Numer. Math. Theor. Meth. Appl. Vol. **5**, No. 1, pp. 1-18
doi: 10.4208/nmtma.2011.m12si01 February 2012

Multigrid Solution of a Lavrentiev-Regularized State-Constrained Parabolic Control Problem

Alfio Borzì[1,*] and Sergio González Andrade[2]

[1] *Institut für Mathematik, Universität Würzburg, Campus Hubland Nord, Emil-Fischer-Str. 30, 97074 Würzburg, Germany.*
[2] *Research Group on Optimization, Departamento de Matemática, Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito, Ecuador.*

Received 01 December 2010; Accepted (in revised version) 05 May 2011

Available 21 December 2011

Abstract. A mesh-independent, robust, and accurate multigrid scheme to solve a linear state-constrained parabolic optimal control problem is presented. We first consider a Lavrentiev regularization of the state-constrained optimization problem. Then, a multigrid scheme is designed for the numerical solution of the regularized optimality system. Central to this scheme is the construction of an iterative pointwise smoother which can be formulated as a local semismooth Newton iteration. Results of numerical experiments and theoretical twogrid local Fourier analysis estimates demonstrate that the proposed scheme is able to solve parabolic state-constrained optimality systems with textbook multigrid efficiency.

AMS subject classifications: 35K10, 49K20, 49J20, 49M05, 65M55, 65C20

Key words: Multigrid methods, Lavrentiev regularization, semismooth Newton methods, parabolic partial differential equations, optimal control theory.

1. Introduction

Optimal control of parabolic systems occurs in many application fields such as chemical reaction simulations and biomedical sciences, among other important fields [3, 4, 20]. These problems require the development of algorithms that are fast and robust with respect to the optimization parameters. Recent developments [1–3, 10] show that a successful framework to develop such algorithms is represented by space-time collective-smoothing multigrid schemes. In fact, Fourier analysis estimates [6, 10] and results of numerical experiments with linear [1] and nonlinear [3, 4, 10] parabolic control problems demonstrate that space-time multigrid schemes provide optimal control solutions with mesh-independent convergence and robustness with respect to the value of the control parameters.

*Corresponding author. *Email addresses:* `alfio.borzi@mathematik.uni-wuerzburg.de` (A. Borzì), `sergio.gonzalez@epn.edu.ec` (S. González Andrade)

Previous contributions to the multigrid solution of parabolic control problems [1, 4, 6, 8, 12] have focused on first-order time discretization, while higher-order space-time discretization and constraints on the control have been considered in [10]. In this contribution, the authors found that the Crank-Nicolson scheme is not a convenient choice while multistep backward differencing schemes are advantageous in the design of very efficient pointwise and linewise smoothers.

In this paper, we contribute to the field of space-time multigrid methods for the case of parabolic optimal control problems with state-constraints and second-order space-time discretization. Our multigrid approach is formulated based on criteria proposed in [1, 2, 10] combining space-time collective smoothing multigrid schemes and Lavrentiev regularization [21,25]. In this framework, the smoothing procedures are pointwise iterative schemes that update the optimization variables collectively and use projection to satisfy the inequality constraints. We show that these iterative schemes can be interpreted as local semismooth Newton methods [13,23,24] applied to the regularized state-constrained problems.

In the next section, we formulate a state-constrained linear parabolic optimal control problem. Further, we obtain a Lavrentiev regularization of the problem and characterize the optimal solutions as solutions of the corresponding regularized optimality system. In Section 3, we discuss a space-time second-order discretization of the optimality system. In Section 4, we illustrate the space-time multigrid framework and focus on the construction of an efficient pointwise smoother. In Section 5, we investigate the proposed smoother using results of twogrid local Fourier analysis to discuss the convergence properties of the multigrid scheme with the pointwise smoother. We obtain smoothing-factor and multigrid convergence-factor estimates that predict typical textbook multigrid efficiency and robustness with respect to the values of the control parameters. Further, we present novel insight that shows that the resulting smoothers can be interpreted as local semismooth Newton schemes. In Section 6, detailed numerical experiments are carried out. The numerical results demonstrate the ability of the proposed multigrid framework to provide efficient solutions to state constrained linear parabolic optimal control problems. Besides, we discuss the application of the receding-horizon methodology [3, 14] to achieve long-time tracking of a desired trajectory with state constraints. A section of conclusion completes this work.

2. A state-constrained parabolic optimal control problem

Optimal control problems are defined for the purpose of determining the optimal way to influence dynamical systems towards a given task. Our optimal control problem consists of a parabolic governing system, a distributed control mechanism, and a criterion defining the cost functional, that models the purpose of the control and describes the cost of its action. The formulation of an optimal control problem is then to minimize the cost functional under the constraint given by the modeling equations. The solution to this problem is characterized by first-order optimality conditions given by the optimality system. In particular, we focus on state-constrained parabolic optimal control problems where the configuration of the controlled system is subject to functional constraints. For a more general and detailed discussion on optimal control problems see, e.g., [17,25].

For the purpose of illustration, consider a plate of a given material which defines a two-dimensional convex domain Ω. Let the state of the material y represent the temperature distribution which is maintained equal to zero along the boundary $\partial\Omega$. With y_0 we denote the initial temperature distribution, and we assume the presence of a given space-time depending heating source f. The control mechanism is also a thermal source u, which is a function of space and time. This system is governed by the following parabolic equation:

$$\begin{cases} -\partial_t y + \Delta y = f & \text{in } Q = \Omega \times (0,T), \\ y = y_0 & \text{on } \Omega \times \{t = 0\}, \\ y = 0 & \text{on } \Sigma = \partial\Omega \times (0,T). \end{cases} \tag{2.1}$$

We consider this parabolic process controlled through source terms with the purpose of tracking a desired trajectory, given by $y_d \in L^2(Q)$, and/or with the objective of reaching a desired terminal state $y_T \in L^2(\Omega)$ at a given final time T. In order to achieve these objectives, we formulate a distributed parabolic optimal control problem, governed by (2.1), as follows

$$\begin{cases} \min\limits_{y,u} J(y,u) := \frac{\alpha}{2}\|y - y_d\|^2_{L^2(Q)} + \frac{\beta}{2}\|y(\cdot,T) - y_T\|^2_{L^2(\Omega)} + \frac{\nu}{2}\|u\|^2_{L^2(Q)}, \\ -\partial_t y + \Delta y = f + u \quad \text{in } Q = \Omega \times (0,T), \\ y = y_0 \quad \text{on } \Omega \times \{t = 0\}, \\ y = 0 \quad \text{on } \Sigma = \partial\Omega \times (0,T). \end{cases} \tag{2.2}$$

Here, $\nu > 0$ stands for the weight of the cost of the control, $\alpha \geq 0$ and $\beta \geq 0$, with $\alpha + \beta > 0$, are control parameters, which allow us to achieve the proposed objectives. Indeed, the case $\alpha = 1$, $\beta = 0$ corresponds to tracking without terminal observation. With $\alpha = 0$, $\beta = 1$, the objective is to reach a given final target configuration without any specification of the trajectory that should be followed. We assume that $f,\ u \in L^2(Q)$, and we select an initial condition $y_0 \in H^1_0(\Omega)$.

