Patch Smoothers for Saddle Point Problems with Applications to PDE-Constrained Optimization Problems

René Simon[1] and Walter Zulehner[2]

[1] SFB 013, Johannes Kepler University Linz, Altenbergerstraße 69, A-4040 Linz, Austria, rene.simon@sfb013.uni-linz.ac.at
[2] Institute of Computational Mathematics, Johannes Kepler University Linz, Altenbergerstraße 69, A-4040 Linz, Austria, zulehner@numa.uni-linz.ac.at

Summary. We consider a multigrid method for solving the discretized optimality system of a PDE-constrained optimization problem. In particular, we discuss the construction of an additive Schwarz-type smoother for a class of elliptic optimal control problems. A rigorous multigrid convergence analysis yields level-independent convergence rates. Numerical experiments indicate that the convergence rates are also independent of the involved regularization parameter.

1 Introduction

In this paper we discuss multigrid methods for solving the discretized optimality system (or Karush-Kuhn-Tucker system, in short KKT system) for optimization problems in function spaces with constraints in form of partial differential equations (PDEs). In particular, we will consider elliptic optimal control problems, see, e.g., [3], and focus on so-called one-shot multigrid methods, see [7], where the multigrid idea is directly applied to the optimality system (instead of a block-wise approach as an alternative).

One of the most important ingredients of such a multigrid method is an appropriate smoother. In this paper we consider patch smoothers: The computational domain is divided into small (overlapping or non-overlapping) sub-domains (patches). One iteration step of the smoothing process consists of solving local problems on each patch one-by-one either in a Jacobi-type or Gauss-Seidel-type manner. This strategy can be seen as an additive or multiplicative Schwarz-type smoother. The technique was successfully used for the Navier-Stokes equations, see [8]. The special case, where each patch consists of a single node of the underlying grid, is usually called a point smoother. Such a smoother was proposed for optimal control problems in [2].

So far, the convergence analysis of multigrid methods with patch smoothers applied to KKT systems of PDE-constrained optimization problems is not as developed as for elliptic PDEs. One line of argument is based on a Fourier analysis, which, strictly speaking, covers only the case of uniform grids with special boundary conditions (and small perturbations of this situation), see [1, 2]. A second and more rigorous strategy exploits the fact that, for certain classes of optimal control problems, the KKT system can be reduced to a compact perturbation of an elliptic system of PDEs. This guarantees the convergence of the multigrid method if the coarse grid is sufficiently fine, see [2]. In [4] the general construction and rigorous analysis of patch smoothers were discussed and applied to the Stokes problem. An extension to KKT systems was presented in [5].

Here we will propose a multigrid method with a patch smoother applied to a reduced system derived from the original KKT system, the same reduced system which was considered in [2]. A rigorous convergence analysis will be presented directly applied to the multigrid method for the reduced system, in contrary what was done in [5]. Compared to the results presented in [5] the numerical experiments show a much better performance of the multigrid method.

In order to keep the notations simple and the strategy transparent the material is presented for a model problem in optimal control only. The extension to more general problems is straight forward.

The paper is organized as follows: In Section 2 the model problem and its discretization are introduced. Section 3 contains the multigrid method, the patch smoother, and the main multigrid convergence result. Finally, in Section 4 some numerical results are presented.

2 An Optimal Control Problem

Let Ω be a bounded convex polygonal domain in $\mathbb{R}^2$. Let $L^2(\Omega)$ and $H^1(\Omega)$ denote the usual Lebesgue space and Sobolev space with norms $\|.\|_{L^2(\Omega)}$ and $\|.\|_{H^1(\Omega)}$, respectively. We consider the following elliptic optimal control problem of tracking type: Find the state $y \in H^1(\Omega)$ and the control $u \in L^2(\Omega)$ such that

$$J(y,u) = \min_{(z,v)\in H^1(\Omega)\times L^2(\Omega)} J(z,v)$$

with cost functional

$$J(z,v) = \frac{1}{2}\,\|z - y_d\|^2_{L^2(\Omega)} + \frac{\gamma}{2}\,\|v\|^2_{L^2(\Omega)}$$

subject to the (weak form of the) state equation

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

where Γ denotes the boundary of Ω, $y_d \in L^2(\Omega)$ is the desired state and $\gamma > 0$ is the weight of the cost of the control (or simply a regularization parameter).

