MULTIGRID METHODS FOR PDE OPTIMIZATION

ALFIO BORZÌ[*] AND VOLKER SCHULZ[†]

Abstract. Research on multigrid methods for optimization problems is reviewed. Optimization problems include shape design, parameter optimization, and optimal control problems governed by partial differential equations of elliptic, parabolic, and hyperbolic type.

Key words. Shape optimization, optimal control problems, partial differential equations, multigrid methods.

AMS subject classifications. 35J55, 35K57, 35L45, 49K20, 65M55, 65N55.

1. Introduction. Multigrid methods and optimization related to partial differential equations (PDE) are two modern fields of research in applied mathematics, both starting in the early seventies with the works of Brandt [36] and Hackbusch [55], and Lions [81], respectively. While research on multigrid methods has been motivated by the need of efficiently solving large application problems, the purpose of optimization has been to define ways of how optimally change or influence real world systems to meet a given target. Eventually, this requires to realize optimization strategies in current practical applications having increasing complexity, that is, to solve large-scale optimization problems in an accurate and computationally efficient way. For this reason, the idea of using multigrid strategies for optimization purposes is receiving increasing attention encouraged by results obtained in the last few years.

It is the purpose of this paper to review some recent efforts and outline present developments in the field of multigrid methods for PDE optimization. The organization of the paper reflects, on the one hand, the classification of optimization problems usual made in the optimization community and, on the other hand, the different ways of how multigrid methods are being used in computational optimization.

An important class of problems in optimization results from optimal control applications. These consist of an evolutionary or equilibrium system including a control mechanism and a functional modeling the purpose of the control. Other important classes of optimization problems are shape design, topology, and parameter optimization. In all these cases the underlying system is given by a set of partial differential equations; see [94] for an exceptional instance.

From the point of view of the multigrid approach, we can roughly distinguish between the direct (one-shot) multigrid method where the optimization problem is implemented within the hierarchy of grid levels and the use of multigrid schemes as inner solvers within an outer optimization loop. We also review recent trends to formulate optimization schemes where the multigrid method defines the outer solver.

In the following section we define optimization problems with PDE constraints using the terminology and notation as usual in the PDE optimization community. The notions of reduced objective functional and Hessian, Lagrangian functions, optimality systems, etc., are introduced. The section is completed by a detailed derivation of the optimality system characterizing a representative optimization problem. In Section 3, examples of steady and time-dependent optimization problems are given. Section 4 provides a self-contained introduction to multigrid methods for linear and nonlinear problems and introduces the related notation. Thereafter a section on Multigrid sequential quadratic programming (SQP) schemes is given. In this case, multigrid schemes are used as inner solvers of given sub-problems. A more direct approach is to apply the multigrid strategy to the solution of the optimality system by defining this system on the entire grid hierarchy. The essential component required to define this solution process is the formulation of appropriate smoothers. We discuss the Schur complement based multigrid smoothers and the collective smoothing

---

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

[†]Department of Mathematics, University of Trier, Universitätsring 15, 54286 Trier, Germany (`volker.schulz@uni-trier.de`).

multigrid approach in Sections 6-7 and 8. In these sections, steady and time-dependent optimization problems are considered and details of implementation of the smoothing schemes are given.

In Section 8.5 a globalization technique that exploits multigrid features is presented. In Section 9 we review recent research on multigrid optimization methods where the multigrid strategy defines the outer optimization iteration. This approach appears to be a promising research field leading to reinterpreting multigrid algorithms from an optimization point of view. A section of conclusions completes this paper.

2. Optimization problems with PDE constraints. Almost all optimization problems to be treated with multigrid methods deal with a (set of) PDE $e(y, u) = 0$ whose solution $y \in Y$ characterizes the state of a system of interest and which can be influenced by certain degrees of freedom $u \in U$. These degrees of freedom should be adjusted in order to reach a certain aim in the form of an objective J to be minimized. Thus we treat the constrained minimization problem

$$(2.1) \qquad\qquad \min J(y, u) \qquad J : Y \times U \to \mathbb{R}$$
$$(2.2) \qquad\qquad \text{s.t.} \quad e(y, u) = 0 \quad \text{where } e_y^{-1} \text{ exists.}$$

Typically, there have to be added further equality and inequality constraints in practical settings. The requirement of the existence of e_y^{-1} enables a clear distinction of variables as states y on the one hand and optimization variables u on the other hand. Furthermore, this requirement also clearly separates this problem class from variational problems, like the classical variational reformulation of elliptic PDE, where the application of multigrid methods is meanwhile common sense.

The dependent state variables y are from a suitable function space. The independent optimization variable u can be a function or a finite dimensional vector. In the latter case, a multigrid structure for u is typically not constructed. An exception occurs, when u can be seen as a finite dimensional projection from a function space, which is performed for manufacturing or regularization reasons. An example, where a multigrid structure is built up also in the finite dimensional case can be found in [105].

If the control u is finite dimensional, the problem under consideration is a parameter optimization problem. If $\dim(u)$ is comparatively small, then these problems can be most efficiently solved by reduced SQP, e.g., [98, 48], where multigrid methods are used for the solution of linear forward and adjoint subproblems, which is discussed in more detail in section 5.

The remainder of the paper is mainly devoted to optimization problems where the independent optimization variable u is from a function space. Often, this leads to ill-conditioning, which has to be coped with by appropriate regularization techniques (regularization by finite dimensional projection, by adding Tichonov-like terms in the objective, by premature stopping of iterations,...). Nevertheless, the resulting problem may not necessarily suffer from ill-conditioning. The determination, which situation occurs in a certain problem, requires some analysis of the reduced Hessian, which can also be viewed as the Hessian operator of the mapping $u \mapsto J(y(u), u)$ with intermediate step in the form

$$u \overset{\text{IFT}}{\mapsto} y(u) \mapsto J(y(u), u),$$

where "IFT" means the application of the implicit function theorem.

The application of the implicit function theorem leads to so-called "black-box" methods. A halfway application of the implicit function theorem leads to Schur-complement-type multigrid methods as considered in section 6. These methods are of particular advantage, if one wants to implement multigrid optimization approaches within a previous existing multigrid code for the state equation. Another direct approach to optimization problems is the collective smoothing multigrid strategy following the one-shot strategy first proposed in [3, 4, 104]. A one-shot multigrid scheme means solving the first-order conditions for a minimum, giving the optimality system for the state, the adjoint, and the control variables, in parallel in the multigrid process.

In the remaining part of this section we provide a detailed derivation of the optimality system characterizing the solution of an optimal control problem. The formulation of optimal

control of systems governed by partial differential equations requires the following terms: 1) The definition of a control (optimization) function u that represents the driving influence of the environment on the system. 2) The set of partial differential equations modeling the controlled system, represented by the state function $y(u)$. 3) The cost functional (objective) that models the purpose of the control on the system.

We denote by u the control function belonging to a 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 $||\cdot||$. The state y 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 $e_y(y, u) \, y'(u, \delta u) + e_u(y, u) \, \delta u = 0$. We assume that second derivative of $u \to y(u)$ is zero.

The cost functional is formally given by

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

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 the reduced cost functional $\hat{J}(u) = J(y(u), u)$.

The optimal control problem is formulated as follows: Find $\hat{u} \in U_{ad}$ such that

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

Of large interest are cost functionals of tracking type as the following

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

where $z \in Z$ is the given target 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 8.1.)

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

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

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

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

where $\hat{y} = y(\hat{u})$.

Introduce $\hat{p} \in Y$ as the unique solution to

$$e_y^*(\hat{y}, \hat{u})\hat{p} = -(\hat{y} - z)$$

where $e_y^* : Y \to Y^*$. $\hat{p}$ is the Lagrange multiplier, also called the adjoint variable. Then via $e_y(y, u) \, y'(u, \delta u) + e_u(y, u) \, \delta u = 0$ and $\delta u = v - \hat{u}$, we have

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

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

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

(2.5)
$$\begin{aligned} e(y, u) &= 0, \\ e_y^*(y, u)p &= -(y - z), \\ -((e_u^* p, v - u)) + \nu(u, v - u) &\geq 0, \end{aligned}$$

for all $v \in U_{ad}$. In this system, the first and second equations are referred to as the state equation and adjoint equation, respectively. The third equation is called the optimality condition.

Another rigorous way to derive the optimality system is also by introducing the Lagrangian function

$$L(y, u, p) = J(y, u) + \langle e(y, u), p \rangle_{Y^*, Y}.$$

By formally equating to zero the Frechét derivatives of L with respect to the triple (y, u, p), we obtain the optimality system; see, e.g., [86].

In general, solutions of optimality systems represent only extremal points and additional second-order conditions must be satisfied to guarantee that they are the minima sought. If J and e are twice continuously differentiable, the second-order sufficient conditions for a minimum require to satisfy the optimality system and the following

$$(2.6) \qquad L_{xx}(y, u, p)(v, v) \geq c_1 \, ||v||^2, \quad c_1 > 0, \text{ for all } v \in \mathcal{N}(e'(y, u)),$$

where $x = (y, u)$ and e' represents the linearized equality constraint; see, e.g., [48]. We assume that the null space $\mathcal{N}(e'(y, u))$ can be represented by $\mathcal{N}(e'(y, u)) = T(y, u) \, U$, where

$$T(y, u) = \left[\begin{array}{c} -e_y^{-1} \, e_u \\ I_u \end{array} \right],$$

and e_y, e_u are evaluated at (y, u). Therefore condition (2.6) becomes

$$(2.7) \qquad H(y, u, p)(w, w) \geq c_2 \, ||w||^2, \quad c_2 > 0,$$

for all $w \in U$. The operator H is the reduced Hessian defined by

$$H(y, u, p) = T(y, u)^* \, L_{xx}(y, u, p) \, T(y, u).$$

That is, H is given by

$$(2.8) \qquad H(y, u, p) = L_{uu}(y, u, p) + C(y, u)^* \, L_{yy}(y, u, p) \, C(y, u),$$

where $C(y, u) = e_y(y, u)^{-1} \, e_u(y, u)$, assuming $e_{yu}(y, u) = 0$ and $J_{yu}(y, u) = 0$.

Notice that H is symmetric. Therefore condition (2.7) requires that, in order to have a minimum, all eigenvalues of the reduced Hessian be positive. This fact is further discussed in Section 8.5 in order to formulate a globalization step within the one-shot multigrid approach.

3. Examples of optimization problems. In this section some representative optimization problems are formulated and the corresponding optimality systems are given. The multigrid solution of these problems is discussed in detail in the sections that follow. Assume that $\Omega \subset \mathbb{R}^2$ is convex or $\partial\Omega$ is $C^{1,1}$ smooth.

Example I: An elliptic distributed optimal control problem.

$$(3.1) \qquad \left\{ \begin{array}{rcll} \min_{u \in U_{ad}} J(y, u), & & & \\ -\Delta y & = & u + g & \text{in } \Omega, \\ y & = & 0 & \text{on } \partial\Omega. \end{array} \right.$$

We assume that the set of admissible controls is the closed convex subset of $L^2(\Omega)$ given by

$$(3.2) \qquad U_{ad} = \{ u \in L^2(\Omega) \, | \, \underline{u}(\mathbf{x}) \leq u(\mathbf{x}) \leq \overline{u}(\mathbf{x}) \text{ a.e. in } \Omega \},$$

where $\underline{u}$ and $\overline{u}$ are elements of $L^\infty(\Omega)$. The objective $J(y, u)$ is given by

$$(3.3) \qquad J(y, u) = \frac{1}{2} ||y - z||^2_{L^2(\Omega)} + \frac{\nu}{2} ||u||^2_{L^2(\Omega)},$$

where $\nu > 0$ and $g, z \in L^2(\Omega)$. In this case u is a distributed control on all of Ω.

