5 A Primal-Dual Active-Set Multigrid Method for Control-Constrained Optimal Control Problems

In this chapter we consider optimal control problems with additional inequality constraints imposed on the control unknown $u$ and for their efficient solution we combine a primal-dual active-set strategy with the multigrid method developed in the previous chapter. Control-constraints are specified by the condition $u \in U_{ad}$, where the set of *admissible controls* $U_{ad} \subset L^2(\Omega)$ is a proper subset of $L^2(\Omega)$ and is assumed to be closed and convex. In particular, we consider the problem

$$\begin{align*}
\text{minimize} & \quad J(y, u) \\
\text{subject to} & \quad C(y, u) = 0 \\
& \quad u \in U_{ad},
\end{align*}$$

where $J$ and $C$ are as in Example 2.12, i.e. $J$ is the tracking-type functional (2.1) and the constraints $C$ are given by the second-order linear elliptic partial differential equation (2.24). The set $U_{ad}$ is defined by so-called *box-constraints*

$$u \in U_{ad} = \{ u \in L^2(\Omega) \mid \alpha \leq u \leq \beta \text{ a.e. in } \Omega \},$$

with given functions $u^{\alpha}, u^{\beta} \in L^\infty(\Omega), \ u^{\alpha} \leq u^{\beta} \text{ a.e. in } \Omega$. Obviously the set $U_{ad}$ defined by (5.1) is closed and convex. Existence and uniqueness of a solution then follows from Theorem 2.4. Theorem 2.8 yields the first-order conditions with the corresponding optimality system (OS), where the optimality condition is given by (2.17).

**Remark 5.1 (Regularity of the Optimal Control).** For $U_{ad} = \{ u \in L^2(\Omega) \mid u \geq 0 \text{ a.e. in } \Omega \}$ one obtains $u \in H^1_0(\Omega)$. In the case of the box-constraints (5.1) one obtains $u \in H^1(\Omega)$ if the bounding functions $u^{\alpha}, u^{\beta}$ are sufficiently regular, in particular $u^{\alpha}, u^{\beta} \in L^\infty(\Omega) \cap H^1(\Omega)$. Without additional assumptions on the regularity of $u^{\alpha}, u^{\beta}$ however, we can only expect $u \in L^2(\Omega)$, cf. [111, Ch. II, Remark 2.3].

5.1 Finite Dimensional Approximation

The approach most common for the finite dimensional approximation of control-constrained optimal control problems is the discretize-then-optimize methodology. To this end, the objective functional $J$, the constraints $C$ and the admissible set $U_{ad}$
viz. the bounding functions $u^\alpha, u^\beta$ are approximated by discrete versions. Existence and uniqueness of a solution $y_h^*, u_h^*$ to the optimization problem in finite dimension immediately follows. We remark that several algorithms can be formulated in an appropriate function space setting and can be applied directly to (OS). A resulting sequence of infinite-dimensional systems then has to be discretized. Convergence can be proved using the interpretation as a semismooth Newton method, we refer to [91, 148] and to [92] for a mesh-independence result. In the present setting however, where we focus on devising an efficient algorithm, a discrete version of (OS) is the appropriate starting point and will be derived now. Discretizing $\mathcal{C}$ as in Section 2.3.2, we obtain the discretized state equation (2.50f) and the corresponding discrete adjoint equation (2.50b). It remains to discretize (5.1) and (2.17), which in the concrete setting is stated in (2.19). To this end, we discretize the bounding functions $u^\alpha, u^\beta$ by piecewise constants and obtain the discrete approximation to the set of admissible controls $\mathcal{U}_{ad}$.

$$\mathcal{U}_{ad,h} = \{ v_h \in U_h \mid u_h^\alpha \leq v_h \leq u_h^\beta \}. \quad (5.2)$$

The $L^2$-inner products appearing in the variational inequality are discretized employing midpoint quadrature, resulting in the same mass matrix $M_h$ as in (2.50f), (2.50b). We then obtain

$$\left( \sigma M_h u_h - M_h p_h \right)^T (v_h - u_h) \geq 0, \quad v_h \in \mathcal{U}_{ad,h} \quad (5.3)$$

as discrete optimality condition. The complete discrete optimality system corresponding to (OS), (2.17) is given by (2.50b), (2.50f) and (5.3). Error estimates have been given in Section 2.3.3.

## 5.2 Multigrid Methods for Variational Inequalities

The two most prominent methods to treat optimization problems with inequality constraints are interior-point methods and active set strategies. Both approaches yield a sequence of equality-constrained quadratic programming problems (QP) and therefore rely on an efficient solution method for such QPs. Currently the state-of-the-art methods discussed in Chapter 3 are frequently employed for their solution and the bulk of numerical methods is made up by their combination with interior-point or active-set methods. Only a few publications consider the application of multigrid within the context of inequality-constrained PDE constrained optimization problems.

