Doctoral Program
Computational Mathematics
Numerical Analysis and Symbolic Computation

Convergence Analysis of Multigrid Methods with Collective Point Smoothers for Optimal Control Problems

Stefan Takacs and Walter Zulehner

DK-Report No. 2011-01

02 2011

A–4040 LINZ, ALTENBERGERSTRASSE 69, AUSTRIA

Supported by

Austrian Science Fund (FWF)

Upper Austria

DK sponsors:

- **Johannes Kepler University Linz** (JKU)
- **Austrian Science Fund** (FWF)
- **Upper Austria**

Convergence Analysis of Multigrid Methods with Collective Point Smoothers for Optimal Control Problems

Stefan Takacs · **Walter Zulehner**

Abstract In this paper we consider multigrid methods for solving saddle point problems. The choice of an appropriate smoothing strategy is a key issue in this case. Here we focus on the widely used class of collective point smoothers. These methods are constructed by a point-wise grouping of the unknowns leading to, e.g., collective Richardson, Jacobi or Gauss-Seidel relaxation methods. Their smoothing properties are well-understood for scalar problems in the symmetric and positive definite case. In this work the analysis of these methods is extended to a special class of saddle point problems, namely to the optimality system of optimal control problems. For elliptic distributed control problems we show that the convergence rates of multigrid methods with collective point smoothers are bounded independent of the grid size and the regularization (or cost) parameter.

Keywords Multigrid methods · Collective point smoothers · Optimal control

1 Introduction

The analysis presented in this work is discussed for the following elliptic distributed control model problem of tracking type: Minimize the cost functional J, given by

$$J(y,u) = \frac{1}{2}\|y-y_D\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)},$$

The work was supported by the Austrian Science Fund (FWF) under grant W1214/DK12.

S. Takacs
Doctoral Program Computational Mathematics, Johannes Kepler University Linz, Austria, E-mail: stefan.takacs@dk-compmath.jku.at

W. Zulehner
Institute of Computational Mathematics, Johannes Kepler University Linz, Austria, E-mail: zulehner@numa.uni-linz.ac.at

subject to the elliptic boundary value problem (BVP)

$$-\Delta y + y = u \text{ in } \Omega \qquad \text{and} \qquad \frac{\partial y}{\partial n} = 0 \text{ on } \partial\Omega, \tag{1}$$

where $y \in H^1(\Omega)$ is the state variable and $u \in L^2(\Omega)$ is the control variable. The function $y_D \in L^2(\Omega)$ is given and $\alpha > 0$ is some regularization or cost parameter. Here Ω is a bounded domain in $\mathbb{R}^d$ for $d \in \{1,2,3\}$ with Lipschitz boundary $\partial\Omega$, the sets $L^2(\Omega)$ and $H^1(\Omega)$ denote the standard Lebesgue and Sobolev spaces with associated standard norms $\|\cdot\|_{L^2(\Omega)}$ and $\|\cdot\|_{H^1(\Omega)}$, respectively.

The main goal of this work is to construct and analyze numerical methods that produce an approximate solution to the optimization problem, where the computational time can be bounded by the number of unknowns times a constant which is independent of the parameter α, in particular for small values of α.

The solution of the optimization problem is characterized by the Karush-Kuhn-Tucker system (KKT system). As we are interested in good approximations of the solution, the discretization of the KKT system leads to a large scale linear system. This linear system will be solved with multigrid methods because they are one of the fastest known methods for such problems. Originally, multigrid methods have been designed and analyzed for elliptic problems. They also work well for saddle point problems (like the KKT systems for PDE-constrained optimization and particularly optimal control problems) and have gained growing interest in this area, see, e.g., [6] and the references cited there.

The unknowns of the discretized KKT system of a PDE-constrained optimization problem can be partitioned into primal and dual variables. In optimal control problems the primal unknowns consist of state and control variables. One approach to solve such problems is to apply multigrid methods in every step of an overall block-structured iterative method to equations in just one of these blocks of variables. Such methods have been proposed, e.g., in [11], [2], [3], [5], [4], [13], [19] and [17].

Another approach, which we will follow here, is to apply the multigrid idea directly to the (reduced or not reduced) KKT system, which is called an all-at-once approach. Such methods have been proposed, e.g., in [23], [1], [24], [7], [21], [6], [14], [22] and [20].

The choice of an appropriate smoother is a key issue in constructing such a multigrid method. For elliptic problems smoothers have been constructed by solving (small) local problems in an additive or multiplicative Schwarz-type manner. If each of the local problems contains just one unknown (typically associated to a point of an underlying grid), this leads to Richardson, Jacobi and Gauss-Seidel relaxation. The extension of such ideas to the case of linear saddle point systems, which result from systems of BVPs, like the KKT systems, can be done in various ways.

One idea is based on splitting the problem into subproblems connected to scalar BVPs, where well known smoothing strategies for positive definite matrices can be applied. Uzawa-type smoothers, which for example have been analyzed in [22] and [20], can be understood in this way. Another approach is based on transforming smoothers, originally introduced in [25], [26] and discussed for optimization problems in [21].

In this paper we focus on collective iteration schemes which are constructed by solving local problems, involving the complete system of BVPs, in an additive or multiplicative Schwarz-type manner. As in the case of elliptic problems the local problems may live on patches or, as in our case, just on single points. Such methods have been proposed, e.g., in [24], [7], [6], [14].

So far, collective point iteration schemes have been analyzed mainly by using Fourier analysis on uniform grids, see, e.g., [24], [7], [6], [14]. In [7] such methods have also been analyzed for general grids based on compactness arguments under the assumption that the coarsest grid is fine enough.

In this paper we present a convergence proof for multigrid methods with special collective point smoothers for general grids, based on the classical splitting of the analysis into smoothing and approximation property, see [12].

This paper is organized as follows. The framework, basic estimates and the proposed multigrid method will be discussed in section 2. The analysis of the smoother and a convergence result are given in section 3. In section 4 we present numerical results which confirm the theoretical results and illustrate the efficiency of the method even in cases which are not covered by the convergence theory.

2 Framework and basic estimates

2.1 The Karush-Kuhn-Tucker system

In this subsection, we will derive the KKT system for the model problem. At first the BVP (1) is written in variational form: Find $y \in H^1(\Omega)$ such that

$$(y,p)_{H^1(\Omega)} = (u,p)_{L^2(\Omega)}$$

holds for all $p \in H^1(\Omega)$. Here $(\cdot,\cdot)_{L^2(\Omega)}$ and $(\cdot,\cdot)_{H^1(\Omega)}$ denote the standard inner products in $L^2(\Omega)$ and $H^1(\Omega)$, respectively. Next the Lagrange functional is introduced by

$$\mathscr{L}(y,u,p) = \frac{1}{2}\|y-y_D\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + (y,p)_{H^1(\Omega)} - (u,p)_{L^2(\Omega)}.$$

Solving the model problem is equivalent to finding a saddle point of the Lagrange functional, which leads to the first order optimality conditions (the KKT system), given by: Find $(y,u,p) \in \hat{X} = H^1(\Omega)\times L^2(\Omega)\times H^1(\Omega)$ such that

$$\begin{aligned}
(y,\widetilde{y})_{L^2(\Omega)} \qquad\qquad\qquad &+ (p,\widetilde{y})_{H^1(\Omega)} = (y_D,\widetilde{y})_{L^2(\Omega)}\\
\alpha\,(u,\widetilde{u})_{L^2(\Omega)} &- (p,\widetilde{u})_{L^2(\Omega)} = 0\\
(y,\widetilde{p})_{H^1(\Omega)} - (u,\widetilde{p})_{L^2(\Omega)} \qquad &\qquad\qquad\quad = 0
\end{aligned}$$

hold for all $(\widetilde{y},\widetilde{u},\widetilde{p}) \in \hat{X}$.

Because $\alpha\,(u,\widetilde{u})_{L^2(\Omega)} = (p,\widetilde{u})_{L^2(\Omega)}$ for all $\widetilde{u} \in L^2(\Omega)$, we obtain $u = \alpha^{-1}p$, which allows us to reduce the KKT system: Find $(y,p) \in X = Y\times P = H^1(\Omega)\times H^1(\Omega)$ such that

$$\begin{aligned}
(y,\widetilde{y})_{L^2(\Omega)} &+ \qquad (p,\widetilde{y})_{H^1(\Omega)} = (y_D,\widetilde{y})_{L^2(\Omega)}\\
(y,\widetilde{p})_{H^1(\Omega)} &- \alpha^{-1}(p,\widetilde{p})_{L^2(\Omega)} = 0
\end{aligned} \tag{2}$$

hold for all $(\widetilde{y},\widetilde{p}) \in X$. The function spaces for the state y and the adjoint state p are both $H^1(\Omega)$. Nevertheless, we use different symbols Y and P for theses spaces since they will be equipped with different norms, see below.