The solution of the optimization problem given above is characterized by the following optimality system

$$
(3.4) \quad
\begin{aligned}
-\Delta y &= u + g & &\text{in } \Omega, \\
y &= 0 & &\text{on } \partial\Omega, \\
-\Delta p &= -(y - z) & &\text{in } \Omega, \\
p &= 0 & &\text{on } \partial\Omega, \\
(\nu u - p, \, v - u) &\geq 0 & &\text{for all } v \in U_{ad}.
\end{aligned}
$$

An example of elliptic boundary optimal control problem with constraints is given by the following. Minimize

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

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

$$
(3.6) \qquad
\begin{aligned}
-\Delta y + y &= g & &\text{in } \Omega, \\
\frac{\partial y}{\partial n} &= u & &\text{on } \partial\Omega.
\end{aligned}
$$

In this case, a possible set of admissible controls is given by

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

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

For the existence of a unique solution to (3.5)-(3.7) we refer to [81]. The solution is characterized by the following optimality system

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

Example II: A parabolic distributed optimal control problem.

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

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

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

where $z \in L^2(Q)$. The optimality system related to this problem is given by

$$
(3.11) \quad
\begin{aligned}
-\partial_t y + \Delta y &= u & &\text{in } Q, \\
y &= 0 & &\text{on } \Sigma, \\
\partial_t p + \Delta p &= -(y - z) & &\text{in } Q, \\
p &= 0 & &\text{on } \Sigma, \\
\nu u - p &= 0 & &\text{in } Q,
\end{aligned}
$$

with initial condition $y(\mathbf{x}, 0) = y_0(\mathbf{x})$ and terminal condition $p(\mathbf{x}, T) = 0$. In fact, notice that the state variable evolves forward in time and the adjoint variable evolves backward in time. A more general formulation of this problem is given in Section 8.2.

Example III: An hyperbolic optimal control problem (optical flow problem) [19, 20].

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

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

$$(3.13) \qquad J(y, \vec{w}) = \frac{1}{2} \int_\Omega \sum_{k=1}^N |y(x_1, x_2, t_k) - Y_k|^2 d\Omega$$

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

Here, it is required that the evolving state y matches as close as possible the sequence $\{Y_k\}_{k=0,N}$ of given functions 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

$$y_t + \vec{w} \cdot \nabla y = 0,$$
$$y(\cdot, 0) = Y_0,$$
$$p_t + \nabla \cdot (\vec{w} p) = \sum_{k=1}^{N-1} [\delta(t - t_k)(y(\cdot, t_k) - Y_k)],$$
$$p(\cdot, T) = -(y(\cdot, T) - Y_N),$$
$$(3.14) \qquad \alpha \frac{\partial^2 u}{\partial t^2} + \beta \nabla \cdot \nabla u + \gamma \frac{\partial}{\partial x_1}(\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_1},$$
$$\alpha \frac{\partial^2 v}{\partial t^2} + \beta \nabla \cdot \nabla v + \gamma \frac{\partial}{\partial x_2}(\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_2},$$

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.

Example IV: An elliptic boundary control problem as a simplified shape optimization problem. We consider a region $\Omega = (0, 1)^2$ and imagine the boundary with $\Gamma_1 := [0, 1] \times \{0\}$ being an outer boundary of a body above Γ_1 whose shape should be optimized so that the flow along the boundary minimizes a difference to a certain pressure distribution. We model the flow in Ω by potential flow and treat only small variations in the geometry. The control u models orthogonal boundary variations. According to [92], we can formulate the following boundary control problem

$$(3.15) \qquad \begin{cases} \min_{u \in H^1(\Gamma_1)} \frac{1}{2} \int_{\Gamma_1} (\frac{\partial y}{\partial x} - P(x))^2 dx, \\ \qquad\qquad -\Delta y = 0 \quad \text{in } \Omega, \\ \qquad\qquad \frac{\partial y}{\partial \vec{n}} = \frac{\partial u}{\partial x} \quad \text{on } \Gamma_1, \\ \qquad\qquad y = 0 \quad \text{on } \partial\Omega \backslash \Gamma_1, \end{cases}$$

The pressure function $P : \Gamma_1 \to \mathbb{R}$ is given data. Notice that in this case u is defined only on Γ_1.

The corresponding optimality system is given by

$$(3.16) \qquad \begin{array}{rcll} -\Delta y & = & 0 & \text{in } \Omega, \\ \frac{\partial y}{\partial \vec{n}} & = & \frac{\partial u}{\partial x} & \text{on } \Gamma_1, \\ y & = & 0 & \text{on } \partial\Omega \backslash \Gamma_1, \\ -\Delta p & = & 0 & \text{in } \Omega, \\ \frac{\partial p}{\partial \vec{n}} & = & -\frac{\partial}{\partial x}(\frac{\partial y}{\partial x} - P(x)) & \text{on } \Gamma_1, \\ p & = & 0 & \text{on } \partial\Omega \backslash \Gamma_1. \end{array}$$

A Fourier analysis computing the symbol of the reduced Hessian shows that the reduced Hessian is a differential operator of second order. Thus, the problem is well posed, although there is no regularization term present.

4. Multigrid methods. Motivation for pursuing the multigrid strategy [30, 37, 61, 106] to solving optimization problems is the achievable optimal computational complexity together with robustness with respect to chosen values of optimization parameters [3, 21,

22, 25, 26, 15, 16, 17, 18, 48, 56, 100]. To illustrate these facts, to make this presentation self-contained, and introduce the related notation, we give in the following an introduction to multigrid methods.

Let us index operators and variables defined on a grid with mesh size $h = h_k$, $k = 1, \dots, L$, where the index k is the level number. Here L denotes the finest level. On each level, a problem of the form

$$(4.1) \qquad \qquad \mathcal{A}_k \mathbf{w}_k = \mathbf{f}_k,$$

is considered where $\mathcal{A}_k$ results from the discretization of a system of coupled partial differential equations for the vector of dependent variables $\mathbf{w}_k = (u_k, v_k, ..)$.

In the case of linear multigrid methods it is convenient to utilize the fact that the solution of (4.1) is equivalent to solving $\mathcal{A}_k \mathbf{w}_k^e = \mathbf{r}_k$, where $\mathbf{w}_k^e = \bar{\mathbf{w}}_k - \mathbf{w}_k$ 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

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

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 $\mathbf{w}_k^{(m_1)} = \mathcal{S}_k^{m_1}(\mathbf{w}_k, \mathbf{f}_k)$, where $\mathcal{S}_k^{m_1}$ is a linear smoothing operator applied m_1 times. For example $\mathcal{S}_k$ can be a pointwise (collective) Gauss-Seidel iteration [106] or a Vanka smoother [110]. One sweep of this iteration is written in the form: $\mathbf{w}_k^{(m_1)} = \mathbf{w}_k^{(m_1-1)} + \mathcal{R}_k \left(\mathbf{f}_k - \mathcal{A}_k \mathbf{w}_k^{(m_1-1)} \right)$ 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,

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

where $\mathbf{w}_{k-1}$ represents on the coarse grid Ω_{k-1} the error $\mathbf{w}_k^e$ on the next finer grid. The operator $\mathcal{I}_k^{k-1} : \mathcal{V}_k \to \mathcal{V}_{k-1}$ 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:

$$(4.4) \qquad \qquad \mathbf{w}_k^{new} = \mathbf{w}_k + \mathcal{I}_{k-1}^k \mathbf{w}_{k-1},$$

where $\mathcal{I}_{k-1}^k : \mathcal{V}_{k-1} \to \mathcal{V}_k$ 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 $\mathbf{w}_k$ that was 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.

ALGORITHM 1 (**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 $\mathbf{f} \in \mathcal{V}_k$ and $\mathbf{q}^0 = 0$.

    1. Set $\mathbf{w}^0 = \tilde{\mathbf{w}} \in \mathcal{V}_k$ (Starting approximation).

    2. Define $\mathbf{w}^l$ for $l = 1, \ldots, m_1$ by

$$\mathbf{w}^l = \mathbf{w}^{l-1} + \mathcal{R}_k(\mathbf{f} - \mathcal{A}_k\,\mathbf{w}^{l-1}).$$

    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

$$\mathbf{q}^i = \mathbf{q}^{i-1} + \mathcal{M}_{k-1}[\mathcal{I}_k^{k-1}(\mathbf{f} - \mathcal{A}_k\,\mathbf{w}^{m_1}) - \mathcal{A}_{k-1}\mathbf{q}^{i-1}].$$

    4. Set $\mathcal{M}_k\mathbf{f} = \mathbf{w}^{m_1+m_2+1}$ 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 [106]. 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 multi-grid algorithm described above must be modified. Among the class of nonlinear multigrid methods [65, 113], the most popular approach is the *full approximation scheme* (FAS) [36]. 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 $\mathbf{w}_{k-1} = \hat{\mathcal{I}}_k^{k-1}\mathbf{w}_k + \mathbf{w}_{k-1}^e$ as the coarse-grid function.

To describe the FAS method consider the following nonlinear discrete problem

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

As in the linear case we need to define a smoothing procedure also denoted by $\mathcal{S}_k$. This may be, for example, a (local) Newton-Gauss-Seidel scheme. Suppose to apply m_1-times this iteration to (4.5) starting with the current approximation $\mathbf{w}_k^{(0)}$ to obtain the approximate solution $\tilde{\mathbf{w}}_k$. Now, the desired correction $\mathbf{w}_k^e$ to $\tilde{\mathbf{w}}_k$ is defined by $\mathcal{A}_k(\tilde{\mathbf{w}}_k + \mathbf{w}_k^e) = \mathbf{f}_k$. This correction can be defined as the solution to

$$\mathcal{A}_k(\tilde{\mathbf{w}}_k + \mathbf{w}_k^e) - \mathcal{A}_k(\tilde{\mathbf{w}}_k) = r_k,$$

where $r_k = \mathbf{f}_k - \mathcal{A}_k(\tilde{\mathbf{w}}_k)$ is the residual associated to $\tilde{\mathbf{w}}_k$.

Next, assume to represent the problem (4.5) on the coarser grid Ω_{k-1}. To represent $\tilde{\mathbf{w}}_k + \mathbf{w}_k^e$ on the coarse grid we write

$$\mathbf{w}_{k-1} = \hat{\mathcal{I}}_k^{k-1}\tilde{\mathbf{w}}_k + \mathbf{w}_{k-1}^e.$$

Now to formulate our problem on the coarse grid, replace $\mathcal{A}_k(\cdot)$ by $\mathcal{A}_{k-1}(\cdot)$, $\tilde{\mathbf{w}}_k$ by $\hat{\mathcal{I}}_k^{k-1}\tilde{\mathbf{w}}_k$, and r_k by $\mathcal{I}_k^{k-1}r_k$. Here $\mathcal{I}_h^H$ is a conveniently chosen restriction operator. We get the following FAS equation

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

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

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

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

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

Assuming that the smoothing iteration has reduced the high frequency components of the error on the finer grid, 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.

ALGORITHM 2 (**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 $\mathbf{f} \in \mathcal{V}_k$ and $\mathbf{q}^0 = 0$.*

1. Set $\mathbf{w}^0 = \tilde{\mathbf{w}} \in \mathcal{V}_k$ (starting approximation).

2. Define $\mathbf{w}^l$ for $l = 1, \ldots, m_1$ by

$$\mathbf{w}^l = \mathbf{w}^{l-1} + \mathcal{R}_k(\mathbf{f} - \mathcal{A}_k\,\mathbf{w}^{l-1}).$$

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

$$\mathbf{q}^i = \mathbf{q}^{i-1} + \mathcal{M}_{k-1}[\mathcal{I}_k^{k-1}(\mathbf{f} - \mathcal{A}_k\,\mathbf{w}^{m_1}) + \mathcal{A}_{k-1}(\hat{\mathcal{I}}_k^{k-1}\mathbf{w}^{m_1}) - \mathcal{A}_{k-1}\mathbf{q}^{i-1}].$$

4. Set $\mathcal{M}_k\mathbf{f} = \mathbf{w}^{m_1+m_2+1}$ 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 this nested iteration technique with the multigrid scheme is called the full multigrid (FMG) scheme. For the sake of completeness, it should be mentioned that there exist also nonlinear variants of the correction scheme due to Hackbusch [61].