The interior-point approach aims at maintaining the complementarity condition at all intermediate steps and introduces (logarithmic) barrier functions. As a result, the QPs can become increasingly ill-conditioned when the interior-point algorithm approaches the solution. The goal of active set methods is to enforce the feasibility of intermediate iterates with respect to the constraints. To this end, at each step the constraints are partitioned into sets of active and inactive constraints with respect to
the current iterate, hence the name of the method. A particular class of active-set
methods is the primal-dual active-set (PDAS) strategy [24, 91]. The PDAS method
is closely related to the projected Newton method of [25], and can be formulated as
a semismooth Newton method in infinite dimension. As such, the PDAS approach
shares important properties with Newton methods, namely superlinear or quadratic
local convergence and a mesh-independence principle [92]. Therefore, we expect an
inner-outer iterative method, defined by the combination of our multigrid method
for the solution of the QPs and the semismooth Newton method to generate the
QPs, to be an efficient and mesh-independent solution method for PDE constrained
optimization problems with inequality constraints on the control. Furthermore, for
control-constrained optimal control problems, in [23] the PDAS approach is shown to
be more efficient than interior-point methods. For the combination of multigrid and
interior-point methods, albeit not in the context of optimal control, we refer to [149].
Combinations of active-set strategies with multigrid have been proposed in [87] and
later in [98] for minimal surface and other obstacle problems. In case the active and
inactive sets generated by the PDAS method and the approach in [98], both outer
iterations yield the same inner linear system. However, in contrast to [98], which is a
primal method only, in [91] dual information is used to predict the active set.

In contrast to inner-outer iterative schemes one can define a variant of a multigrid
method which treats the inequality constraints within the smoothing iteration. In the
context of a semismooth Newton method, this amounts to a local linearization pro-
cess in the smoother instead of a global one in an outer iteration. Thus, this multigrid
approach is related to the Full Storage Approximation (FAS) scheme [40], which was
mainly developed to treat nonlinear problems. In [42], the projected FAS (PFAS)
has been developed and applied to complementarity formulations of free-boundary
problems. In [33, 35], the collective Gauss-Seidel smoother of the CGSM multigrid
approach (cf. Section 4.2) has been augmented by a local projection step $u_h = P_{\Pi_{U^0}}\tilde{u}_h$
to allow for inequality constraints on $u_h$. To our knowledge, this is the only published
application of a PFAS-like method in the context of PDE constrained optimization.
Related approaches such as the monotone multigrid method [107], which also em-
ploy projected Gauss-Seidel smoothing, are applied mostly to variational inequalities
derived from obstacle- and contact-problems in elasticity and mechanics.

A major advantage of treating the inequality constraints in an outer iteration is as
follows: The resulting inner subsystem and in particular the smoother of a multigrid
method applied there does not need to take the constraints into account. Thus, the
inner systems to be solved are strictly linear and equality-constrained and the multigrid
method of Chapter 4 can be applied as QP subsolver with only minor modifications.
The connection to semismooth Newton methods lets us expect superlinear convergence
and thus a fast detection of the active sets within the outer iteration.\footnote{In this respect the PDAS method differs significantly from primal active-set methods like the Simplex algorithm since many constraints per iteration can be identified.}
next-neighbor related identification of active nodes within a projected Gauss-Seidel smoother can lead to slower identification of the active set.

5.3 The Primal-Dual Active Set Method

In this section we briefly describe the primal-dual active set strategy that will be used as outer iteration to handle the control-constraints [23, 24, 91]. To this end, we note that for the discrete variational inequality (5.3), an equivalent formulation is given by

$$\sigma M_h u_h - M_h p_h + \lambda = 0,$$

$$\lambda = \max(\lambda + c(u_h - u^\beta_h), 0) + \min(\lambda + c(u_h - u^\alpha_h), 0), \quad c > 0.$$  

Here, the unknowns $\lambda$ are the Lagrange multipliers associated with the inequality constraints. They satisfy the Karush-Kuhn-Tucker conditions

$$\lambda \leq 0 \quad \text{on} \quad T_h, A^- \quad \text{on} \quad \{i \mid i \in A^- \} \quad \text{on} \quad \{i \mid i \in A^- \},$$

$$\lambda \geq 0 \quad \text{on} \quad T_h, A^+ \quad \text{on} \quad \{i \mid i \in A^+ \} \quad \text{on} \quad \{i \mid i \in A^+ \},$$

$$\lambda = 0 \quad \text{on} \quad T_h, I \quad \text{on} \quad \{i \mid u^\alpha_h < u_h < u^\beta_h \ \text{on} \ T_i \} \quad \text{on} \quad \{i \mid u^\alpha_h < u_h < u^\beta_h \ \text{on} \ T_i \}.$$ 

Here, $A^-$ and $A^+$ are the active sets and $I$ is the inactive set at the (discrete) optimal solution $u^*_h$. To unburden the notation, active and inactive sets are not designated with the discretization index $h$, however, they always refer to the discrete unknowns.

The primal-dual active set strategy is an iterative algorithm that makes use of (5.4) to predict the active and inactive sets and treats an associated equality constrained optimization problem at each step. This leads to Algorithm 7. Note that for the case without control constraints, we have $A^- = A^+ = \emptyset$ and the overall algorithm reduces to just the solution of (EQP), which in turn reduces to the saddle point system (2.51). This concludes the description of the PDAS method. For details and convergence properties we refer to [24]. In particular, it is proved there that, if Algorithm 7 stops due to the criterion in line 4, the solution of (EQP) is the solution of the original optimality system. No additional stopping criterion was required in our implementation, in our numerical experiments Algorithm 7 always stopped due to the rule in line 4.