Obviously, this problem can also be interpreted as one single variational equation: Find $x \in X$ such that

$$\mathscr{B}(x,\widetilde{x}) = \mathscr{F}(\widetilde{x}) \tag{3}$$

holds for all $\widetilde{x} \in X$ with

$$\begin{aligned}
\mathscr{B}(x,\widetilde{x}) &= (y,\widetilde{y})_{L^2(\Omega)} + (p,\widetilde{y})_{H^1(\Omega)} + (y,\widetilde{p})_{H^1(\Omega)}\\
&\quad - \alpha^{-1}(p,\widetilde{p})_{L^2(\Omega)},\\
\mathscr{F}(\widetilde{x}) &= (y_D,\widetilde{y})_{L^2(\Omega)}
\end{aligned}$$

for $x = (y,p)$ and $\widetilde{x} = (\widetilde{y},\widetilde{p})$.

In [20] it was shown that the problem is well-posed in the norm given by

$$\|x\|_X = \|(y,p)\|_X = \left(\|y\|_Y^2 + \|p\|_P^2\right)^{1/2}, \tag{4}$$

where

$$\|y\|_Y = \left(\|y\|^2_{L^2(\Omega)} + \alpha^{1/2}\|y\|^2_{H^1(\Omega)}\right)^{1/2} \tag{5}$$

and

$$\|p\|_P = \left(\alpha^{-1}\|p\|^2_{L^2(\Omega)} + \alpha^{-1/2}\|p\|^2_{H^1(\Omega)}\right)^{1/2}, \tag{6}$$

more precisely, there are constants $\underline{C} > 0$ and $\overline{C}$ (independent of α) such that

$$\underline{C}\,\|x\|_X \le \sup_{0\neq\widetilde{x}\in X} \frac{\mathscr{B}(x,\widetilde{x})}{\|\widetilde{x}\|_X} \le \overline{C}\,\|x\|_X \tag{7}$$

holds for all $x \in X$. This implies that for every right-hand-side $\mathscr{F} \in X^*$ the problem (3) has a unique solution, which depends continuously on the data $\mathscr{F} \in X^*$. Here, X^* is the dual space of X.

The discretization is done by standard techniques. For the model problem we use a family of meshes which is obtained based on some coarsest triangular mesh (grid level $k = 0$) and uniform refinement. For $k \in \{0,\dots,K\}$ we denote the size of the largest edge of the triangulation by h_k. Due to the fact that we have uniform refinement $h_k = 2^{-k}h_0$ holds.

The space of discretized functions $X_k = Y_k \times P_k$ is constructed by the Courant element: $Y_k = P_k$ is the set of continuous and piecewise linear functions.

The discretized problem is also well posed in the norm given by (4) – (6), see [20].

Using the standard nodal basis, we can rewrite the optimality system (2) in matrix-vector notation as follows:

$$\begin{pmatrix} M_k & K_k \\ K_k & -\alpha^{-1}M_k \end{pmatrix}\begin{pmatrix} \underline{y}_k \\ \underline{p}_k \end{pmatrix} = \begin{pmatrix} \underline{g}_k \\ 0 \end{pmatrix} \tag{8}$$

with mass matrix M_k and stiffness matrix K_k. The symbols $\underline{y}_k$ and $\underline{p}_k$ denote the coefficient vectors of the corresponding functions y_k and p_k with respect to the nodal basis.

2.2 Multigrid solvers for saddle point problems

The main focus of this subsection is the construction of a multigrid method for solving saddle point problems of the following form. Find $x \in X = Y \times P$ such that

$$\mathscr{B}(x,\widetilde{x}) = \mathscr{F}(\widetilde{x}) \qquad \text{holds for all } \widetilde{x} \in X,$$

where

$$\mathscr{B}(x,\widetilde{x}) = a(y,\widetilde{y}) + b(p,\widetilde{y}) + b(y,\widetilde{p}) - \alpha^{-1}c(p,\widetilde{p})$$

with $x = (y,p)$ and $\widetilde{x} = (\widetilde{y},\widetilde{p})$. Here, Y and P are Hilbert spaces with the same set of members, a, b and c are symmetric and non-negative bilinear forms and $\mathscr{F} \in X^*$. We assume that there is a sequence of grids for $k \in \{0,\dots,K\}$ which induces a sequence of nested subspaces $Y_k = P_k$ of $Y = P$.

Using a basis, we can rewrite the discretized problem: Find $x_k \in X_k = Y_k \times P_k$ such that

$$\mathscr{B}(x_k,\widetilde{x}_k) = \mathscr{F}(\widetilde{x}_k) \qquad \text{holds for all } \widetilde{x}_k \in X_k, \tag{9}$$

in matrix-vector notation as follows:

$$\underbrace{\begin{pmatrix} A_k & B_k \\ B_k & -\alpha^{-1}C_k \end{pmatrix}}_{\mathscr{A}_k =}\underbrace{\begin{pmatrix} \underline{y}_k \\ \underline{p}_k \end{pmatrix}}_{\underline{x}_k =} = \underbrace{\begin{pmatrix} \underline{g}_k \\ \underline{q}_k \end{pmatrix}}_{\underline{f}_k =}, \tag{10}$$

where the symmetric and positive semidefinite matrices A_k, B_k and $C_k \in \mathbb{R}^{N_k\times N_k}$ represent the bilinear forms a, b and c, respectively.

Between two consecutive grid levels $k-1$ and k we need intergrid-transfer operators I^k_{k-1} and I^{k-1}_k. For the prolongation operator I^k_{k-1} we choose the matrix representation of the canonical embedding of the associated finite element subspaces, and its adjoint for the restriction operator I^{k-1}_k.

Now we can introduce the multigrid iteration for solving the discretized equation (10) on grid level k. Starting from an initial approximation $\underline{x}^{(0)}_k$, one step of the iteration is given in the following way:

- Apply ν smoothing steps:

  $$\underline{x}^{(0,m)}_k = \underline{x}^{(0,m-1)}_k + \tau\,\hat{\mathscr{A}}^{-1}_k(\underline{f}_k - \mathscr{A}_k\,\underline{x}^{(0,m-1)}_k) \tag{11}$$

  for $m \in \{1,\dots,\nu\}$ with $\underline{x}^{(0,0)}_k = \underline{x}^{(0)}_k$. The choice of τ and $\hat{\mathscr{A}}_k$ will be discussed below.

- Apply the coarse-grid correction, i.e.:
  - Compute the defect and restrict it to the coarser grid:

    $$\underline{r}^{(1)}_{k-1} = I^{k-1}_k(\underline{f}_k - \mathscr{A}_k\,\underline{x}^{(0,\nu)}_k).$$

  - Solve (approximatively) the linear system

    $$\mathscr{A}_{k-1}\underline{w}^{(1)}_{k-1} = \underline{r}^{(1)}_{k-1}, \tag{12}$$

    living on the coarser grid.
  - Prolongate the result and add it to the last iterate:

    $$\underline{x}^{(1)}_k = \underline{x}^{(0,\nu)}_k + I^k_{k-1}\underline{w}^{(1)}_{k-1}.$$

If the problem (12) is solved exactly, we obtain

$$\underline{x}^{(1)}_k = \underline{x}^{(0,\nu)}_k + I^k_{k-1}\mathscr{A}^{-1}_{k-1}I^{k-1}_k(\underline{f}_k - \mathscr{A}_k\,\underline{x}^{(0,\nu)}_k)$$

for the next iterate (two-grid method). In practice the solution of (12) is approximated by applying one step (V-cycle) or two steps (W-cycle) of the multigrid method, recursively. Just on grid level $k = 0$ the problem is solved exactly.