By introducing the adjoint state $p \in H^1(\Omega)$ we get the following equivalent optimality system, see e.g., [3]:

1. The adjoint state equation:

$$-\Delta p + p = -(y - y_d) \quad \text{in } \Omega, \qquad \frac{\partial p}{\partial n} = 0 \quad \text{on } \Gamma.$$

2. The control equation:

$$\gamma u - p \;=\; 0 \quad \text{in } \Omega.$$

3. The state equation:

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

The control equation yields a simple algebraic relation between the control u and the adjoint state p, which is used to eliminate the control in the state equation. After multiplying by γ we obtain from the state equation:

$$p - \gamma(-\Delta y + y) = 0 \quad \text{in } \Omega, \qquad \frac{\partial y}{\partial n} = 0 \quad \text{on } \Gamma.$$

The weak formulation of the reduced problem in p and y leads to a mixed variational problem: Find $p \in Q = H^1(\Omega)$ and $y \in Y = H^1(\Omega)$ such that

$$\begin{aligned} a(p,q) + b(q,y) \;&=\; \langle F, q\rangle \quad &&\text{for all } q \in Q,\\ b(p,z) \;-\; \gamma a(y,z) \;&=\; 0 &&\text{for all } z \in Y \end{aligned}$$

with

$$a(p,q) = (p,q)_{H^1(\Omega)}, \quad b(q,z) = (q,z)_{L^2(\Omega)}, \quad \langle F, q\rangle = (y_d, q)_{L^2(\Omega)}.$$

Here $(.,.)_H$ denotes the standard scalar product in a Hilbert space H and $\langle .,. \rangle$ is used for the duality product of linear functionals from the dual space H^* and elements in H.

The mixed variational problem can also be written as a variational problem on $Q \times Y$: Find $(p,y) \in Q \times Y$ such that

$$\mathcal{B}((p,y),(q,z)) = \mathcal{F}(q,z) \quad \text{for all } (q,z) \in Q \times Y$$

with the bilinear form

$$\mathcal{B}((p,y),(q,z)) = a(p,q) + b(q,y) + b(p,z) - \gamma a(y,z)$$

and the linear functional

$$\mathcal{F}(q,z) = \langle F, q\rangle.$$

Let $(\mathcal{T}_k)$ be a sequence of triangulations of Ω, where $\mathcal{T}_{k+1}$ is obtained by dividing each triangle into four smaller triangles by connecting the midpoints of the edges of the triangles in $\mathcal{T}_k$. The quantity $\max\{\operatorname{diam} T : T \in \mathcal{T}_k\}$ is denoted by h_k.

We consider the following discretization by continuous and piecewise linear finite elements:

$$Q_k = Y_k = \{w \in C(\bar{\Omega}) : w|_T \in P_1 \text{ for all } T \in \mathcal{T}_k\},$$

where P_1 denotes the polynomials of total degree less or equal to 1. Then we obtain the following discrete variational problem: Find $p_k \in Q_k$ and $y_k \in Y_k$ such that

$$\begin{aligned} a(p_k,q) + b(q,y_k) &= \langle F,q\rangle \quad \text{for all } q \in Q_k, \\ b(p_k,z) - \gamma a(y_k,z) &= 0 \quad\quad\quad \text{for all } z \in Y_k. \end{aligned}$$

The discrete mixed variational problem can also be written as a discrete variational problem on $Q_k \times Y_k$: Find $(p_k,y_k) \in Q_k \times Y_k$ such that

$$\mathcal{B}((p_k,y_k),(q,z)) = \mathcal{F}(q,z) \quad \text{for all } (q,z) \in Q_k \times Y_k. \tag{1}$$

By introducing the standard nodal basis for Q_k and Y_k, we finally obtain the following saddle point problem in matrix-vector notation: Find the coefficient vectors $(\underline{p}_k,\underline{y}_k) \in \mathbb{R}^{N_k} \times \mathbb{R}^{N_k}$ such that

$$\mathcal{K}_k \begin{pmatrix} \underline{p}_k \\ \underline{y}_k \end{pmatrix} = \begin{pmatrix} \underline{f}_k \\ 0 \end{pmatrix} \quad \text{with} \quad \mathcal{K}_k = \begin{pmatrix} K_k & M_k \\ M_k & -\gamma K_k \end{pmatrix}. \tag{2}$$

Here N_k denotes the number of nodes of the triangulation $\mathcal{T}_k$, M_k is the mass matrix representing the $L^2(\Omega)$ scalar product on $Y_k = Q_k$, and K_k is the stiffness matrix representing the $H^1(\Omega)$ scalar product on $Y_k = Q_k$.