For problems which are highly non-regular (i.e. discontinuous coefficients, singularities, etc.) and in the case where complex computational domains are considered, the design of geometric multigrid schemes is more involved. In these cases more sophisticated techniques are required to design intergrid transfer operators and smoothers. To resolve this difficulty Algebraic multigrid (AMG) solvers were developed [27, 39, 40, 43, 95, 102]. AMG schemes resemble the geometric multigrid process described above by utilizing only information contained in the algebraic system to be solved. To define a hierarchy of algebraic problems, the AMG algorithm starts by first applying a splitting procedure to select a subset of coarse (C) points among all fine (F) points representing the unknown variables. Based on this splitting a restriction operator $\mathcal{I}_k^{k-1}$ is constructed which has full rank. This immediately defines an interpolation operator $\mathcal{I}_{k-1}^k = (\mathcal{I}_k^{k-1})^T$, from the set of coarse points to the set of fine points, and a coarse matrix of coefficients $\mathcal{A}_{k-1}$ is given by the Galerkin formula

$$\mathcal{A}_{k-1} = \mathcal{I}_k^{k-1}\,\mathcal{A}_k\,\mathcal{I}_{k-1}^k.$$

It can be proved that if $\mathcal{A}_k$ is symmetric and positive definite and $\mathcal{I}_k^{k-1}$ is any restriction operator having full rank, then the AMG scheme is a convergent iteration. For a detailed review of AMG see [102].

In the sections that follow, we describe various multigrid approaches to optimization problems.

5. Multigrid aspects in reduced SQP approaches to parameter optimization. We call optimization problems of the form (2.1)–(2.2) parameter optimization problems, if the space U for the influence variables is finite dimensional, i.e. $u \in U = \mathbb{R}^{n_u}$, and there is no further multigrid structure conceived within U. Nevertheless, the constraint (2.2) is thought of as a PDE. This situation happens frequently in simulation-based optimization problems as, e.g., parameter identification problems or shape optimization with an a-priori defined shape parametrization. Typically, the number of parameters, i.e. n_u, is low— often below 100. In this framework, the method of choice is a reduced SQP approach as demonstrated in [98, 97, 48]. Some detail has to be discussed, if the PDE in (2.2) is solved by a multigrid method. A straight forward implementation of a reduced SQP method within the separability framework applied to problem (2.1)–(2.2) iterates over the following steps:

(0) initialize $\ell = 0, y_0, u_0$

(1) solve the adjoint problem $e_y^*(y_\ell, u_\ell)p_\ell = -J_y(y_\ell, u_\ell)$
and build the reduced gradient $\gamma_\ell = J_u^\top + e_u^*(y_\ell, u_\ell)p_\ell$

(2) build some approximation $B_\ell \approx H(y_\ell, u_\ell, p_\ell)$ of the reduced Hessian, e.g., by Quasi-Newton update formulae

(3) solve $\Delta u = \arg \min\limits_{u \in \mathcal{L}U(u_\ell)} \frac{1}{2}u^\top B_\ell u + \gamma_\ell^\top u$, where $\mathcal{L}U(u_\ell)$ denotes the linearization of U in u_ℓ

(4) solve the linear problem $e_y(y_\ell, u_\ell)\Delta y = -(e_u(y_\ell, u_\ell)\Delta u + e(y_\ell, u_\ell))$

(5) update $(y_{\ell+1}, u_{\ell+1}) = (y_\ell, u_\ell) + \tau \cdot (\Delta y, \Delta u)$, where τ is some line-search updating factor in the early iterations.

In steps (4) and (1), linear systems have to be solved, in our context by application of multigrid methods. Since these systems are adjoint to each other, the natural question arises whether this fact should be reflected in the respective multigrid solvers, as well. This fact is particularly important, if the linear systems are not solved exactly but only up to some numerical error left over after finitely many iterations. Classical reduced SQP convergence theory can only guarantee convergence if we know that the reduced gradient γ can be interpreted as a derivative, i.e., we need the consistency condition

$$(5.1) \qquad \gamma_\ell^\top = \frac{\partial}{\partial u}J(y_\ell - Ae_u(y_\ell, u_\ell)u, u_\ell + u)$$

where $A \approx e_y(y_\ell, u_\ell)^{-1}$ is the approximation to $e_y(y_\ell, u_\ell)^{-1}$ defined by the multigrid algorithm for the forward problem. If we do not satisfy this consistency condition, the necessary condition $\gamma_\ell \approx 0$ for optimality cannot be used as an indication for the optimal solution and therefore as a stopping rule for the iterations, if we only want to use a comparatively coarse accuracy in the forward solver. Then, this inconsistent γ might even give a direction, which is not a descent direction.

In [97] a proof is given for the fact that the condition (5.1) leads to the following requirements for the construction of grid transfer operators and the smoothing operator

$$_{\mathcal{A}}\mathcal{I}_k^{k-1} = \left(_{\mathcal{F}}\mathcal{I}_{k-1}^k \right)^*, \quad _{\mathcal{A}}\mathcal{S} = (_{\mathcal{F}}\mathcal{S})^*, \quad _{\mathcal{A}}\mathcal{I}_{k-1}^k = \left(_{\mathcal{F}}\mathcal{I}_k^{k-1} \right)^*$$

where $_{\mathcal{A}}\mathcal{I}$ and $_{\mathcal{F}}\mathcal{I}$ with indices mean the transfer operators for the adjoint and forward problems, respectively. Here $_{\mathcal{A}}\mathcal{S}$ and $_{\mathcal{F}}\mathcal{S}$ represent the respective smoothing operators. For efficiency reasons, the accuracy in the respective linear systems can be adapted in the style of inexact reduced SQP methods as in [68], where the accuracy of the linear subproblems is continuously increased, when zooming in to the solution. This is not required when using approximate reduced SQP methods, as in [98], where the optimization problem is reformulated so that a stagnation point of the resulting approximate algorithm is always an optimal solution, regardless whether the accuracy of the linear subproblems is increased during the nonlinear iterations or not.

Often, additional constraints are to be satisfied, either formulated in the description of the set U or as a finite number of state constraints. These additional constraints can be efficiently taken care of within the set-up above in so-called partially reduced SQP methods, as introduced in [98].

6. Schur complement based multigrid smoothers. The earliest multigrid optimization approaches and many later ones have been based on a smoothing concept which can be interpreted as a Schur-complement splitting of the KKT-matrix. Considering a Newton-iteration for the necessary conditions to problem (2.1)-(2.2), we obtain an incremental iteration where the increments in all variables

$$\mathbf{w} = \left(\begin{array}{c} \Delta y \\ \Delta u \\ \Delta p \end{array} \right)$$

are the solution to the equation

$$(6.1) \qquad \mathcal{A}\mathbf{w} = \left(\begin{array}{c} -\nabla_y L(y, u, p) \\ -\nabla_u L(y, u, p) \\ -e(y, u) \end{array} \right) =: \mathbf{f}.$$

The function $L(y, u, p)$ is the Lagrangian of the optimization problem and the operator matrix

$$(6.2) \qquad \mathcal{A} = \begin{bmatrix} L_{yy} & L_{yu} & e_y^* \\ L_{uy} & L_{uu} & e_u^* \\ e_y & e_u & 0 \end{bmatrix}$$

is the Karush-Kuhn-Tucker matrix, i.e. matrix of second order derivatives of the Lagrangian of the optimization problem. All variants of SQP methods for nonlinear problems play with variable approximations of the matrix $\mathcal{A}$ above, since the system (6.1) can be viewed as a linear-quadratic optimization problem.

For the linear-quadratic problem of Example I, the operator matrix is constant

$$\mathcal{A} = \begin{bmatrix} I & 0 & -\Delta \\ 0 & \nu \cdot I & -I \\ -\Delta & -I & 0 \end{bmatrix}$$

where I is the identity operator in the interior of the domain Ω, and Δ represents the Laplacian with Dirichlet boundary conditions (c.f. equation (3.4)).

Schur-complement smoothing approaches that can still be written in the form

$$\mathbf{w}^l = \mathbf{w}^{l-1} + \mathcal{R}(\mathbf{f} - \mathcal{A}\,\mathbf{w}^{l-1}).$$

aim at maintaining a high degree of modularity in the implementation of a multigrid optimization method. If, for example, one has a fast Poisson-solver for inverting $-\Delta$, one aims at iterative methods which use this fast solver and thus at a modular method. This is the starting point of the early multigrid optimization methods in [56]. Before going into more details, we briefly refer the basic Schur-complement approach.

A Schur-decomposition of a general 2×2-block matrix

$$K = \begin{bmatrix} A & B^\top \\ B & D \end{bmatrix}$$

with symmetric blocks A and D, and A invertible, is an explicit reformulation of a block-Gauss-decomposition, i.e.

$$K \begin{bmatrix} I & -A^{-1}B^\top \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ B & S \end{bmatrix}$$

where $S = D - BA^{-1}B^\top$ is the so-called Schur-complement. Obviously, in Schur-complement approaches, on needs the inverses of the blocks A and S or at least approximations of them, thus defining iterative methods rather than factorization methods. Iterative Schur-complement solvers are based on the scheme

$$(6.3) \qquad \mathbf{w}^l = \mathbf{w}^{l-1} + \begin{bmatrix} I & -\tilde{A}^{-1}B^\top \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{A} & 0 \\ B & \tilde{S} \end{bmatrix}^{-1} (\mathbf{f} - K\,\mathbf{w}^{l-1}).$$

where $\tilde{A}$ and $\tilde{S}$ are approximations to A and S.

If we want to employ this technique, we first have to match the blocks in the 3×3-matrix $\mathcal{A}$ (6.2) with the blocks in the 2×2-matrix K. A possible approach is the identification

$$A = \begin{bmatrix} L_{yy} & L_{yu} \\ L_{uy} & L_{uu} \end{bmatrix}$$

and B and D are chosen accordingly. The factorization is a so-called range space factorization. In many cases, the A-block thus defined may not be invertible which is a limiting factor for the method. Therefore, this arrangement is not well suited for PDE constrained optimization problems, unlike to variational problems like Stokes or Navier-Stokes [28, 112].

Interchanging the 2nd and 3rd row and column in the matrix $\mathcal{A}$ and identifying

$$A = \begin{bmatrix} L_{yy} & e_y^* \\ e_y & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} L_{uy} & e_u^* \end{bmatrix}, \qquad D = L_{uu}$$

leads to a so-called nullspace decomposition. In the iterative version of this approach, the A and S blocks are inverted only approximatively. With this decomposition the Schur-complement reads

$$S = L_{uu} - L_{uy}e_y^{-1}e_u - e_u^* e_y^{*-1}L_{yu} + e_y^{-1}e_u L_{yy}e_u^* e_y^{*-1}$$

that is also the otherwise called reduced Hessian that characterizes the optimization problem. Coercivity of the reduced Hessian guarantees well-posedness of the overall optimization problem.