Next we construct the smoother (11) based on the idea of collective iteration schemes. For simplicity we concentrate on collective Jacobi relaxation. Standard Jacobi relaxation, which can be used as a smoother for a linear system $A_k \underline{x}_k = \underline{f}_k$, where $A_k \in \mathbb{R}^{N_k \times N_k}$ is symmetric and positive definite, reads as

$$x_i^{(0,m+1)} = x_i^{(0,m)} + \tau a_{ii}^{-1} \left(f_i - \sum_{j=1}^{N_k} a_{ij} x_j^{(0,m)} \right),$$

where $x_i^{(0,m)}$, f_i and a_{ij} are the components of the vectors $\underline{x}_k^{(0,m)}$ and $\underline{f}_k$ and the matrix A_k, respectively. This iteration scheme can be carried over to saddle point problems of the form (10), which leads to collective Jacobi relaxation, which reads as

$$\mathbf{x}_i^{(0,m+1)} = \mathbf{x}_i^{(0,m)} + \tau \mathscr{A}_{ii}^{-1} \left(\mathbf{f}_i - \sum_{j=1}^{N_k} \mathscr{A}_{ij} \mathbf{x}_j^{(0,m)} \right),$$

where $\mathbf{x}_i^{(0,m)} = (y_i^{(0,m)}, p_i^{(0,m)})^T$, $\mathbf{f}_i = (g_i, 0)^T$ and

$$\mathscr{A}_{ij} = \begin{pmatrix} a_{ij} & b_{ij} \\ b_{ij} & -\alpha^{-1} c_{ij} \end{pmatrix}.$$

Here $y_i^{(0,m)}$, $p_i^{(0,m)}$, g_i, a_{ij}, b_{ij} and c_{ij} are the components of $\underline{y}_k^{(0,m)}$, $\underline{p}_k^{(0,m)}$, $\underline{g}_k$, A_k, B_k and C_k, respectively.

Collective Richardson and Gauss-Seidel relaxation are constructed analogously. Of course, such iteration schemes can be represented in the compact notation (11) using the preconditioning matrix

$$\hat{\mathscr{A}}_k = \begin{pmatrix} \hat{A}_k & \hat{B}_k \\ \hat{B}_k & -\alpha^{-1} \hat{C}_k \end{pmatrix},$$

where $\hat{A}_k$, $\hat{B}_k$ and $\hat{C}_k$ are preconditioning matrices for A_k, B_k and C_k, respectively. In particular:

- In the case of collective Jacobi relaxation $\hat{A}_k$, $\hat{B}_k$ and $\hat{C}_k$ are the diagonals of A_k, B_k and C_k, respectively, and the damping parameter τ is chosen to be in $(0,1)$.
- In the case of collective Richardson relaxation we have $\hat{A}_k = a_k I$, $\hat{B}_k = b_k I$ and $\hat{C}_k = c_k I$, where for some $C > 0$

  $\frac{1}{2}\lambda_{\max}(A_k) \le a_k \le \frac{C}{2}\lambda_{\max}(A_k)$,
  $\frac{1}{2}\lambda_{\max}(B_k) \le b_k \le \frac{C}{2}\lambda_{\max}(B_k)$ and
  $\frac{1}{2}\lambda_{\max}(C_k) \le c_k \le \frac{C}{2}\lambda_{\max}(C_k)$

  holds. The damping parameter τ is chosen to be in $(0,1)$.
- In the case of collective Gauss-Seidel iteration $\hat{A}_k$, $\hat{B}_k$ and $\hat{C}_k$ are the left-lower trigonal part (including the diagonal) of A_k, B_k and C_k, respectively, and the damping parameter τ is chosen to be 1.

Each of these three iteration schemes can be realized efficiently if it is implemented analogously to standard Richardson, Jacobi or Gauss-Seidel relaxation, see e.g. [14].

2.3 Elliptic boundary control model problem

The multigrid method proposed in subsection 2.2 is not just applicable to the distributed control model problem but also to several other control problems. One particular example is the elliptic boundary control model problem which reads as follows. Minimize the functional J, given by

$$J(y,u) = \frac{1}{2}\|y - y_D\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\partial\Omega)},$$

subject to the elliptic boundary value problem

$$-\Delta y + y = 0 \text{ in } \Omega \qquad \text{and} \qquad \frac{\partial y}{\partial n} = u \text{ on } \partial\Omega,$$

where $y \in H^1(\Omega)$ is the state variable and $u \in L^2(\partial\Omega)$ is the control variable. The function $y_D \in L^2(\Omega)$ is given and $\alpha > 0$ is some regularization or cost parameter.

As it was done in subsection 2.1, we can set up the KKT system, reduce it to a 2-by-2-formulation and discretize it by standard techniques, which leads to

$$\begin{pmatrix} M_k & K_k \\ K_k & -\alpha^{-1} M_{\Gamma\Gamma,k} \end{pmatrix} \begin{pmatrix} \underline{y}_k \\ \underline{p}_k \end{pmatrix} = \begin{pmatrix} \underline{g}_k \\ 0 \end{pmatrix},$$

with $M_{\Gamma\Gamma,k} = ((\varphi_{k,i}, \varphi_{k,j})_{L^2(\partial\Omega)})_{i,j=1,\ldots,N_k}$ (boundary mass matrix). Here, $(\varphi_{k,i})_{i=1,\ldots,N_k}$ denotes the nodal basis. M_k and K_k are the standard mass and stiffness matrix, respectively.

This system fits into the general framework of subsection 2.2 with $A_k = M_k$, $B_k = K_k$ and $C_k = M_{\Gamma\Gamma,k}$. So, the same classes of collective point smoother are available for this saddle point problem. The construction of other smoothers is typically more involved. For example in [20] for the construction of an Uzawa type smoother the norm $\|\cdot\|_X$, which we have introduced in (4), was used. Unfortunately such a norm is not available for the boundary control model problem. Such insight is not necessary for constructing collective point smoothers.

3 Convergence theory

In this section we give a convergence proof for the proposed multigrid method that is based on the classical splitting of the analysis into smoothing and approximation property, introduced in [12]. Such an approach is applicable to general grids.

To achieve a robust convergence result for the problems of our interest, we have to choose an appropriately scaled L^2-like norm, say $\|\cdot\|_{0,k}$, and an associated residual norm

$$\|x_k\|_{2,k} = \sup_{0 \neq \widetilde{x}_k \in X_k} \frac{\mathscr{B}(x_k, \widetilde{x}_k)}{\|\widetilde{x}_k\|_{0,k}}. \tag{13}$$

For convergence it is sufficient to show the following two conditions:

– Smoothing property:
  There is some function η with $\lim_{\nu\to\infty}\eta(\nu) = 0$ such that

$$\|x_k^{(0,\nu)} - x_k^*\|_{2,k} \leq \eta(\nu)\|x_k^{(0)} - x_k^*\|_{0,k} \quad (14)$$

  holds for all $\nu \in \mathbb{N}$.
– Approximation property:

$$\|x_k^{(1)} - x_k^*\|_{0,k} \leq C_A\|x_k^{(0,\nu)} - x_k^*\|_{2,k} \quad (15)$$

  holds for some constant $C_A > 0$.

Here, x_k^* is the exact solution of (9). $x_k^{(n,m)}$ and $x_k^{(n)}$ are those functions in X_k which correspond to the iterates of the two-grid method.

The combination of both estimates, (14) and (15), implies that the two-grid method converges for ν large enough. Due to standard arguments the convergence of the two-grid method implies the convergence of the W-cycle multigrid method under weak assumptions. Hence analyzing smoothing and approximation property stated above, is of our particular interest.

3.1 An algebraic smoothing theorem

In theorem 1 we give an estimate that just relies on algebraic relations between the involved matrices which allows us to show the smoothing property for the elliptic distributed control model problem in the next subsection.

Theorem 1 *Consider the block-matrix $\mathscr{A}_k$, which is given by*

$$\mathscr{A}_k = \begin{pmatrix} A_k & B_k \\ B_k & -\alpha^{-1}A_k \end{pmatrix},$$

where $A_k, B_k \in \mathbb{R}^{N_k\times N_k}$ are symmetric matrices. Let the preconditioning matrix $\hat{\mathscr{A}}_k$ be given by

$$\hat{\mathscr{A}}_k = \begin{pmatrix} \hat{A}_k & \hat{B}_k \\ \hat{B}_k & -\alpha^{-1}\hat{A}_k \end{pmatrix}.$$

Here, $\hat{A}_k, \hat{B}_k \in \mathbb{R}^{N_k\times N_k}$ are preconditioning matrices such that

$$\rho(I - \hat{A}_k^{-1}A_k) \leq 1 \qquad \text{and} \qquad \rho(I - \hat{B}_k^{-1}B_k) \leq 1 \quad (16)$$

holds, where ρ denotes the spectral radius. Moreover we assume that there is a symmetric positive definite matrix $\hat{D}_k$ such that $\hat{A}_k = a_k\hat{D}_k$ and $\hat{B}_k = b_k\hat{D}_k$, where $a_k > 0$ and $b_k > 0$ are scalars.