In this paper, we are interested in state-constrained control problems. Thus, we require the state y to satisfy additional criteria. A representative problem is given by bilateral pointwise state constraints, which leads us to consider the following restrictions

$$\underline{y}(x,t) \leq y(x,t) \leq \overline{y}(x,t), \text{ a.e. in } Q, \tag{2.3}$$

where $\underline{y}$ and $\overline{y}$ are elements of $C(Q)$ with $\underline{y} < \overline{y}$.

The solution approach to state-constrained optimal control problems through Lagrange multipliers associated with the state constraint leads to technical difficulties [19, 21, 25]. For instance, if we consider functions $\underline{y}$ and $\overline{y}$ such that the admissible set is not empty, then it is possible to show the existence of an optimal control $u^* \in L^2(Q)$ with corresponding optimal state $y^* \in W(0,T)$; see [21, 25]. However, in order to have a Lagrange multiplier rule of Karush-Kuhn-Tucker type, we need y to be continuous and to satisfy a Slater condition. In this case, we require continuity of the constraints $\underline{y}$ and $\overline{y}$, and still we need to assume further restrictions on the dimension of Ω; see [21].

The main issue is that the Lagrange multipliers associated with the state constraints are only regular Borel measures. This fact prevents us from using known approximation techniques. In fact, to the best of our knowledge (at least in a finite difference context), no approximation methods for such a class of functions is available. Therefore, we need to discuss a regularization procedure for problem (2.2); see [19,21] and references given there.

In this paper, we consider the Lavrentiev-type regularization approach because it elegantly accommodates in our framework. The Lavrentiev-type regularization consists in approximating the pointwise state constraints $\underline{y}(x,t) \leq y(x,t) \leq \overline{y}(x,t)$ by the following

$$\underline{y}(x,t) \;\leq\; y(x,t) - \lambda u(x,t) \;\leq\; \overline{y}(x,t), \qquad \text{a.e. in } Q,$$

where $\lambda > 0$ is a small parameter. As a result of this procedure, we do not need to consider further restrictions on the structure nor on the dimension of Ω to obtain sufficiently regular Lagrange multipliers for the regularized resulting problem. In fact, these associated multipliers can be assumed to be functions in $L^2(Q)$; see [19,21].

Once we have relaxed the constraints, we introduce the auxiliary variable $v := y - \lambda u$ and express the control function u in terms of v. Consequently, the state-constrained optimal control problem becomes

$$\begin{cases} \min\limits_{y,v} J(y,v) := \frac{\alpha}{2}\|y-y_d\|^2_{L^2(Q)} + \frac{\beta}{2}\|y(\cdot,T)-y_T\|^2_{L^2(\Omega)} + \frac{\nu}{2\lambda^2}\,\|y-v\|^2_{L^2(Q)}, \\ -\partial_t y - \frac{y}{\lambda} + \Delta y = f - \frac{v}{\lambda} \quad \text{in } \; Q = \Omega\times(0,T), \\ y = y_0 \quad \text{on } \; \Omega\times\{t=0\}, \\ y = 0 \quad \text{on } \; \Sigma = \partial\Omega\times(0,T), \end{cases} \tag{2.4}$$

with the following bilateral pointwise restriction on v

$$\underline{y}(x,t) \;\leq\; v(x,t) \;\leq\; \overline{y}(x,t), \text{ a.e. in } Q.$$

Notice that after the transformation, a regularized optimal control problem is obtained having a "control-constrained" structure. Thanks to this substitution technique, we can guarantee the existence of solutions for the problem (2.4), for every $\lambda > 0$. Moreover, we are able to characterize the solution to this problem by the following optimality system (*e.g.* [21, Th. 2.2])

$$-\partial_t y + \Delta y - \frac{y}{\lambda} + \frac{v}{\lambda} = f \quad \text{in } \; Q, \tag{2.5a}$$

$$y = 0 \quad \text{on } \; \Sigma, \tag{2.5b}$$

$$y = y_0 \quad \text{on } \; \Omega\times\{t=0\}, \tag{2.5c}$$

$$\partial_t p + \Delta p - \frac{p}{\lambda} + \alpha(y-y_d) + \eta(y-v) = 0 \qquad \text{in } \; Q, \tag{2.5d}$$

$$p = 0 \qquad \text{on } \; \Sigma, \tag{2.5e}$$

$$p = \beta(y-y_T) \quad \text{on } \; \Omega\times\{t=T\}, \tag{2.5f}$$

$$\left(\frac{p}{\lambda} - \eta(y-v), w-v\right) \geq 0 \ \text{ in } Q, \tag{2.5g}$$

where $\eta = \nu/\lambda^2$ and the inequality (2.5g) must hold for all $w \in V_{ad}$ where

$$V_{ad} = \{v \in L^2(Q) : \underline{y}(x,t) \leq v(x,t) \leq \overline{y}(x,t), \text{ a.e. in } Q\}.$$

The convergence analysis of regularized solutions for vanishing Lavrentiev parameter λ requires further regularity assumptions, that is out of the scope of this paper. However, by following [19, 21], we state that it is possible to prove that the regularized solutions of (2.4) converge to admissible solutions for (2.2) as $\lambda \to 0$. Further, numerical experience demonstrates that whenever the ratio $\eta = \nu/\lambda^2$ is kept constant while reducing the value of λ the computational performance of our multigrid solution procedure does not deteriorate. This fact was also shown in [2] and deserves further investigation.

As stated before, the main advantage of using the Lavrentiev regularization is the fact that we can avoid Borel measures in the optimality system. The associated Lagrange multipliers are functions in $L^2(Q)$ and the regularized systems can be solved using standard discretization schemes. See [15] for an alternative approach.

3. Second-order space-time discretization

We consider second-order discretization schemes discussed in [10]. These schemes use second-order backward differentiation formula (BDF2) together with the Crank–Nicolson (CN) method in order to obtain a second-order time discretization scheme of the optimality system.