For the purpose of illustration, we elaborate on the above expression in the case of Example I. The reduced Hessian is

$$S = \nu \cdot I - 0 \cdot \Delta^{-1} \cdot I - \cdot I \cdot \Delta^{-1} \cdot 0 + I \cdot \Delta^{-1} \cdot I \cdot \Delta^{-1} \cdot I = \nu \cdot I + (\Delta^{-1})^2$$

which is the compact operator $(\Delta^{-1})^2$ perturbed by $\nu \cdot I$. In [56], Hackbusch uses this insight in proposing a multigrid smoother for integral Fredholm operators of the second kind, operating on the controls u only.

$$u^l = \frac{1}{\nu}(\gamma - (\Delta^{-1})^2\, u^{l-1}) = u^{l-1} + \frac{1}{\nu}\left(\gamma - S\, u^{l-1}\right).$$

where $\gamma = \Delta^{-1}z$ is the reduced gradient at zero. In the nullspace Schur-complement setting, this corresponds to choosing

$$\tilde{A} = \begin{bmatrix} 0 & e_y^* \\ e_y & 0 \end{bmatrix}, \qquad \tilde{S} = \nu \cdot I$$

Notice that because in Example I we have $L_{uy} = 0$ and $L_{yu} = 0$, certain terms in the iteration (6.3) vanish, so that indeed only one exact solution with e_y and one exact solution with e_y^* is to be performed. Here, a fast Poisson-solver at hand is used. But still, the forward system and the adjoint system are solved exactly. Similar ideas lead to the generalization to parabolic optimal control problems in [59, 2].

If one wants to save effort and so does not perform a full solution of the forward and adjoint systems in each smoothing step but rather successive smoothing steps for the forward and adjoint system, the resulting iteration (6.3) is not longer a smoothing step of a multigrid method of second kind. In [99, 84, 100, 48] iteration (6.3) is interpreted as a transforming smoothing iteration and successfully applied in various practical problems. The choices for the algorithmic blocks are

$$\tilde{A} = \begin{bmatrix} L_{yy} & \tilde{e}_y^* \\ \tilde{e}_y & 0 \end{bmatrix} \qquad \text{and} \qquad \tilde{S} = \nu \cdot I$$

where $\tilde{e}_y$ is some approximation to e_y useful for smoothing, e.g., just the diagonal of e_y in Jacobi-smoothing. The block $\tilde{S}$ mostly consists of the regularizing part but in numerical experiments it has been shown that a deterioration of the algorithmic performance for $\nu \to 0$ can be avoided by a small number of cg-iterations for the Schur-complement system. The same iteration is used as an iterative solver in [71] and in [54] as a preconditioner for Krylov methods for the optimality conditions.

Each smoothing step of the approximate null-space iterations for the solution of system (6.1) runs through the following steps:

1) compute defects $\begin{pmatrix} d_a \\ d_d \\ d_e \end{pmatrix} := \mathcal{A}\mathbf{w}^l - \mathbf{f}$

2) $\tilde{d}_e := \tilde{e}_y^{-1}d_e$

3) $\tilde{d}_a := \tilde{e}_y^{-*}(d_a - L_{yy}\tilde{d}_e)$

4) $\Delta u := -\tilde{S}^{-1}(d_d + L_{uy}^*\tilde{d}_e + e_u^*\tilde{d}_a)$

5) $\Delta y := \tilde{d}_e + \tilde{e}_y^{-1}e_u\Delta u$

6) $\Delta\lambda := \tilde{d}_a + \tilde{e}_y^{-*}(L_{yu}\Delta u - L_{yy}\Delta y)$

7) $\mathbf{w}^{l+1} = \mathbf{w}^l + \begin{pmatrix} \Delta y \\ \Delta u \\ \Delta \lambda \end{pmatrix}$

Convergence of the resulting multigrid methods is shown in [101].

The Schur complement based smoothing approaches discussed above decouple the smoothing of the forward and adjoint equation from the smoothing of the design equation or Schur complement equation. In this way, the smoothing algorithm for the overall optimization system is still a highly modular algorithm. The price for this high degree of modularity is the necessity to deal with the null-space Schur complement. Typically, only the easily accessible part stemming from regularization is used as an approximation, i.e., $\tilde{S} = L_{uu}$. If the regularization parameter (here ν) tends towards zero, this strategy runs into trouble. In [2], this problem is resolved by a more refined analysis of the reduced Hessian for inverse problems [47]. In Example I, additional accuracy with respect to the Schur complement system

$$\bar{S}u^l = -\gamma, \qquad \text{where} \qquad \bar{S} = \nu \cdot I + e_u^*(\tilde{c}_y^*)^{-1}\tilde{c}_y^{-1}e_u$$

can be achieved by a small number of conjugate gradient steps as demonstrated in [101]. It should be noted, that this approximate Schur complement is formed with the approximations $\tilde{c}_y \approx e_y$ which are cheaply inverted because they are used for smoothing the forward and the adjoint system.

A variation of the null-space Schur complement iteration is presented in [78, 67], where the $D = L_{uu}$-block is used as a pivot instead of the A-block. The resulting Schur complement is then

$$S = \begin{bmatrix} L_{yy} & e_y^* \\ e_y & 0 \end{bmatrix} - \begin{bmatrix} L_{yu}L_{uu}^{-1}L_{uy} & L_{yu}L_{uu}^{-1}e_u^* \\ e_u L_{uu}^{-1}L_{uy} & e_u L_{uu}^{-1}e_u^* \end{bmatrix}$$

In particular in cases similar to Example I, where $L_{yu} = 0$ and $L_{uu} = \nu \cdot I$, we see that

$$S = \begin{bmatrix} L_{yy} & e_y^* \\ e_y & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & \frac{1}{\nu}e_u e_u^* \end{bmatrix}$$

This means that

$$\begin{bmatrix} L_{yy} & e_y^* \\ e_y & 0 \end{bmatrix}^{-1} \cdot S = I - \text{compact operator}$$

which makes the above mentioned multigrid methods of second kind applicable. So far, this has been used only in the form of two-level cascadic methods.

Schur complement multigrid approaches for parabolic problems have not been widely used. They can be found, e.g., in [60], where the forward and the adjoint systems are solved exactly. Schur complement approaches are advantageous in hyperbolic optimization problems because of the non-diffusive nature of the forward problem, where an exact solver can be provided. This is shown in more detail in the next section in the form of an application to optical flow problems.

7. Multigrid for the optimal control in the coefficients of an hyperbolic problem. Multigrid methods for optimal control problems governed by hyperbolic systems are a much less investigated field of research, partly because multigrid methods aren't a natural choice for the solution of hyperbolic equations. Therefore, it is natural to use multigrid in a decoupled form as discussed in the previous section on Schur complement approaches. In fact, in many cases an elliptic nullspace Schur complement results, making the use of multigrid methods appropriate.

In this context, an interesting application problem results from the optimal control formulation of the optical flow problem as proposed in [19, 20]. An optical flow is the field of apparent velocities in a sequence of images; see [6, 53, 111]. 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.

The forward model is based on the assumptions 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. This approach leads to the following *optical flow constraint* (OFC) equation

$$(7.1) \qquad \frac{\partial y}{\partial t} + u \frac{\partial y}{\partial x_1} + v \frac{\partial y}{\partial x_2} = 0,$$

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.

Now consider a sequence of image frames $\{Y_k\}_{k=0,N}$ on Ω sampled at increasing time steps, $t_k \in [0, T]$, $k = 0, 2, \ldots, N$, where $t_0 = 0$ and $t_N = T$. In the optimal control formulation we require to estimate $\vec{w}$ such that the resulting $y(\cdot, t_k, \vec{w})$ approximates Y_k at the sampling times. This means to solve

$$(7.2) \qquad \begin{cases} y_t + \vec{w} \cdot \nabla y = 0, & \text{in } Q = \Omega \times (0, T], \\ y(\cdot, 0) = Y_0, \end{cases}$$

and minimize the cost functional

$$(7.3) \quad J(y, \vec{w}) = \frac{1}{2} \int_\Omega \sum_{k=1}^{N} |y(x_1, x_2, t_k) - Y_k|^2 d\Omega$$
$$+ \frac{\alpha}{2} \int_Q \Phi(|\frac{\partial \vec{w}}{\partial t}|^2) dq + \frac{\beta}{2} \int_Q \Psi(|\nabla u|^2 + |\nabla v|^2) dq + \frac{\gamma}{2} \int_Q |\nabla \cdot \vec{w}|^2 dq.$$

Here, α, β, and γ are predefined nonnegative weights. The term with Φ provides bounded variation type regularization across edges and corners of images, where $\nabla \vec{w}$ is large; see, e.g., [74]. The last term in (7.3) improves the filling-in properties of the optimal control solution; see [19, 19] for details.

The optimal solution is characterized by the following optimality system

$$y_t + \vec{w} \cdot \nabla y = 0, \text{ with } y(\cdot, 0) = Y_0,$$

$$p_t + \nabla \cdot (\vec{w}p) = \sum_{k=1}^{N-1} \left[\delta(t - t_k)(y(\cdot, t_k) - Y_k) \right], \text{ with } p(\cdot, T) = -(y(\cdot, T) - Y_N),$$

$$(7.4) \qquad \alpha \frac{\partial^2 u}{\partial t^2} + \beta \nabla \cdot [\Psi'(|\nabla u|^2 + |\nabla v|^2) \nabla u] + \gamma \frac{\partial}{\partial x_1} (\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_1},$$

$$\alpha \frac{\partial^2 v}{\partial t^2} + \beta \nabla \cdot [\Psi'(|\nabla u|^2 + |\nabla v|^2) \nabla v] + \gamma \frac{\partial}{\partial x_2} (\nabla \cdot \vec{w}) = p \frac{\partial y}{\partial x_2},$$

where δ denotes the Dirac δ-function. The interpretation of the second equation in (7.4) is

$$(7.5) \qquad p_t + \nabla \cdot (\vec{w}p) = 0, \text{ on } t \in (t_{k-1}, t_k), \text{ for } k = 1, \ldots, N,$$
$$(7.6) \qquad p(\cdot, t_k^+) - p(\cdot, t_k^-) = y(\cdot, t_k) - Y_k, \text{ for } k = 1, \ldots N - 1.$$

The last two equations are nonlinear elliptic equations representing the optimality condition. As boundary conditions for $\vec{w}$ one can choose homogeneous Dirichlet boundary conditions on the spatial boundary and natural boundary conditions at the temporal boundaries of Q.

To solve (7.4), an explicit time-marching second-order TVD scheme for the forward-backward hyperbolic subsystem and a FAS multigrid method for the elliptic control equations is proposed in [19, 20]. The proposed method is summarized as follows

Loop for solving the optimal control problem (7.4).
1. Apply the Horn & Schunck [53] 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.

Results presented in [19, 20] show that the optimal control approach allows accurate and robust determination of optical flow also in the limit case where only two image frames are given.

8. The collective smoothing multigrid approach. A collective smoothing multigrid (CSMG) approach means solving the optimality system for the state, the adjoint, and the control variables simultaneously in the multigrid process by using collective smoothers for the optimizations variables. The CSMG approach is in contrast to the sequential solving of the state, the adjoint, and control equations. This approach requires that the uncontrolled state equations be solvable, and thus it cannot be applied to singular optimal control problems [82] where the uncontrolled system may not have a solution or blowup in finite time. A CSMG based scheme aims at realizing the tight coupling in the optimality system along the hierarchy of grids. By employing collective smoothing, that is, by realizing the coupling in the optimality system at the smoothing step level, robustness and typical multigrid efficiency is achieved. This fact is sustained by results of multigrid convergence analysis [25] obtained by means of local Fourier analysis [106] and in the framework of [34].