Then, for all $\tau \in (0,1)$, there is a constant $C_S > 0$ such that

$$\|\mathscr{L}_k^{-1/2}\mathscr{A}_k(I - \tau\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\mathscr{L}_k^{-1/2}\|_{\ell^2} \leq \frac{C_S}{\sqrt{\nu}} \quad (17)$$

holds for all grid levels $k \in \{0,\dots,K\}$, for all choices of $\alpha > 0$ and for all $\nu \in \mathbb{N}$. Here, the matrix $\mathscr{L}_k$ is given by

$$\mathscr{L}_k = \begin{pmatrix} (\hat{A}_k^2 + \alpha\hat{B}_k^2)^{1/2} & \\ & \alpha^{-1}(\hat{A}_k^2 + \alpha\hat{B}_k^2)^{1/2} \end{pmatrix}. \quad (18)$$

The matrix $\mathscr{L}_k$, given in (18), is symmetric and positive definite and induces therefore a vector norm

$$\|\underline{x}_k\|_{\mathscr{L}_k} = \|\mathscr{L}_k^{1/2}\underline{x}_k\|_{\ell^2}$$

and a matrix norm

$$\|M\|_{\mathscr{L}_k} = \|\mathscr{L}_k^{1/2}M\mathscr{L}_k^{-1/2}\|_{\ell^2}.$$

The property of power boundedness is of our particular interest. We say that a matrix M is power bounded (with respect to a certain norm $\|\cdot\|$) if there is a constant C (independent of h_k and α) such that for all $\nu \in \mathbb{N}$

$$\|M^\nu\| \leq C$$

holds.

To prove theorem 1, we use a variant of Reusken's lemma, see [18] for the original work.

Lemma 1 *Let $\mathscr{L}_k$ be a symmetric positive definite matrix and let $\mathscr{M}_k$ be a matrix that is power bounded with respect to $\|\cdot\|_{\mathscr{L}_k}$. Then for every choice of the damping parameter $\tau \in (0,1)$ there is a constant C (independent of h_k and α) such that*

$$\|(I - \mathscr{M}_k)((1-\tau)I + \tau\mathscr{M}_k)^\nu\|_{\mathscr{L}_k} \leq \frac{C}{\sqrt{\nu}}$$

holds for all $\nu \in \mathbb{N}$.

Proof The proof was given in [10] for the case $\|\mathscr{M}_k\|_{\mathscr{L}_k} \leq 1$ and can easily be extended to the case that $\mathscr{M}_k$ is power bounded. □

Due to Reusken's lemma, we have to show that the iteration matrix of the (non-damped) iteration scheme is power bounded. This will be done in the next two lemmas.

Lemma 2 *Using the notations of theorem 1, the identity*

$$\|(I - \hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\|_{\mathscr{L}_k} = \|\widetilde{Z}_k^\nu\|_{\ell^2}$$

holds for all $\nu \in \mathbb{N}$, where $\widetilde{Z}_k$ is given by

$$\widetilde{Z}_k = (\hat{A}_k^2 + \alpha\hat{B}_k^2)^{1/4}Z_k(\hat{A}_k^2 + \alpha\hat{B}_k^2)^{-1/4}$$

with

$$Z_k = (\hat{A}_k + \sqrt{\alpha}\hat{B}_k\mathrm{i})^{-1}(\Delta A_k + \sqrt{\alpha}\Delta B_k\mathrm{i}).$$

Proof One easily verifies that

$$I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k=\hat{\mathscr{A}}_k^{-1}(\hat{\mathscr{A}}_k-\mathscr{A}_k)=\begin{pmatrix} X_k & Y_k \\ -\alpha Y_k & X_k \end{pmatrix}$$

with

$$X_k=(\alpha\hat{A}_k^{-1}\hat{B}_k+\hat{B}_k^{-1}\hat{A}_k)^{-1}(\alpha\hat{A}_k^{-1}\Delta B_k+\hat{B}_k^{-1}\Delta A_k)$$
$$Y_k=(\alpha\hat{A}_k^{-1}\hat{B}_k+\hat{B}_k^{-1}\hat{A}_k)^{-1}(\hat{B}_k^{-1}\Delta B_k-\hat{A}_k^{-1}\Delta A_k),$$

where $\Delta A_k=\hat{A}_k-A_k$ and $\Delta B_k=\hat{B}_k-B_k$.

A similarity transformation with the matrix

$$\mathscr{N}_k=\begin{pmatrix} \mathrm{i}I & -\mathrm{i}I \\ \sqrt{\alpha}I & \sqrt{\alpha}I \end{pmatrix},$$

leads to a block-diagonal matrix $\mathscr{M}_k$:

$$\begin{aligned}\mathscr{M}_k&=\mathscr{N}_k^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)\mathscr{N}_k\\ &=\begin{pmatrix} X_k-\mathrm{i}\sqrt{\alpha}Y_k & \\ & X_k+\mathrm{i}\sqrt{\alpha}Y_k \end{pmatrix}\end{aligned}$$

with

$$X_k-\mathrm{i}\sqrt{\alpha}Y_k=(\hat{A}_k+\sqrt{\alpha}\hat{B}_k\mathrm{i})^{-1}(\Delta A_k+\sqrt{\alpha}\Delta B_k\mathrm{i})$$
$$X_k+\mathrm{i}\sqrt{\alpha}Y_k=(\hat{A}_k-\sqrt{\alpha}\hat{B}_k\mathrm{i})^{-1}(\Delta A_k-\sqrt{\alpha}\Delta B_k\mathrm{i}).$$

It is easy to see that

$$\mathscr{N}_k\mathscr{N}_k^H=2\begin{pmatrix} I & \\ & \alpha I \end{pmatrix},$$

where $\mathscr{N}_k^H$ denotes the Hermitian transpose of $\mathscr{N}_k$. We introduce

$$\widetilde{\mathscr{N}_k}=\frac{1}{2}\begin{pmatrix} (\hat{A}_k^2+\alpha\hat{B}_k^2)^{-1/4} & \\ & (\hat{A}_k^2+\alpha\hat{B}_k^2)^{-1/4} \end{pmatrix}\mathscr{N}_k$$

and obtain $(\widetilde{\mathscr{N}_k}\widetilde{\mathscr{N}_k^H})^{-1}=\mathscr{L}_k$. The matrix

$$\widetilde{\mathscr{M}_k}=\widetilde{\mathscr{N}_k}^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)\widetilde{\mathscr{N}_k}$$

is block diagonal with (1,1)-block $\widetilde{Z}_k$. The (2,2)-block is the conjugate complex of the (1,1)-block. Therefore obviously

$$\|\widetilde{\mathscr{M}_k^\nu}\|_{\ell^2}=\|\widetilde{Z}_k^\nu\|_{\ell^2}$$

holds. Since

$$\begin{aligned}\|\widetilde{\mathscr{M}_k^\nu}\|_{\ell^2}^2&=\|(\widetilde{\mathscr{N}_k}^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)\widetilde{\mathscr{N}_k})^\nu\|_{\ell^2}^2\\ &=\|\widetilde{\mathscr{N}_k}^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\widetilde{\mathscr{N}_k}\|_{\ell^2}^2\\ &=\rho(\widetilde{\mathscr{N}_k^H}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^{\nu T}\widetilde{\mathscr{N}_k}^{-H}\widetilde{\mathscr{N}_k}^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\widetilde{\mathscr{N}_k})\\ &=\rho(\mathscr{L}_k^{-1}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^{\nu T}\mathscr{L}_k(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu)\\ &=\|\mathscr{L}_k^{1/2}(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\mathscr{L}_k^{-1/2}\|_{\ell^2}^2\\ &=\|(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\|_{\mathscr{L}_k}^2,\end{aligned}$$

the proof is completed. □

Lemma 3 *Under the assumptions and notations of theorem 1 the matrix $I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k$ is power bounded with constant* 2*, i.e.,*

$$\|(I-\hat{\mathscr{A}}_k^{-1}\mathscr{A}_k)^\nu\|_{\mathscr{L}_k}\le 2$$

holds for all $\nu\in\mathbb{N}$.

Proof It is sufficient to show that $\widetilde{Z}_k$, given in lemma 2, is power bounded (with constant 2). We will show that

$$r(\widetilde{Z}_k)\le 1 \qquad (19)$$

holds, where

$$r(\widetilde{Z}_k)=\sup_{0\ne\underline{x}_k\in\mathbb{C}^{N_k}}\left|\frac{(\widetilde{Z}_k\underline{x}_k,\underline{x}_k)_{\ell^2}}{(\underline{x}_k,\underline{x}_k)_{\ell^2}}\right|$$

is the numerical radius of the matrix $\widetilde{Z}_k$.