3 The Multigrid Method

Next we describe the multigrid algorithm: One iteration step for solving (1) at level k is given in the following form:

Let $(p_k^{(0)},y_k^{(0)}) \in Q_k \times Y_k$ be a given approximation of the exact solution $(p_k,y_k) \in Q_k \times Y_k$ to (1). Then the iteration proceeds in two stages:

1. Smoothing: For $j = 0,1,\ldots,m-1$ compute $(p_k^{(j+1)},y_k^{(j+1)}) \in Q_k \times Y_k$ by an iterative procedure of the form
$$(p_k^{(j+1)},y_k^{(j+1)}) = \mathcal{S}_k\,(p_k^{(j)},y_k^{(j)}).$$
2. Coarse grid correction: Set
$$\tilde{\mathcal{F}}(q,z) = \mathcal{F}(q,z) - \mathcal{B}\big((p_k^{(m)},y_k^{(m)}),(q,z)\big)$$
for $(q,z) \in Q_{k-1} \times Y_{k-1}$ and let $(\tilde{s}_{k-1},\tilde{r}_{k-1}) \in Q_{k-1} \times Y_{k-1}$ satisfy
$$\mathcal{B}((\tilde{s}_{k-1},\tilde{r}_{k-1}),(q,z)) = \tilde{\mathcal{F}}(q,z) \quad \text{for all } (q,z) \in Q_{k-1} \times Y_{k-1}. \tag{3}$$
If $k = 1$, compute the exact solution of (3) and set $(s_{k-1},r_{k-1}) = (\tilde{s}_{k-1},\tilde{r}_{k-1})$.
If $k > 1$, compute approximations (s_{k-1},r_{k-1}) by applying $\mu \geq 2$ iteration steps of the multigrid algorithm applied to (3) on level $k-1$ with zero starting values.
Set
$$(p_k^{(m+1)},y_k^{(m+1)}) = (p_k^{(m)},y_k^{(m)}) + (s_{k-1},r_{k-1}).$$

3.1 The Patch Smoother

We will now define a space decomposition of $\mathbb{R}^{N_k} \times \mathbb{R}^{N_k}$ into N_k subspaces in terms of prolongation matrices $P_{k,i}$ and $Q_{k,i}$, $i = 1,\dots,N_k$, for the variables p and y, respectively: For each $i \in \{1,\dots,N_k\}$ representing a node of the triangulation, let $\mathcal{N}_{k,i}$ be the set of all indices consisting of i and the indices of all neighboring nodes (all nodes which are connected to the node with index i by an edge of the triangulation). Then, for each $i \in \{1,\dots,N_k\}$, the associated local patch consists of all unknowns of $\underline{p}_k$ which are associated to nodes with indices from $\mathcal{N}_{k,i}$ and of the unknown of $\underline{y}_k$ which is associated to the node with index i, see Fig. 1 for an illustration of a local patch. The corresponding canonical embeddings for the variables p and y from the local patches into $\mathbb{R}^{N_k}$ are denoted by $\widehat{P}_{k,i}$ and $Q_{k,i}$, respectively.

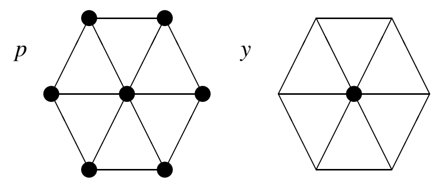

Fig. 1. Local patches

Observe that all entries in $\widehat{P}_{k,i}$ and $Q_{k,i}$ are either 0 or 1. A single component of $\underline{y}_k$ belongs to exactly one patch, while a single component of $\underline{p}_k$ belongs, in general, to more than one patch. Let $d_{k,j}$ be the local overlap depth at the node with index j, i.e., the number of all indices l with $j \in \mathcal{N}_{k,l}$, for $j = 1,\dots,N_k$. Let $\mathcal{D}_k$ be the $N_k \times N_k$ diagonal matrix whose diagonal entries are $d_{k,j}$, $j = 1,\dots,N_k$. The prolongation matrices $P_{k,i}$ are given by:

$$P_{k,i} = \mathcal{D}_k^{-1/2}\,\widehat{P}_{k,i}.$$

Now we can describe the smoothing procedure: Starting from some approximations $\underline{p}_k^{(j)}$ and $\underline{y}_k^{(j)}$ of the exact solutions $\underline{p}_k$ and $\underline{y}_k$ of (2) we consider iterative methods of form:

$$\underline{p}_k^{(j+1)} = \underline{p}_k^{(j)} + \omega \sum_{i=1}^{N_k} P_{k,i}\,\underline{s}_{k,i}, \quad \underline{y}_k^{(j+1)} = \underline{y}_k^{(j)} + \omega \sum_{i=1}^{N_k} Q_{k,i}\,\underline{r}_{k,i},$$

where $(\underline{s}_{k,i},\underline{r}_{k,i})$ solves a small local saddle point problem of the form

$$\widehat{\mathcal{K}}_{k,i} \begin{pmatrix} \underline{s}_{k,i} \\ \underline{r}_{k,i} \end{pmatrix} = \begin{pmatrix} P_{k,i}^T\big[\underline{f}_k - K_k\,\underline{p}_k^{(j)} - M_k\,\underline{y}_k^{(j)}\big] \\ Q_{k,i}^T\big[-M_k\,\underline{p}_k^{(j)} + \gamma K_k\,\underline{y}_k^{(j)}\big] \end{pmatrix} \quad \text{for all } i = 1,\dots,N_k.$$

The local matrix $\widehat{\mathcal{K}}_{k,i}$ is given by

$$\widehat{\mathcal{K}}_{k,i} = \begin{pmatrix} \widehat{K}_{k,i} & M_{k,i}^T \\ M_{k,i} & M_{k,i}\widehat{K}_{k,i}^{-1}M_{k,i}^T - \widehat{S}_{k,i} \end{pmatrix},$$

where

$$\widehat{K}_{k,i} = \widehat{P}_{k,i}^T \widehat{K}_k \widehat{P}_{k,i} \quad \text{with} \quad \widehat{K}_k = \frac{1}{\sigma}\,\mathrm{diag}\,K_k, \quad M_{k,i} = Q_{k,i}^T M_k \mathcal{D}_k^{1/2}\,\widehat{P}_{k,i}$$

and

$$\widehat{S}_{k,i} = \frac{1}{\tau}\left[\gamma Q_{k,i}^T K_k Q_{k,i} + M_{k,i}\widehat{K}_{k,i}^{-1} M_{k,i}^T\right].$$

The positive parameters σ and τ have to be chosen such that

$$\widehat{K}_k \geq K_k \quad \text{and} \quad \widehat{S}_k \geq \gamma K_k + M_k \widehat{K}_k^{-1} M_k \tag{4}$$

with

$$\widehat{S}_k = \Big(\sum_{i=1}^{N_k} Q_{k,i}\widehat{S}_{k,i}^{-1} Q_{k,i}^T\Big)^{-1}.$$

Observe that there is an additional relaxation factor ω in the smoothing procedure.

For the proposed multigrid method the following convergence result can be shown, see [6]:

Theorem 1. *Let $\omega \in (0,2)$. Then there exists a constant $C > 0$ such that*

$$\|(p_k^{(m+1)} - p_k, y_k^{(m+1)} - y_k)\| \leq C\, m^{-1/2}\, \|(p_k^{(0)} - p_k, y_k^{(0)} - y_k)\|,$$

where (p_k, y_k) is the solution of the discrete problem (1), $(p_k^{(0)}, y_k^{(0)})$ is the initial guess, $(p_k^{(m+1)}, y_k^{(m+1)})$ is the result of one multigrid iteration, and the norm is given by

$$\|(q,z)\| = \big(\|q\|^2_{L^2(\Omega)} + \|z\|^2_{L^2(\Omega)}\big)^{1/2}.$$

Therefore, the W-cycle multigrid method (i.e. $\mu = 2$) is a contraction with contraction number bounded away from one, independent of the grid level k, if the number m of smoothing steps is sufficiently large.