Development strategies of collective smoothers for optimality system appears well established now also in an algebraic multigrid context [17, 18]. This is also true for problems with control- or state-constraints [13, 14, 23]. In the control-constrained case, the CSMG approach 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 [14, 23] it is possible to show the phenomenon of 'chattering control' [9] for elliptic systems which appears to be a problem which received little attention.

The results above concern multigrid methods for elliptic optimality systems. First works concerning multigrid solution of parabolic optimal control problems are [59, 60]. Recent contributions to this field are given in [15, 16, 22]. Parabolic optimality systems are characterized by a set of parabolic partial differential equations with opposite time orientation. The starting point for the recent developments is represented by space-time parabolic multigrid methods [58, 109] and the approach presented in [57]. 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 the smoothing scheme as proposed in [15, 22]. This iterative scheme has been successfully applied in combination with different coarsening strategies to solve singular parabolic optimal control problems. In fact, the collective smoothing scheme proposed in [14, 15, 16, 22, 23] implements the coupling between state and adjoint variables at each time step, thus avoiding to solve the state equation without control. With this approach, optimal convergence factors and robustness with respect to changes of the values of the optimization parameters are obtained. These facts are confirmed by the results of the multigrid convergence analysis presented in [16]. CSMG multigrid methods for optimal control problems governed by hyperbolic partial differential equations haven't been considered yet.

In the next three subsections we provide additional details on the CSMG approach. In particular, we focus on the smoothing strategy within the nonlinear FAS multigrid framework. We start considering the case of elliptic constrained optimal control problems.

8.1. Multigrid schemes for elliptic constrained optimal control problems. We discuss the design of a collective smoothing iteration for an optimal control problem with control constraints [14, 23]. 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 the CSMG algorithm a useful tool to investigate bang-bang type control phenomena [50].

Consider Example I (3.1) where we require that the set of admissible controls is the closed convex subset of $L^2(\Omega)$ given by

$$(8.1) \qquad U_{ad} = \{u \in L^2(\omega) \,|\, \underline{u}(\mathbf{x}) \le u(\mathbf{x}) \le \overline{u}(\mathbf{x}) \text{ a.e in } \omega \subset \Omega\},$$

where $\underline{u}$ and $\overline{u}$ are elements of $L^\infty(\Omega)$ and ω is a subset of Ω.

Existence of a unique solution to (3.1)–(8.1) and its characterization are well known

[23, 81, 85]. We have that the solution is characterized by the following optimality system

(8.2)
$$
\begin{aligned}
-\Delta y &= u + g & &\text{in } \Omega, \\
y &= 0 & &\text{on } \partial\Omega, \\
-\Delta p &= -(y - z) & &\text{in } \Omega, \\
p &= 0 & &\text{on } \partial\Omega, \\
(\nu u - p,\, v - u) &\geq 0 & &\text{for all } v \in U_{ad}.
\end{aligned}
$$

Notice that the last equation in (8.2) giving the optimality condition is equivalent to (see [81, 85])

(8.3)
$$
u = \max\{\underline{u}, \min\{\overline{u}, \tfrac{1}{\nu} p(u)\}\} \text{ in } \omega, \quad \text{if } \nu > 0.
$$

The unique solution u to (3.1)–(8.1) with $\nu = 0$ satisfies (8.2) with the optimality condition given by

(8.4)
$$
p = \min\{0, p + u - \underline{u}\} + \max\{0, p + u - \overline{u}\} \quad \text{in } \Omega.
$$

To define a smoothing iteration for the present problem, consider the discrete optimality system at $\mathbf{x} \in \Omega_h$, where $\mathbf{x} = (ih, jh)$ and i, j index the grid points lexicographically. We have

(8.5)
$$
-(y_{i-1\,j} + y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1}) + 4y_{ij} - h^2\, u_{ij} = h^2 g_{ij} + h^2\, f_{ij}^{(y)},
$$

(8.6)
$$
-(p_{i-1\,j} + p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1}) + 4p_{ij} + h^2\, y_{ij} = h^2\, z_h + h^2\, f_{ij}^{(p)},
$$

(8.7)
$$
(\nu u_{ij} - p_{ij}) \cdot (v_{ij} - u_{ij}) \geq 0 \text{ for all } v_h \in U_{adh},
$$

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

A collective smoothing 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

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

and

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

Replacing these two constants in (8.5) and (8.6), we obtain y_{ij} and p_{ij} as functions of u_{ij} as follows

(8.8)
$$
y_{ij} = (e_y + h^2\, u_{ij})/4
$$

and

(8.9)
$$
p_{ij} = (4\, e_p - h^2\, e_y + 4 h^2\, z_{ij} - h^4\, u_{ij})/16.
$$

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

(8.10)
$$
\tilde{u}_{ij} = \frac{1}{16\,\nu + h^4}\,(4\, e_p - h^2\, e_y + 4 h^2\, z_{ij}).
$$

Here $\tilde{u}_{ij}$ is defined as the solution to the optimality condition equation without constraints, i.e. $\nabla \hat{J}(u) = \nu u - p(u) = 0$. Then, the new value for u_{ij} resulting from the smoothing step is given by

(8.11)
$$
u_{ij} = \begin{cases}
\overline{u}_{ij} & \text{if } \tilde{u}_{ij} > \overline{u}_{ij} \\
\tilde{u}_{ij} & \text{if } \underline{u}_{ij} \leq \tilde{u}_{ij} \leq \overline{u}_{ij} \\
\underline{u}_{ij} & \text{if } \tilde{u}_{ij} < \underline{u}_{ij}
\end{cases}
$$

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 description of the collective smoothing step. It satisfies the inequality constraint; see [23]. Further, in case $\nu = 0$ the smoothing iteration defined above satisfies (8.4). Because of (8.11) we can consider the present iteration belongs to the class of projected iterative schemes [41].

The approach described above can be easily extended to the case of boundary optimal control problem with constraints as formulated by (3.5)-(3.7) and (3.8). To implement the control on the boundary, one can approximate $\frac{\partial}{\partial n}$ by centered differences quotient and combine the Neumann boundary conditions with the stencil of the differential operator considered at the boundary. The application of the collective smoothing iteration follows along the same lines as described above; see [23]. In the presence of nonlinearities in the state equation it is not possible to solve explicitly the optimality system at the grid point level. Instead, a collective local Newton iteration can be applied, resulting in a CSMG scheme that provides convergence factors that are almost independent of ν and of the mesh size; see [21] for more details.

8.2. A space-time multigrid scheme for parabolic optimal control problems. In this section, we describe space-time multigrid schemes for the solution of parabolic optimal control problems in the whole space-time cylinder [15, 16, 22]. With this approach we are able to implement time coupling in the optimality system consisting of parabolic partial differential equations with opposite time orientation. For this purpose, appropriate collective smoothing schemes are defined. The space-time CSMG strategy results in fast solvers whose convergence factors are mesh independent and do not deteriorate as the optimization weights tends to be small.

Consider the following optimal control problem

$$(8.12) \quad \begin{cases} \min_{u \in L^2(Q)} J(y, u), \\ \quad -\partial_t y + \sigma \, \Delta y \;=\; u & \text{in } Q = \Omega \times (0, T), \\ \quad\quad\quad\quad y(\mathbf{x}, 0) \;=\; y_0(\mathbf{x}) & \text{in } \Omega \text{ at } t = 0, \\ \quad\quad\quad\quad y(\mathbf{x}, t) \;=\; 0 & \text{on } \Sigma = \partial\Omega \times (0, T), \end{cases}$$

where we take $y_0(\mathbf{x}) \in H_0^1(\Omega)$. Control may be required to track a desired trajectory given by $y_d(\mathbf{x}, t) \in L^2(Q)$ or to reach a desired terminal state $y_T(\mathbf{x}) \in L^2(\Omega)$. For this purpose we choose a cost functional of the tracking type given by

$$(8.13) \quad J(y, u) = \frac{\alpha}{2} ||y - y_d||^2_{L^2(Q)} + \frac{\beta}{2} ||y(\cdot, T) - y_T||^2_{L^2(\Omega)} + \frac{\nu}{2} ||u||^2_{L^2(Q)}.$$

Then there exists a unique solution to the optimal control problem above; see [81]. Here, $\nu > 0$ is the weight of the cost of the control and $\alpha \geq 0$, $\beta \geq 0$, $\alpha + \beta > 0$ are optimization parameters. For example, the case $\alpha = 1$, $\beta = 0$ corresponds to tracking without terminal observation.

The solution to (8.12) is characterized by the following optimality system

$$(8.14) \quad\quad\quad\quad\quad -\partial_t y + \sigma \, \Delta y = u,$$

$$(8.15) \quad\quad\quad\quad\quad \partial_t p + \sigma \, \Delta p + \alpha \, (y - z) = 0,$$

$$(8.16) \quad\quad\quad\quad\quad \nu u - p = 0,$$

with initial condition $y(\mathbf{x}, 0) = y_0(\mathbf{x})$ for the state equation (evolving forward in time) and terminal condition

$$(8.17) \quad\quad\quad\quad\quad p(\mathbf{x}, T) = \beta(y(\mathbf{x}, T) - y_T(\mathbf{x})),$$

for the adjoint equation (evolving backward in time).

Now, we discuss the design of two robust collective smoothing schemes for solving (8.14)–(8.16) discretized by finite differences and backward Euler scheme. For simplicity of illustration, we eliminate the control variable by means of the optimality condition $\nu u_h^m - p_h^m = 0$. We have

$$-[1 + 4\sigma\gamma]\, y_{i\,j\,m} + \sigma\gamma\, [y_{i+1\,j\,m} + y_{i-1\,j\,m} + y_{i\,j+1\,m} + y_{i\,j-1\,m}] + y_{i\,j\,m-1}$$

$$(8.18) \quad\quad\quad\quad - \frac{\delta t}{\nu}\, p_{i\,j\,m} = 0, \quad 2 \leq m \leq N_t + 1,$$

$$- [1 + 4\sigma\gamma]\, p_{i\,j\,m} + \sigma\gamma\, [p_{i+1\,j\,m} + p_{i-1\,j\,m} + p_{i\,j+1\,m} + p_{i\,j-1\,m}] + p_{i\,j\,m+1}$$

(8.19)
$$+ \,\delta t\,\alpha\,(y_{i\,j\,m} - y_{d\,i\,j\,m}) = 0, \quad 1 \le m \le N_t.$$

Let us define a collective iteration step which is applied at any space-time grid point to update $w_{i\,j\,m} = (y_{i\,j\,m}, p_{i\,j\,m})$. For this purpose consider (8.18) and (8.19) for the two variables $y_{i\,j\,m}$ and $p_{i\,j\,m}$ at the grid point $i\,j\,m$. We can refer to the left-hand sides of (8.18) and (8.19) as the negative of the residuals $r_y(w_{i\,j\,m})$ and $r_p(w_{i\,j\,m})$, respectively. A step of a collective smoothing iteration at this point consists of a local update given by

$$(8.20)\quad \begin{pmatrix} y \\ p \end{pmatrix}_{i\,j\,m}^{(1)} = \begin{pmatrix} y \\ p \end{pmatrix}_{i\,j\,m}^{(0)} + \begin{bmatrix} -(1 + 4\sigma\gamma) & -\delta t/\nu \\ \delta t\,\alpha & -(1 + 4\sigma\gamma) + \delta t \end{bmatrix}_{i\,j\,m}^{(0)\;-1} \begin{pmatrix} r_y \\ r_p \end{pmatrix}_{i\,j\,m}$$

where r_y and r_p denote the residuals at $i\,j\,m$ prior to the update. While a sweep of this smoothing iteration can be performed in any ordering of i, j, the problem of how to proceed along time direction arises.

To solve this problem the first vector component of (8.20) marching in the forward direction is used to update the state variable and the adjoint variable p is being updated using the second component of (8.20) marching backwards in time. In this way a robust iteration is obtained given by the following algorithm [21, 22, 15].