Observe that

$$Z_k=(a_k+\sqrt{\alpha}b_k\mathrm{i})^{-1}\hat{D}_k^{-1}(\Delta A_k+\sqrt{\alpha}\Delta B_k\mathrm{i})$$

and, therefore,

$$\begin{aligned}\widetilde{Z}_k&=(\hat{A}_k^2+\alpha\hat{B}_k^2)^{1/4}Z_k(\hat{A}_k^2+\alpha\hat{B}_k^2)^{-1/4}=\hat{D}_k^{1/2}Z_k\hat{D}_k^{-1/2}\\ &=(a_k+\sqrt{\alpha}b_k\mathrm{i})^{-1}\hat{D}_k^{-1/2}(\Delta A_k+\sqrt{\alpha}\Delta B_k\mathrm{i})\hat{D}_k^{-1/2}.\end{aligned}$$

Hence we obtain

$$\begin{aligned}r(\widetilde{Z}_k)&=\sup_{0\ne\underline{x}_k\in\mathbb{C}^{N_k}}\left|\frac{(\widetilde{Z}_k\underline{x}_k,\underline{x}_k)_{\ell^2}}{(\underline{x}_k,\underline{x}_k)_{\ell^2}}\right|\\ &=\sup_{0\ne\underline{x}_k\in\mathbb{C}^{N_k}}\left|\frac{((\Delta A_k+\sqrt{\alpha}\Delta B_k\mathrm{i})\underline{x}_k,\underline{x}_k)_{\ell^2}}{(a_k+\sqrt{\alpha}b_k\mathrm{i})(\hat{D}_k^{1/2}\underline{x}_k,\hat{D}_k^{1/2}\underline{x}_k)_{\ell^2}}\right|\\ &=\sup_{0\ne\underline{x}_k\in\mathbb{C}^{N_k}}\left|\frac{(\Delta A_k\underline{x}_k,\underline{x}_k)_{\ell^2}+\sqrt{\alpha}(\Delta B_k\underline{x}_k,\underline{x}_k)_{\ell^2}\mathrm{i}}{(\hat{A}_k\underline{x}_k,\underline{x}_k)_{\ell^2}+\sqrt{\alpha}(\hat{B}_k\underline{x}_k,\underline{x}_k)_{\ell^2}\mathrm{i}}\right|\\ &=\sup_{0\ne\underline{x}_k\in\mathbb{C}^{N_k}}\sqrt{\frac{(\Delta A_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2+\alpha(\Delta B_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2}{(\hat{A}_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2+\alpha(\hat{B}_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2}}.\end{aligned}$$

The last equation holds because all involved scalar products have real values. We know that numerical radius is bounded by 1, if we can show that $(\Delta A_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2\le(\hat{A}_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2$ and $(\Delta B_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2\le(\hat{B}_k\underline{x}_k,\underline{x}_k)_{\ell^2}^2$ holds for all $\underline{x}_k\in\mathbb{C}^{N_k}$.

This property can be shown: The estimate (16) implies that

$$((\hat{A}_k^{-1/2}A_k\hat{A}_k^{-1/2}-I)\underline{x}_k,\underline{x}_k)_{\ell^2}\le(\underline{x}_k,\underline{x}_k)_{\ell^2}$$

holds for all vectors $\underline{x}_k\in\mathbb{C}^{N_k}$, since $\hat{A}_k$ is symmetric and positive definite. Using $\Delta A_k=A_k-\hat{A}_k$, this implies

$$(\Delta A_k\underline{x}_k,\underline{x}_k)_{\ell^2}\le(\hat{A}_k\underline{x}_k,\underline{x}_k)_{\ell^2}. \qquad (20)$$

Since A_k is symmetric and positive definite, we have moreover

$$-(\hat{A}_k\underline{x}_k,\underline{x}_k)_{\ell^2}\le(\Delta A_k\underline{x}_k,\underline{x}_k)_{\ell^2}. \qquad (21)$$

Combining (20) and (21) shows that

$$(\Delta A_k \underline{x}_k, \underline{x}_k)_{\ell^2}^2 \leq (\hat{A}_k \underline{x}_k, \underline{x}_k)_{\ell^2}^2$$

holds for all $\underline{x}_k \in \mathbb{C}^{N_k}$. The argument for B_k is completely analogous.

Hence we have shown (19). Using the power inequality for the numerical radius, see e.g. in [16], we obtain that

$$r(\widetilde{Z}_k^\nu) \leq 1$$

holds for all $\nu \in \mathbb{N}$. Using the fact, that $\|M\|_{\ell^2} \leq 2\,r(M)$ holds for all matrices, we know that

$$\|\widetilde{Z}_k^\nu\|_{\ell^2} \leq 2$$

holds for all $\nu \in \mathbb{N}$, which finishes the proof. □

Additionally, we need that the preconditioning matrix $\hat{\mathscr{A}}_k$ can be bounded from above using the matrix $\mathscr{L}_k$:

Lemma 4 *Under the assumptions and notations of theorem 1 we have*

$$\|\mathscr{L}_k^{-1/2} \hat{\mathscr{A}}_k \mathscr{L}_k^{-1/2}\|_{\ell^2} = 1.$$

Proof Using the definition $\hat{Z}_k = (\hat{A}_k^2 + \alpha \hat{B}_k^2)^{1/4}$, we observe that $\hat{Z}_k = (a_k^2 + \alpha b_k^2)^{1/4} \hat{D}_k^{1/4}$. Therefore the desired result immediately follows:

$$\begin{aligned}
&\|\mathscr{L}_k^{-1/2} \hat{\mathscr{A}}_k \mathscr{L}_k^{-1/2}\|_{\ell^2} \\
&\quad = \left\| \begin{pmatrix} \hat{Z}_k^{-1} \hat{A}_k \hat{Z}_k^{-1} & \hat{Z}_k^{-1} \alpha^{1/2} \hat{B}_k \hat{Z}_k^{-1} \\ \hat{Z}_k^{-1} \alpha^{1/2} \hat{B}_k \hat{Z}_k^{-1} & -\hat{Z}_k^{-1} \hat{A}_k \hat{Z}_k^{-1} \end{pmatrix} \right\|_{\ell^2} \\
&\quad = (a_k^2 + \alpha b_k^2)^{-1/2} \left\| \begin{pmatrix} a_k I & \alpha^{1/2} b_k I \\ \alpha^{1/2} b_k I & -a_k I \end{pmatrix} \right\|_{\ell^2} = 1.
\end{aligned}$$

□

We combine lemmas 1 – 4 to prove theorem 1 as follows:

Proof of theorem 1 Let $\mathscr{M}_k = I - \hat{\mathscr{A}}_k^{-1} \mathscr{A}_k$. Lemma 3 states that

$$\|\mathscr{M}_k^\nu\|_{\mathscr{L}_k} \leq 2$$

holds. Using lemma 1 we conclude

$$\|(I - \mathscr{M}_k)((1-\tau)I + \tau \mathscr{M}_k)^\nu\|_{\mathscr{L}_k} \leq \frac{C}{\sqrt{\nu}}.$$

By plugging in for $\mathscr{M}_k$, we obtain

$$\|\mathscr{L}_k^{1/2} \hat{\mathscr{A}}_k^{-1} \mathscr{A}_k (I - \tau \hat{\mathscr{A}}_k^{-1} \mathscr{A}_k)^\nu \mathscr{L}_k^{-1/2}\|_{\ell^2} \leq \frac{C}{\sqrt{\nu}}.$$

Using the sub-multiplicativity of norms, we obtain

$$\begin{aligned}
&\|\mathscr{L}_k^{-1/2} \mathscr{A}_k (I - \tau \hat{\mathscr{A}}_k^{-1} \mathscr{A}_k)^\nu \mathscr{L}_k^{-1/2}\|_{\ell^2} \\
&\quad \leq \frac{C}{\sqrt{\nu}} \|\mathscr{L}_k^{-1/2} \hat{\mathscr{A}}_k \mathscr{L}_k^{-1/2}\|_{\ell^2},
\end{aligned}$$

which finishes the proof, as we know from lemma 4 that $\|\mathscr{L}_k^{-1/2} \hat{\mathscr{A}}_k \mathscr{L}_k^{-1/2}\|_{\ell^2} = 1$. □

