SpezialForschungsBereich F 32

Karl–Franzens Universität Graz
Technische Universität Graz
Medizinische Universität Graz

On the use of state constraints in optimal control of singular PDEs

Christian Clason   Barbara Kaltenbacher

SFB-Report No. 2011–030   December 2011

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the
Austrian Science Fund (FWF)

SFB sponsors:

- **Austrian Science Fund** (FWF)
- **University of Graz**
- **Graz University of Technology**
- **Medical University of Graz**
- **Government of Styria**
- **City of Graz**

Das Land Steiermark

Stadt GRAZ Wissenschaft

ON THE USE OF STATE CONSTRAINTS IN OPTIMAL CONTROL OF SINGULAR PDES

Christian Clason* Barbara Kaltenbacher†

December 27, 2011

We consider optimal control of nonlinear partial differential equations involving potentially singular solution-dependent terms. Singularity can be prevented by either restricting controls to a closed admissible set for which well-posedness of the equation can be guaranteed, or by explicitly enforcing pointwise bounds on the state. By means of an elliptic model problem, we constrast the requirements for deriving existence of solutions and first order optimality conditions for both the control-constrained and the state-constrained formulation. Our theoretical findings as well as numerical tests illustrate that control constraints lead to severe restrictions on the attainable states, which is not the case for state constraints.

1 INTRODUCTION

Many nonlinear models in science and engineering involve solution-dependent terms that can become singular, signaling failure of the model or actual physical blow-up. This introduces difficulties for the optimal control of such equations, since existence of a solution to the state equation can be guaranteed a priori only for controls which are small in some norm. Ensuring well-posedness of the control-to-state mapping thus requires imposing such a bound as a control constraint. The purpose of this work is to demonstrate that this is an overly restrictive approach that can prevent even approximate attainment of the prescribed target, and that this can be avoided by incorporating explicit pointwise bounds as state constraints into the optimal control problem.

*Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria (christian.clason@uni-graz.at)

†Institute of Mathematics, Alpen-Adria Universität Klagenfurt, Universitätsstrasse 65–67 A-9020 Klagenfurt, Austria (barbara.kaltenbacher@uni-klu.ac.at)

We focus on a simple model problem which allows explicit calculation of bounds (and is of practical relevance in itself), although the approach is applicable for a wide range of nonlinear models with potential degeneracy. Consider

$$\begin{cases} -\Delta y = -\dfrac{\beta u}{(1+y)^2} & \text{on } \Omega, \\ \quad y = 0 & \text{on } \partial\Omega, \end{cases} \tag{1.1}$$

for $\Omega \subseteq \mathbb{R}^n$, $n \in \{1, 2, 3\}$ (typically $n = 2$), which models the deflection of an idealized electrostatically actuated microelectromechanical systems (MEMS), where y is the mechanical displacement, u is the reciprocal of the dielectric constant and β is a dimensionless number proportional to the applied voltage [11], [12]. The case $y(x) = -1$ corresponds to the "pull-in" instability where the membrane touches the ground plate. Motivated by the problem of optimal design of dielectric properties, we consider the control problem

$$\begin{cases} \min\limits_{y,u} \dfrac{1}{2} \|y - y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2} \|u\|^2_{L^2(\Omega)} \\ -\Delta y = -\dfrac{\beta u}{(1+y)^2}, \quad y|_{\partial\Omega} = 0, \end{cases}$$

where y_d is a prescribed displacement under a given voltage. To prevent the "pull-in" instability, we contrast the use of two different constraints:

- *control constraints*: $\|u\|_{L^2(\Omega)} \leqslant M_u$, where M_u is chosen small enough so that $\|y\|_{C(\Omega)} < 1$,
- *state constraint*: $-y(x) \leqslant M_y$ for $M_y < 1$ and $x \in \Omega$.

While the former approach allows application of standard techniques since the control-to-state mapping $u \mapsto y$ is well-defined for all admissible u, this is not the case for the latter. The main difficulties are that (1.1) is not of monotone type, and that the reduced formulation is not well-defined for all $u \in \mathcal{U}$, due to potential degeneracy of the equation. Indeed it has been shown (Theorem 3.1 in [11]) that (1.1) is not solvable for sufficiently large controls βu; hence the control or state constraints are vital for guranteeing that the PDE is solvable at all. Correspondingly, the crucial step for deriving optimality conditions in both approaches is proving well-posedness of the linearized equation, for which the condition in the state-constrained case will turn out to be less restrictive.

Optimal control of elliptic partial differential equations subject to control constraints is by now classical (cf., eg., [17] and the literature cited therein). The literature on state constraints is now similarly extensive, and we only refer to (besides the aforementioned monograph) the standard works [1, 4, 2]. Constraints for discontinuous and possibly unbounded states were considered in [15], although their objective is different from ours (missing regularity of the state as opposed to pointwise singularity of the equation here) and they consider linear PDE constraints only.