The main computational effort in this algorithm has to be spent for the solution of (EQP). In most publications this is done by solving the associated reduced system for the control unknowns $u_h^k$ by a conjugate gradient method, i.e. the methods discussed in Chapter 3 are applied to (EQP). In our context it is natural to employ the multigrid method developed in the preceding chapter for the solution of (EQP). To this end, we modify the system (EQP) in such a way that it can be formulated as a KKT system (2.51). We proceed as follows: First, we partition the control unknowns according to $u_h^k = [u_h^{T_k} \quad u_h^{A^-} \quad u_h^{A^+}]$. The same partitioning applies to the Lagrange multipliers $\lambda^k = [\lambda^{T_k} \quad \lambda^{A^-} \quad \lambda^{A^+}]$. Note that this partitioning induces corresponding $3 \times 3$
5.3 The Primal-Dual Active Set Method

The Outer PDAS Iteration

1. Choose initial values $y^0_h, u^0_h, p^0_h, \lambda^0$ and set $k = 1$
2. while not converged do
3. predict $A^k_-, A^k_+, I^k$ as follows:
   \[ A^k_- = \{ i \mid u^k_h - 1 < u^k_h \sigma < u^k_h \alpha \text{ on } T_i \} \] \hspace{1cm} (5.6a)
   \[ A^k_+ = \{ i \mid u^k_h - 1 > u^k_h \beta \text{ on } T_i \} \] \hspace{1cm} (5.6b)
   \[ I^k = \{ i \mid i \notin A^k_- \cup A^k_+ \} \] \hspace{1cm} (5.6c)
4. if $k \geq 2$ and $A^k_- = A^{k-1}_-, A^k_+ = A^{k-1}_+, I^k = I^{k-1}$ then
5. converged = true
6. else
7. solve the equality-constrained problem
   \[ M_h y^k_h + L_h p^k_h = M_h \bar{y}_h \]
   \[ \sigma M_h u^k_h - M_h^T p^k_h + \lambda^k = 0 \]
   \[ L_h y^k_h - M_h u^k_h = M_h f_h \] \hspace{1cm} (EQP)
   \[ \lambda^k = 0 \text{ on } T_h, \bar{T}_h \]
   \[ u^k_h = u^\alpha_h \text{ on } T_h, A^k_- \]
   \[ u^k_h = u^\beta_h \text{ on } T_h, A^k_+ \]
8. $k = k + 1$

Algorithm 7: The Primal-Dual Active-Set Strategy as outer iteration.

block, $3 \times 1$ column and $1 \times 3$ row block partitions of the mass matrix $M_h$. Then, the system given by the first three lines in (EQP) can be written as

\[
\begin{bmatrix}
M_h & L_h^T & 0 \\
\sigma M^T_h \bar{T} & \sigma M^T_h A^- & \sigma M^T_h A^+ & -M^T_h \bar{\lambda} \\
\sigma M^T_h A^- & \sigma M^T_h A^+ & -M^T_h \bar{\lambda} & -M^T_h \bar{\lambda} \\
L_h & -M_h \bar{T} & -M_h A^- & -M_h A^+
\end{bmatrix}
\begin{bmatrix}
y^k_h \\
u^k_h \\
\bar{\lambda} \\
p^k_h
\end{bmatrix}
= \begin{bmatrix}
M_h \bar{y}_h \\
-M_h \bar{\lambda} \\
-M \bar{\lambda} \\
M_h f_h
\end{bmatrix}.
\] \hspace{1cm} (5.7)

Now we utilize the last three equations in (EQP) to reduce (5.7) to a system for
set \( y^k_h, u^k_h, p^k_h \), i.e. we eliminate \( u^A_h, u^A_h \) and we consider the controls \( u^k_h \) only on the inactive set \( \mathcal{I}^k \). The solution of (EQP) then proceeds in two steps: First, the saddle point system

\[
K_h^T x_h^k = r_h^k
\]

has to be solved, where

\[
K_h^T = \begin{bmatrix}
M_h & \sigma M_h^{T_2} & L_h^T \\
\sigma M_h^{T_2} & -M_h^{T_2,a} & \sigma M_h^{T_2,a}
\end{bmatrix}, \quad r_h^k = \begin{bmatrix}
M_h y_h \\
-\sigma M_h^{T_2,a} u_h^a - \sigma M_h^{T_2,a} u_h^\beta
\end{bmatrix},
\]

and the vector of unknowns is given by \( x_h^T = [y_h^k u_h^k p_h^k] \). Note that the KKT operator \( K_h^T \) and the right hand side vector \( r_h^T \) depend on the index \( k \) of the outer iteration. In the second step, the Lagrange multipliers \( \lambda^k \) are computed by

\[
\lambda^{T_2} = M_h^{A^k, p} \sigma M_h^{A^k} u_h^{T_2} - \sigma M_h^{A^k, u} u_h^{T_2} - \sigma M_h^{A^k, A^k} u_h^{T_2} u_h^\beta,
\]

compare the lines 3 and 4 in (5.7). On the inactive set, we just set \( \lambda^{T_2} = 0 \). Note again that, for the case without control constraints, the system (5.8) reduces to (2.51), i.e. we just have \( K_h^T = K_h \), \( r_h^T = b_h \) and \( x_h^T = x_h \). Finally, we remark that in the context of a projected Newton method the operator \( K_h^T \) corresponds to the projected\(^2\) Hessian [25].