3.2 Convergence analysis for distributed control model problem

In this subsection we give an overall convergence result for the multigrid method using the collective Richardson relaxation to approximate the solution of the elliptic distributed control model problem. In this case we can apply theorem 1 and obtain a result with respect to the norm $\|\cdot\|_{\mathscr{L}_k}$, which reads for the collective Richardson relaxation as follows:

$$\|(\underline{y}_k, \underline{p}_k)\|_{\mathscr{L}_k}^2 = (a_k^2 + \alpha\, b_k^2)^{1/2} \left(\|\underline{y}_k\|_{\ell^2}^2 + \alpha^{-1} \|\underline{p}_k\|_{\ell^2}^2 \right),$$

where $\frac{1}{2}\lambda_{\max}(A_k) \leq a_k \leq \frac{C}{2}\lambda_{\max}(A_k)$ and $\frac{1}{2}\lambda_{\max}(B_k) \leq b_k \leq \frac{C}{2}\lambda_{\max}(B_k))$ holds for some constant $C > 0$. For the model problem A_k is the mass matrix and B_k is the stiffness matrix, therefore standard scaling arguments (for quasi uniform grids) imply that the norm $\|\cdot\|_{\mathscr{L}_k}$ is equivalent to the norm $\|\cdot\|_{0,k}$, given by

$$\|(y_k, p_k)\|_{0,k}^2 = (1 + \alpha^{1/2} h_k^{-2}) \left(\|y_k\|_{L^2(\Omega)}^2 + \alpha^{-1} \|p_k\|_{L^2(\Omega)}^2 \right),$$

in the sense that there are constants $\underline{C}$ and $\overline{C}$ (independent of h_k and α) such that

$$\underline{C}\, \|\underline{x}_k\|_{\mathscr{L}_k} \leq \|x_k\|_{0,k} \leq \overline{C}\, \|\underline{x}_k\|_{\mathscr{L}_k} \tag{22}$$

holds, where again $\underline{x}_k$ is the coefficient vector representing some function $x_k \in X_k$. Observe that the norm $\|\cdot\|_{0,k}$ is obtained by replacing $\|\cdot\|_{H^1(\Omega)}$ by $h_k^{-1}\|\cdot\|_{L^2(\Omega)}$ in the definition of the norm $\|\cdot\|_X$.

Using the equivalence relation (22), the smoothing property can be shown for the model problem:

Theorem 2 (Smoothing property) *For the elliptic distributed control model problem* (8) *the collective Richardson relaxation satisfies for all choices of the damping parameter* $\tau \in (0,1)$ *the smoothing property* (14) *with rate* $\eta(\nu) = \frac{C_S}{\sqrt{\nu}}$ *and a constant* $C_S > 0$ *independent of* h_k *and* α*, i.e.,*

$$\|x_k^{(0,\nu)} - x_k^*\|_{2,k} \leq \frac{C_S}{\sqrt{\nu}} \|x_k^{(0)} - x_k^*\|_{0,k}$$

holds for all $\nu \in \mathbb{N}$.

Proof Using the error-propagation operator of the smoothing iteration $\mathscr{M}_k = I - \tau \hat{\mathscr{A}}_k^{-1} \mathscr{A}_k$ and the initial error $\underline{w}_k^{(0)} = \underline{x}_k^{(0)} - \underline{x}_k^*$, we can express the error after ν smoothing steps:

$$\underline{w}_k^{(0,\nu)} = \underline{x}_k^{(0,\nu)} - \underline{x}_k^* = \mathscr{M}_k^\nu \underline{w}_k^{(0)}.$$

The norm of $\mathscr{M}_k^\nu \underline{w}_k^{(0)}$ can be bounded using theorem 1, as the collective Richardson relaxation is covered by the theorem: There is a constant $\widetilde{C}_S > 0$ such that

$$\|\mathscr{L}_k^{-1/2} \mathscr{A}_k \mathscr{M}_k^\nu \mathscr{L}_k^{-1/2}\|_{\ell^2} \leq \frac{\widetilde{C}_S}{\sqrt{\nu}}$$

holds, which implies

$$\|\mathscr{L}_k^{-1/2}\mathscr{A}_k\mathscr{M}_k^{\nu}\underline{w}_k^{(0)}\|_{\ell^2} \leq \frac{\widetilde{C}_S}{\sqrt{\nu}}\|\mathscr{L}_k^{1/2}\underline{w}_k^{(0)}\|_{\ell^2},$$

$$\|\mathscr{L}_k^{-1/2}\mathscr{A}_k\underline{w}_k^{(0,\nu)}\|_{\ell^2} \leq \frac{\widetilde{C}_S}{\sqrt{\nu}}\|\underline{w}_k^{(0)}\|_{\mathscr{L}_k}$$

and

$$\sup_{\underline{\widetilde{w}}_k\in\mathbb{R}^{2N_k}} \frac{(\mathscr{A}_k\underline{w}_k^{(0,\nu)},\underline{\widetilde{w}}_k)_{\ell^2}}{\|\underline{\widetilde{w}}_k\|_{\mathscr{L}_k}} \leq \frac{\widetilde{C}_S}{\sqrt{\nu}}\|\underline{w}_k^{(0)}\|_{\mathscr{L}_k}.$$

Using the equivalence relation (22), we obtain the desired result. □

Theorem 3.1 in [20] shows that for this model problem also the approximation property (15) holds, if the following regularity result holds:

Assumption 1 *There is a constant $c > 0$ such that for all $f \in L^2(\Omega)$ the solution $v_f \in H^1(\Omega)$ of the variational equation*

$$(v_f, w)_{H^1(\Omega)} = (f, w)_{L^2(\Omega)} \qquad \text{for all } w \in H^1(\Omega)$$

satisfies: $v_f \in H^2(\Omega)$ and $\|v_f\|_{H^2(\Omega)} \leq c\|f\|_{L^2(\Omega)}$.

By combining theorem 2 (smoothing property) and theorem 3.1 in [20] (approximation property) we obtain:

Theorem 3 (Two-grid-convergence) *Assume that assumption 1 is satisfied. For the elliptic distributed control model problem* (8) *the two-grid method proposed in subsection 2.2 with the collective Richardson relaxation converges for all choices of the damping parameter $\tau \in (0,1)$ if sufficiently many smoothing steps ν are applied, i.e., the estimate*

$$\|x_k^{(1)} - x_k^*\|_{0,k} \leq q\|x_k^{(0)} - x_k^*\|_{0,k}$$

holds, where the convergence rate $q = \frac{C_A C_S}{\sqrt{\nu}}$ is independent of h_k and α.

Remark 1 Assumption 1 holds for domains which have a sufficiently smooth boundary (see, e.g., [15]) or which are polygonal or polyhedral (see, e.g., [8] or [9]).

Remark 2 One can extend the convergence result stated in theorem 3 to the *W-cycle multigrid method*, see e.g. [12].

Theorem 3 is a convergence result in the non-standard mesh-dependent norm $\|\cdot\|_{0,k}$ but this result implies also convergence (with the same rate) in the standard norm $\|\cdot\|_{L^2(\Omega)}$.

Corollary 1 *Under the notations and assumptions of theorem 3 there is a constant $C > 0$ and a convergence rate $q < 1$, both independent of h_k and α, such that the L^2-convergence result*

$$\|x_k^{(n)} - x_k^*\|_{L^2(\Omega)} \leq Cq^n\|y_D\|_{L^2(\Omega)}$$

holds for all $n \in \mathbb{N}$, all $k \in \{0,\ldots,K\}$ and all $\alpha \in (0,1]$, provided $y_k^{(0)} = p_k^{(0)} = 0$.

Proof Theorem 3 implies

$$\|x_k^{(n)} - x_k^*\|_{0,k} \leq q^n\|x_k^{(0)} - x_k^*\|_{0,k},$$

which is equivalent to

$$\begin{aligned}(\|y_k^{(n)} - y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^{(n)} - p_k^*\|^2_{L^2(\Omega)})^{1/2} \\ \leq q^n(\|y_k^{(0)} - y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^{(0)} - p_k^*\|^2_{L^2(\Omega)})^{1/2}.\end{aligned}$$

Assuming $y_k^{(0)} = p_k^{(0)} = 0$ implies

$$\begin{aligned}(\|y_k^{(n)} - y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^{(n)} - p_k^*\|^2_{L^2(\Omega)})^{1/2} \\ \leq q^n(\|y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^*\|^2_{L^2(\Omega)})^{1/2},\end{aligned}$$

The right-hand-side is bounded from above by $q^n\|x_k^*\|_X$. Using the inf-sup condition (7), we obtain

$$\|x_k^*\|_X \leq \underline{C}^{-1} \sup_{0\neq\widetilde{x}_k\in X_k} \frac{\mathscr{B}(x_k^*,\widetilde{x}_k)}{\|\widetilde{x}_k\|_X} = \underline{C}^{-1} \sup_{0\neq\widetilde{x}_k\in X_k} \frac{\mathscr{F}(\widetilde{x}_k)}{\|\widetilde{x}_k\|_X}.$$