To illustrate our approach, we use the framework in [9, 11] and assume that the space domain Ω is a square and $\overline{\Omega}_h$ is a uniform space mesh, where h is the mesh size, and Ω_h defines the set of interior mesh-points, $(x_i, y_j) = ((i-1)h, (j-1)h)$, $2 \leq i, j \leq N_x$. On this mesh, $-\Delta_h$ denotes the second-order accurate negative Laplacian approximated by the common five-point stencil including homogeneous Dirichlet boundary conditions. For grid functions v_h and w_h defined on Ω_h, we have the discrete $L^2(\Omega)$-scalar product

$$(v_h, w_h)_{L^2_h(\Omega_h)} = h^2 \sum_{x \in \Omega_h} v_h(x)\, w_h(x),$$

with associated norm $|v_h| = (v_h, v_h)^{1/2}_{L^2_h(\Omega_h)}$. Further, let $\delta t = T/N_t$ be the time-step size and define the following space-time mesh

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

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

On the $Q_{h,\delta t}$ grid, y_h^m and p_h^m denote grid functions at time level m. The action of the one-step backward and forward time-discretization operator on these functions is defined as follows

$$\partial^+ y_h^m := \frac{y_h^m - y_h^{m-1}}{\delta t} \quad \text{and} \quad \partial^- p_h^m := -\frac{p_h^m - p_h^{m+1}}{\delta t}.$$

The action of the BDF2 time-difference operators is as follows

$$\partial_{BD}^+ y_h^m := \frac{3y_h^m - 4y_h^{m-1} + y_h^{m-2}}{2\delta t} \quad \text{and} \quad \partial_{BD}^- p_h^m := -\frac{3p_h^m - 4p_h^{m+1} + p_h^{m+2}}{2\delta t}.$$

The coefficients in the last two expressions above are given by the classical BDF2 formula (see, e.g., [9]) while the minus sign in the second operator allows us to discretize the adjoint variable taking into account its backward evolution in time. For simplicity, we assume sufficient regularity of the data, y_d, y_T, and f, such that these functions are properly approximated by their values at grid points.

With this setting, the following discrete optimality system is obtained

$$-\partial_{BD}^+ y_h^m + \Delta_h y_h^m - \frac{y_h^m}{\lambda} + \frac{v_h^m}{\lambda} = f_h^m \qquad \text{in } Q, \tag{3.1a}$$

$$\partial_{BD}^- p_h^m + \Delta_h p_h^m - \frac{p_h^m}{\lambda} + \alpha(y_h^m - y_{d,h}^m) + \eta(y_h^m - v_h^m) = 0 \qquad \text{in } Q, \tag{3.1b}$$

$$\left(\frac{p_h^m}{\lambda} - \eta(y_h^m - v_h^m), w_h^m - v_h^m\right) \geq 0 \qquad \text{in } Q, \tag{3.1c}$$

where the inequality must hold for all $w \in Y_{ad}^h$ and Y_{ad}^h is a grid function approximation of the admissible set Y_{ad}.

As stated before, at $t = \delta t$, represented by $m = 2$, and at $t = T - \delta t$, given by $m = N_t$, we combine the multistep BDF2 method with the CN method.

4. The space-time multigrid framework

In this section, we discuss the extension of the space-time collective-smoothing multigrid strategy for parabolic optimal control problems [1–3,10] to the case of state-constraints and second-order space-time discretization. We recall the space-time multigrid scheme that belongs to the class of nonlinear full approximation storage (FAS) methods [7].

Consider L grid levels indexed by $k = 1, \cdots, L$, where $k = L$ refers to the finest grid. The mesh of level k is denoted by $Q_k = Q_{h_k,\delta t_k}$ where $h_k = h_1/2^{k-1}$ and $\delta t_k = \delta t$, that corresponds to semicoarsening in space. Any operator and variable defined on the discrete space-time cylinder Q_k is indexed by k. The optimality system at level k with given initial, terminal, and boundary conditions is represented by the following equation

$$A_k(w_k) = f_k, \qquad w_k = (y_k, u_k, p_k). \tag{4.1}$$

As well known [7,26], the multigrid strategy combines two complementary schemes. The high-frequency components of the solution error are reduced by a smoothing iteration, denoted by S_k and defined in the following subsection, while the low-frequency error components are effectively reduced by a coarse-grid correction method as defined below.

The action of one multigrid cycle applied to (4.1) is illustrated in the following algorithm. Starting with an initial approximation $w_k^{(0)}$, the result of one multigrid cycle for solving $A_k(w_k) = f_k$ is given by $w_k^{(\nu_1+\nu_2+1)}$ which is computed as follows.

Algorithm 4.1. *Space-Time Multigrid (STMG) Cycle*

1. *Set the starting approximation* $w_k^{(0)}$;
2. *If* $k = 1$ *solve* $A_k(w_k) = f_k$ *exactly;*
3. *Pre-smoothing steps:* $w_k^{(l)} = S(w_k^{(l-1)}, f_k)$, $l = 1, \cdots, \nu_1$;
4. *Computation of the residual:* $r_k = f_k - A_k(w_k^{(\nu_1)})$;
5. *Restriction of the residual:* $r_{k-1} = I_k^{k-1} r_k$;
6. *Set* $w_{k-1} = \hat{I}_k^{k-1} w_k^{(\nu_1)}$;
7. *Set* $f_{k-1} = r_{k-1} + A_{k-1}(w_{k-1})$
8. *Call* μ *times the STMG scheme to solve* $A_{k-1}(w_{k-1}) = f_{k-1}$;
9. *Coarse-grid correction:* $w_k^{(\nu_1+1)} = w_k^{(\nu_1)} + I_{k-1}^k (w_{k-1} - \hat{I}_k^{k-1} w_k^{(\nu_1)})$;
10. *Post-smoothing steps:* $w_k^{(l)} = S(w_k^{(l-1)}, f_k)$, $l = \nu_1 + 2, \cdots, \nu_1 + \nu_2 + 1$;
11. *End.*

Notice that we can perform μ two-grid iterations at each working level. For $\mu = 1$ we have a $V(\nu_1, \nu_2)$-cycle and for $\mu = 2$ we have a $W(\nu_1, \nu_2)$-cycle; μ is called the cycle index [26].

In our implementation, we choose I_k^{k-1} to be the full-weighted restriction operator in space with no averaging in the time direction [26]. The prolongation I_{k-1}^k is defined by bilinear interpolation in space, and we choose $\hat{I}_k^{k-1}$ to be straight injection. The intergrid transfer operators do not involve time since we are using semicoarsening.

4.1. A pointwise space-time smoothing scheme

A key component in the STMG Algorithm 4.1, is the smoothing scheme S_k. This smoothing process must be efficient in solving high-frequency error components and robust with respect to the control parameters. We discuss a pointwise smoothing scheme for state-constrained control problems.

In order to develop the smoothing schemes, we write the discretized optimality system (3.1) in expanded form, for a space-time grid point $(i\,j\,m)$. For this purpose, let $h = h_k$ and $x \in \Omega_h$, where $x = (ih, jh)$ and (i,j) index the grid points, e.g., lexicographically. Denote with ω_{ij} the set of grid index pairs s,t of the stencil of Δ_h centered at (i,j) excluding the pair (i,j) itself. Using this notation, we can express the action of Δ_h on the function v_h in the following compact form

$$\Delta_h v_h|_{ij} = \frac{1}{h^2}\left(\sum_{s,t\in\omega_{ij}} v_{st} - 4\,v_{ij}\right).$$

We next illustrate the construction of the smoothing algorithm for the BDF2 discretization, *i.e.*, we consider the optimality system (3.1) at the space-time grid points (ijm) for $m = 3,\cdots,N_t - 1$. However, we point out that a similar discussion follows for $m = 1$ and $m = N_t$, where the BDF2 and the CN discretization both appear in the optimality system.