### 5.4 A PDAS Multigrid Method for the Solution of Control-Constrained Optimal Control Problems

It remains to obtain the solution of (5.8) in a fast and efficient fashion. To this end, we proceed by modifying the multigrid method of Chapter 4. For a fixed index \( k \) we set \( \mathcal{I} = \mathcal{I}^k \) and define the smoothing iteration by the modification

\[
w_j^{T,r+1} = w_j^{T,r} + (B_{K_j}^T)^{-1}(b_j^{T} - K_j^{T_2} w_j^{T_2}).
\]

of iteration (4.23). The inexact constraint preconditioner (4.24) now takes the form

\[
B_{K_j}^T = \begin{bmatrix}
\hat{L}_j^T & \bar{H}_j^{T_2} \\
\hat{L}_j & -M_j^{T^2, T_2}
\end{bmatrix},
\]

\(^2\)In the literature, the name reduced Hessian is often used and in this context is meant in the sense of reducing to inactive constraints, not to be confused with the reduced Hessian \( H_Z \) in the sense of Chapters 3 and 4.
with the modified reduced Hessian given by

\[ \hat{H}_{I_j}^T = M_j^{-1} L_j^{-T} M_j L_j^{-1} M_j^{-1} + \sigma M_j^{-1} \]

(5.13)

instead of (4.26). Using (5.11)–(5.13), the smoothing step then follows from Algorithms 5 and 6 and is denoted in compact form as

\[ \tilde{w}_j^T = (S_{j,\alpha,\beta}^T)^\nu (w_j^T, b_j^T) \]

(5.14)

with the same parameters \( \nu, \alpha, \beta \) as for (4.27).

From the two-step solution of (EQP) it follows that the Lagrange multipliers \( \lambda \) as well as the bounding functions \( u^\alpha \) and \( u^\beta \) need to be discretized on the finest grid level only. However, for the discretization of \( x_{h_j}^T, b_{h_j}^T \) and the operator \( K_{h_j}^T \) in (5.8) on a grid level \( j < J \), it is evident that we have to approximate the inactive set \( I \) on that grid level. This has to be done in each PDAS iteration, after the inactive and active set on the finest level \( J \) have been detected by the algorithm and before the solution of the system (5.8). In the context of node-based discretizations for obstacle problems, two different strategies have been previously used. In [87], the inactive set shrinks when reduced to coarser grids and [98] later followed the same strategy. In [42], the size of the inactive set increases with reduced level index. The difficulty of representing \( I \) for levels \( j < J \) has been circumvented altogether in [30] by employing cascadic multigrid. Here, the iteration starts on the coarsest level \( j = 0 \), and never returns to the coarser grids. On each grid level, the active and inactive sets are determined based on the current approximation. This leads to a fast detection of \( I \) due to the nested iteration approach, however the convergence speed of cascadic multigrid suffers from the non-existent coarse grid correction. A further strategy is the modification of coarse grid basis functions which is employed within monotone multigrid.

Consistent with the cell-centered discretization, here we proceed as follows: The third step of the PDAS algorithm yields index sets \( \mathcal{I}, \mathcal{A}_- \) and \( \mathcal{A}_+ \) on the finest level \( J \) and we denote the set of grid cells corresponding to \( \mathcal{I} \) with \( \mathcal{T}_J \). Now, for given \( \mathcal{T}_j \) we define \( \mathcal{T}_{j-1} \) as the set of all coarser grid cells for which at least one fine-grid subcell is contained in \( \mathcal{T}_j \), i.e.

\[ \mathcal{T}_{j-1} = \{ T_i \in \mathcal{T}_{j-1} \mid T_i^s \in \mathcal{T}_j \text{ for } s \in \{1,2,3,4\} \}, \quad j = J, \ldots, 1. \]

(5.15)

In our numerical experiments, this “outer” approximation of \( \mathcal{I} \) on coarser levels has shown to yield faster convergence than the alternative “inner” approximation. Most likely the difference is due to the coarse grid correction: when \( \mathcal{I} \) shrinks with decreasing \( j \), no coarse grid correction occurs for the current iterate on \( \mathcal{T}_j \setminus \mathcal{T}_{j-1} \), and the error on \( \mathcal{T}_j \setminus \mathcal{T}_{j-1} \) has to be improved exclusively by the smoothing iteration.

In any case, the coarsest grid has to be fine enough to avoid \( \mathcal{A}_{-,0} = \mathcal{A}_{+,0} = \emptyset \), since otherwise the bounds \( u^\alpha, u^\beta \) would be completely removed from the coarse grid correction process which necessarily leads to degradation of convergence. Note that
no representation of the active sets $A_-$ and $A_+$ is needed on coarser levels. From
the sequence of meshes given by (5.15) we then obtain a sequence of operators $K^{j}$ as
before by direct discretization.

The intergrid transfer operators follow from (4.60) with minor modifications with
respect to the control component $u^{j}$. They are given by

$$
R^{j-1} = \begin{pmatrix}
R^{j-1}_{1} & R^{j-1,1}_{1} \\
R^{j-1}_{2} & R^{j-1}_{2}
\end{pmatrix}, \quad P^{j-1}_{j} = \begin{pmatrix}
P^{j}_{j-1} & P^{j,1}_{j-1} \\
P^{j}_{j-1} & P^{j}_{j-1}
\end{pmatrix}.
$$

(5.16)

In (5.16), the symbol $R^{j-1,1}_{1}$ denotes the four-point average operator (4.12) giving
values only for grid cells $T_{i} \in T_{j-1}$. When applying $R^{j-1,1}_{1}$ to obtain a coarse grid
value of the $u^{j-1}$-component, fine grid values on the active set $T_{j} \setminus T_{j-1}$ could be needed.

However, on $A_{-}$ and $A_{+}$ the solution is fixed to the constraints $u^{\alpha}, u^{\beta}$, respectively
and the corresponding residuals vanish. Thus, active nodes should provide no contri-
bution here and consistently the corresponding stencil entries are set to zero. Similar
considerations apply to the prolongation $P^{j,1}_{j-1}$. Here, coarse grid values on the active
set $T_{j-1} \setminus T_{j}$ could enter the prolongation stencil when computing the correction for
a fine grid value of $u^{j}$. However again the corrections due to active nodes should be
zero and the corresponding stencil entries are set to zero. With these modifications,
the multigrid Algorithm 3 can be applied for the solution of (5.8), which concludes
the description of the PDAS-multigrid method.