Using

$$\begin{aligned}\mathscr{F}(\widetilde{y}_k,\widetilde{p}_k) = (y_D,\widetilde{y}_k)_{L^2(\Omega)} \leq \|y_D\|_{L^2(\Omega)}\|\widetilde{y}_k\|_{L^2(\Omega)} \\ \leq \|y_D\|_{L^2(\Omega)}\|(\widetilde{y}_k,\widetilde{p}_k)\|_X,\end{aligned}$$

we obtain $\|x_k^*\|_X \leq \underline{C}^{-1}\|y_D\|_{L^2(\Omega)}$ and further

$$\begin{aligned}(\|y_k^{(n)} - y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^{(n)} - p_k^*\|^2_{L^2(\Omega)})^{1/2} \\ \leq \underline{C}^{-1}q^n\|y_D\|_{L^2(\Omega)}.\end{aligned}$$

For $\alpha \leq 1$, we have

$$\begin{aligned}(\|y_k^{(n)} - y_k^*\|^2_{L^2(\Omega)} + \|p_k^{(n)} - p_k^*\|^2_{L^2(\Omega)})^{1/2} \\ \leq (\|y_k^{(n)} - y_k^*\|^2_{L^2(\Omega)} + \alpha^{-1}\|p_k^{(n)} - p_k^*\|^2_{L^2(\Omega)})^{1/2},\end{aligned}$$

which completes the proof. □

4 Numerical results

In this section we present some numerical experiments for the proposed model problems on the domain $\Omega = (0,1)^2$. The mesh at the coarsest grid level ($k = 0$) consists of 2 triangles which are constructed by connecting the points $(0,0)$ and $(1,1)$. Without loss of generality, we choose homogeneous data $y_D = 0$, where the exact solution is of course given by $x_k^* = 0$. We did numerical experiments for the distributed control model problem and the boundary control model problem. The tests have been done for collective Jacobi relaxation with damping parameter $\tau = 1/2$, while the collective Gauss-Seidel iteration was not relaxed. The number of smoothing steps ν is written as $\nu = \nu_{pre} + \nu_{post}$ for ν_{pre} and ν_{post} pre- and post- smoothing steps, respectively. In all cases a W-cycle multigrid method was applied.

Tables 1 and 2 show the convergence rates q and the number of iterations n needed to reduce the initial error $\|x_k^{(0)} - x_k^*\|_{0,k}$ by a factor of $\varepsilon = 10^{-8}$ for the distributed control model problem. The starting values $\underline{x}_k^{(0)}$ were chosen randomly. We can see that the number of iterations does not depend on the grid size which confirms optimal complexity. If we compare table 1 with table 2, we see that the number of iterations decays if the number of smoothing steps ν is increased. Moreover we see that the number of iterations is robust in the cost or regularization parameter α.

In table 3, we give the convergence results for the collective Gauss-Seidel smoother for the distributed control model problem, which is not covered by the theory. The numerical experiment shows very fast convergence (again optimal and robust in the parameter α).

Table 4 shows that the good behavior which we have seen for the distributed control model problem carries also over to the boundary control model problem, which is not covered by our theory. We have also in that case optimal convergence results that are again robust in the parameter α.

5 Conclusions and outlook

We have seen that collective point smoothers, especially collective Gauss-Seidel iteration, allow to construct all-at-once multigrid methods for some classes of optimal control problems (and of course other saddle point problems which allow the construction of collective point smoothers as done in this work). The construction of these methods is easy and also possible without knowing stability results for the solution.

We could prove that the smoothing property of these smoothers holds just using algebraic relations between the involved matrices. So the convergence theory can be applied to all kinds of saddle point problems which fit into our framework.

The limitations on the construction of the smoothing procedure which we have seen in our paper are just due to the convergence proof and they cannot be seen in the numerical results. The extension of the convergence analysis to more general smoothers or to a more general class of problems is a challenging task for further work.

References

1. Arian, E., and Ta'asan, S.: Multigrid one-shot methods for optimal control problems: Infinite dimensional control. (1994)
2. Battermann, A., and Heinkenschloss, M.: Preconditioners for Karush-Kuhn-Tucker matrices arising in the optimal control of distributed systems. In: Desch, W. (ed.) et al. (ed.) Control and estimation of distributed
3. Battermann, A., and Sachs, E. W.: Block preconditioners for KKT systems in PDE-governed optimal control problems. In: Hoffmann, Karl-Heinz (ed.) et al. (ed.) Fast solution of discretized optimization problems. Workshop at the Weierstrass Inst. for Applied Analysis and Stochastics, Berlin, Germany, May 8-12, 2000, 1 – 18. Basel: Birkhäuser. ISNM, Int. Ser. Numer. Math. 138 (2001)
4. Biros, G., and Ghattas, O.: Parallel Lagrange-Newton–Krylov-Schur methods for PDE-constrained optimization. II: The Lagrange-Newton solver and its application to optimal control of steady viscous flows. SIAM J. Sci. Comput. **27**(2), 714 – 739 (2005)
5. Biros, G., and Ghattas, O.: Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. I: The Krylov-Schur solver. SIAM J. Sci. Comput. **27**(2), 687 – 713 (2005)
6. Borzi, A., and Schulz, V.: Multigrid Methods for PDE Optimization. SIAM Review **51**, 361 – 395 (2009)
7. Borzi, A., Kunisch, K. and Kwak, D. Y.: Accuracy and convergence properties of the finite difference multigrid solution of an optimal control optimality system. SIAM J. Control Optim **41(5)**, 1477 – 1497 (2003)
8. Dauge, M.: Elliptic boundary value problems on corner domains. Smoothness and asymptotics of solutions. Lecture Notes in Mathematics, 1341. Berlin etc.: Springer-Verlag (1988)
9. Dauge, M.: Neumann and mixed problems on curvilinear polyhedra. Integral Equations Oper. Theory **15**, 227 – 261 (1992)
10. Ecker, A., and Zulehner, W.: On the Smoothing Property of Multi-Grid Methods in Non-Symmetric Case. Numer. Lin. Alg. Appl. **3(2)**, 161 – 172 (1996)
11. Hackbusch, W.: Fast solution of elliptic control problems. J. Optimization Theory Appl. **31**, 565 – 581 (1980)
12. Hackbusch, W.: Multi-Grid Methods and Applications. Springer, Berlin (1985)
13. Hazra, S. B., and Schulz, V.: Simultaneous pseudo-timestepping for PDE-model based optimization problems. BIT **44**(3), 457 – 472 (2004)
14. Lass, O., Vallejos, M., Borzi, A., and Douglas, C. C.: Implementation and analysis of multigrid schemes with finite elements for elliptic optimal control problems. Computing **84**(1 – 2), 27 – –48 (2009)
15. Necas, J.: Les méthodes directes en théorie des équations elliptiques. Masson, Paris (1967). 351 p.
16. Pearcy, C.: An Elementary Proof of the Power Inequality for the Numerical Radius. Michigan Math. Journal **13(3)**, 289 – 291 (1966)
17. Rees, T., Dollar, S., and Wathen, A.: Optimal solvers for pde-constrained optimization. SIAM Journal on Scientific Computing **32**(1), 271 – 298 (2010)
18. Reusken, A.: A new lemma in multigrid convergence theory. Report-RANA **91-07** (1991)
19. Schöberl, J., and Zulehner, W.: Symmetric indefinite preconditioners for saddle point problems with applications to pde-constrained optimization problems. SIAM J. Matrix Anal. Appl. **29**(3), 752 – 773 (2007)
20. Schöberl, J., Simon, R., and Zulehner, W.: A Robust Multigrid Method for Elliptic Optimal Control Problems. NuMa-Report **1** (2010)
21. Schulz, V., and Wittum, G.: Transforming smoothers for PDE constrained optimization problems. Computing and Visualization in Science **11**(4-11), 207–219 (2008)
22. Simon, R., and Zulehner, W.: On Schwarz-type smoothers for saddle point problems with applications to PDE-constrained optimization problems. Numer. Math. **111**, 445 – 468 (2009)
23. Ta'asan, S.: "One-shot" methods for optimal control of distributed parameter systems I: The finite dimensional case. ICASE-Report 91-2, NASA Langley Research Center, Hampton VA (1991)
24. Trottenberg, U., Oosterlee, C., and Schüller, A.: Multigrid. Academic Press, London (2001)
25. Wittum, G.: Multi-grid methods for Stokes and Navier-Stokes equations. Transforming smoothers: Algorithms and numerical results. Numer. Math. **54**(5), 543 – 563 (1988)
26. Wittum, G.: On the convergence of multi-grid methods with transforming smoothers. Numer. Math. **57**(1), 15 – 38 (1990)

