On the solution of some PDE control problems in the framework of the Pontryagin's maximum principle

Alfio Borzì

Institut für Mathematik, Universität Würzburg
Chair Mathematik IX - Scientific Computing

Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Opening remarks

The Maximum Principle (MP) was developed by L.S. Pontryagin, V.G. Boltyanskii and R.V. Gamkrelidze in 1956–58 and, after the publication of the book by Pontryagin, Boltyanskii, Gamkrelidze and Mischchenko (1961), an enormous and ever lasting effort in the further development and application of the MP to control and extremum problems took place, with fundamental early contribution from A.Ya. Dubovitskii, A.A. Milyutin, and many other mathematicians.

The question of obtaining MP optimality conditions in optimal control problems with ODEs seems now completely solved. However, this seems not the case for PDE control problems and, in this realm, we need to transform the MP into a working tool to be applied to concrete problems.

- Study the MP for PDE control problems;
- Develop numerical strategies based on the MP.

Needle variation

To obtain optimality conditions in optimal control problems different classes of variations have been considered. In particular,

a) Uniformly small variations;
b) Needle-type variations.

The MP corresponds to minima including both the uniformly small and needle variations of the control (intermediate between the classical weak and strong minima).

- Uniformly small variations characterize the Lagrange framework.
- Needle-type variations characterize the MP framework.

One of the purposes of our work is to exploit the needle-variation framework at the numerical level.

Two controlled PDE evolution models

Consider the Cauchy problem $\dot{x} = b(x, t, u)$, $x(0) = x_0$.
The corresponding Liouville equation is given by

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}\left(b(x,t,u)\,\rho(x,t)\right) = 0, \qquad \rho(x,0) = \rho_0(x).$$

This problem describes the ensemble of trajectories of the ODE model for a density of initial conditions ρ_0. It also describes the transport of a density of non-interacting identical particles.
In both cases, the control u is in the (coefficient) drift b.

We also consider the heat equation that models a diffusion process

$$\frac{\partial}{\partial t}y(x,t) - D\,\Delta y(x,t) = u, \qquad y(x,0) = y_0(x).$$

In this model, the control u represents a (distributed) source term.

A Liouville control problem

A transport-type Liouville control problem can be formulated as follows

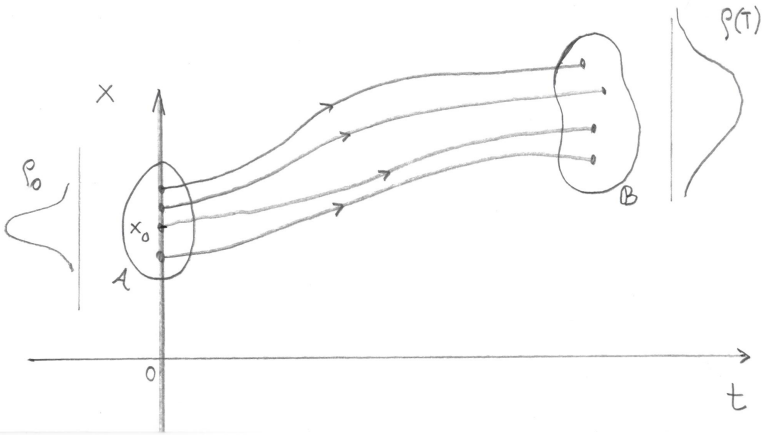
$$\max J(\rho, u) := \int_{\mathcal{B}} \rho(x, T)\ dx,$$

such that

$$\rho_t + \nabla \cdot (b(x, t, u)\, \rho) = 0,$$
$$\rho(x, 0) = \rho_0,$$

where $x \in \mathbb{R}^n$ and $b : \mathbb{R}^n \times \mathbb{R} \times U \to \mathbb{R}^n$ and $\rho_t = \frac{\partial \rho}{\partial t}$, ∇ is the Cartesian gradient w.r.t. x and $\nabla\cdot$ denotes divergence.

The initial density $\rho_0 \geq 0$ is normalized to 1 and has compact support in $\mathcal{A} \subset \mathbb{R}^n$. We also assume that $u(t) \in U$, $U \subset \mathbb{R}^m$, $m \leq n$. The set of values of the control U the target set $\mathcal{B} \subset \mathbb{R}^n$ are compact and nonempty.



Control setting and flow field

We consider the following set of admissible controls

$$\mathcal{U} = \{u = (u_1, \dots u_m) \in L^\infty(0, T; \mathbb{R}^m),\ u(t) \in U \text{ for all } t \in [0, T]\}.$$

The drift b satisfies the following Assumption A1

The map $b : \mathbb{R}^n \times [0, T] \times U \to \mathbb{R}^n$ is continuous;

The map $b(\cdot, t, u) \in C^k(\mathbb{R}^n)$, $k > 2$;

There are constants L, C such that, for all $x, x' \in \mathbb{R}^n$, $t \in [0, T]$, $u \in \mathcal{U}$:

$|b(x, t, u) - b(x', t, u)| \leq L\,|x - x'|$, and $|b(x, t, u)| \leq C\,(1 + |x|)$.

With this assumption and $u \in \mathcal{U}$, the Cauchy problem $\dot{x} = b(x, t, u)$, $x(0) = x_0$, admits a unique absolutely continuous solution, $x : [0, T] \to \mathbb{R}^n$; see Carathéodory theorem.

The unique solution $s \mapsto V_t^s(x)$ to the Cauchy problem $\dot{y}(s) = b(s, y(s))$, $y(t) = x$, defines the map $(s, t, x) \mapsto V_t^s(x)$, which is called the flow of the vector field b.

Existence of an optimal control

Theorem: Let b satisfies the conditions of Assumption A1, it is measurable in t for all fixed $x \in \mathbb{R}^n$, and $\sup_{x\in\mathbb{R}^d} |b(x,t,u)| \leq \beta(t)$ a.e. for some positive function $\beta \in L^1([0,T])$, then the Liouville problem with $\rho_0 \in \mathcal{D}'^0(\mathbb{R}^n))$ admits a unique solution $\rho \in AC([0,T];\mathcal{D}'^0(\mathbb{R}^n))$, where $\mathcal{D}'^k(\mathbb{R}^n)$, $k \in \mathbb{N}$, $k \geq 0$, denotes the subspace of distributions of order k.

Theorem [N.I. Pogodaev, 2016]: Let $b = (b^1, \ldots, b^n)$ has the form

$$b(x,t,u) = b_0(x,t) + \sum_{j=1}^{m} \Phi_j(u_j(t))\, b_j(x,t), \tag{1}$$

where $b_0 = (b_0^1, \ldots, b_0^n)$ and $b_j = (b_j^1, \ldots, b_j^n)$, $j = 1, \ldots, m$, satisfy Assumption A1, and the Φ_j are convex functions. Further, assume that the target set $\mathcal{B}$ be closed. Then the Liouville control problem has a solution in $\mathcal{U}$.