Let us then start by setting

$$A_{ijm} = \sum_{s,t\in\omega_{ij}} y^m_{st},\ B_{ijm} = \sum_{s,t\in\omega_{ij}} p^m_{st},$$

and introducing the following notations

$$\begin{aligned}
S_{ijm} &:= \gamma A_{ijm} + 2y_{ijm-1} - \tfrac{1}{2}y_{ijm-2} - \delta t\, f_{ijm},\\
R_{ijm} &:= \gamma B_{ijm} + 2p_{ijm+1} - \tfrac{1}{2}p_{ijm+2} - \delta t\,\alpha\, y_{d\,ijm},
\end{aligned}$$

where $\gamma := \delta t/h^2$. Notice that A_{ijm} and B_{ijm} are considered constant during the update of the variables at (ijm).

Next, by calling $a := \left(\frac{3}{2} + 4\gamma + \frac{\delta t}{\lambda}\right)$, we can write the optimality system (3.1) at $(i,\,j,\,m)$ as follows

$$-\,a\,y_{ijm} + S_{ijm} + \frac{\delta t}{\lambda}v_{ijm} = 0, \tag{4.2a}$$

$$-\,a\,p_{ijm} + R_{ijm} + \delta t\,\left(\alpha + \eta\right)\,y_{ijm} - \delta t\eta\,v_{ijm} = 0, \tag{4.2b}$$

$$\left(\frac{1}{\lambda}p_{ijm} - \eta\,(y_{ijm} - v_{ijm})\right)(w_{ijm} - v_{ijm}) \geq 0,\ \ \text{for all } w_h \in V_{adh}, \tag{4.2c}$$

where $V_{adh} = \{v \in L^2_h(Q_h)\,|\,\underline{y} \leq v \leq \overline{y} \text{ in } Q_h\}$. This is a nonlinear problem that includes an inequality constraint. To solve this problem, we adapt the scheme proposed in [10] to the case of state-constrained problems. This scheme is constructed by using a projection procedure to satisfy the inequality constraint (4.2c). Moreover, in the next section, we show that this approach can be interpreted as a local semismooth Newton method [13,23, 24].

Consider the equations (4.2a) and (4.2b). The Jacobian of these two equations and its inverse are given by

$$J_{ijm} := \begin{pmatrix} -a & 0\\ \delta t\,(\alpha+\eta) & -a\end{pmatrix} \text{ and } J^{-1}_{ijm} = \frac{1}{a^2}\begin{pmatrix} -a & 0\\ -\delta t\,(\alpha+\eta) & -a\end{pmatrix}, \tag{4.3}$$

respectively.

Now, for a given v_{ijm}, a classical local Newton update for the state and adjoint variables $\widehat{y}_{ijm}$ and $\widehat{p}_{ijm}$ is given by

$$\begin{pmatrix} \widehat{y} \\ \widehat{p} \end{pmatrix}_{ijm} = \begin{pmatrix} y \\ p \end{pmatrix}_{ijm} + J_{ijm}^{-1} \begin{pmatrix} r_y(v) \\ r_p(v) \end{pmatrix}_{ijm}, \tag{4.4}$$

where $r_y(v)$ and $r_p(v)$ denote the residuals of the discrete state and adjoint equations, which are given by the negative of (4.2a) and (4.2b), respectively. We point out that these residuals are functions of the control.

Based on the Newton update, we can write $\widehat{y}_{ijm}$ and $\widehat{p}_{ijm}$ as functions of v_{ijm} as follows

$$\begin{aligned} \widehat{y}_{ijm}(v_{ijm}) &= y_{ijm} - \frac{r_y(v_{ijm})}{a}, \\ \widehat{p}_{ijm}(v_{ijm}) &= p_{ijm} + \frac{-\delta t(\alpha+\eta)\, r_y(v_{ijm}) - a\, r_p(v_{ijm})}{a^2}. \end{aligned} \tag{4.5}$$

Next, we compute the value of an auxiliary control variable that corresponds to the control update in the case where no constraints on the control are present. We denote this auxiliary control function with $\tilde{v}_{ijm}$, which is defined as the solution to

$$\frac{1}{\lambda} \widehat{p}_{ijm}(\tilde{v}_{ijm}) - \eta\, (\widehat{y}_{ijm}(\tilde{v}_{ijm}) - \tilde{v}_{ijm}) = 0. \tag{4.6}$$

Now recall that the update to v_{ijm} must satisfy the pointwise constraint $\underline{y} \le v \le \overline{y}$. Therefore a feasible update is obtained by projection as follows

$$v_{ijm} = \begin{cases} \overline{y}_{ijm} & \text{if } \tilde{v}_{ijm} \ge \overline{y}_{ijm}, \\ \tilde{v}_{ijm} & \text{if } \underline{y}_{ijm} < \tilde{v}_{ijm} < \overline{y}_{ijm}, \\ \underline{y}_{ijm} & \text{if } \tilde{v}_{ijm} \le \underline{y}_{ijm}. \end{cases} \tag{4.7}$$

With this updated control, we obtain new values for the adjoint and state variables using the Newton mapping (4.5) as $p_{ijm} = \widehat{p}_{ijm}(v_{ijm})$ and $y_{ijm} = \widehat{y}_{ijm}(v_{ijm})$, respectively.

In order to describe the smoothing algorithm, note that, by using (4.5), the expression (4.6) can be rewritten as

$$Q_{ijm}\, \tilde{v}_{ijm} + \frac{R_{ijm}}{a\,\lambda} - \frac{\eta}{a} S_{ijm} + \frac{\delta t(\alpha+\eta)}{a^2 \lambda} S_{ijm} = 0, \tag{4.8}$$

where Q_{ijm} is given by

$$Q_{ijm} := -\frac{\eta\,\delta t}{a\,\lambda} + \eta + \frac{\delta t}{a\,\lambda} \left(\frac{\delta t}{a\,\lambda}(\alpha+\eta) - \eta \right).$$

Next, we define the following auxiliary magnitude

$$T_{ijm} := -\eta(\widehat{y}_{ijm} - \tilde{v}_{ijm}^{(0)}) + \frac{\widehat{p}_{ijm}}{\lambda}$$

and calculate the update for $\widetilde{v}_{ijm}$ as follows

$$\widetilde{v}_{ijm}^{(1)} = \widetilde{v}_{ijm}^{(0)} - Q_{ijm}^{-1} T_{ijm}. \tag{4.9}$$

By using these definitions, we propose the following Collective Gauss-Seidel smoothing algorithm.