Some additional remarks are in order when using the full multigrid Algorithm 4 for
the solution of (5.8). First we note that now the bounds $u^{\alpha}, u^{\beta}$ need to be discretized
on all levels $0 \leq j \leq J$ for two reasons: first the right hand side vector $r^{j}_{1}$ (5.9) needs to
be constructed on all levels $j < J$, and second, the FMG prolongation $P^{j-1}_{j-1}$ transfers
the solution $u^{j}_{1}$ which requires that the actual values of $u_{j}$ on $A_{-}$ and $A_{+}$ need to
be taken into account. For these reasons, additionally the active sets $A_{-}, A_{+}$ need to
be represented on each grid level. To this end, instead of proceeding as in (5.15), we
restrict $A_{-}$ and $A_{+}$ according to

$$
T_{A_{\pm},j-1} = \{ T_{i} \in T_{j-1} \mid \cup T_{s}^{s} \in T_{A_{\pm},j} \text{ for } s \in \{1, 2, 3, 4\} \}, \quad j = J, \ldots, 1,
$$

(5.17)

where $A_{\pm}$ stands for $A_{-}$ or $A_{+}$. After the restriction step (5.17) we set

$$
A_{j-1} = A_{-j-1} \cup A_{+j-1} \text{ and } T_{j-1} = T_{j} \setminus T_{A_{j-1}}.
$$

(5.18)

Recall that for the conventional multigrid cycle only the sequence $T_{j}, j = J, \ldots, 0$ was
needed.

We note that this strategy differs from other nested iteration approaches to obstacle
problems in the following sense: Here, we employ the full multigrid with the purpose of
solving (5.8) for fixed sets $I_{j}$ and $A_{-j}, A_{+j}$, as provided by the outer PDAS iteration
on level $J$. From results presented in Section 4.5.1 we can expect that this is achieved
with optimal complexity. We do not employ the nested iteration strategy to change the current predictions of \( I_j \) and \( A_{-j}, A_{+j} \). The latter strategy would be natural when using an FAS-based method.

## 5.5 Numerical Results

In this section we conduct several numerical experiments to test the convergence and efficiency of the proposed PDAS-multigrid method. We consider the model problem (LQP \( h \)) with different lower and upper bounds \( u_\alpha^J, u_\beta^J \). The convergence factor of the inner multigrid iteration is measured analogously to Section 4.5.1. Here, \( \varrho^m \) is based on the fine grid residual \( res_{J}^{Tk,m} \) of the EQP (5.8) which is defined as

\[
res_{J}^{Tk,m} = r_{J}^{Tk} - K_{J}^{Tk} x_{J}^{Tk,m}
\]

for each index \( k \) of the outer PDAS iteration. As before, the superscript \( m \) is the index of the multigrid iteration. With respect to the PDAS iteration, we use the fact that the optimal control \( u^* \) satisfies the projection formula

\[
u^* = \Pi_{U_{ad}}(\frac{1}{\sigma^p}) (5.20)
\]

to construct an exact solution \( u^* \) and measure the error \( e^k_{u_j} \) in the discrete \( L^2 \)-norm, cf. (3.24). Additionally we consider the violation of the bounds

\[
e^k_{a} = \max_{T \in T_{J}}(u^a_{J} - u^k_{J}), \quad e^k_{\beta} = \max_{T \in T_{J}}(u^k_{J} - u^\beta_{J}). \quad (5.21)
\]

For the multigrid solution of (5.8) we employ the smoothing iteration \( S_{J,1,1}^{\nu} \) and use GS-LEX as constraint smoother. Unless noted otherwise, the inner iterations are stopped as soon as \( res_{J}^{Tk,m} \leq \max(1_{-16}res_{J}^{Tk,0}, 1_{-12}) \) or \( m = 20 \). Furthermore, the coarse grid size is set to \( h_0 = 1/8 \).

In [23] different strategies are discussed for the initialization of the PDAS algorithm. It was noted that the algorithm is rather insensitive to the initial value \( x_0^J \). A special initialization proposed requires the solution of the state and adjoint equation for a given feasible initial control \( u_0^J \). The second strategy consists of solving the unconstrained problem and we will adopt this approach. To this end, we use a general initial value \( x_0^J \) (cf. Section 4.5.1) and set

\[
T_{I,J} = T_{J}, \quad A^0 = \emptyset, \quad \lambda \equiv 0. \quad (5.22)
\]

With these choices, the first PDAS iteration yields the solution of the unconstrained problem since (5.8) reduces to (2.51).
5 A PDAS Multigrid Method for Constrained Optimal Control

Figure 5.1: Computed optimal control $u^*_J$ (left) and corresponding active set $A^*_J$ (shaded region, right) on a mesh with $h_J = 2^{-8}$ for (LQP$_h$) with control-constraints (5.23).

Figure 5.2: Discrete $L^2$-error of the control $u_J$ (left) and error $e^\beta_u$ (right) for the outer PDAS iteration and different number of levels $J$ with $h_J = 2^{-(J+3)}$.