Remark 1. Observe that the norm used in the last theorem is the L^2-norm, which is weaker than the H^1-norm one would normally expect for the state and the adjoint state.

4 Numerical Experiments

Next we present some numerical results for the domain $\Omega = (0,1)\times(0,1)$ and homogeneous data $y_d = 0$. The initial grid consists of two triangles by connecting the nodes $(0,0)$ and $(1,1)$. For the first series of experiments the regularization parameter γ was set equal to 1. The dependence of the convergence rate on the regularization parameter γ was investigated subsequently.

Randomly chosen starting values for $\underline{p}_k^{(0)}$ and $\underline{y}_k^{(0)}$ for the exact solution $\underline{p}_k = 0$ and $\underline{y}_k = 0$ were used. The discretized problem was solved by a multigrid iteration with a W-cycle ($\mu = 2$) and $m/2$ pre- and $m/2$ post-smoothing steps. The multigrid iteration was performed until the Euclidean norm of the solution was reduced by a factor $\varepsilon = 10^{-8}$. All tests were done with $\sigma = \tau = 0.5$ in order to guarantee (4) and with $\omega = 1.6$, which is motivated by a Fourier analysis on uniform grids.

Table 1 contains the (average) convergence rates q depending on the level k, the total number of unknowns $2N_k$ and the number of smoothing steps, written in the form $m/2$ + $m/2$ (for $m/2$ pre- and $m/2$ post-smoothing steps). It shows a typical multigrid convergence behavior, namely the independence of the grid level and the expected improvement of the rates with an increasing number of smoothing steps.

Table 1. Convergence rates

level k	$2N_k$	1+1	2+2	3+3	5+5
5	2 178	0.301	0.127	0.067	0.023
6	8 450	0.302	0.128	0.066	0.024
7	33 282	0.302	0.135	0.067	0.024
8	132 098	0.302	0.135	0.067	0.024
9	526 338	0.302	0.135	0.068	0.024

Table 2 shows the convergence rates obtained at grid level 7 with 1 pre- and 1 post-smoothing step for values of γ ranging from 1 down to 10^{-6}. Although the analysis presented here does not predict convergence rates that are robust in γ, the numerical experiments indicate robustness with respect to the regularization parameter.

Table 2. Dependence on the regularization parameter γ

γ	1	10^{-2}	10^{-4}	10^{-6}
q	0.302	0.302	0.302	0.302

In summary, the numerical experiments confirm the theoretical results of level-independent convergence rates for the multigrid method with the proposed patch smoother. The convergence rates are much better than in [5] and comparable with the rates presented in [2]. Moreover, they strongly support the conjecture that the convergence rates are also independent of the regularization parameter, as already stated in [2] for the point smoother on the basis of a Fourier analysis on uniform grids.

Acknowledgement. The work was supported by the Austrian Science Fund (FWF) under grant SFB 013/F1309.

References

[1] Arian, E., Ta'asan, S.: Multigrid one-shot methods for optimal control problems: Infinite dimensional control. ICASE-Report 94-52, NASA Langley Research Center, Hampton VA, 1994.

[2] Borzì, A., Kunisch, K., 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 (electronic), 2002.

[3] Lions, J.-L.: *Optimal control of systems governed by partial differential equations.* Translated from the French by S. K. Mitter. Die Grundlehren der mathematischen Wissenschaften, 170. Springer, New York, 1971.

[4] Schöberl, J., Zulehner, W.: On Schwarz-type smoothers for saddle point problems. *Numer. Math.*, 95(2):377–399, 2003.

[5] Simon, R., Zulehner, W.: On Schwarz-type Smoothers for Saddle Point Problems with Applications to PDE-constrained Optimization Problems. *Numer. Math.*, 111:445 – 468, 2009.

[6] Simon, R.: *Multigrid solvers for saddle point problems in PDE-constrained optimization*. PhD thesis, Johannes Kepler University Linz, Austria, 2008.

[7] 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.

[8] Vanka, S.P.: Block-implicit multigrid solution of Navier-Stokes equations in primitive variables. *J. Comput. Phys.*, 65(1):138–158, 1986.