Algorithm 4.2. *Projected Time-Splitted Collective Gauss-Seidel Iteration (P-TS-CGS)*

1. *Set the starting approximation:* $w_{ijm}^{(0)} := (y_{ijm}^{(0)}, p_{ijm}^{(0)})$, *with* $y_{ij1}^{(0)} := y_0$.
2. *For* ij *in, e.g., lexicographic order do*
3. *For* $m = 2$ $(t = \delta t)$, *calculate* $y_{ij2}^{(1)}$, $p_{ij2}^{(1)}$ *and* $v_{ij2}^{(1)}$, *by using the corresponding CN-BDF2 versions of* (4.5), (4.7) *and* (4.9).
4. *For* $m = 3, \cdots, N_t - 1$: *compute* $(r_y)_{ijm}$ *and* $(r_p)_{ijm}$, *and by using* (4.5) *calculate* $\widehat{y}_{ijm}$ *and* $\widehat{p}_{ijm}$. *Next, obtain* $\widetilde{v}_{ijm}^{(1)}$ *according with* (4.9) *and calculate the update for the control* $v_{ijm}^{(1)}$ *by projection* (4.7). *Thus, calculate the update for the state variable by:*
$$y_{ijm}^{(1)} = y_{ijm}^{(0)} - \frac{(\widehat{r}_y)_{ijm}}{a},$$
where $(\widehat{r}_y)_{ijm} := a\, y_{ijm}^{(0)} - S_{ijm} - \frac{\delta t}{\lambda} v_{ijm}^{(1)}$.
5. *For* $k = N_t - 1, \cdots, 3$ *(backwards): compute* $(r_y)_{ijk}$ *and* $(r_p)_{ijk}$, *and by using* (4.5) *calculate* $\widehat{y}_{ijk}$ *and* $\widehat{p}_{ijk}$. *Next, obtain* $\widetilde{v}_{ijk}^{(1)}$ *according with* (4.9) *and calculate the update for the control* $v_{ijk}^{(1)}$ *by projection* (4.7). *Thus, calculate the update for the adjoint variable by:*
$$p_{ijk}^{(1)} = p_{ijk}^{(0)} + \frac{-\delta t(\alpha+\eta)(\widehat{r}_y)_{ijk} - a\,(\widehat{r}_p)_{ijk}}{a^2},$$
where $(\widehat{r}_y)_{ijk} := a\, y_{ijk}^{(0)} - S_{ijk} - \frac{\delta t}{\lambda} v_{ijk}^{(1)}$ *and* $(\widehat{r}_p)_{ijk} = a\, p_{ijk}^{(0)} - R_{ijk} - \delta t\ (\alpha+\eta)\ y_{ijk}^{(0)} + \delta t \eta\, v_{ijk}^{(1)}$.
6. *For* $m = N_t$ $(t = T - \delta t)$, *calculate* $y_{ijN_t}^{(1)}$, $p_{ijN_t}^{(1)}$ *and* $v_{ijN_t}^{(1)}$, *by using the corresponding BDF2-CN versions of* (4.5), (4.7) *and* (4.9).
7. *For* $m = N_t + 1$ $(t = T)$, *calculate* $y_{ijN_{t+1}}^{(1)}$, $p_{ijN_{t+1}}^{(1)}$ *and* $v_{ijN_{t+1}}^{(1)}$, *by using the terminal condition* $p_{ijN_{t+1}} = \beta(y_{ijN_{t+1}} - y_T)$, (4.5), (4.7) *and* (4.9).
8. *End.*

5. Smoothing and convergence analysis

In this section, we use local Fourier analysis (LFA) [1, 26] to estimate the smoothing factor $\mu(S_k)$ of the smoothing scheme described above, and the convergence factor $\eta(TG_k^{k-1})$ for the corresponding multigrid scheme. For this analysis, we consider the linear parabolic equation resulting from the Lavrentiev regularization. However, we require that the state-constraints are not active since the LFA framework applies only to linear problems. This means that the optimality condition is satisfied in the sense that we can eliminate the variable v using the following

$$v = y - \frac{\lambda}{\nu} p.$$

Using this equation in the state and adjoint equations, we obtain the optimality system of the parabolic unconstrained case. Therefore in this case the LFA analysis provides the same estimates obtained in [10], that we report in Table 1. We have that the smoothing and convergence factors are almost independent of the value of the weight ν and of the discretization parameter γ and obviously (because the constraints are considered inactive) they do not depend on the Lavrentiev regularization parameter.

Table 1: Smoothing factor $\mu(S_k)$ and convergence factor $\eta(TG_k^{k-1})$ for the TS-CGS multigrid scheme ($\nu_1 = \nu_2 = 1$). Parameters: $\delta t = 1/64$, $\alpha = 1$ and $\beta = 0$.

STMG	γ \ ν	10^{-8}	10^{-6}	10^{-4}	10^{-2}
$\mu(S_k)$	32	0.2289	0.4843	0.4516	0.4493
	48	0.3317	0.4737	0.4502	0.4486
	64	0.4056	0.4677	0.4494	0.4483
$\eta(TG_k^{k-1})$	32	0.0427	0.1317	0.1361	0.1347
	48	0.0822	0.1352	0.1354	0.1344
	64	0.1147	0.1368	0.1350	0.1342

5.1. Smoothing iteration as a SSN method

Here, we analyze the application of a (local) semismooth Newton (SSN) method to (4.2) to show that the resulting iterative scheme is equivalent to the P-TS-CGS scheme. Notice that equation (4.2c) is equivalent to the following pointwise relation in Q [18]

$$v(x,t) = \max\left\{\underline{y}(x,t),\, \min\left\{\overline{y}(x,t),\, y(x,t) - \frac{\lambda}{\nu} p(x,t)\right\}\right\}, \text{ for } \nu > 0. \tag{5.1}$$

We consider equation (5.1) at a grid point $(i\,j\,m)$, and we rewrite (4.2) as the following operator equation

$$\Phi(y_{ijm}, p_{ijm}, u_{ijm}) := \begin{bmatrix} -a\, y_{ijm} + S_{ijm} + \frac{\delta t}{\lambda} v_{ijm} \\ -a\, p_{ijm} + R_{ijm} + \delta t(\alpha+\eta) y_{ijm} - \delta t \eta\, v_{ijm} \\ v_{ijm} - \max\left\{\underline{y}_{ijm},\, \min\left\{\overline{y}_{ijm},\, y_{ijm} - \frac{\lambda}{\nu} p_{ijm}\right\}\right\} \end{bmatrix} = 0. \tag{5.2}$$