Although the results on optimal control of the MEMS model problem are new in themselves, the use of state constraints for the purpose of preventing singularity of the state equation is the main novel idea of this paper. It is expected to have many practically relevant applications for the control of potentially singular PDEs, such as the Westervelt and the Kuznetsov equation in nonlinear acoustics [6], Euler equations with general pressure laws in isentropic gas dynamics [5], or the Richards equation modelling flow in unsaturated porous media [13], to name just a few examples.

Remark 1.1. Although in the considered application, the physically relevant regime requires $\beta \geqslant 0$ and $1 \geqslant u > 0$, we do not enforce these constraints in order to focus on the use of control or state constraints to prevent singularity of the state equation, since the main purpose of this model is to illustrate and contrast the general approaches.

This paper is organized as follows. In the remainder of this section, we introduce some common notation for both approaches. In Section 2, we discuss the control-constrained optimal control problem, for which we derive sufficient conditions for well-posedness and differentiability of the control-to-state mapping in 2.1. These are used in 2.2 to obtain first order optimality conditions. The case of state constraints is considered in Section 3, where we consider the state equation as a nonlinear equality constraint, for which we derive continuity and differentiability in Section 2.1. The corresponding optimality system is discussed in Section 3.2. Section 4 contrasts both approaches for a simple numerical example.

In the following, we denote the unconstrained control space by

$$\mathcal{U} := L^2(\Omega)$$

and the unconstrained state space by

$$\mathcal{Y} := H^2(\Omega) \cap H^1_0(\Omega) \hookrightarrow C_0(\Omega),$$

where the embedding is continuous and compact. We will make use of the lifting from $L^2(\Omega)$ to $C_0(\Omega)$ via the inverse Laplacian with Dirichlet conditions, which is a continuous operator, i.e., there exists a $C_L > 0$ such that for all $v \in L^2(\Omega)$,

$$\left\|(-\Delta)^{-1} f\right\|_{C(\Omega)} \leqslant C_L \left\|f\right\|_{L^2(\Omega)} \tag{1.2}$$

holds.

For given $M_u > 0$, we define the admissible control space

$$\mathcal{U}_M := \left\{u \in \mathcal{U} : \|u\|_{L^2(\Omega)} \leqslant M_u\right\}$$

and for $M_y < 1$, the admissible state space

$$\mathcal{Y}_M := \{y \in \mathcal{Y} : -y(x) \leqslant M_y \text{ for all } x \in \Omega\}.$$

Remark 1.2. The results below remain valid for any cost functional $J(u, y)$ that is bounded from below, coercive with respect to u (this condition may be omitted in the control-constrained case), weakly lower semicontinuous, and differentiable on $\mathcal{U} \times \mathcal{Y}$.

2 CONTROL CONSTRAINTS

For the sake of comparison, we first consider the control-constrained optimal control problem

$$(\mathcal{P}_{cc}) \qquad \begin{cases} \min\limits_{y\in\mathcal{Y},u\in\mathcal{U}} \dfrac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)} \\ -\Delta y = -\dfrac{\beta u}{(1+y)^2}, \\ \|u\|^2_{L^2(\Omega)} \leqslant M_u^2, \end{cases}$$

where M_u is chosen sufficiently small to obtain well-posedness of the state equation.

2.1 STATE EQUATION

We first derive sufficient conditions on u such that (1.1) has a unique solution. The proof technique is chosen to yield explicit bounds which can be used as control constraints.

Theorem 2.1. *There exists a constant $M_u > 0$ such that for all $u \in \mathcal{U}_M$, there exists a unique solution $y \in \mathcal{Y}$ to (1.1).*

Proof. We use Banach's Fixed Point Theorem with the fixed point operator $T : W \to W$, $v \mapsto Tv := y$, where y solves

$$\begin{cases} -\Delta y = -\dfrac{\beta u}{(1+v)^2} & \text{on } \Omega\,, \\ y = 0 & \text{on } \partial\Omega, \end{cases}$$

and

$$W = \left\{ v \in \mathcal{Y} : \|v\|_{C(\Omega)} \leqslant \overline{m} \right\} \tag{2.1}$$

for some $\overline{m} \in (0,1)$. To show that T is a self-mapping, consider an arbitrary $v \in W$. Then we have

$$\|Tv\|_{C(\Omega)} \leqslant C_L|\beta| \left\| \frac{u}{(1+v)^2} \right\|_{L^2(\Omega)} \leqslant \frac{C_L|\beta|}{(1-\overline{m})^2}\|u\|_{L^2(\Omega)}\,,$$

where C_L is the constant in (1.2), and the right hand side is not larger than $\overline{m}$ if

$$\|u\|_{L^2(\Omega)} \leqslant \frac{\overline{m}(1-\overline{m})^2}{C_L|\beta|}\,. \tag{2.2}$$