5.5.1 A Model Problem

First we consider the unilaterally constrained problem with

$$u \leq u^\beta = 0.$$  

(5.23)

The computed optimal control $u^*_J$ and the corresponding active set $A^*$ are depicted in Figure 5.1 on a mesh with $h_J = 2^{-8}$. Figure 5.2 shows the iteration history of the outer PDAS iteration for different levels $J$ with a mesh size of the fine grid given by $h_J = 2^{-(J+3)}$. On the left, we plot the discrete $L^2$-error of the control $u_J$ and on the right we show the error $e^\beta_u$. Note that for $k = 4$ the computed solution satisfies $e^\beta_u = 0$. Further recall that in the first iteration, the unconstrained problem is solved. Thus, $I^*$ and $A^*$ are detected in 3 iterations, independent of the level number $J$ (disregarding that one less iteration is needed for $J = 4$, i.e. the lowest resolution of
5.5 Numerical Results

The mark ◦ indicates the respective initial residual $\text{res}^{T_{k,0}}_J$.

Table 5.1: Average convergence rate of the $V_{1,1}$-multigrid cycle for the solution of (5.8) at each step $k$ of the PDAS iteration.

<table>
<thead>
<tr>
<th>$J$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>8.99_{-2}</td>
<td>1.08_{-1}</td>
<td>1.06_{-1}</td>
<td>1.09_{-1}</td>
</tr>
<tr>
<td>7</td>
<td>8.86_{-2}</td>
<td>1.08_{-1}</td>
<td>1.04_{-1}</td>
<td>9.48_{-2}</td>
</tr>
<tr>
<td>6</td>
<td>8.93_{-2}</td>
<td>1.10_{-1}</td>
<td>1.05_{-1}</td>
<td>9.36_{-2}</td>
</tr>
<tr>
<td>5</td>
<td>9.17_{-2}</td>
<td>1.10_{-1}</td>
<td>1.07_{-1}</td>
<td>9.49_{-2}</td>
</tr>
<tr>
<td>4</td>
<td>9.51_{-2}</td>
<td>1.09_{-1}</td>
<td>1.01_{-1}</td>
<td>—</td>
</tr>
</tbody>
</table>

the fine mesh). The PDAS convergences at a superlinear rate, furthermore, for the final error obtained in iteration $k = 4$ we have $e_J^k \sim O(h_J^2)$ (this will be confirmed shortly below). In Figure 5.3 the reduction of $\|\text{res}^{T_{k,m}}_J\|$ is shown for each $k = 1, 2, 3, 4$. The symbol ◦ denotes the begin of each PDAS step $k$. The corresponding average convergence rates for different $J$ and in each PDAS iteration are given in Table 5.1. We observe that the convergence of the inner multigrid iteration is independent of the outer iteration and for each system (5.8) corresponds to the rates which have been obtained for the unconstrained model problem in Section 4.5.1. The rather minuscule variations in $\varrho_{\text{avg}}$ are due the changing $\mathcal{I}$, which affects smoothing and more so the coarse grid correction. In Table 5.2 we present the discrete $L^2$-error $e_J^\sigma$ obtained in the final PDAS step. In the second and third column, we give the errors and associated ratios which have been obtained by solving (5.8) with the standard multigrid solver, i.e. iterations with the $V_{1,1}$-cycle have been performed until the stopping criterion applied. Clearly second-order convergence is observed. In the fourth and fifth column we give the same data which has been computed by using just one iteration of the full multigrid for each PDAS step. Within the FMG, the same $V_{1,1}$-cycle has been used.
The absolute error is roughly one order of magnitude larger than for the fully converged multigrid solution, but it still reduces at the same rate. In the last two columns, we again present the same data, however this time the FMG has been followed by one additional \( V_{1,1} \)-cycle. This reduces the error to the same order of magnitude as that of the conventional multigrid solver. These results reflect the situation for scalar elliptic problems. Under the assumption that the convergence rate of the employed multigrid cycle is smaller than 1/6, one FMG iteration yields an approximate solution with an error of \( (5/2)ch^2 \) and one additional multigrid cycle reduces that error below \( (1/2)ch^2 \). Here, \( c \) is the constant from the error estimate \( \|u_j^* - u^*\| \leq ch^2 \). Comparison of the relative performance in terms of wall-clock time will be given below for a different example.

As the second test case let us consider the bilaterally constrained problem with

\[
u^a = \begin{cases} 
-0.75 & \text{for } y \leq 0.5 \\
-0.9 & \text{for } y > 0.5
\end{cases} \quad \text{and } u^b = y^3 - 0.5.
\] (5.24)

The computed optimal control \( u_j^* \) and \( A^* = A_\alpha^+ \cup A_+^\alpha \) are depicted in Figure 5.4 left and right, respectively, computed on a mesh with \( h_J = 2^{-8} \). In Figure 5.5 we show the discrete inactive and active set on levels \( J = 0, 1, 2, 3 \) as generated by the coarsening process (5.15). In Figure 5.6 the iteration history of the outer PDAS iteration is shown. The inner systems are solved with the FMG followed by an additional multigrid cycle. On the left, we show the error \( e_j^u \), on the right, we show \( e_j^v \). The error \( e_{j}^{u^k} \) vanishes already in the second iteration and is not displayed here. Again, superlinear convergence of the outer iteration is clearly visible. The final error obtained for \( k = 4 \) again is second-order convergent with respect to \( h_J \). The corresponding data is given in Table 5.2, where we also list the computing time in seconds. The numbers confirm the optimal complexity \( O(3h_J^{-2}) \) of the FMG solver.