MP characterization of the optimal control

Theorem [N.I. Pogodaev, 2016]: Let $\mathcal{B}$ be a compact set with the interior ball property, $\rho_0 \in C^1(\mathbb{R}^n)$ and b satisfies all conditions of Assumption A1. Let u^* be an optimal control for the Liouville control problem, and ρ^* be the corresponding density function. Then, for almost every $t \in [0, T]$, the following holds

$$\int_{\partial \mathcal{B}^t} \rho^*(x,t)\, b(x,t,u^*(t)) \cdot \eta^*_{\mathcal{B}^t}(x)\; d\sigma(x)$$
$$= \min_{w \in U} \int_{\partial \mathcal{B}^t} \rho^*(x,t)\, b(x,t,w) \cdot \eta^*_{\mathcal{B}^t}(x)\; d\sigma(x),$$

where $\mathcal{B}^t = \left(\bar{V}^t_T\right)^* (\mathcal{B})$, with $\left(\bar{V}\right)^*$ being the optimal (adjoint) flow corresponding to the vector field $(x,t) \mapsto b(x,t,u^*(t))$, $\eta_{\mathcal{B}^t}(x)$ is the measure-theoretic outer unit normal to $\mathcal{B}^t$ at x, σ is $(n-1)$-dimensional Hausdorff measure.

If a $\partial\mathcal{B}$ is a C^2 surface, then $\mathcal{B}$ satisfies the interior ball condition. This is also true for domains with $C^{1,1}$ boundary. In these cases, $\eta_{\mathcal{B}}(x)$ is the usual outer unit normal to $\mathcal{B}$.

Lagrange characterization of the optimal control

If all Φ_j are differentiable w.r.t. u, we have

Theorem: The first-order optimality condition for the Liouville control problem is given by

$$-\int_0^T \int_{\mathbb{R}^n} \left\{ \nabla \cdot \left(\frac{\partial b}{\partial u}(x,t,u^*(t))\, \rho^*(x,t) \right) \right\} q^*(x,t) \cdot (u(t) - u^*(t))\, dx dt \leq 0,$$

where u^* denotes the optimal control, $\rho^* = \rho(u^*)$ represents the solution to the Liouville problem with $u = u^*$. Further, $q^* = q(u^*)$ denotes the solution to the adjoint Liouville equation

$$\begin{aligned} q_t + b(x,t,u) \cdot \nabla q &= 0, \\ q(x,T) &= \chi_{\mathcal{B}}(x), \end{aligned}$$

with $u = u^*$, where $\chi_{\mathcal{B}}$ denotes the characteristic function of the set $\mathcal{B}$.

A reformulation of the MP condition

We use the following

$$\int_{\mathbb{R}^n} (\nabla \cdot v(x))\, \chi_{\mathcal{B}}(x)\, dx = \int_{\partial \mathcal{B}} (v(x) \cdot \eta_{\mathcal{B}}(x))\; d\sigma(x),$$

and notice that $\chi^*_{\mathcal{B}^t}$ coincides with the optimal adjoint function $q^*(x,t)$.
We obtain the following reformulation of the MP condition

$$\int_{\mathbb{R}^n} \nabla \cdot (b(x,t,u^*(t))\, \rho^*(x,t))\, q^*(x,t)\; dx$$
$$= \min_{w \in U} \int_{\mathbb{R}^n} \nabla \cdot (b(x,t,w)\, \rho^*(x,t))\, q^*(x,t)\; dx.$$

If the drift b has the structure (1), we obtain

$$u^*(t) = \operatorname*{argmin}_{w \in U} \sum_{j=1}^{m} \Phi_j(w_j) \int_{\mathbb{R}^n} \nabla \cdot (b_j(x,t)\, \rho^*(x,t))\, q^*(x,t)\; dx.$$

Opening remarks for numerical optimization

Notice that the integral term in

$$u^*(t) = \underset{w \in U}{\operatorname{argmin}} \sum_{j=1}^{m} \Phi_j(w_j) \int_{\mathbb{R}^n} \nabla \cdot (b_j(x,t)\, \rho^*(x,t))\, q^*(x,t)\, dx.$$

does not depend on u (i.e. w). However, we do not know ρ^* and q^*, which are determined by u^*. So we should better write

$$u^*(t) = \underset{w \in U}{\operatorname{argmin}} \sum_{j=1}^{m} \Phi_j(w_j) \int_{\mathbb{R}^n} \nabla \cdot (b_j(x,t)\, \rho(u^*)(x,t))\, q(u^*)(x,t)\, dx.$$

This fact suggests that we have to take into account the dependence of ρ and q on the control u, namely $\rho = \rho(u)$ and $q = q(u)$.

For this reason, we consider the following optimization problem

$$\min_{w \in U} \sum_{j=1}^{m} \Phi_j(w_j) \int_{\mathbb{R}^n} \nabla \cdot (b_j(x,t^k)\, \rho(w)(x,t^k))\, q(w)(x,t^k)\, dx, \qquad \text{(P)}$$

at each t^k

Numerical local control-to-state and control-to-adjoint maps

The straightforward implementation of (P) seems not advantageous for a numerical scheme. However, considering local numerical approximations of the Liouville equation and its adjoint, we have the maps $\rho_h^k = \rho_h^k(w)$ and $q_h^k = q_h^k(w)$ at the time step t^k on the mesh Ω_h.

Illustration: we choose $m = 1$, $b_0 = 0$, $b_1 = 1$, $\Phi_1(u) = u$ in (1) and assume that U is such that $w \geq 0$. Further, assume that the control is piecewise constant in the intervals (t^k, t^{k+1}), $t^k = k\delta t$, $\delta t = T/N$. Denote the control's values in these intervals with $u^{k+1/2}$, $k = 0, \ldots, N-1$.
Let $x_i = ih$, and use first-order upwind schemes

$$\rho_i^{k+1}(w) := \rho_i^{k+1} = \rho_i^k - w\,\Delta t\, D_x^- \rho_i^k.$$

Similarly for the adjoint variable, we have

$$q_i^k(w) := q_i^k = q_i^{k+1} + w\,\Delta t\, D_x^+ q_i^{k+1}.$$

where $D_x^+ v_i = (v_{i+1} - v_i)/h$ and $D_x^- v_i = (v_i - v_{i-1})/h$.