Table 1 Distributed control model problem $\nu = 1+1$, Collective Jacobi relaxation, $\tau = 1/2$

grid level	$\alpha = 1$ n	q	$\alpha = 10^{-2}$ n	q	$\alpha = 10^{-4}$ n	q	$\alpha = 10^{-6}$ n	q	$\alpha = 10^{-8}$ n	q	$\alpha = 10^{-10}$ n	q
$k = 4$	28	0.5133	28	0.5154	28	0.5135	21	0.4124	27	0.5014	29	0.5234
$k = 5$	28	0.5149	28	0.5166	28	0.5152	28	0.5075	18	0.3477	28	0.5149
$k = 6$	29	0.5197	29	0.5195	28	0.5158	29	0.5200	25	0.4736	25	0.4719
$k = 7$	29	0.5197	29	0.5205	29	0.5199	29	0.5201	28	0.5163	18	0.3501
$k = 8$	29	0.5206	29	0.5209	29	0.5210	29	0.5208	29	0.5202	27	0.4996
$k = 9$	29	0.5209	29	0.5209	29	0.5210	29	0.5212	29	0.5206	29	0.5198

Table 2 Distributed control model problem, $\nu = 2+2$, collective Jacobi relaxation, $\tau = 1/2$

grid level	$\alpha = 1$ n	q	$\alpha = 10^{-2}$ n	q	$\alpha = 10^{-4}$ n	q	$\alpha = 10^{-6}$ n	q	$\alpha = 10^{-8}$ n	q	$\alpha = 10^{-10}$ n	q
$k = 4$	15	0.2807	15	0.2799	15	0.2761	11	0.1633	14	0.2529	15	0.2749
$k = 5$	15	0.2821	15	0.2826	15	0.2812	14	0.2616	9	0.1205	15	0.2709
$k = 6$	15	0.2841	15	0.2866	15	0.2853	15	0.2832	13	0.2293	13	0.2245
$k = 7$	15	0.2867	15	0.2869	15	0.2868	15	0.2865	15	0.2794	9	0.1240
$k = 8$	15	0.2880	15	0.2879	15	0.2878	15	0.2880	15	0.2868	14	0.2549
$k = 9$	15	0.2880	15	0.2883	15	0.2882	15	0.2879	15	0.2878	15	0.2841

Table 3 Distributed control model problem, $\nu = 1+1$, collective Gauss-Seidel iteration

grid level	$\alpha = 1$ n	q	$\alpha = 10^{-2}$ n	q	$\alpha = 10^{-4}$ n	q	$\alpha = 10^{-6}$ n	q	$\alpha = 10^{-8}$ n	q	$\alpha = 10^{-10}$ n	q
$k = 4$	8	0.0907	8	0.0863	9	0.1082	12	0.1940	8	0.0872	8	0.0756
$k = 5$	9	0.1017	9	0.1045	9	0.1018	10	0.1373	11	0.1818	8	0.0740
$k = 6$	9	0.1055	9	0.1067	9	0.1039	9	0.1117	11	0.1691	9	0.1068
$k = 7$	9	0.1063	9	0.1066	9	0.1052	9	0.1067	9	0.1265	12	0.1963
$k = 8$	9	0.1067	9	0.1072	9	0.1069	9	0.1069	9	0.1102	10	0.1512
$k = 9$	9	0.1072	9	0.1076	9	0.1073	9	0.1071	9	0.1075	9	0.1207

Table 4 Boundary control model problem, $\nu = 2+2$, collective Jacobi relaxation, $\tau = 1/2$

grid level	$\alpha = 1$ n	q	$\alpha = 10^{-2}$ n	q	$\alpha = 10^{-4}$ n	q	$\alpha = 10^{-6}$ n	q	$\alpha = 10^{-8}$ n	q	$\alpha = 10^{-10}$ n	q
$k = 4$	15	0.2806	15	0.2810	16	0.2958	16	0.3028	16	0.3141	16	0.3152
$k = 5$	15	0.2821	15	0.2828	15	0.2865	17	0.3299	16	0.3157	17	0.3180
$k = 6$	15	0.2866	15	0.2866	15	0.2878	17	0.3200	16	0.3000	17	0.3174
$k = 7$	15	0.2881	15	0.2880	15	0.2877	15	0.2918	17	0.3174	16	0.3074
$k = 8$	15	0.2876	15	0.2875	15	0.2876	15	0.2900	19	0.3757	16	0.2950
$k = 9$	15	0.2879	15	0.2878	15	0.2879	15	0.2884	16	0.2975	16	0.3153

Technical Reports of the Doctoral Program
"Computational Mathematics"

2011

2011-01 S. Takacs, W. Zulehner: *Convergence Analysis of Multigrid Methods with Collective Point Smoothers for Optimal Control Problems* February 2011. Eds.: U. Langer, J. Schicho

2010

2010-01 S. Radu, J. Sellers: *Parity Results for Broken k-diamond Partitions and (2k+1)-cores* March 2010. Eds.: P. Paule, V. Pillwein

2010-02 P.G. Gruber: *Adaptive Strategies for High Order FEM in Elastoplasticity* March 2010. Eds.: U. Langer, V. Pillwein

2010-03 Y. Huang, L.X.Châu Ngô: *Rational General Solutions of High Order Non-autonomous ODEs* June 2010. Eds.: F. Winkler, P. Paule

2010-04 S. Beuchler, V. Pillwein, S. Zaglmayr: *Sparsity optimized high order finite element functions for H(div) on simplices* September 2010. Eds.: U. Langer, P. Paule

2010-05 C. Hofreither, U. Langer, C. Pechstein: *Analysis of a non-standard finite element method based on boundary integral operators* September 2010. Eds.: B. Jüttler, J. Schicho

2010-06 M. Hodorog, J. Schicho: *A symbolic-numeric algorithm for genus computation* September 2010. Eds.: B. Jüttler, R. Ramlau

2010-07 M. Hodorog, J. Schicho: *Computational geometry and combinatorial algorithms for the genus computation problem* September 2010. Eds.: B. Jüttler, R. Ramlau

2010-08 C. Koukouvinos, V. Pillwein, D.E. Simos, Z. Zafeirakopoulos: *A Note on the Average Complexity Analysis of the Computation of Periodic and Aperiodic Ternary Complementary Pairs* October 2010. Eds.: P. Paule, J. Schicho

2010-09 V. Pillwein, S. Takacs: *Computing smoothing rates of collective point smoothers for optimal control problems using symbolic computation* October 2010. Eds.: U. Langer, P. Paule

2010-10 T. Takacs, B. Jüttler: *Existence of Stiffness Matrix Integrals for Singularly Parameterized Domains in Isogeometric Analysis* November 2010. Eds.: J. Schicho, W. Zulehner

2010-11 Y. Huang, L.X.Châu Ngô: *Rational Solutions of a Rational System of Autonomous ODEs: Generalization to Trivariate Case and Problems* November 2010. Eds.: F. Winkler, P. Paule

2010-12 S. Béla, B. Jüttler: *Approximating Algebraic Space Curves by Circular Arcs* November 2010. Eds.: J. Schicho, F. Winkler

2010-13 C. Hofreither: *L_2 Error Estimates for a Nonstandard Finite Element Method on Polyhedral Meshes* December 2010. Eds.: U. Langer, V. Pillwein

2010-14 I. Georgieva, C. Hofreither, R. Uluchev: *Interpolation in the unit disk based on Radon projections and function values* December 2010. Eds.: U. Langer, J. Schicho

2010-15 C. Aistleitner: *Relations between Gröbner bases, differential Gröbner bases, and differential characteristic sets* December 2010. Eds.: F. Winkler, W. Schreiner

The complete list since 2009 can be found at
`https://www.dk-compmath.jku.at/publications/`

Doctoral Program
“Computational Mathematics”

Director:

Prof. Dr. Peter Paule
Research Institute for Symbolic Computation

Deputy Director:

Prof. Dr. Bert Jüttler
Institute of Applied Geometry

Address:

Johannes Kepler University Linz
Doctoral Program “Computational Mathematics”
Altenbergerstr. 69
A-4040 Linz
Austria
Tel.: ++43 732-2468-7174

E-Mail:

office@dk-compmath.jku.at

Homepage:

http://www.dk-compmath.jku.at

Submissions to the DK-Report Series are sent to two members of the Editorial Board who communicate their decision to the Managing Editor.