<table>
<thead>
<tr>
<th>( J )</th>
<th>( V_{1,1} )</th>
<th>FMG</th>
<th>FMG + ( V_{1,1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( e^4_{u_j} )</td>
<td>( e^4_{u_j} )</td>
<td>( e^4_{u_j} )</td>
</tr>
<tr>
<td>3</td>
<td>2.9225_5</td>
<td>—</td>
<td>3.1906_4</td>
</tr>
<tr>
<td>4</td>
<td>7.3051_6</td>
<td>2.49_1</td>
<td>8.3699_5</td>
</tr>
<tr>
<td>5</td>
<td>1.8262_7</td>
<td>2.50_1</td>
<td>2.1282_5</td>
</tr>
<tr>
<td>6</td>
<td>4.5655_7</td>
<td>2.50_1</td>
<td>5.3510_6</td>
</tr>
<tr>
<td>7</td>
<td>1.1414_7</td>
<td>2.50_1</td>
<td>1.3403_6</td>
</tr>
<tr>
<td>8</td>
<td>2.8534_8</td>
<td>2.50_1</td>
<td>3.3527_7</td>
</tr>
</tbody>
</table>
Figure 5.4: Computed optimal control $u^*_J$ (left) and corresponding active set $\mathcal{A}^* = \mathcal{A}^*_+ \cup \mathcal{A}^*_-$ (shaded region, right) on a mesh with $h_J = 2^{-8}$ for (LQP$_h$) with control-constraints (5.24).

Figure 5.5: Inactive set $I^*$ and active set $\mathcal{A}^*_\pm$ (shaded) generated by coarsening (5.15) on levels $J = 0, 1, 2, 3$ for (LQP$_h$) with control-constraints (5.24).
A PDAS Multigrid Method for Constrained Optimal Control

Figure 5.6: PDAS iteration history for (LQP\(_h\)) and constraints (5.24). The inner system (5.8) was solved with FMG followed by one \(V_{1,1}\)-cycle. Shown are the discrete \(L^2\)-error of the control (left) and the violation of the upper bound (right).

Table 5.3: Discrete \(L^2\)-error of the control \(u_J^k\) for (LQP\(_h\)) with control-constraints (5.23) and wall-clock time for the solution with FMG plus one additional \(V_{1,1}\)-cycle.

<table>
<thead>
<tr>
<th>(J)</th>
<th>(e_{u,J}^k)</th>
<th>Ratio</th>
<th>time [s]</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.4246(_{-6})</td>
<td>—</td>
<td>1.2723(_{+1})</td>
<td>—</td>
</tr>
<tr>
<td>6</td>
<td>8.5909(_{-7})</td>
<td>2.51(_{-1})</td>
<td>5.9187(_{+1})</td>
<td>4.65</td>
</tr>
<tr>
<td>7</td>
<td>2.1570(_{-7})</td>
<td>2.51(_{-1})</td>
<td>2.6605(_{+2})</td>
<td>4.49</td>
</tr>
<tr>
<td>8</td>
<td>5.4015(_{-8})</td>
<td>2.50(_{-1})</td>
<td>1.1127(_{+3})</td>
<td>4.18</td>
</tr>
<tr>
<td>9</td>
<td>1.3530(_{-8})</td>
<td>2.50(_{-1})</td>
<td>4.5630(_{+3})</td>
<td>4.10</td>
</tr>
</tbody>
</table>

In the context of Newton-like methods, a different strategy is commonly employed to optimize the efficiency with respect to computing time. In inexact Newton methods (cf. Section 6.3.1), the accuracy requirement for the solution of the inner systems is coupled to the progress of the outer iteration. Here, we test how a fixed number of multigrid iterations, e.g. just one or two cycles per outer PDAS step, performs. This is a specific truncated (semismooth) Newton method. Naturally, we expect an increase in the number of PDAS iterations and at least for \(m = 1\), we can only expect linear convergence. This is confirmed in Figure 5.7, where the error \(e_{u,J}^k\) is plotted against the wall-clock time measured in seconds. Marks indicate a new PDAS step. However, measured in total units of multigrid cycles both truncated approaches are very competitive and the overall performance equals that of the full multigrid approach.

For fixed \(\sigma\), the convergence of the outer PDAS iteration does not depend on \(h_J\). Let us now consider the dependence on the regularization parameter \(\sigma\). To this end, we consider the first test case with the unilateral bound (5.23) and vary \(\sigma\) between \(1_{-2}\) and \(1_{-5}\). Figure 5.8 shows the discrete \(L^2\)-error \(e_{u,J}^k\) on the left, and the bound
Figure 5.7: Performance comparison of FMG, FMG +V₁,₁, and 1 and 2 V₁,₁-cycles per PDAS iteration, respectively.

Figure 5.8: Convergence of the outer PDAS iterations for different values of the regularization parameter $\sigma$. $L^2$-error $e_β^k$ (left) and bound violation $e_β^k$ (right).

violation measured with $e_β^k$ on the right. We observe a slight increase in the number of iterations for decreasing $\sigma$. This is in accordance with results reported in [24] and is therefore not related to method used to solve (EQP). The given heuristic explanation is that for smaller $\sigma$, the constraints can act stronger and $A^*$ is larger for smaller $\sigma$. Depending on the initial guess, more iterations are required until $A^*$ is fully resolved. The size of the active set and the growth in each PDAS iteration is reported in Table 5.4 for $\sigma = 1_{-2}, 1_{-5}$ and $J = 6, 7, 8$. The $k$-th column contains the data for the $k$-th PDAS iteration. In the first column, we give the size of $A^1$, each following column contains the growth, i.e. the increase in cell numbers, from $A^k_1$ to $A^{k+1}_1$. Furthermore, we report the error $e_β^k$. 