A discrete MP optimization problem

We discretize (P) and obtain the following numerical MP (NMP) optimization problem

$$\min_{w\in U} w \sum_{x_i\in\Omega_h} \frac{h}{2}\left[D_x^+\left(\rho_i^k+\rho_i^{k+1}(w)\right)\right]\frac{1}{2}[q_i^k(w)+q_i^{k+1}], \qquad \text{(Ph)}$$

Notice that in this illustrative example a cubic polynomial in w is obtained. If $\Phi(u)=|u|$, a cubic polynomial in $|w|$ is obtained. A similar result is obtained in two dimensions and we argue that this is true in all dimensions with a multivariate cubic polynomial, and the optimal solution $(u^*)^{k+1/2}$ may belong to the interior of U or to its boundary.

To solve (Ph), we consider a uniform grid of values in U and find the minimum by fast direct search.

A more sophisticated numerical setting

We consider the Sanders' TVD finite-volume scheme.
This scheme is the only known TVD-FV scheme preserving higher-order accuracy even at smooth extrema of the solution. This scheme uses the TVD property of the reconstructing polynomial rather than that of the solution.

Theorem: The scheme of Sanders can be written in a conservative form.

Theorem[R. Sanders, 1988]: The Sanders scheme is positive in the sense that $\rho_0 \geq 0 \Rightarrow \rho_j^k \geq 0$ under the CFL condition $\max_{x\in\Omega} |b'(x,t,u)|\lambda < 1$, $\lambda = \Delta t/h$, $\forall t \in [0,T]$.

Theorem: The Sanders' scheme is second-order accurate in the L^1-norm as follows

$$\|\rho_h(x,T) - \rho(x,T)\|_1 \leq D(T)\, h^2,$$

under the given CFL condition.

Also with Sanders' scheme, our NMP requires to optimize a cubic polynomial at each time step.

A new MP optimization scheme

Input: Number of iter. M, initial guess for the control, $u = u_{(0)}$: compute the corresponding ρ and q.

1. for $r = 1, \ldots, M$ (outer iteration loop)
2. for $k = 0, \ldots, N-1$ (first inner iteration loop; forward update)
3. Solve the NMP to obtain $u^{k+1/2}$, and update ρ^{k+1} using the discrete Liouville equation.
4. end
5. for $k = N-1, \ldots, 0$ (second inner iteration loop; backward update)
6. Solve NMP to compute $u_{(r)}^{k+1/2} := u^{k+1/2}$, and update q^k using the discrete adjoint Liouville equation.
7. end
8. break if convergence criteria is fulfilled: $\|u_{(r)} - u_{(r-1)}\|_{L^2_{\delta t}(0,T)} < \epsilon$.
9. If $r < M$ repeat the outer iteration.

Numerical experiment I: Setting

Case 1: we consider $b(x,t,u) = u$. Let $U = [-1, 2.5]$ and $\mathcal{B} = [r_T, s_T]$. The optimal control without constraints is given by $\bar{u} = (r_T + s_T)/(2T)$. We chose $T = 2$ and $r_T = 2$, $s_T = 3$, $N = 100$. The domain $\Omega = (-8, 8)$ is discretized with 100 subintervals.
For a given $u = u(t)$, we have $\mathcal{B}^t = [r_t, s_t]$, where

$$r_t = r_T + \int_T^t u(\tau)d\tau, \qquad s_t = s_T + \int_T^t u(\tau)d\tau.$$

Further, we have $\eta_{\mathcal{B}^t}(r_t) = -1$ and $\eta_{\mathcal{B}^t}(s_t) = 1$.
The initial density is given by

$$\rho_0(x) = \begin{cases} 1, & -c^{1/2} < x + 2 < c^{1/2}, \\ 0, & \text{otherwise}, \end{cases}$$

where $c = (3/4)^{2/3}$. Hence $\mathcal{A} = [-2 - c^{1/2}, -2 + c^{1/2}]$.

Numerical experiment I: Results

We apply the MP scheme to solve the problem above with the initial guess $u(t) = 1.5$. Our MP iteration converges to the optimal solution with just 2 iterations.

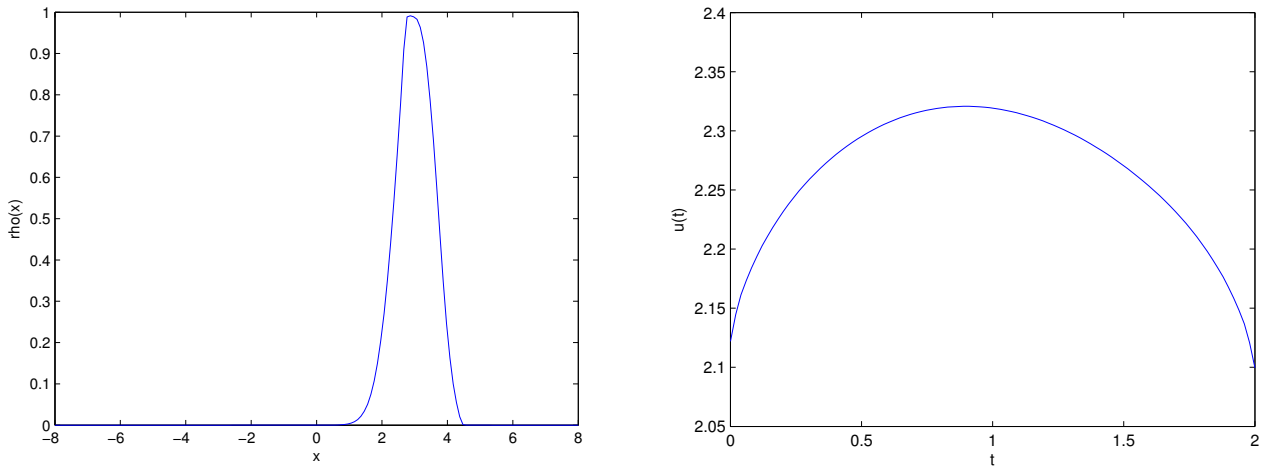


Figure: Case 1: MP solution ρ (left) at $t = T$. Right: the optimal control.

Changing the initial guess for u and the discretization parameters does not change the solution.
A similar problem is solved in the Lagrange framework using a Projected-NCG scheme. In this case, a mollification of the initial condition and ca. 10 times more CPU time are required.

Numerical experiment II: Setting & Results

Case 2: we consider $b(x,t,u) = -2 + 4u\sin(\pi t)$. Let $U = [-2,2]$ and $\mathcal{B} = [r_T, s_T]$ with $T = 2$ and $r_T = 3.5$, $s_T = 4.5$. Otherwise as Case 1. We obtain a bang-bang control with switching at $t = 1$.