We can state that both the max and min functions involved in (5.2) are semismooth. Indeed, it is well known (see [13, Lem. 3.1]) that the mappings $w \mapsto \max(0,w)$ and $w \mapsto \min(0,w)$, from $\mathbb{R}$ to $\mathbb{R}$, are Newton differentiable with Newton derivatives given by

$$\Gamma_{max} := \begin{cases} 1 & \text{if } w \geq 0, \\ 0 & \text{if } w < 0, \end{cases} \quad \text{and} \quad \Gamma_{min} := \begin{cases} 1 & \text{if } w \leq 0, \\ 0 & \text{if } w > 0, \end{cases} \tag{5.3}$$

respectively. Thus, [24, Th. 4.6] implies that the third equation in (5.2) is Newton differentiable, with respect to all variables (y,p,v), and its partial Newton derivatives are given by

$$\Gamma_y := \chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-}, \qquad \Gamma_p := -\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-}\frac{\lambda}{\nu}, \qquad \Gamma_v := 1,$$

where $\chi_{\mathscr{A}_+}$ and $\chi_{\mathscr{A}_-}$ are defined by

$$\chi_{\mathscr{A}_+} := \begin{cases} 1 & \text{if } \min\{\overline{y}_{ijm}, y_{ijm} - \frac{\lambda}{\nu} p_{ijm}\} \geq \underline{y}_{ijm}, \\ 0 & \text{if } \min\{\overline{y}_{ijm}, y_{ijm} - \frac{\lambda}{\nu} p_{ijm}\} < \underline{y}_{ijm}, \end{cases} \tag{5.4a}$$

$$\chi_{\mathscr{A}_-} := \begin{cases} 1 & \text{if } y_{ijm} - \frac{\lambda}{\nu} p_{ijm} \leq \overline{y}_{ijm}, \\ 0 & \text{if } y_{ijm} - \frac{\lambda}{\nu} p_{ijm} > \overline{y}_{ijm}. \end{cases} \tag{5.4b}$$

Consequently, we obtain the semismooth Newton step applied to the operator equation (5.2) as follows

$$\begin{bmatrix} -a & 0 & \frac{\delta t}{\lambda} \\ \delta t(\alpha+\eta) & -a & -\delta t \eta \\ -\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-} & \chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-}\frac{\lambda}{\nu} & 1 \end{bmatrix}_{ijm} \begin{pmatrix} \delta_y \\ \delta_p \\ \delta_v \end{pmatrix}_{ijm} = \begin{pmatrix} r_y \\ r_p \\ r_v \end{pmatrix}_{ijm}. \tag{5.5}$$

From this system we obtain the following update for the state and adjoint variables

$$\begin{aligned} \begin{pmatrix} y \\ p \end{pmatrix}^{(1)}_{ijm} &= \begin{pmatrix} y \\ p \end{pmatrix}^{(0)}_{ijm} \\ &+ \begin{bmatrix} -a + \frac{\delta t}{\lambda}\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-} & -\frac{\delta t}{\nu}\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-} \\ \delta t(\alpha+\eta) - \delta t\,\eta\,\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-} & -a + \frac{\delta t}{\lambda}\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-} \end{bmatrix}^{(0)\,-1}_{ijm} \begin{pmatrix} \widehat{r}_y \\ \widehat{r}_p \end{pmatrix}^{(0)}_{ijm}, \end{aligned} \tag{5.6a}$$

where

$$\begin{aligned} (\widehat{r}_y)_{ijm} &:= (a\,y - S - \frac{\delta t}{\lambda} V)^{(0)}_{ijm}, \\ (\widehat{r}_p)_{ijm} &:= (a\,p - R - \delta t(\alpha+\eta)y + \delta t \eta V)^{(0)}_{ijm}, \\ V^{(0)}_{ijm} &:= \max\left\{\underline{y}_{ijm}, \min\left\{\overline{y}_{ijm}, y^{(0)}_{ijm} - \frac{\lambda}{\nu} p^{(0)}_{ijm}\right\}\right\}. \end{aligned}$$

The update for the control v_{ijm} results as follows

$$v_{ijm}^{(1)} = V_{ijm}^{(0)} + \chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-}(\delta_y)_{ijm} - \frac{\lambda}{\nu}\chi_{\mathscr{A}_+}\chi_{\mathscr{A}_-}(\delta_p)_{ijm}. \tag{5.6b}$$

Now, we show that one iteration step given by (5.6) is equivalent to one iteration of the smoothing step described in the previous section. Also in the semismooth Newton approach, the local SSN iteration must be performed in the forward time-direction to calculate the updates for y_{ijm} and in the backwards time-direction to calculate the updates for p_{ijm}.

Consider the three possible cases arising in (5.6b).

(i) $y_{ijm} - \frac{\lambda}{\nu}p_{ijm} > \overline{y}_{ijm}$. Here, we have that $\chi_{\mathscr{A}_-} = 0$, and we obtain that $V_{ijm}^{(0)} := \overline{y}_{ijm}$. Therefore, from (5.6b), we obtain that $v_{ijm}^{(1)} = \overline{y}_{ijm}$ and, from (5.6a), the following updates for y_{ijm} and p_{ijm}:

$$y_{ijm}^{(1)} = y_{ijm}^{(0)} + \frac{(\widehat{r}_y)_{ijm}}{-a},\ \ p_{ijm}^{(1)} = p_{ijm}^{(0)} + \frac{-\delta t(\alpha+\eta)(\widehat{r}_y)_{ijm} - a(\widehat{r}_p)_{ijm}}{a^2}, \tag{5.7}$$

respectively, where

$$\begin{aligned}(\widehat{r}_y)_{ijm} &:= a\,y_{ijm}^{(0)} - S_{ijm} - \frac{\delta t}{\lambda}\overline{y}_{ijm},\\ (\widehat{r}_p)_{ijm} &:= a\,p_{ijm}^{(0)} - R_{ijm} - \delta t(\alpha+\eta)y_{ijm}^{(0)} + \delta t\eta\,\overline{y}_{ijm}.\end{aligned}$$

(ii) $y_{ijm} - \frac{\lambda}{\nu}p_{ijm} < \underline{y}_{ijm}$. In this case, we have that $\chi_{\mathscr{A}_+} = 0$, since $\min\{\overline{y}_{ijm}, y_{ijm} - \frac{\lambda}{\nu}p_{ijm}\} < \underline{u}_{ijm}$. Hence, we have that $V_{ijm}^{(0)} = \underline{y}_{ijm}$ and (5.6b) implies that $v_{ijm}^{(1)} = \underline{y}_{ijm}$. Further, (5.6a) gives the following updates for y_{ijm} and p_{ijm}:

$$y_{ijm}^{(1)} = y_{ijm}^{(0)} + \frac{(\widehat{r}_y)_{ijm}}{-a},\ \ p_{ijm}^{(1)} = p_{ijm}^{(0)} + \frac{-\delta t(\alpha+\eta)(\widehat{r}_y)_{ijm} - a(\widehat{r}_p)_{ijm}}{a^2}, \tag{5.8}$$

respectively, where

$$\begin{aligned}(\widehat{r}_y)_{ijm} &:= a\,y_{ijm}^{(0)} - S_{ijm} + \frac{\delta t}{\lambda}\underline{y}_{ijm},\\ (\widehat{r}_p)_{ijm} &:= a\,p_{ijm}^{(0)} - R_{ijm} - \delta t(\alpha+\eta)y_{ijm}^{(0)} + \delta t\eta\,\underline{y}_{ijm}.\end{aligned}$$