Table 5.4: Size of the active set $A_k^+$ and error $e^k_\beta$ for $J = 6, 7, 8$ and regularization parameters $\sigma = 1.0e-2$ and $\sigma = 1.0e-5$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>$J$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>124952</td>
<td>+6096</td>
<td>+24</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>6.843_{-2}</td>
<td>4.149_{-4}</td>
<td>0.0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1_{-2}</td>
<td>7</td>
<td>499670</td>
<td>+24474</td>
<td>+144</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>6.848_{-2}</td>
<td>4.800_{-4}</td>
<td>0.0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>1998192</td>
<td>+98254</td>
<td>+706</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>6.856_{-2}</td>
<td>4.820_{-4}</td>
<td>0.0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>101450</td>
<td>+20528</td>
<td>+7744</td>
<td>+1330</td>
<td>+20</td>
</tr>
<tr>
<td></td>
<td>2.338_{-1}</td>
<td>9.117_{-2}</td>
<td>8.090_{-3}</td>
<td>3.379_{-4}</td>
<td>0.0</td>
<td>—</td>
</tr>
<tr>
<td>1_{-5}</td>
<td>7</td>
<td>405590</td>
<td>+82330</td>
<td>+30916</td>
<td>+5342</td>
<td>+110</td>
</tr>
<tr>
<td></td>
<td>2.326_{-1}</td>
<td>8.836_{-2}</td>
<td>9.050_{-3}</td>
<td>3.848_{-4}</td>
<td>0.0</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>1622360</td>
<td>+329360</td>
<td>+123534</td>
<td>+21352</td>
<td>+546</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>2.331_{-1}</td>
<td>9.109_{-2}</td>
<td>1.055_{-2}</td>
<td>3.960_{-4}</td>
<td>0.0</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 5.5: Number of cells in the final inactive set $I^*$ (fraction of total cell number in parentheses) for different values of $\sigma$.

<table>
<thead>
<tr>
<th>$\sigma$</th>
<th>1.0</th>
<th>1_{-1}</th>
<th>1_{-2}</th>
<th>1_{-3}</th>
<th>1_{-4}</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>I^*</td>
<td>$</td>
<td>100288 (3.83_{-1})</td>
<td>15616 (5.96_{-2})</td>
<td>1152 (4.4_{-3})</td>
</tr>
</tbody>
</table>

### 5.5.2 Example: A Bang-Bang Control Problem

This test case is an example for a so-called bang-bang control. Such controls are almost everywhere equal to the bounding functions $u^\alpha_h, u^\beta_h$. We prescribe the target state

$$\bar{y} = 128 \pi^2 \sin(4\pi x_1) \sin(4\pi x_2),$$

(5.25)

the lower and upper bounds are given by

$$u^\alpha = -1, \text{ and } u^\beta = 1,$$

(5.26)

respectively. In Figure 5.9 we show the computed optimal state $y_h^*\sigma$ for $\sigma = 1_{-4}$ on a mesh with $h = 2^{-9}$. Figure 5.10 shows the computed optimal controls $u^\alpha_h, u^\beta_h$ on the same mesh for a decreasing sequence of values for $\sigma$. For $\sigma = 1_{-4}$, $u^\alpha_h$ everywhere attains the values of the bounds $u^\alpha_h, u^\beta_h$. Correspondingly, the size of the inactive set, $|I^*|$, is zero and the bound constraints are active in every cell. In Table 5.5 we list the number of cells and the fraction of the total number of cells (in parentheses) in $|I^*|$ for the different values of $\sigma$. The given values confirm that the size of $I^*$ shrinks with decreasing $\sigma$. The active set for this example was always detected in two PDAS iterations.
5.5 Numerical Results

Figure 5.9: Computed optimal state $y_h^*$ for the example problem with an optimal control $u_h^*$ of bang-bang type and $\sigma = 1_{-4}$ on the finest mesh with $h_J = 2^{-9}$.

(a) $\sigma = 1_{0}$ (left) and $\sigma = 1_{-1}$ (right).

(b) $\sigma = 1_{-2}$ (left) and $\sigma = 1_{-4}$ (right).

Figure 5.10: Computed optimal controls $u_h^*$ for the finest grid with a mesh size $h_J = 2^{-9}$ and different regularization parameters $\sigma$. 
Summary

We considered the numerical solution of linear-quadratic PDE constrained optimization problems with additional pointwise inequality constraints imposed on the control function. The control constraints are treated in an outer iteration, which here is given by an implementation of the primal-dual active-set strategy (PDAS). The PDAS method generates a sequence of equality constrained problems. In order to solve these problems efficiently, we extended the multigrid approach devised for equality-constrained problems (cf. Chapter 4). The required modifications have been described in detail. We also adapted the full multigrid method in order to solve the PDAS subproblems with optimal complexity.

Several numerical examples have been discussed, including constant and non constant upper and lower bounds. It has been demonstrated that the PDAS-multigrid method yields an efficient solver for discrete optimality systems with pointwise inequality constraints on the control. For both the outer PDAS iteration and the inner multigrid method the convergence does not depend on the mesh size $h_J$ of the finest discretization level. It also has been demonstrated that the solution of the subproblems is achieved with optimal complexity, employing the full multigrid method. Overall, the resulting PDAS-multigrid method is a fast solver for linear-quadratic programming problems. A natural application for such methods is as solver for the quadratic subproblems which arise within a sequential quadratic programming (SQP) algorithm. This will be the topic of the next chapter.