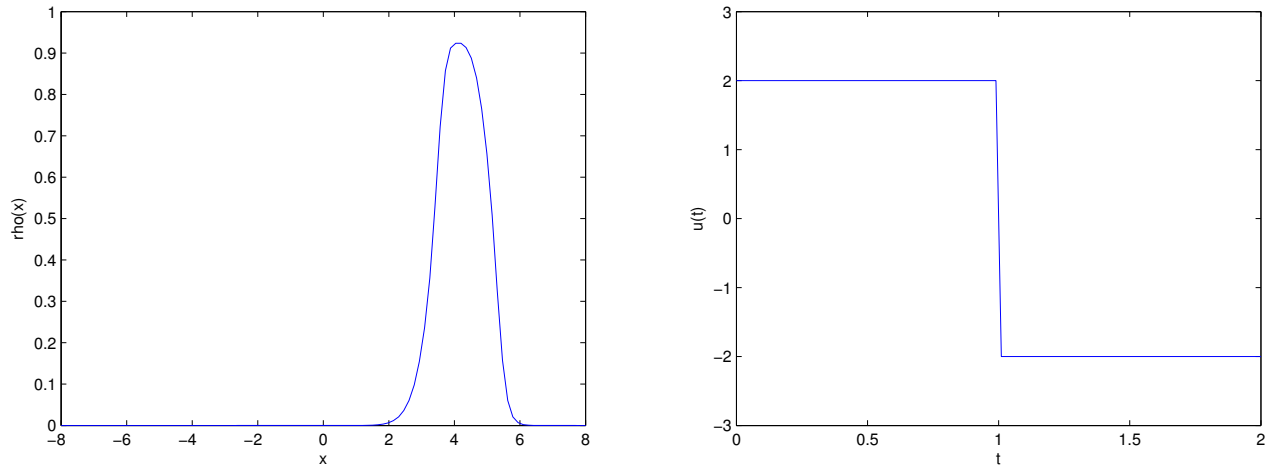


Figure: Case 2: MP bang-bang solution ρ (left) at $t = T$. Right: the optimal control.

Numerical experiment III: Setting & Results

Case 3: we consider a vector field b that is a non-differentiable function of u, namely $b(x,t,u) = -2 + 4|u|\sin(\pi t)$. Otherwise as Case 2.
As in Case 2, we obtain a switch at $t = 1$. However, because of the different control mechanism, the control becomes zero after the switching point.

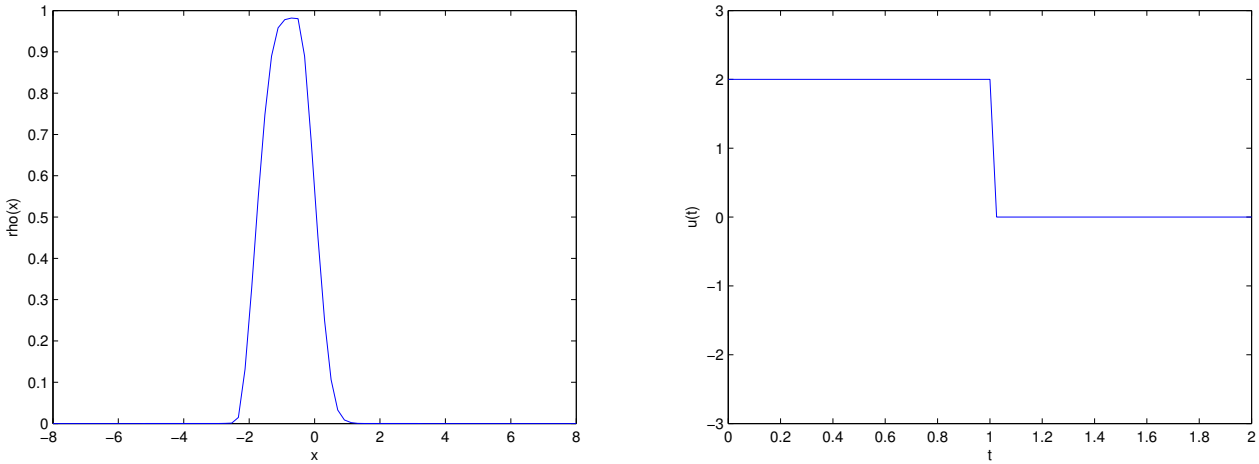


Figure: Case 3: MP bang-bang solution ρ (left) at $t = T$. Right: the optimal control.

A parabolic optimal control problem

We discuss a parabolic control problem in the space-time cylinder $Q = \Omega \times (0, T]$, $\Omega \subset \mathbb{R}^2$, where Ω is a bounded convex domain with C^2 boundary. The purpose of the optimal control is to

$$\min J(y, u) := \frac{1}{2}||y - y_d||^2_{L^2(Q)} + \frac{\alpha}{2}||u||^2_{L^2(Q)} + G(u), \qquad \alpha \geq 0,$$

such that

$$\begin{aligned} y_t(x,t) - D\,\Delta y(x,t) &= u(x,t), && \text{in } Q \\ y(x,0) &= y_0(x), && \text{in } \Omega \times \{t = 0\} \\ y(x,t) &= 0, && \text{on } \Omega \times (0, T) \end{aligned}$$

where $y_0 \in L^\infty(\Omega) \cap H_0^1(\Omega)$ denotes the initial condition, and $u \in U_{ad}$; $D > 0$.

A discontinuous cost functional

The cost functional J includes a 'classical' differentiable part, including a tracking term with a desired trajectory $y_d \in L^\infty(Q)$, and an L^2 cost of the control.
It contains the following discontinuous functional

$$G(u) := \gamma \int_Q g_{d,s}(u(x,t))\, dxdt, \qquad \gamma > 0,$$

where $g_{d,s} : \mathbb{R} \to \mathbb{R}$ is the following non-negative lower semi-continuous function

$$g_{d,s}(u) := \begin{cases} |u-d| & \text{if } |u-d| > s, \\ 0 & \text{otherwise .} \end{cases}$$

Notice that $G(u)$ measures zero costs as far as the control u is in the L^1 closed ball centered in $d \in \mathbb{R}$ with radius $s > 0$. If u is in the complement of this ball, then the cost given by G is of L^1 type.

Existence of an optimal control

The proof of existence of an optimal control is a delicate issue. Since our functional is (not weakly) lower semi-continuous, we need to consider admissible control sets that are compact. This is not the case for

$$U_{ad} := \left\{ u \in L^2(Q) \,|\, u(x,t) \in K_U, \text{ a.e. in } Q \right\}. \qquad (K_U = [u_a, u_b].)$$

However, let for example V_c be given by $V_c := \{v \in W^{1,2}(Q) \,|\, \|v\|_{W^{1,2}(Q)} \leq c\}$ where c is a chosen positive constant. Then $U_{ad} \cap V_c$ is compact. Another possible choice is the set of jump-continuous functions in Q with values in $[u_a, u_b]$.

Assuming $u \in L^2(Q)$, we have that there exists a unique solution $y \in L^2\left(0, T; H^2(\Omega)\right) \cap L^\infty\left(0, T; H_0^1(\Omega)\right)$. We denote this solution with $y = S(u)$, S being the continuous control-to-state map. Therefore we can define $\hat{J}(u) := J(S(u), u)$.