ALGORITHM 3. *Time-Splitted Collective Gauss-Seidel Iteration (TS-CGS)*
1. *Set the starting approximation.*
2. *For $m = 2, \ldots, N_t$ do*
3. *For ij in, e.g., lexicographic order do*

$$y_{i\,j\,m}^{(1)} = y_{i\,j\,m}^{(0)} + \frac{[-(1 + 4\sigma\gamma)]\, r_y(w) + \frac{\delta t}{\nu}\, r_p(w)}{[-(1 + 4\sigma\gamma)]^2 + \frac{\delta t^2}{\nu}\alpha}\Big|_{i\,j\,m}^{(0)},$$

$$p_{i\,j\,N_t-m+2}^{(1)} = p_{i\,j\,N_t-m+2}^{(0)} + \frac{[-(1 + 4\sigma\gamma)]\, r_p(w) - \delta t\,\alpha\, r_y(w)}{[-(1 + 4\sigma\gamma)]^2 + \frac{\delta t^2}{\nu}\alpha}\Big|_{i\,j\,N_t-m+2}^{(0)};$$

4. *end.*

The TS-CGS scheme applies with few modifications to the case of boundary control; see [12] for details. Results of local Fourier analysis show that the TS-CGS scheme has good smoothing properties, independently of the value of ν. This is also confirmed by results of numerical experiments.

In the regime of small σ (or γ), however, the TS-CGS iteration cannot provide robust smoothing because of lack of strong coupling in the space directions. To overcome this problem, block-relaxation of the variables that are strongly connected should be performed. For small σ (or γ) this means solving for the pairs of state and adjoint variables along the time-direction for each space coordinate. This type of smoothing belongs to the class of block relaxation [106].

To describe this procedure, consider the discrete optimality system (8.18)–(8.19) at any i, j and for all time steps. Thus for each spatial grid point i, j a block-tridiagonal system is obtained, where each block is a 2×2 matrix corresponding to the pair (y, p). This block-tridiagonal system has the following form

$$(8.21)\qquad M = \begin{bmatrix} A_2 & C_2 & & & & \\ B_3 & A_3 & C_3 & & & \\ & B_4 & A_4 & C_4 & & \\ & & & \ddots & & \\ & & & & & C_{N_t} \\ & & & & B_{N+1} & A_{N+1} \end{bmatrix}.$$

Centered at t_m, the entries B_m, A_m, C_m refer to the variables (y, p) at t_{m-1}, t_m, and t_{m+1}, respectively. The block A_m, $m = 2, \ldots, N_t$, is given by

$$(8.22)\qquad A_m = \begin{bmatrix} -(1 + 4\sigma\gamma) & -\frac{\delta t}{\nu} \\ \\ \delta t\,\alpha & -(1 + 4\sigma\gamma) \end{bmatrix},$$

where all functions within the brackets [ ] are evaluated at t_m. Correspondingly, the B_m and e_m blocks are given by

$$(8.23) \qquad\qquad B_m = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ and } C_m = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Clearly, for each time step, the variables neighboring the point ij are taken as constant and contribute to the right-hand side of the system.

It remains to discuss the block A_{N_t+1} for $\beta \neq 0$. At $t_m = T$, we have the terminal condition (8.17) which we rewrite as

$$\beta\left(y_h^m - y_{Th}^m\right) - p_h^m = 0, \quad m = N_t + 1.$$

Thus, the block A_{N_t+1} is given by

$$(8.24) \qquad\qquad A_{N_t+1} = \begin{bmatrix} -(1 + 4\sigma\gamma) & -\frac{\delta t}{\nu} \\ \beta & -1 \end{bmatrix}.$$

For each i, j we have to solve a tridiagonal system $Mw = r$ where $w = (y_h^2, p_h^2, \ldots, y_h^{N_t+1}, p_h^{N_t+1})$ and $r = (r_y(w^2), r_p(w^2), \ldots, r_y(w^{N_t+1}), r_p(w^{N_t+1}))$. In particular we have $r_p(w^{N_t+1}) = p_h^{N_t+1} - \beta\left(y_h^{N_t+1} - y_{Th}^{N_t+1}\right)$. Block-tridiagonal systems can be solved efficiently with $\mathcal{O}(N_t)$ effort. A block-tridiagonal solver is given in [25]. Summarizing our collective t-line relaxation is given by the following algorithm [25, 26]

ALGORITHM 4. *Time-Line Collective Gauss-Seidel Iteration (TL-CGS)*
1. *Set the starting approximation.*
2. *For ij in, e.g., lexicographic order do*

$$\begin{pmatrix} y \\ p \end{pmatrix}_{ij}^{(1)} = \begin{pmatrix} y \\ p \end{pmatrix}_{ij}^{(0)} + M^{-1} \begin{pmatrix} r_y \\ r_p \end{pmatrix}_{ij};$$

3. *end.*

Also in this case r_y and r_p denote the residuals at i, j and for all m prior to the update. Since the solution in time is exact, no time splitting is required.

8.3. Receding horizon approach. It is possible to combine multigrid schemes with receding horizon techniques [73] to develop an efficient optimal control algorithm for tracking a desired trajectory over very long time intervals. In the following, we sketch the implementation of the multigrid receding horizon scheme.

Consider the optimal control problem of tracking y_d for $t \geq 0$. Define time windows of size Δt. In each time window, an optimal control problem with tracking ($\alpha = 1$) and terminal observation ($\beta = 1$) is solved. The resulting optimal state at $n\Delta t$ defines the initial condition for the next optimal control problem defined in $(n\Delta t, (n+1)\Delta t)$ with desired terminal state given by $y_T(\mathbf{x}) = y_d(\mathbf{x}, (n+1)\Delta t)$. The following algorithm results.

ALGORITHM 5. *Multigrid Receding Horizon (MG-RH) Scheme*
1. *Set $y(\mathbf{x}, 0) = y_0(\mathbf{x})$ and $n = 0$.*
2. *Set $y_T(\mathbf{x}) = y_d(\mathbf{x}, (n+1)\Delta t)$.*
3. *CSMG Solve (8.14)–(8.16) in $(n\Delta t, (n+1)\Delta t)$.*
4. *Update $n := n + 1$, set $y_0(\mathbf{x}) = y(\mathbf{x}, n\Delta t)$ and goto 2.*

8.4. A multigrid schemes for a shape optimization problem. In Example IV, a simplified shape optimization problem is formulated as an elliptic boundary control problem. Here the shape should be optimized so that a flow along the boundary approaches a certain pressure distribution.

Let us assume that $\Omega = (0, 1)^2$ and Γ_1 consists of the boundary points with $x_2 = 0$, the

bottom boundary. The optimality system is given by

$$
\begin{aligned}
-\Delta y &= 0 && \text{in } \Omega, \\
\frac{\partial y}{\partial \bar{n}} &= \frac{\partial u}{\partial x} && \text{on } \Gamma_1, \\
y &= 0 && \text{on } \partial\Omega\backslash\Gamma_1, \\
-\Delta p &= 0 && \text{in } \Omega, \\
\frac{\partial p}{\partial \bar{n}} &= -\frac{\partial}{\partial x}\left(\frac{\partial y}{\partial x} - P(x)\right) && \text{on } \Gamma_1, \\
p &= 0 && \text{on } \partial\Omega\backslash\Gamma_1.
\end{aligned}
\tag{8.25}
$$

As shown in [104], an appropriate choice for determining the control is to set $u = \frac{\partial p}{\partial x}$.

Denote first-order backward and forward partial derivatives of v_h in the x_i direction by ∂_i^- and ∂_i^+, respectively. They are given by

$$
\partial_i^- v_h(\mathbf{x}) = \frac{v_h(\mathbf{x}) - v_h(\mathbf{x} - \hat{i}\,h)}{h} \qquad \text{and} \qquad \partial_i^+ v_h(\mathbf{x}) = \frac{v_h(\mathbf{x} + \hat{i}\,h) - v_h(\mathbf{x})}{h}
$$

where $\hat{i}$ denotes the i coordinate direction vector and v_h is extended by 0 on grid points outside of Ω; see [62]. We have the five-point Laplacian $\Delta_h = \partial_1^+\partial_1^- + \partial_2^+\partial_2^-$.

In this framework, the discretization of (8.25) gives

$$
\begin{aligned}
-\Delta_h y &= 0 && \text{in } \Omega, \\
\partial_h^n y &= \partial_1^+\partial_1^- p_h && \text{on } \Gamma_1, \\
y &= 0 && \text{on } \partial\Omega\backslash\Gamma_1, \\
-\Delta_h p &= 0 && \text{in } \Omega, \\
\partial_h^n p &= -\partial_1^+\partial_1^- y_h + P_x && \text{on } \Gamma_1, \\
p &= 0 && \text{on } \partial\Omega\backslash\Gamma_1.
\end{aligned}
\tag{8.26}
$$

where $P_x = \frac{\partial P}{\partial x}$, and $\partial_h^n v_h = -(v_{i\,j+1} - v_{i\,j-1})/2h$ at Γ_1.

Notice that to solve (8.26) we need to realize the coupling on Γ_1 while in the interior of the domain we have two Laplace equations. For this purpose we combine the Neumann boundary conditions with the stencil of the discrete elliptic operators considered at the boundary. For $\mathbf{x} = (ih,\, jh)$ being a boundary grid point on the side $x_2 = 0$ we have

$$
-(2y_{i+1\,j} + y_{i\,j-1} + y_{i\,j+1} - 4y_{i\,j}) - \frac{2}{h}(p_{i+1\,j} - 2p_{i\,j} + p_{i-1\,j}) = 0,
$$

$$
-(2p_{i+1\,j} + p_{i\,j-1} + p_{i\,j+1} - 4p_{i\,j}) + \frac{2}{h}(y_{i+1\,j} - 2y_{i\,j} + y_{i-1\,j}) = 2hP_x.
$$

The equations obtained in this way have the same structure as (8.5) and (8.6) (without constraints) and the application of the collective Gauss-Seidel iteration follows along the same lines as above. Hence we can apply the FAS multigrid method previously described. Notice that bilinear prolongation and full-weighting restriction have to be used to guarantee the right scaling for the coarse-grid problem formulation at the boundary; see the discussion in [106]. Clearly, on the boundary these transfer operators are mirrored.

To numerically validate the FAS multigrid algorithm for solving (8.26), consider the desired profile given by $P_x = 1$ Results for this case are reported in Table 8.1 for a V-cycle with two pre- and post-smoothing steps.

TABLE 8.1
Results of experiments for a shape optimization problem

Mesh	129×129	257×257	517×517
ρ	0.08	0.08	0.08

8.5. Globalization for the one-shot multigrid method. In a convex setting where the optimal control solution is unique, solving the optimality system is equivalent to solving the optimal control problem. However, in general, these solutions represent only extremal points and additional conditions must be satisfied to guarantee that they are the minima sought or otherwise escape the critical point. The starting point to formulate these conditions is to recall the second-order sufficient conditions for a minimum.