Contractivity of T follows from the fact that for any $v, w \in W$, there holds

$$\begin{aligned} \|Tv - Tw\|_{C(\Omega)} &\leqslant C_L|\beta| \left\| \left(\frac{1}{(1+v)^2} - \frac{1}{(1+w)^2} \right) u \right\|_{L^2(\Omega)} \\ &\leqslant C_L|\beta| \left\| \left(\frac{1}{(1+v)^2(1+w)} + \frac{1}{(1+w)^2(1+v)} \right)(v-w)u \right\|_{L^2(\Omega)} \\ &\leqslant \frac{2C_L|\beta|}{(1-\overline{m})^3}\|u\|_{L^2(\Omega)}\|v-w\|_{C(\Omega)}\,. \end{aligned}$$

Hence contractivity holds if

$$\|u\|_{L^2(\Omega)} < \frac{(1-\overline{m})^3}{2C_L|\beta|}\,. \tag{2.3}$$

The maximum over $\overline{m} \in (0,1)$ of the minimum of the two right hand sides in (2.2), (2.3) can be found by equilibrating $\frac{\overline{m}(1-\overline{m})^2}{C_L|\beta|} = \frac{(1-\overline{m})^3}{2C_L|\beta|}$, which yields the optimal bound

$$\overline{m} = \tfrac{1}{3}\,.$$

Therefore, a solution to (1.1) exists if

$$\|u\|_{L^2(\Omega)} \leqslant M_u := \frac{4}{27C_L|\beta|}. \tag{2.4}$$

□

Remark 2.2. For $\Omega = (0,1)^2$ and $\beta = 1$, we can explicitly calculate

$$M_u = \frac{4}{27} \approx 0.14815.$$

We can now introduce for sufficiently small M_u the control-to-state mapping

$$S : \mathcal{U}_M \to \mathcal{Y}, \qquad u \mapsto y \text{ solving (1.1)}.$$

By Theorem 2.1, this mapping is well-defined. We next show continuity of S.

Theorem 2.3. *There exists a constant $M_u > 0$ such that the control-to-state mapping $S : \mathcal{U}_M \to \mathcal{Y}$ is Lipschitz continuous and weakly continuous.*

Proof. The difference $\tilde{y} - y$ between two solutions $\tilde{y} = S(\tilde{u})$ and $y = S(u)$ satisfies

$$\begin{cases} -\Delta(\tilde{y}-y) - \dfrac{\beta\tilde{u}(2+\tilde{y}+y)}{(1+y)^2(1+\tilde{y})^2}(\tilde{y}-y) = -\dfrac{\beta}{(1+y)^2}(\tilde{u}-u) & \text{on } \Omega, \\ \tilde{y}-y = 0 & \text{on } \partial\Omega, \end{cases}$$

hence

$$\|\Delta(\tilde{y}-y)\|_{L^2(\Omega)} \leqslant \frac{2|\beta|}{(1-\overline{m})^3}\|\tilde{u}\|_{L^2(\Omega)}\|\tilde{y}-y\|_{C(\Omega)} + \frac{|\beta|}{(1-\overline{m})^2}\|\tilde{u}-u\|_{L^2(\Omega)}$$

where we have used the fact that $\tilde{y}, y \in W$, with W defined as in (2.1). Thus we arrive at

$$\|\Delta(\tilde{y}-y)\|_{L^2(\Omega)} \leqslant L(\overline{m}, M_u)\,\|\tilde{u}-u\|_{L^2(\Omega)}\,,$$

where

$$L(\overline{m}, M_u) := \left\{1 - \frac{2C_L|\beta|M_u}{(1-\overline{m})^3}\right\}^{-1}\frac{|\beta|}{(1-\overline{m})^2}\|\tilde{u}-u\|_{L^2(\Omega)}\,,$$

provided (2.3) holds, which implies that the expression in braces is positive. This shows continuity with Lipschitz constant $L(\overline{m}, M_u)$.