Theorem: Let $g_{d,s}$ be a non-negative lower semi-continuous function with $(g_{d,s} \circ u)$ bounded for all $u \in U_{ad}$ and consider the set of controls $U_{ad} \cap V_c$. Then the parabolic optimal control problem admits an optimal solution $u^* \in U_{ad} \cap V_c$.

Sketch of proof of existence

The functional J is non-negative and we can construct a minimizing sequence with $\lim_{n\to\infty} \hat{J}(u_n) = \bar{J}$ where $\bar{J} = \inf_{u\in U_{ad}\cap V_c} \hat{J}(u)$. Notice that $W^{\mathbf{1},q}(Q)$ is a reflexive Banach space for $1 < q < \infty$, and the set $U_{ad} \cap V_c$ is convex, closed and bounded, then $U_{ad} \cap V_c$ is weakly sequentially compact. Therefore, there exists a subsequence which converges weakly $u_n \rightharpoonup \bar{u}$ in $U_{ad} \cap V_c$.
Further, $W^{\mathbf{1},q}(Q)$ is compactly embedded in $C(\bar{Q})$ equipped with the maximum norm, since the boundary of Q is locally Lipschitz. Therefore there exists a minimizing subsequence, also denoted with $(u_n)_{n\in\mathbb{N}}$, which converges strongly in $C(\bar{Q})$. This means that $u_n(x,t) \to \bar{u}(x,t)$ for $n \to \infty$ for all $(x,t) \in Q$.
We have

$$\bar{J} = \liminf_{n\to\infty} \left(\hat{J}_c(u_n) + G_{d,s}(u_n)\right) \geq \liminf_{n\to\infty} \hat{J}_c(u_n) + \liminf_{n\to\infty} G_{d,s}(u_n).$$

Further, notice that $\hat{J}_c$ is lower semi-continuous and $\liminf_{n\to\infty} \hat{J}_c(u_n) \geq \hat{J}_c(\bar{u})$.
Now, notice that the composition $(g_{d,s} \circ u)$ is lower semi-continuous and bounded for all $u \in U_{ad}$. We can apply Fatou's Lemma as follows

$$\begin{aligned}\liminf_{n\to\infty} G_{d,s}(u_n) &\geq \liminf_{n\to\infty} \gamma \int_Q g_{d,s}(u_n(x,t))\,dxdt \\ &\geq \gamma \int_Q \liminf_{n\to\infty} g_{d,s}(u_n(x,t))\,dxdt \geq \gamma \int_Q g_{d,s}(\bar{u}(x,t))\,dxdt.\end{aligned}$$

Therefore we have $\bar{J} \geq \hat{J}_c(\bar{u}) + G_{d,s}(\bar{u})$. Thus, $\bar{u} \in U_{ad} \cap V_c$ is an optimal control.

Similarly, one can prove existence of an optimal control in $U_{ad} \cap V_c$ for the case where a $L^{\mathbf{0}}$-cost of the control appears

$$J(y,u) := J_c(y,u) + \gamma \int_Q \zeta(u(x,t))\,dxdt, \qquad \zeta(u(x,t)) := \begin{cases} 1 & \text{if } u(x,t) \neq 0 \\ 0 & \text{else.} \end{cases}$$

The adjoint parabolic equation

The adjoint equation corresponding to our parabolic optimal control problem is given by (in weak form)

$$\begin{aligned}
\left(-p'\left(\cdot,t\right),v\right)+D\left(\nabla p\left(\cdot,t\right),\nabla v\right)&=\left(y\left(\cdot,t\right)-y_d\left(\cdot,t\right),v\right)\ \text{ in } Q\\
p\left(\cdot,T\right)&=0 \qquad \text{on } \Omega\times\left\{T=0\right\}\\
p&=0 \qquad \text{on } \partial\Omega.
\end{aligned}$$

This problem is similar to the forward parabolic problem after a transformation of the time variable $t := T-\tau$ and noticing that $y-y_d \in L^2\left(0,T;L^2\left(\Omega\right)\right)$. Hence, there exists a unique solution $p \in L^2\left(0,T;H^2\left(\Omega\right)\right)\cap L^\infty\left(0,T;H_0^1\left(\Omega\right)\right)$ and $p' \in L^2\left(0,T;L^2\left(\Omega\right)\right)$.

The MP Hamiltonian

We apply the results of [Raymond and Zidani, 1999] to our optimal control problem with a discontinuous cost functional. For this purpose, we introduce the (weak) Hamiltonian

$$\begin{aligned} H_Q(x,t,y,u,p) &:= \frac{1}{2}\left(y(x,t)-y_d(x,t)\right)^2+\frac{\alpha}{2}u^2(x,t)+\gamma\, g_{d,s}\left(u(x,t)\right) \\ &+ p(x,t)\,u(x,t)-D\sum_{i=1}^{n} y_{x_i}(x,t)\,p_{x_i}(x,t)\,. \end{aligned}$$

Further, we define

$$F(x,t,y,u) := \frac{1}{2}\left(y(x,t)-y_d(x,t)\right)^2+\frac{\alpha}{2}u^2(x,t)+\gamma\, g_{d,s}\left(u(x,t)\right).$$

Notice that $J(y,u)=\int_Q F(x,t,y,u)\,dxdt$. We do not require continuity of $F(x,t,y,\cdot)$.

The intermediate adjoint and a lemma

We define the intermediate adjoint equation

$$\left(-\tilde{p}'\left(\cdot,t\right),v\right)+D\left(\nabla\tilde{p}\left(\cdot,t\right),\nabla v\right)=\left(\frac{1}{2}\left(y_1\left(\cdot,t\right)+y_2\left(\cdot,t\right)\right)-y_d\left(\cdot,t\right),v\right),$$

$$\tilde{p}\left(\cdot,T\right)=0,$$

where $y_1 = S(u_1)$ and $y_2 = S(u_2)$, $u_1, u_2 \in U_{ad}$.

We have the following

Lemma [Raymond and Zidani, 1999]: The following equation holds

$$J\left(y_1,u_1\right)-J\left(y_2,u_2\right)=\int_Q\left(H_Q\left(x,t,y_2,u_1,\tilde{p}\right)-H_Q\left(x,t,y_2,u_2,\tilde{p}\right)\right)dxdt,$$

where $y_1 = S(u_1)$ and $y_2 = S(u_2)$ and $\tilde{p}$ is the solution to the intermediate adjoint equation.