(iii) $\underline{y}_{ijm} \le y_{ijm} - \frac{\lambda}{\nu}p_{ijm} \le \overline{y}_{ijm}$. In this case, we have that $\chi_{\mathscr{A}_+} = \chi_{\mathscr{A}_-} = 1$. Thus, $V_{ijm}^{(0)} = y_{ijm}^{(0)} - \frac{\lambda}{\nu}p_{ijm}^{(0)}$ and (5.6b) yields that

$$\begin{aligned}v_{ijm}^{(1)} &= y_{ijm}^{(0)} - \frac{\lambda}{\nu}p_{ijm}^{(0)} + (\delta_y)_{ijm} - \frac{\lambda}{\nu}(\delta_p)_{ijm}\\ &= y_{ijm}^{(0)} - \frac{\lambda}{\nu}p_{ijm}^{(0)} + (y_{ijm}^{(1)} - y_{ijm}^{(0)}) - \frac{\lambda}{\nu}(p_{ijm}^{(1)} - p_{ijm}^{(0)})\\ &= y_{ijm}^{(1)} - \frac{\lambda}{\nu}p_{ijm}^{(1)}.\end{aligned} \tag{5.9}$$

From system (5.6a), we obtain the following updates for y_{ijm} and p_{ijm}

$$y_{ijm}^{(1)} = y_{ijm}^{(0)} + \frac{(\widehat{r}_y)_{ijm}}{-a}, \;\; p_{ijm}^{(1)} = p_{ijm}^{(0)} + \frac{-\delta t(\alpha+\eta)(\widehat{r}_y)_{ijm} - a(\widehat{r}_p)_{ijm}}{a^2}, \tag{5.10}$$

respectively, where

$$\begin{aligned}(\widehat{r}_y)_{ijm} &:= a\, y_{ijm}^{(0)} - S_{ijm} + \frac{\delta t}{\lambda} v_{ijm}^{(1)},\\ (\widehat{r}_p)_{ijm} &:= a\, p_{ijm}^{(0)} - R_{ijm} - \delta t(\alpha+\eta) y_{ijm}^{(0)} + \delta t \eta\, v_{ijm}^{(1)}.\end{aligned}$$

Thus, since $\widetilde{v}_{ijm} = y_{ijm} - \frac{\lambda}{\nu} p_{ijm}$, the equivalence between the SSN iteration and the P-TS-CGS iteration is totally established by comparing the auxiliary residuals $(\widehat{r}_y)_{ijm}$ and $(\widehat{r}_p)_{ijm}$ constructed in the two iterative process, and by considering the cases arising in the update given by the projection procedure (4.7) and in the update given by (5.6b).

6. Numerical experiments

We report results of numerical experiments to validate our multigrid solution to state-constrained parabolic control problems. We use W-cycles [7] with two pre- and post-smoothing steps of the symmetric (i.e. a forward and a backward sweeps) version of the smoothers described above with lexicographic ordering. For details regarding coarsening of the variational inequality see [5, 16, 22]. W-cycles appear to be superior in the case of active constraints [5, 10]; for the present isotropic problems space-ordering is not essential for convergence. Let $\Omega = (0,1)^2$ and a finest mesh given by $257 \times 257 \times 257$. We will use five levels in all the experiments. All unknown variables are initialized to be zero and we choose $f = 0$.

First, we consider the desired state $y_d(x_1, x_2) = \sin(2\pi t)\sin(3\pi x_1)\sin(\pi x_2)$ and the constraints $\underline{y}(x) = -1/2$ and $\overline{y}(x) = 1/2$. We solve the tracking problem in the time interval $(0, 1)$. In Fig. 1, the calculated state for the choice $\nu = 10^{-10}$ and $\lambda = 10^{-5}$, at $t = T/4$ and $t = 3T/4$, are depicted. Due to the form of the desired state, the constraints are active in all the considered instants of time.

In Fig. 2, we report the convergence history for the smoothing scheme standalone and for the multigrid scheme depicting the value of the sum of the L^2-norm of the residuals. We notice that during the first few iterations the smoothing scheme provides fast convergence comparable to the first few multigrid cycles. This is symptomatic of initial solution errors with many high-frequencies. Afterwards, we observe a slowdown that is also due to the setting up of the active sets in the whole space-time domain. Further on, while the smoothing scheme convergence becomes flat we obtain fast convergence behavior of the multigrid with a typical superlinear convergence.

In Table 2 we report the observed convergence factors. We obtain the expected multigrid convergence factors, considering that we are taking values for ν ranging from 10^{-6} to 10^{-10} and $\nu = \lambda^2$. With ν held fixed and decreasing λ the resulting convergence factors worsen.

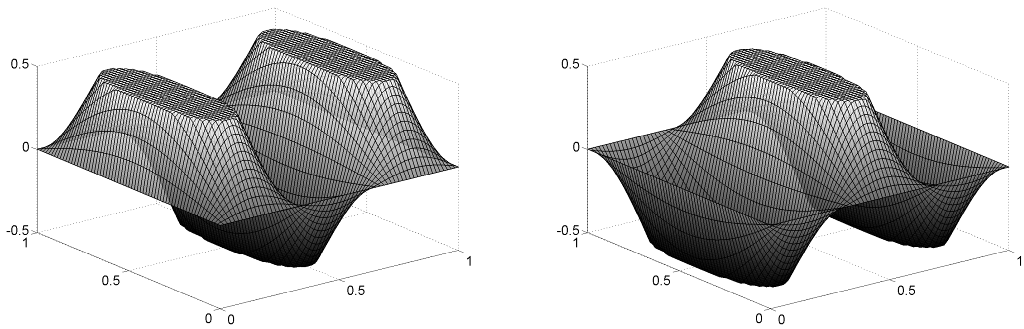

Figure 1: State-constrained problem. The state at $t = T/4$ (left) an at $t = 3T/4$ (right). Parameters: $\alpha = 1$, $\beta = 0$, $\nu = 10^{-10}$ and $\lambda = 10^{-5}$.

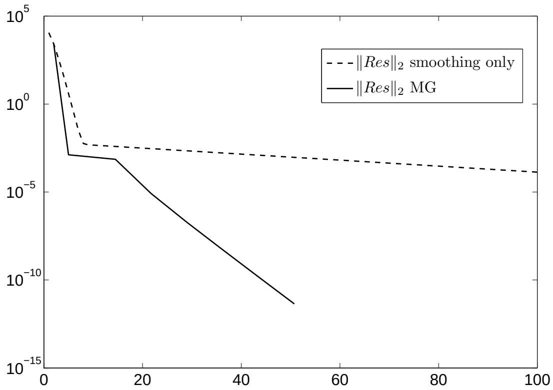

Figure 2: Convergence history for smoothing only (dashed) and multigrid $W(2,2)$-cycle (solid line). Parameters: $\nu = 10^{-10}$ and $\lambda = 10^{-5}$.