Let us recall the Lagrangian function

$$L(y, u, p) = J(y, u) + \langle e(y, u), p \rangle_{Y^*, Y}.$$

By equating to zero the Frechét derivatives of L with respect to the triple (y, u, p), we obtain the optimality system. If J and e are twice continuously differentiable, the second-order sufficient conditions for a minimum have to satisfy the optimality system and the following

$$(8.27) \qquad L_{xx}(y, u, p)(v, v) \geq c_1 \, ||v||^2, \quad c_1 > 0, \text{ for all } v \in \mathcal{N}(e'(y, u)),$$

where $x = (y, u)$ and e' represents the linearized equality constraint; see, e.g., [48]. We assume that the null space $\mathcal{N}(e'(y, u))$ can be represented by $\mathcal{N}(e'(y, u)) = T(y, u)\, U$, where

$$T(y, u) = \left[\begin{array}{c} -e_y^{-1}\, e_u \\ I_u \end{array} \right],$$

and e_y, e_u are evaluated at (y, u). Therefore condition (8.27) becomes

$$(8.28) \qquad H(y, u, p)(w, w) \geq c_2 \, ||w||^2, \quad c_2 > 0,$$

for all $w \in U$. The operator H is the reduced Hessian defined by

$$H(y, u, p) = T(y, u)^* \, L_{xx}(y, u, p)\, T(y, u).$$

That is, H is given by

$$(8.29) \qquad H(y, u, p) = L_{uu}(y, u, p) + C(y, u)^* \, L_{yy}(y, u, p)\, C(y, u),$$

where $C(y, u) = e_y(y, u)^{-1}\, e_u(y, u)$, assuming $e_{yu}(y, u) = 0$.

Notice that H is symmetric. Therefore condition (8.28) requires that, in order to have a minimum, all eigenvalues of the reduced Hessian be positive. The occurrence of nonpositive eigenvalues indicates the presence of possible maxima or saddle points.

Clearly, in an infinite dimensional setting, the analysis of the spectrum of H is not a viable task. Solving the eigenvalue problem may be computationally more expensive than solving the optimality system. However, assuming that the spectral properties of the reduced Hessian are well represented on the hierarchy of grids, we can define a globalization step based on the spectral properties of the Hessian H on the (appropriately chosen) coarsest grid.

In the case negative eigenvalues of the reduced Hessian are detected, one can use the eigenvector corresponding to the smallest eigenvalue to determine an escape direction. This direction of negative curvature [87] is given by the eigenvector corresponding to the negative eigenvalue with largest absolute value. Then the corresponding normalized eigenvector ϕ_h is used to perform the following globalization step

$$(8.30) \qquad u_h^{new} = u_h - \sigma\, \phi_h.$$

With appropriately chosen $|\sigma|$ and the sign of σ is such that $\sigma\, \phi_h \cdot (\nu g'(u_h) + e_u^* \, p_h(u_h)) \geq 0$.

To guarantee that once we escape the undesired critical point, the multigrid algorithm does not return to this critical point, the smoothing and the coarse-grid correction must produce minimizing updates; see [24]. In the next section, we review recent developments focusing on the fact that, under appropriate conditions, the coarse-grid correction provides a descent direction giving the possibility to accommodate robust globalization strategies.

9. Multigrid as outer optimization scheme. In some recent papers, Nash [88] and Lewis and Nash [80], propose a multigrid approach to optimization problems, called MG/OPT, which closely resembles the well known full approximation storage (FAS) scheme [36] and is similar to the nonlinear multigrid (NMGM) methods discussed in [61]. Extension of the multigrid strategy to optimization problems is a quite new field of research with increasing interest from optimization community; see also, e.g., the recent contribution [11, 52, 91].

In [80, 88] it is emphasized that under appropriate assumptions, the multigrid coarse-grid correction provides a descent direction and, therefore, combining this fact with a line search procedure and a minimizing 'smoothing' iteration, a globally convergent algorithm is obtained. Numerical experiments, e.g. [88], demonstrate that MG/OPT greatly improves the efficiency of the underlying optimization scheme used as 'smoother', suggesting that the MG/OPT scheme may be beneficial in combination with well known optimization algorithms. This claim appears to be true as far as a line search along the coarse-grid correction is performed. Also in [88] it is reported that MG/OPT without line search diverges in some cases. Therefore a line search appears to be necessary for convergence. Results on an a-priori coarse grid correction step-length, which guarantees a globally convergent MG/OPT iteration are given in [11].

9.1. The MG/OPT method for optimization. Consider the following (locally) convex optimization problem

$$(9.1) \qquad \min_{y_k} f_k(y_k)$$

where $k = 1, 2, \ldots, L$, is the resolution or discretization parameter, L denotes the finest resolution, and y_k is the (unconstrained) optimization variable in the space V_k. For variables defined on V_k we introduce the inner product $(\cdot, \cdot)_k$ with associated norm $\|y\|_k = (y, y)_k^{1/2}$. Among spaces V_k, restriction operators $I_k^{k-1} : V_k \to V_{k-1}$ and prolongation operators $I_{k-1}^k : V_{k-1} \to V_k$ are defined. We require that $(I_k^{k-1}y, v)_{k-1} = (y, I_{k-1}^k v)_k$ for all $y \in V_k$ and $v \in V_{k-1}$.

On each space, denote with S_k an optimization algorithm. For example the truncated Newton scheme used in [88]. Given an initial approximation y_k^0 to the solution of (9.1), the application of S_k results in $f_k(S_k(y_k^0)) < f_k(y_k^0)$.

The MG/OPT scheme is an iterative method. One cycle of this method is defined as follows. Let y_k^0 be the starting approximation at resolution k.

ALGORITHM 6 (**MG/OPT (k) Algorithm**). *If $k = 1$ (coarsest resolution) solve (9.1)*

exactly.
Else if $k > 1$:
  1. Pre-optimization. Define $y_k^1 = S_k(y_k^0)$.
  2. setup and solve a coarse-grid minimization problem. Define $y_{k-1}^1 = I_k^{k-1} y_k^1$ and $\tau_{k-1} = \nabla f_{k-1}(y_{k-1}^1) - I_k^{k-1} \nabla f_k(y_k^1)$. The coarse-grid minimization problem is given by

$$(9.2) \qquad \min_{y_{k-1}} \left(f_{k-1}(y_{k-1}) - \tau_{k-1}^T y_{k-1} \right).$$

  Apply one cycle of MG/OPT(k-1) to (9.2) to obtain y_{k-1}^2.
  3. Line-search and coarse-grid correction. Perform a line search in the $I_{k-1}^k(y_{k-1}^2 - y_{k-1}^1)$ direction to obtain α_k. The coarse-grid correction is given by

$$y_k^2 = y_k^1 + \alpha_k \, I_{k-1}^k (y_{k-1}^2 - I_k^{k-1} y_k^1)$$

  4. Post-optimization. Define $y_k^3 = S_k(y_k^2)$.

Roughly speaking, the essential guideline for constructing f_k on coarse levels is that it must sufficiently well approximate the convexity properties of the functional at finest resolution. In addition we have that the gradient of the coarse-grid functional at $y_{k-1}^1 = I_k^{k-1} y_k^1$ equals the restriction of the gradient of the fine-grid functional at y_k^1. In fact

REMARK 1. By adding the term $-\tau_{k-1}^T y_{k-1}$ in Step 2. we have that

$$\nabla \left(f_{k-1}(y_{k-1}) - \tau_{k-1}^T y_{k-1} \right) |_{y_{k-1}^1} = I_k^{k-1} \nabla f_k(y_k^1).$$

The following lemma states that the coarse-grid correction provides a descent direction.

LEMMA 9.1. *Take* $y_k \in V_k$ *and define* $\tilde{y}_{k-1} = I_k^{k-1} y_k$. *Denote with* $\hat{f}_{k-1}(y_{k-1}) = f_{k-1}(y_{k-1}) - \tau_{k-1}^T y_{k-1}$ *where* $\tau_{k-1} = \nabla f_{k-1}(I_k^{k-1} y_k) - I_k^{k-1} \nabla f_k(y_k)$. *Let* y_{k-1} *be such that* $\hat{f}_{k-1}(y_{k-1}) < \hat{f}_{k-1}(I_k^{k-1} y_k)$ *and define* $e_k = I_{k-1}^k (y_{k-1} - I_k^{k-1} y_k)$. *Then*

$$(\nabla f_k(y_k), e_k)_k < 0.$$

Proof. From

$$\hat{f}_{k-1}(y+z) = \hat{f}_{k-1}(y) + (\nabla \hat{f}_{k-1}(y), z)_{k-1} + \frac{1}{2} \int_0^1 (\nabla^2 \hat{f}_{k-1}(x+tz)z, z)_{k-1} \, dt$$

we obtain

$$(\nabla \hat{f}_{k-1}(y_{k-1}), y_{k-1} - I_k^{k-1} y_k)_{k-1} \le \hat{f}_{k-1}(y_{k-1}) - \hat{f}_{k-1}(I_k^{k-1} y_k) < 0.$$

Now we have

$$(\nabla f_k(y_k), e_k)_k = (\nabla f_k(y_k), I_{k-1}^k (y_{k-1} - I_k^{k-1} y_k))_k = (I_k^{k-1} \nabla f_k(y_k), y_{k-1} - I_k^{k-1} y_k)_{k-1}$$
$$= (\nabla \hat{f}_{k-1}(I_k^{k-1} y_k), y_{k-1} - I_k^{k-1} y_k)_{k-1} < 0.$$

For the last equality recall Remark 1. □

10. Conclusions. In this paper we presented a review on multigrid methods for PDE optimization, emphasizing new methodological achievements exploiting the multigrid strategy beyond its use as a solver module. To better illustrate these advances in computational optimization, we gave a detailed description of some multigrid optimization techniques. In particular, we discussed Schur-complement and collective-smoothing based multigrid schemes. On the other hand, we outlined recent approaches to optimization where the multigrid strategy is the guideline to the design of outer loop optimization procedures. These are all quite new concepts with very promising perspectives in computation and optimization disciplines. This review is timely, considering the numerous scientific meetings and granted projects on the field of scientific computing and optimization.

REFERENCES

[1] N.U. AHMED AND K.L. TEO, *Optimal control of distributed parameter systems*, Elsevier, New York, 1981.

[2] V. AKCELIK, G. BIROS, A. DRAGANESCU, O. GHATTAS, J. HILL, AND B. VAN BLOEMEN WAANDERS, *Dynamic data-driven inversion for terascale simulations: real-time identification of airborne contaminants*, ACM/IEEE SCXY conference series, 2005

[3] 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.

[4] E. ARIAN AND S. TA'ASAN, *Multigrid one shot methods for optimal control problems: Infinite dimensional control*, ICASE Report No. 94-52, 1994.

[5] H.T. BANKS AND K. KUNISCH, *Estimation Techniques for Distributed Parameter Systems*, Birkhäuser, Boston, 1989.

[6] J.L. BARRON, D.J. FLEET, AND S.S. BEAUCHEMIN, *Performance of optical flow techniques*, IJCV, 12 (1994), pp. 43–77.

[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, *On the convergence of the MG/OPT method*, PAMM, Proceedings GAMM Annual Meeting - Luxembourg 2005, 5(1)(2005), pp. 735–736.

[12] A. BORZÌ, *Space-time multigrid methods for solving unsteady optimal control problems*, to appear in the Proceedings of the Second Sandia Workshop on PDE-Constrained Optimization: Toward Real-time and Online PDE-constrained Optimization, May 19-21, 2004, Bishop's Lodge, Santa Fe, New Mexico. Workshop organizing committee: Larry Biegler, Omar Ghattas, Matthias Heinkenschloss, David Keyes, Bart van Bloemen Waanders.

[13] A. BORZÌ, *Smoothers for control- and state-constrained optimal control problems*, to appear in Computing and Visualization in Science.

[14] A. BORZÌ, *High-order discretization and multigrid solution of elliptic nonlinear constrained optimal control problems*, to appear in Journal of Computational and Applied Mathematics.

[15] A. BORZÌ, *Fast multigrid methods for parabolic optimal control problems*, Proceedings of the 18th GAMM-Seminar Leipzig, 2002, pp. 1–10.

[16] A. BORZÌ, *Multigrid methods for parabolic distributed optimal control problems*, J. Comp. Appl. Math., 157 (2003), pp. 365–382.