The needle variation

The classical approach to prove the MP is the method of needle variation [Dmitruk and Osmolovskii, 2016]. Let $S_k(x_0, t_0)$ be an open ball centered at $(x_0, t_0) \in Q$ with radius $s^k_{x_0,t_0}$ such that $\lim_{k\to\infty} |S_k(x_0, t_0)| = 0$. We define the needle variation at (x_0, t_0) of an admissible control $\bar{u} \in U_{ad}$ as follows

$$u_k(x,t) := \begin{cases} \bar{u}(x,t) & \text{on } Q \backslash S_k(x_0,t_0) \\ u & \text{in } S_k(x_0,t_0) \cap Q, \end{cases}$$

where $u \in K_U$.

Lemma: Let $\bar{u} \in U_{ad}$ be an admissible control and $u \in K_U$. Furthermore, let u_k be defined as above and $y_k = S(u_k)$. Then,

$$\lim_{k\to\infty} \frac{1}{|S_k(x,t)|} \left(J(y_k,u_k) - J(\bar{y},\bar{u})\right)$$
$$= H_Q(x,t,\bar{y}(x,t),u,\bar{p}(x,t)) - H_Q(x,t,\bar{y}(x,t),\bar{u}(x,t),\bar{p}(x,t)),$$

for almost all $(x,t) \in Q$, and $\bar{y} = S(\bar{u})$ and $\bar{p}$ is the solution to adjoint equation with $y \leftarrow \bar{y}$.

The MP optimality condition

Theorem: Let (y^*, u^*, p^*) be an optimal solution to the parabolic optimal control problem where $y^* = S(u^*)$ and p^* is the solution to the adjoint problem with $y \leftarrow y^*$. Then, the following holds

$$H_Q\left(x,t,y^*\left(x,t\right),u^*\left(x,t\right),p^*\left(x,t\right)\right) = \min_{u\in K_U} H_Q\left(x,t,y^*\left(x,t\right),u,p^*\left(x,t\right)\right)$$

for almost every $(x,t) \in Q$.

Recall the previous lemma and the construction of the needle variation of u^*. Notice that $\left(J\left(y_k,u_k\right) - J\left(y^*,u^*\right)\right) \geq 0$. Then, we have the following

$$\begin{aligned} 0 &\leq \lim_{k\to\infty} \frac{1}{|S_k\left(x_0,t_0\right)|}\left(J\left(y_k,u_k\right) - J\left(y^*,\bar{u}\right)\right) \\ &= H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u,p^*\left(x_0,t_0\right)\right) - H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u^*\left(x_0,t_0\right),p^*\left(x_0,t_0\right)\right), \end{aligned}$$

for almost all $(x_0,t_0) \in Q$. Therefore for almost every point of Q, we have

$$H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u^*\left(x_0,t_0\right),p^*\left(x_0,t_0\right)\right) \leq H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u,p^*\left(x_0,t_0\right)\right),$$

for all $u \in K_U$. Consequently, we obtain

$$H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u^*\left(x_0,t_0\right),p^*\left(x_0,t_0\right)\right) = \min_{u\in K_U} H_Q\left(x_0,t_0,y^*\left(x_0,t_0\right),u,p^*\left(x_0,t_0\right)\right),$$

for almost every point of Q.

A Hamiltonian system

Within the MP framework, we can define the strong Hamiltonian

$$\begin{aligned}\hat{H}(x,t,y,u,p) &:= \frac{1}{2}\left(y(x,t)-y_d(x,t)\right)^2+\frac{\alpha}{2}u^2(x,t)+\gamma g_{d,s}(u(x,t))\\ &+ p(x,t)\,u(x,t)+D\,p(x,t)\,\Delta y(x,t),\end{aligned}$$

and its 'adjoint'

$$\begin{aligned}\tilde{H}(x,t,y,u,p) &:= \frac{1}{2}\left(y(x,t)-y_d(x,t)\right)^2+\frac{\alpha}{2}u^2(x,t)+\gamma g_{d,s}(u(x,t))\\ &+ p(x,t)\,u(x,t)+D\,y(x,t)\,\Delta p(x,t).\end{aligned}$$

Then the strong formulation of the state equation is given by

$$\frac{\partial}{\partial t}y=\frac{\partial}{\partial p}\hat{H},$$

where $y(\cdot,0)=y_0$ and zero boundary conditions, and the strong formulation of the adjoint equation is given by

$$\frac{\partial}{\partial t}p=-\frac{\partial}{\partial y}\tilde{H},$$

where $p(\cdot,T)=0$ and zero boundary conditions.

Opening remarks for numerical optimization

Consider a discretized setting for our space-time cylinder $Q = \Omega \times (0, T)$ with $\Omega = (a, b)$. We have

$$Q_{h,\triangle t} := \{(x_i, t_m)\,, \mid x_i = a + ih \in \Omega_h,\ t_m = m\triangle t,\ \}\,,$$

The space and time mesh-sizes are given by $h := \frac{b-a}{N}$, $\triangle t := \frac{T}{N_t}$. We assume that the grid points $(x_{i_1 \ldots i_n}, t_m)$ and $t_m = m\triangle t$ are ordered lexicographically.

We approximate state and adjoint equations using the implicit Euler scheme and finite differences. y_i^m and p_i^m denote the approximations to $y\,(a + ih, m\triangle t)$ and $p\,(a + ih, m\triangle t)$, respectively, and $(y_d)_i^m = y_d\,(a + ih, m\triangle t)$.

Our purpose is to investigate an iterative numerical needle variation procedure.

Local minimization of the Hamiltonian and updates

The basic idea is to minimize H_Q pointwise. In order to calculate the element of K_U which minimizes the Hamiltonian H_Q in a given point of $Q_{h,\triangle t}$, we discretize K_U and choose the corresponding minimizing element (by array search) in K_U to update the control u^* at this point.

The need arises to update y^* and p^*. Indeed, one could proceed recalculating these functions after every control update, but this approach requires a very large computational effort.

For this reason, we choose an integer $recalc \in \mathbb{N}$ (a fraction of the total number of grid points) and proceed as follows. We test a number of grid points equal to $recalc \in \{1, \ldots, N_t\,(N-1)\}$ and, if the control is updated at least once, we re-compute the state variable y and the value of the cost functional J.

If the value of J does not represent an improvement towards the minimum, then we discard the control updates and go to the next *recalc* grid points.

However, if the control update results in a reduction of the cost functional, we keep the changes in u and y and correspondingly update the adjoint variable p and go to the next *recalc* grid points.

