Verlag, New York, 1992.
[62] W. Hackbusch, *Iterative Solution of Large Sparse Systems of Equations*, Springer-Verlag, New York, 1994.
[63] W. Hackbusch and M. Griebel (Org.), *Multigrid and related methods for optimization problems*, 18^{th} GAMM-Seminar Leipzig, 2002.
[64] W. Hackbusch and G. Wittum (Org.), *Seventh European MultiGrid Conference*, Pforzheim, 2002.
[65] M. Heinkenschloss, R.H.W. Hoppe, and V. Schulz (Org.), *Numerical Techniques for Optimization Problems with PDE Constraints*, Mathematisches Forschungsinstitut Oberwolfach, 2003.
[66] K.H. Hoffmann, R. Hoppe, and V. Schulz (Eds.), *Fast Solution of Discretized Optimization Problems*, International Series on Numerical Mathematics, Vol. 138, Birkhäuser, 2001.
[67] K.H. Hoffmann, I. Lasiecka, G. Leugering, J. Sprekels, and F. Tröltzsch (Eds.), *Optimal Control of Complex Structures*, International Series on Numerical Mathematics, Vol. 139, Birkhäuser, 2002.
[68] G. Horton and S. Vandewalle, *A space-time multigrid method for parabolic partial differential equations*, SIAM J. Sci. Comput., 16(4) (1995), pp. 848–864.
[69] K. Ito and K. Kunisch, *Asymptotic properties of receding horizon optimal control problems*, SIAM J. Control Optim., 40(5) (2002), pp. 1585–1610.
[70] K. Ito and K. Kunisch, *Optimal control of the solid fuel model with* H^1*-cost*, SIAM J. Control Optim., 40(5)(2002), pp. 1455-1472.
[71] K. Ito and K. Kunisch, *BV-type regularization methods for convoluted objects with edge, flat and grey scales*, Inverse Problems, 16 (2000), pp. 909–928.
[72] A. Kauffmann, *Optimal Control of the Solid Fuel Ignition Model*, PhD Thesis, Technical University of Berlin, Berlin 1998.
[73] C.T. Kelley, *Iterative methods for optimization*, SIAM, Frontiers in applied mathematics, SIAM, Philadelphia, 1999.
[74] C.T. Kelley and E.W. Sachs, *Multilevel algorithms for constrained compact fixed point problems*, SIAM J. Sci. Stat. Comput., 15 (1994), pp. 645–667.
[75] A. Kumar and A.R. Tannenbaum, and G.J. Balas, *Optical flow: A curve evolution approach*, IEEE Trans. Image Processing, 5 (1996), pp. 598–610.
[76] Do Y. Kwak, *V-Cycle multigrid for cell-centered finite differences*, SIAM J. Sci. Comput., 11 (1999), 552–564.
[77] Do Y. Kwak, *A general multigrid framework for a class of Perturbed Problems*, presented at the AMS-IMS-SIAM Joint Summer Research Conference, Mt. Holyoke College, South Hadley, MA, 2001.
[78] R.M. Lewis and S.G. Nash, *A multigrid approach to the optimization of systems governed by differential equations*, AIAA-2000-4890, 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, (2000).
[79] J.L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, Berlin, 1971.
[80] J.L. Lions, *Control of Distributed Singular Systems*, Gauthier-Villars, Paris, 1985.
[81] D.G. Luenberger, *Optimization by vector space methods*, Wiley, New York, 1969.
[82] K. Malanowski, *Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems*, Appl. Math. Optim., 8 (1981), pp. 69–95.
[83] J. Mandel, *Local approximation estimators for algebraic multigrid*. To appear in ETNA.
[84] P. Neittaanmäki and D. Tiba, *Optimal control of nonlinear parabolic systems*, Marcel Dekker, New York, 1994.
[85] W. Niethammer, *Relaxation bei nichtsymmetrischen Matrizen* (in German), Math. Zeitschr., 85 (1964), pp. 319–327.
[86] D. Oeltz, *An algebraic multigrid method for linear elasticity*, 7th Copper Mountain Conference on Iterative Methods, Co, USA, March 2002.