Table 2: Convergence factors for $\nu = \lambda^2$.

mesh	$\lambda = 10^{-3}$	$\lambda = 10^{-4}$	$\lambda = 10^{-5}$
$64 \times 64 \times 64$	0.057	0.085	0.028
$128 \times 128 \times 128$	0.050	0.094	0.012
$257 \times 257 \times 257$	0.045	0.056	0.115

When studying parabolic control problems, it is of particular interest to track a desired trajectory over long-time intervals. Following results in [3, 10], we propose the combination of our multigrid method with receding-horizon techniques to solve this concern. Therefore, we use the receding-horizon algorithm developed in [3, Sect. 3.2] in order to show the ability of this approach to track over long-time intervals also in the presence of state-constraints.

We test the receding-horizon algorithm by solving the state-constrained problem with the following desired trajectory $y_d(x_1, x_2, t) := t\sin(2\pi t)\sin(\pi x_1)\sin(\pi x_2)$ and pointwise constraints given by $\underline{y} = -1$ and $\overline{y} = 1$. We study the tracking of the given trajectory, over the time interval $(0, 4)$, considering 8 time windows of size $\Delta t = 0.5$.

In this case, the optimal control problem is solved to the required tolerance, on a grid with $\gamma = 64$, by 8 STMG-W$(2,2)$-cycles, in each time window. We take $\alpha = 1$ and $\beta = 10^{-4}$. Taking bigger values for β results in a worsening of the convergence behavior. In Fig. 3, the time evolution of the state variable compared to the desired trajectory is depicted.

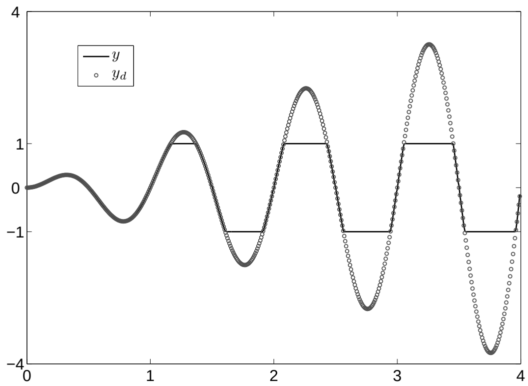

Figure 3: Receding-horizon technique for a state-constrained problem. Time evolution of the state y (solid line) and the desired trajectory y_d (dots) at $(x_1, x_2) = (0.5, 0.5)$. Parameters: $\alpha = 1$, $\beta = 10^{-4}$, $\nu = 10^{-6}$ and $\lambda = 10^{-3}$.

7. Conclusions

A multigrid scheme to solve state-constrained linear parabolic optimal control problems was presented. This approach defines collective smoothing steps that combine local Newton update with projection through the gradient of the reduced cost functional to overcome the lack of differentiability due to the presence of inequality constraints. Besides, we have proved that these collective-smoothing iterations can be formulated as local semismooth Newton methods.

Results of numerical experiments were presented to show that the resulting multigrid schemes provide typical multigrid computational efficiency whenever the value of the weight of the cost of the control ν is of the same order as λ^2. Deterioration of convergence behavior may occur when the Lavrentiev regularization parameter λ becomes too small with respect to ν. This convergence behavior suggests that there is a theoretical relationship between ν and λ which leads the iteration to fast convergence.

References

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

[2] A. Borzì, *Smoothers for control- and state-constrained optimal control problems*, Computing and Visualization in Science, 11 (2008), pp. 59–66.

[3] A. Borzì, *Space-time multigrid methods for solving unsteady optimal control problems*, (Chapter 5) in L.T. Biegler, O. Ghattas, M. Heinkenschloss, D. Keyes and B. van Bloemen Waanders (Eds.), Real-Time PDE-Constrained Optimization, Computational Science and Engineering, Vol. 3, SIAM, Philadelphia, 2007.

[4] A. Borzì and R. Griesse, *Distributed optimal control of lambda-omega systems*, Journal of Numerical Mathematics, 14 (2006), pp. 17–40.

[5] A. Borzì and K. Kunisch, *A multigrid scheme for elliptic constrained optimal control problems*, Comput. Optim. Appl., 31 (2005), pp. 309–333.

[6] A. Borzì and G. von Winckel, *Multigrid methods and sparse-grid collocation techniques for parabolic optimal control problems with random coefficients*, SIAM J. Sci. Comp., 31 (2009), pp. 2172–2192.

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

[8] Th. Dreyer, B. Maar and V. Schulz, *Multigrid Optimization in Applications*, J. Comput. Appl. Math., 120 (2000), pp. 67–84.

[9] E. Emmrich, *Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations*, Computational Methods in Applied Mathematics, 9 (2009), pp. 37–62.

[10] S. González Andrade and A. Borzì, *Multigrid second-order accurate solution of parabolic control-constrained problems*, Comput. Optim. Appl., 2010, pp. 1–32. DOI: 10.1007/s10589-010-9358-y.

[11] W. Hackbusch, *Elliptic Differential Equations*, Springer-Verlag, Berlin, 1992.

[12] W. Hackbusch, *Numerical solution of linear and nonlinear parabolic optimal control problems*, Lecture Notes in Control and Information Science, Springer-Verlag, 1981.

[13] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semi-smooth Newton method*, SIAM J. Opt., 13 (2003), pp. 865–888.

[14] K. Ito and K. Kunisch, *Asymptotic properties of receding horizon optimal control problems*, SIAM J. Control Optim., 40 (5) (2002), pp. 1585–1610.

[15] K. Ito and K. Kunisch, *Semi-smooth Newton methods for state-constrained optimal control problems*, Systems and Control Letters, 50 (2003), pp. 221–228.

[16] R. Kornhuber, *On constrained Newton linearization and multigrid for variational inequalities*, Numer. Math, 91 (2002), pp. 699–721.

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

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

[19] C. Meyer, A. Rösch and F. Tröltzsch, *Optimal control of PDEs with regularized pointwise state constraints*, Comput. Optim. Appl., 33 (2006), 209–228.

[20] Ch. Nagaiah, K. Kunisch and G. Plank, *Numerical solution for optimal control of the reaction-diffusion equations in cardiac electrophysiology*, Comput. Optim. Appl., 2010, pp. 1–30. DOI:10.1007/s10589-009-9280-3.

[21] I. Neitzel and F. Tröltzsch, *On regularization methods for the numerical solution of parabolic control problems with pointwise state constraints*, ESAIM: Control, Optimization and Calculus of Variations, 15 (2009), pp. 426–453.

[22] C. Reisinger and G. Wittum, *On multigrid for anisotropic equations and variational inequalities pricing multi-dimensional european and american options*, Comput. Vis. Sci., 4 (2004), pp. 189–197.

[23] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993),

pp. 353–368.
[24] G. Stadler, *Semismooth Newton and augmented Lagrangian methods for a simplified friction problem*, SIAM J. Optim., 15 (2004), pp. 39–62.
[25] F. Tröltzsch, *Optimal Control of Partial Differential Equations*, AMS, USA, 2010.
[26] U. Trottenberg, C. Oosterlee, and A. Schüller, *Multigrid*, Academic Press, London, 2001.