An iterative needle-variation (INV) scheme

1. Discretize Q and set $recalc \in \{1, \ldots, N_t\,(N-1)^n\}$, $\epsilon_J, \epsilon_H > 0$, $k \leftarrow 0$.
2. Choose $u^{\mathbf{0}} \in U_{ad}$ and set $\hat{u} \leftarrow u^{\mathbf{0}}$; compute $y^{\mathbf{0}}$ with $u \leftarrow u^{\mathbf{0}}$ and $p^{\mathbf{0}}$ with $y \leftarrow y^{\mathbf{0}}$.
3. Set $J_{old} \leftarrow J\left(y^{k-\mathbf{1}}, u^{k-\mathbf{1}}\right)$ and $H_{old} \leftarrow \int_Q H_Q\left(x, t, y^{k-\mathbf{1}}, u^{k-\mathbf{1}}, p^{k-\mathbf{1}}\right) dxdt$
4. Set $counter \leftarrow 0$, $converged \leftarrow 1$, $u_{changed} \leftarrow 0$ and $i \leftarrow 0$
5. For $m = 0, 1, \ldots, N_t - 1$, $i_j = 1, \ldots, N-1$, $j \in \{1, \ldots, n\}$; $i \leftarrow i+1$;
   - 5.1 Choose $u \in K_U$ such that $u = \text{argmin}_{\tilde{u} \in K_U} H_Q\left(x_i, t_m, \left(y^k\right)^m_{i_{\mathbf{1}} \ldots i_n}, \tilde{u}, \left(p^k\right)^m_{i_{\mathbf{1}}, \ldots, i_n}\right)$
   - 5.2 Set $counter \leftarrow counter + 1$, $\quad \hat{u}^m_{i_{\mathbf{1}} \ldots i_n} \leftarrow u$;
     If $\hat{u}^m_{i_{\mathbf{1}} \ldots i_n} \neq (u^k)^m_{i_{\mathbf{1}} \ldots i_n}$, then $u_{changed} \leftarrow 1$
   - 5.3 If $\left((counter \geq recalc \text{ or } i = m\,(N-1)^n) \text{ and } u_{changed} = 1\right)$, then
     Compute $\hat{y}$ for $u \leftarrow \hat{u}$
     If $J(\hat{y}, \hat{u}) - J\left(y^k, u^k\right) < 0$, then
     Set $k \leftarrow k+1$, $\quad$ Set $y^k \leftarrow \hat{y}$ and $u^k \leftarrow \hat{u}$
     Compute p^k with $y \leftarrow y^k$
     $converged \leftarrow 0$
     Else
     Set $\hat{u} \leftarrow u^{k-\mathbf{1}}$
     End
     Set $counter \leftarrow 0$
     End
   - 5.4 If $counter \geq recalc$, then $counter \leftarrow 0$ and $u_{changed} \leftarrow 0$

   End
6. If $converged = 1$ or $|J\left(y^k, u^k\right) - J_{old}| < \epsilon_J$ and
   $|\int_Q H_Q\left(x, t, y^k(x,t), u^k(x,t), p^k(x,t)\right) dxdt - H_{old}| < \epsilon_H$, then stop, else go to 3.

Numerical experiment I: Setting

In our numerical experiments, we consider $\Omega = (0,1)$ and $T = 1$. The initial guess of the control is the zero function.
We choose $recalc = \frac{N_t(N-1)}{N_{re}}$ where $N_{re} = 50$ represents the maximum number of times the variable y and p are re-computed on all the grid points. The numerical parameters are set as follows, $N = 100$, $N_t = 200$, $D = \frac{1}{5}$, and if not otherwise stated $\alpha = 10^{-5}$, $\gamma = 10^{-1}$. Furthermore, we have, $K_U = [0, 10]$ discretized in steps of $\frac{1}{100}$, and the desired trajectory

$$y_d(x,t) = \begin{cases} 5 & \text{if } \bar{x}(t) - c \leq x \leq \bar{x}(t) + c \\ 0 & \text{else} \end{cases}$$

where $\bar{x}(t) := x_0 + \frac{2}{5}(b-a)\sin\left(2\pi\frac{t}{T}\right)$, $x_0 = \frac{b+a}{2}$, and $c = \frac{7}{100}(b-a)$.
In J, we set $d = 0$, $s = 1$.
Further, ϵ_J and ϵ_H are taken equal to machine precision.

Numerical experiment I: Results

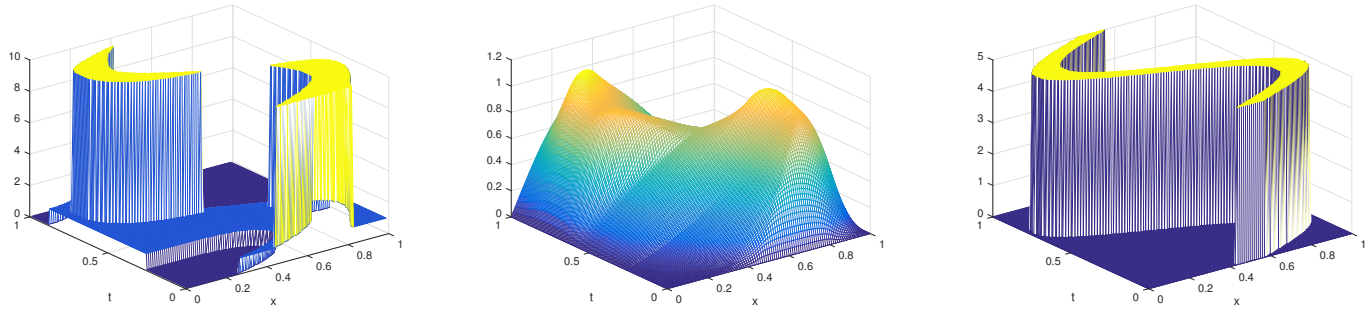


Figure: Optimal solution for the first experiment; from left to right: u, y, and y_d.

The INV algorithm converges in 13 steps and we obtain the state and control functions depicted in the Figure. The plot of the control function shows the action of the discontinuous cost of the control given by $g_{0,1}$ and the presence of the control's upper bound at 10.

Numerical experiment I: Robustness

We investigate the computational performance of the INV algorithm with respect to different choices of the optimization parameters.

α	γ	k	CPU time/s	J	$\|\|y - y_d\|\|$
10^{-1}	10^{-5}	12	8.5	1.63	1.77
10^{-3}	10^{-5}	28	20.0	1.33	1.62
10^{-5}	10^{-5}	28	19.0	1.31	1.62
0	10^{-5}	16	10.9	1.31	1.62
0	0	20	13.1	1.31	1.62
10^{-5}	0	17	11.8	1.32	1.62
10^{-5}	10^{-3}	19	13.5	1.32	1.62
10^{-5}	10^{-2}	18	12.1	1.34	1.62
10^{-5}	10^{-1}	13	9.0	1.51	1.66