[87] C.V. Pao, *Nonlinear Parabolic and Elliptic Equations*, Plenum Press, New York, 1992.
[88] C.V. Pao, *Monotone iterative methods for finite difference reaction diffusion equations*, Numer. Math., 46 (1985), pp. 571–586.
[89] A.P. Peirce, M.A. Dahleh, and H. Rabitz, *Optimal control of quantum-mechanical systems: Existence, numerical approximation, and applications*, Phys. Rev. A, 37 (1988), pp. 4950–4964.
[90] A. Rösch, *Error estimates for linear-quadratic control problems with control constraints*, submitted

to Optimization Methods and Software.
[91] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268.
[92] J.W. Ruge and K. Stüben, *Algebraic Multigrid (AMG)*, In "Multigrid Methods" (S. McCormick, ed.), Frontiers in Applied Mathematics, Vol. 5, SIAM, Philadelphia 1986.
[93] Y. Saad, *SPARSKIT: A basic tool kit for sparse matrix computations*, Rep. No. 90-20, Research Institute for Advanced Computer Science, NASA Ames Research Center, Moffet Field, CA, 1990.
[94] D.H. Sattinger, *Monotone methods in nonlinear elliptic and parabolic boundary value problems*, Indiana Univ. Math. Journal, 21 (1972), pp. 979–1000.
[95] V. Schulz and G. Wittum, *Multigrid optimization methods for stationary parameter identification problems in groundwater flow*. In W. Hackbusch, G. Wittum (eds.): Multigrid Methods V, Lecture Notes in Computational Science and Engineering 3, pp. 276-288, Springer, 1998.
[96] K. Stüben, *Algebraic Multigrid (AMG): An Introduction with Applications*, GMD Report 53, March 1999.
[97] E. Süli, *Convergence of finite volume schemes for Poisson's equation on nonuniform meshes*, SIAM J. Numer. Anal., 28 (1991), 1419–1430.
[98] S. Ta'asan, *Optimal multigrid method for inviscid flows*. In P.W. Hemker and P. Wesseling (Eds.), Multigrid methods, IV (Amsterdam, 1993), 309–320, Internat. Ser. Numer. Math., 116, Birkhäuser, Basel, 1994.
[99] S. Ta'asan, *Inverse Design and Optimisation Methods*, von Karman Institute Lecture Notes, VKI LS 1997-05.
[100] U. Trottenberg, C. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.
[101] S. Vandewalle and G. Horton, *Fourier mode analysis of the multigrid waveform relaxation and time-parallel multigrid methods*, Computing, 54(4) (1995), pp. 317–330.
[102] S. Vandewalle and R. Piessens, *Efficient parallel algorithms for solving initial-boundary value and time-periodic parabolic partial differential equations*, SIAM J. Sci. Stat. Comput., 13(6) (1992), pp. 1330–1346.
[103] S. Vandewalle, *Parallel Multigrid Waveform Relaxation for Parabolic Problems*, B.G. Teubner Stuttgart, 1993.
[104] P. Vanek, M. Brezina, and Jan Mandel, *Convergence of algebraic multigrid based on smoothed aggregation*, Numer. Math., 88(3) (2001), pp. 559–579.
[105] S. Vanka, *Block-implicit multigrid calculation of two-dimensional recirculating flows*, Comp. Meth. Appl. Mech. Eng., 59(1) (1986), 29–48.
[106] M. Vanmaele, K.W. Morton, E. Süli, and A. Borzì, *Analysis of the cell vertex finite volume method for the Cauchy-Riemann equations*, SIAM J. Num. Anal., 34(5) (1997), pp. 2043-2062.
[107] J. Weickert and C. Schnörr, *Variational optic flow computation with a spatio-temporal smoothness constraint*, J. Math. Imaging and Vision, 14 (2001), pp. 245–255.
[108] J. Weickert, *On discontinuity-preserving optical flow*. In S. Orphanoudakis, P. Trahanias, J. Crowley, N. Katevas (Eds.), Proc. Computer Vision and Mobile Robotics Workshop (CVMR '98, Santorini, 1998), pp. 115-122, 1998.