[17] A. BORZÌ AND G. BORZÌ, *An algebraic multigrid method for a class of elliptic differential systems*, SIAM J. Sci. Comp., 25(1) (2003), pp. 302–323.

[18] A. BORZÌ AND G. BORZÌ, *An efficient algebraic multigrid method for solving optimality systems*, Computing and Visualization in Science, 7(3/4) (2004), pp. 183–188.

[19] A. BORZÌ, K. ITO, AND K. KUNISCH, *Optimal control formulation for determining optical flow*, SIAM J. Sci. Comp., 24(3) (2002), pp. 818-847.

[20] 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.

[21] 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.

[22] 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.

[23] A. BORZÌ AND K. KUNISCH, *A multigrid scheme for elliptic constrained optimal control problems*, Computational Optimization and Applications, 31 (2005), pp. 309–333.

[24] A. BORZÌ AND K. KUNISCH, *A globalization strategy for the multigrid solution of elliptic optimal control problems*, Optimization Methods and Software, 21(3) (2006), 445-459.

[25] 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.

[26] 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.

[27] D. BRAESS, *Towards algebraic multigrid for elliptic problems of second order*, Computing 55, (1995), pp. 379–393.

[28] D. BRAESS, R. SARAZIN, *An efficient smoother for the Stokes problem*, Appl. Numer. Math. 23, No.1(1997), pp. 3–19

[29] J.H. BRAMBLE, *Error estimates for difference methods in forced vibration problems*, J. SIAM Numer. Anal., 3 (1966), pp. 1–12.

[30] J.H. BRAMBLE, *Multigrid Methods*, Pitman Research Notes in Mathematics Series, Essex, 1993.

[31] 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.

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

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

[34] 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.

[35] 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.

[36] A. BRANDT, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390.

[37] A. BRANDT, *Multigrid Techniques: 1984 Guide with Applications to Fluid Dynamic*, GMD-Studien no. 85, St. Augustin, Germany, 1984.

[38] 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.

[39] A. BRANDT, *Algebraic multigrid theory: The symmetric case*, Proc. Int. Multigrid Conf., Copper Mountain, Colorado, 1983; Appl. Math. Comp., 19 (1986), pp. 23–56.

[40] A. BRANDT, *General highly algebraic coarsening*, ETNA, 10 (2000), pp. 1–20.

[41] 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.

[42] 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.

[43] 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.

[44] R. BULIRSCH, A. MIELE, J. STOER, AND K.H. WELL (EDS.), *Optimal Control*, International Series on

Numerical Mathematics, Vol. 111, Birkhäuser, 1993.

[45] A.G. BUTKOVSKIY AND YU.I. SAMOILENKO, *Control of Quantum-Mechanical Processes and Systems*, Kluwer Acad. Publ., 1990.

[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] A. DRAGANESCU, *Two investigation in numerical analysis: Monotonicity preserving finite element methods, and multigrid methods for inverse parabolic problems*, PhD thesis, University of Chicago, 2004.

[48] TH. DREYER, B. MAAR, V. SCHULZ, *Multigrid optimization in applications*, J. Comput. Appl. Math., 120 (2000), pp. 67–84.

[49] H.O. FATTORINI, *Infinite dimensional optimization and control theory*, Cambridge Univ. Press, 1999.

[50] K. GLASHOFF AND E. SACHS, *On theoretical and numerical aspects of the bang-bang principle*, Numer. Math., 29 (1977), pp. 93–113.

[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] S. GRATTON, A. SARTENAER, AND PH.L. TOINT, *Second-order convergence properties of trust-region methods using incomplete curvature information, with an application to multigrid optimization*, to appear in Journal of Computational and Applied Mathematics.

[53] B.K.P. HORN AND B.G. SCHUNCK, *Determining optical flow*, AI, 17 (1981), pp. 185–204.

[54] E. HABER, U. ASCHER AND D. OLDENBURG, *Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach*, Geophysics 69 (2004), pp. 1216–1228.

[55] W. HACKBUSCH, On the multi-grid method applied to difference equations, Computing, 20 (1978), pp. 291–306.

[56] W. HACKBUSCH, *Fast solution of elliptic control problems*, Journal of Optimization Theory and Application, 31 (1980), 565–581.

[57] W. HACKBUSCH, *A numerical method for solving parabolic equations with opposite orientations*, Computing, 20 (1978), pp. 229–240.

[58] W. HACKBUSCH, *Parabolic multigrid methods*, In R. Glowinski and J.-L. Lions, Computing Methods in Applied Sciences and Engineering VI, North-Holland, Amsterdam, 1984.

[59] W. HACKBUSCH, *On the fast solving of parabolic boundary control problems*, SIAM J. Control and Optim., 17 (1979), pp. 231–244.

[60] W. HACKBUSCH, *Numerical solution of linear and nonlinear parabolic optimal control problems*, Lecture Notes in Control and Information Science 30, Springer Verlag, Berlin, 1981.

[61] W. HACKBUSCH, *Multi-grid Methods and Applications*, Springer-Verlag, New York, 1985.

[62] W. HACKBUSCH, *Elliptic Differential Equations*, Springer-Verlag, New York, 1992.

[63] W. HACKBUSCH, *Iterative Solution of Large Sparse Systems of Equations*, Springer-Verlag, New York, 1994.

[64] W. HACKBUSCH AND M. GRIEBEL (ORG.), *Multigrid and related methods for optimization problems*, 18th GAMM-Seminar Leipzig, 2002.

[65] W. HACKBUSCH AND G. WITTUM (ORG.), *Seventh European MultiGrid Conference*, Pforzheim, 2002.

[66] M. HEINKENSCHLOSS, R.H.W. HOPPE, AND V. SCHULZ (ORG.), *Numerical Techniques for Optimization Problems with PDE Constraints*, Mathematisches Forschungsinstitut Oberwolfach, 2003.

[67] M. HEINKENSCHLOSS, *The numerical solution of a control problem governed by a phase field model*, Journal of Optimization Theory and Applications, Vol. 94 (1997), pp. 237–309.

[68] M. HEINKENSCHLOSS AND L.N. VICENTE, *Analysis of Inexact Trust-Region SQP Algorithms*, SIAM J. Optimization, Vol. 12 (2001), No. 2, pp. 283-302.

[69] K.H. HOFFMANN, R. HOPPE, AND V. SCHULZ (EDS.), *Fast Solution of Discretized Optimization Problems*, International Series on Numerical Mathematics, Vol. 138, Birkhäuser, 2001.

[70] 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.

[71] R.H.W. HOPPE, S.I. PETROVA AND V.H. SCHULZ, *A primal-dual Newton-type interior-point method for topology optimization*, J.Optim.Theory Appl., Vol.114, No.3 (2002), pp. 545–571.

[72] 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.

[73] K. ITO AND K. KUNISCH, *Asymptotic properties of receding horizon optimal control problems*, SIAM J. Control Optim., 40(5) (2002), pp. 1585–1610.

[74] 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.

[75] 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.

[76] A. KAUFFMANN, *Optimal Control of the Solid Fuel Ignition Model*, PhD Thesis, Technical University of Berlin, Berlin 1998.

[77] C.T. KELLEY, *Iterative methods for optimization*, SIAM, Frontiers in applied mathematics, SIAM, Philadelphia, 1999.

[78] C.T. KELLEY AND E.W. SACHS, *Fast algorithms for compact fixed point problems with inexact function evaluations*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 725–742.

[79] 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.

[80] R.M. LEWIS AND S.G. NASH, *A multigrid approach to the optimization of systems governed by dif-*

ferential equations, AIAA-2000-4890, 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, (2000).

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

[82] J.L. LIONS, *Control of Distributed Singular Systems*, Gauthier-Villars, Paris, 1985.

[83] D.G. LUENBERGER, *Optimization by vector space methods*, Wiley, New York, 1969.

[84] B. MAAR AND V. SCHULZ, *Interior Point Multigrid Methods for Topology Optimization*, Structural Optimization, vol. 19, No. 3 (2000), pp. 214–224.

[85] K. MALANOWSKI, *Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems*, Appl. Math. Optim., 8 (1981), pp. 69–95.

[86] H. MAURER AND J. ZOWE, *First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems*, Mathematical Programming, 16, (1979) pp. 98–110, .

[87] S.G. NASH, *Newton-type minimization via the Lanczos method*, SIAM J. Numer. Anal., 21(4) (1984), pp. 770–788.

[88] S.G. Nash, A multigrid approach to discretized optimization problems, Optimization Methods and Software, 14 (2000), pp. 99–116.

[89] P. NEITTAANMÄKI AND D. TIBA, *Optimal control of nonlinear parabolic systems*, Marcel Dekker, New York, 1994.

[90] D. OELTZ, *An algebraic multigrid method for linear elasticity*, 7th Copper Mountain Conference on Iterative Methods, Co, USA, March 2002.

[91] S. OH, C.A. BOUMAN, AND K.J. WEBB, *Multigrid Tomographic Inversion with Variable Resolution Data and Image Spaces*, IEEE Trans. on Image Processing, vol. 15(9) (2006), pp. 2805–2819.

[92] O. PIRONNEAU, Optimal Shape Design for Elliptic Systems, Springer, 1984.

[93] A. RÖSCH, *Error estimates for linear-quadratic control problems with control constraints*, submitted to Optimization Methods and Software.

[94] H. REGLER AND U. RUEDE, *Layout optimization with algebraic multigrid methods (AMG)*, In Proceedings of the Sixth Copper Mountain Conference on Multigrid Methods, Copper Mountain, April 4-9, 1993, Conference Publication. NASA, 1993.

[95] J.W. RUGE AND K. STÜBEN, *Algebraic Multigrid (AMG)*, In "Multigrid Methods" (S. McCormick, ed.), Frontiers in Applied Mathematics, Vol. 5, SIAM, Philadelphia 1986.

[96] 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.

[97] V. SCHULZ *Numerical optimization of the cross-sectional shape of turbine blades*, Z. Angew. Math. Mech. 76, Suppl. 1, 207-210 (1996)

[98] V. SCHULZ, *Solving discretized optimization problems by partially reduced SQP methods*, Comput. Vis. Sci. 1, No.2, 83-96, 1998

[99] V. SCHULZ, A. BARDOSSY, AND R. HELMIG, *Conditional Statistical Inverse Modeling in Groundwater Flow by Multigrid Methods*, Computational Geosciences 3 (1999), pp. 49–68.

[100] 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.

[101] V. SCHULZ AND G. WITTUM, *Transforming smoothers for optimization saddle point problems*. submitted to Computing and Visualization in Science, 2006.

[102] K. STÜBEN, *Algebraic Multigrid (AMG): An Introduction with Applications*, GMD Report 53, March 1999.

[103] 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.

[104] S. TA'ASAN, *Introduction to Shape Design and Control, Inverse Design and Optimisation Methods*, von Karman Institute Lecture Notes, VKI LS 1997-05.

[105] S. TA'ASAN AND G. KURUVILA AND M.D. SALAS, *Aerodynamic design and optimization in one shot*, 30th Aerospace Sciences Meeting, Reno, NV, AIAA Paper 92-0025, 1992.

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

[107] 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.

[108] 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.

[109] S. VANDEWALLE, *Parallel Multigrid Waveform Relaxation for Parabolic Problems*, B.G. Teubner Stuttgart, 1993.

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

[111] 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.

[112] G. WITTUM, *Multigrid methods for Stokes and Navier-Stokes equations. Transforming smoothers: Algorithms and numerical results*, Numer. Math. 54, No.5 (1988), pp. 543–563.

[113] I. YAVNEH, *A Multilevel Nonlinear Method*, SIAM J. Sci. Comput., 28(1) (2006), pp. 24–46.