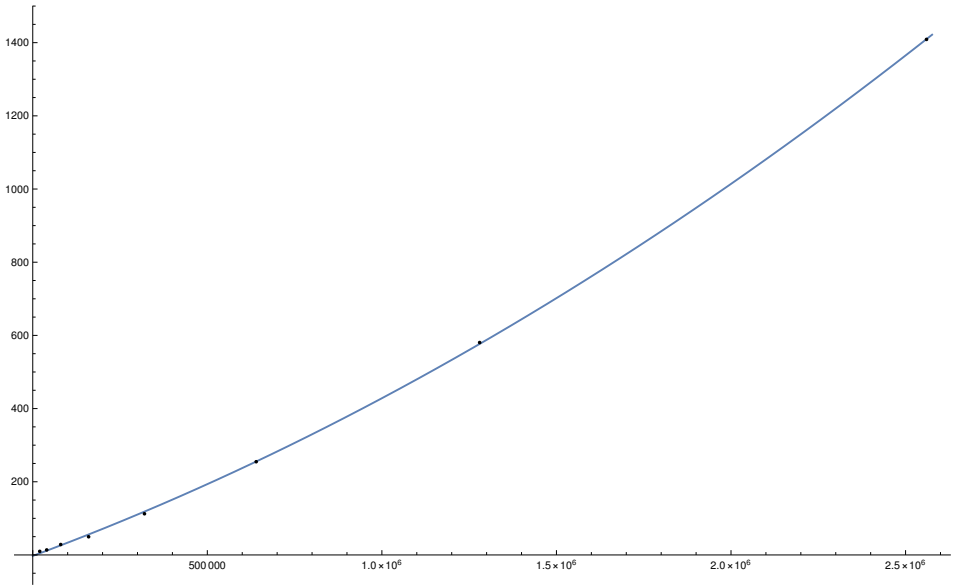
Numerical experiment I: Complexity

We argue that the complexity of the INV algorithm satisfies

$$C\left(N_{gp}\right) \leq N_{gp}\left(c_1 + c_2 N_{gp}\right),$$

for $c_1, c_2(N_{re}) > 0$; $N_{gp} = N_t N$. To validate this estimate, we solve the same optimization problem as above using different scales of discretization.

$\frac{N}{100} \times \frac{N_t}{100}$	1×2	2×2	2×4	4×4	4×8	8×8	8×16	16×16
CPU time/s	9.8	13.7	28.5	49.7	112.5	254.8	580.5	1408.9



Numerical experiment I ($\gamma = 0$): INV vs. projected gradient & Armijo linesearch

α	$N_{gp} = N \times N_t$	INV		pGM	
		CPU/s	n. iter	CPU/s	n. iter
10^{-1}	200×400	2.9	13	3.1	103
10^{-1}	400×800	8.8	13	9.4	103
10^{-1}	800×1600	32.4	13	33.8	103
10^{-2}	200×400	7.0	30	27.0	927
10^{-2}	400×800	19.8	30	82.2	927
10^{-2}	800×1600	76.8	32	307.1	927
10^{-3}	200×400	6.3	30	190.2	6541
10^{-3}	400×800	18.3	29	568.4	6516
10^{-3}	800×1600	69.0	29	2125.2	6509

Numerical experiment I ($\gamma = 0$): INV vs. projected NCG

α	$N_{gp} = N \times N_t$	INV		pNCG	
		CPU/s	n. iter	CPU/s	n. iter
10^{-1}	200×400	2.9	13	1.4	15
10^{-1}	400×800	8.8	13	3.5	15
10^{-1}	800×1600	32.4	13	14.0	15
10^{-3}	200×400	6.3	30	1.0	8
10^{-3}	400×800	18.3	29	2.7	8
10^{-3}	800×1600	69.0	29	9.3	8
10^{-5}	200×400	9.4	47	1.0	7
10^{-5}	400×800	18.7	29	2.5	7
10^{-5}	800×1600	69.2	29	55.4	84
10^{-7}	200×400	3.2	15	1.0	7
10^{-7}	400×800	11.8	19	2.5	7
10^{-7}	800×1600	66.3	32	8.4	7

Numerical experiment II: Setting & Results

In this experiment, we consider a parabolic optimal control problem with a L^0-cost functional. We choose

$$y_d(x,t) = 5\sin\left(2\pi\frac{t}{T}\right).$$

Further, we have $K_U = [-10, 10]$, $\alpha = 10^{-5}$, $N = 800$ and $N_t = 800$. The results for the case with L^0 costs and $\gamma = 0.1$ are reported in the Figure. Similar results are obtained with $g(u) = \sqrt{|u|}$.

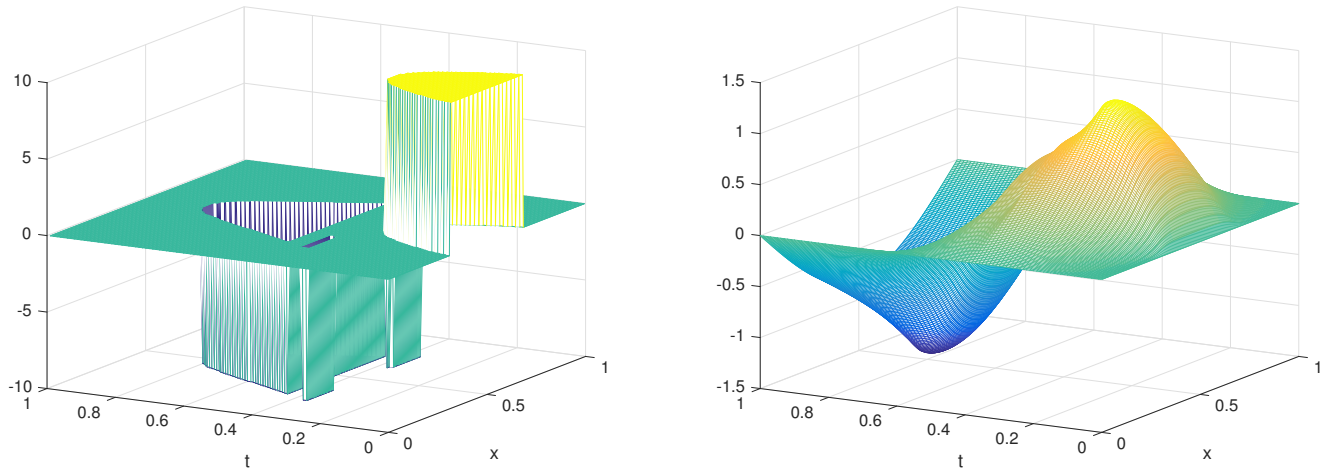


Figure: Optimal solution for the second experiment; left: the control; right: the state function.

Closing remarks

The Liouville model allows to directly 'lift' ODE control problems in PDE control problems and thus help in implementing the MP in the latter case.

The idea of needle variation seems the most appropriate to develop a 'good' MP numerical framework. Collective updates and multigrid methods make this approach efficient and robust.

This work was done in collaboration with

Figure: Dr. Souvik Roy and Mr. Tim Breitenbach

Formulation and Numerical Solution of Quantum Control Problems
A. Borzì, G. Ciaramella, M. Sprengel, SIAM Publications, Philadelphia, 2017