To see weak continuity we use a subsequence-subsequence argument. Let $(u_n)_{n\in\mathbb{N}} \subseteq \mathcal{U}_M$ be an arbitrary sequence converging weakly to u^* in $L^2(\Omega)$. Then $u^* \in \mathcal{U}_M$ and the sequence $(S(u_n))_{n\in\mathbb{N}} \subseteq \mathcal{Y}$ is bounded in $\mathcal{Y}$ (by $|\beta|(1-\overline{m})^{-2} \sup_{n\in\mathbb{N}} \|u_n\|_{L^2(\Omega)}$ with $\overline{m}$ as in the proof of Theorem 2.1). Therefore any subsequence of $(S(u_n))_{n\in\mathbb{N}}$ has a weakly convergent subsequence. Moreover, the limit y^* of any weakly convergent subsequence of $(S(u_n))_{n\in\mathbb{N}}$ – which by compactness of the embedding $\mathcal{Y} \to C_0(\Omega)$ contains a subsequence converging strongly in $C_0(\Omega)$ to y^* – has to solve (1.1) with $u = u^*$. Therefore, by uniqueness in Theorem 2.1, y^* coincides with $S(u^*)$. □

To show differentiability, we need the well-posedness of the linearized state equation at $\overline{u} \in \mathcal{U}_M, \overline{y} \in \mathcal{Y}$:

$$\begin{cases} -\Delta y - \dfrac{2\beta\overline{u}}{(1+\overline{y})^3} y = f & \text{on } \Omega, \\ \qquad\qquad\qquad y = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.5}$$

for given $f \in L^2(\Omega)$.

Lemma 2.4. *For all $\overline{u} \in \mathcal{U}_M$, $\overline{y} = S(\overline{u})$, and all $f \in L^2(\Omega)$, the linearized state equation (2.5) has a unique solution $y \in \mathcal{Y}$, which depends continuously on f.*

Proof. Note that the coefficient $a := -\frac{2\beta\overline{u}}{(1+\overline{y})^3}$ typically has the wrong sign and is not in $L^\infty(\Omega)$, so the Lax-Milgram Lemma does not apply directly to (2.5), but rather a fixed point argument similar to the one in the proof of Theorem 2.1 has to be used. We refer to [7] but nevertheless provide the full proof for the setting relevant here.

Consider the fixed point operator mapping v to $\tilde{T}v := y$ defined as the solution to

$$\begin{cases} -\Delta y = -av + f & \text{on } \Omega\,, \\ \quad y = 0 & \text{on } \partial\Omega. \end{cases}$$

Due to the estimate

$$\left\|\Delta \tilde{T} v\right\|_{L^2(\Omega)} \leqslant C_L \|a\|_{L^2(\Omega)} \|\Delta v\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)}\,, \tag{2.6}$$

the operator $\tilde{T}$ is a self-mapping on

$$\tilde{W} = \left\{ v \in \mathcal{Y} : \|\Delta v\|_{L^2(\Omega)} \leqslant R \right\}$$

for any fixed $R > 0$, as long as

$$C_L \|a\|_{L^2(\Omega)} < 1 \tag{2.7}$$

is satisfied (which can be guaranteed by $\|\overline{u}\|_{L^2(\Omega)} \leqslant M_u$ as in (2.3)) and $\|f\|_{L^2(\Omega)} \leqslant (1 - C_L \|a\|_{L^2(\Omega)})R =: r$. Condition (2.7) also yields contractivity via the estimate

$$\left\|\Delta(\tilde{T}v - \tilde{T}w)\right\|_{L^2(\Omega)} \leqslant C_L \|a\|_{L^2(\Omega)} \|\Delta(v-w)\|_{L^2(\Omega)}\,,$$

hence $\tilde{T}$ has a unique fixed point $y \in W$. Using estimate (2.6) with y in place of v and $\tilde{T}v$, we deduce that the operator $\tilde{S}$ mapping f to a solution of (2.5) is well-defined and bounded from $L^2(\Omega)$ to $\mathcal{Y}$, on the ball of radius r, and hence by linearity on $L^2(\Omega)$. □

Finally, we address differentiability of the control-to-state mapping.

Theorem 2.5. *The control-to state mapping* $S : \mathcal{U}_M \to \mathcal{Y}$ *is Fréchet differentiable, with derivative* $S'(u)(\tilde{u}-u)$ *given by the solution* $w \in \mathcal{Y}$ *to*

$$\begin{cases} -\Delta w - \dfrac{2\beta u}{(1+y)^3} w = -\dfrac{\beta}{(1+y)^2}(\tilde{u}-u) & \text{on } \Omega, \\ w = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.8}$$

where $y = S(u)$.

Proof. By Lemma 2.4, a unique solution $w \in \mathcal{Y}$ of (2.8) exists for any $u \in \mathcal{U}_M$. To show Fréchet differentiability of S, we use the fact that $z := \tilde{y} - y - w$ solves

$$\begin{cases} -\Delta z - \dfrac{2\beta u}{(1+y)^3} z = \dfrac{2\beta}{(1+y)^3}(\tilde{u}-u)(\tilde{y}-y) \\ \qquad\qquad - \dfrac{\beta\tilde{u}(3+y-\tilde{y})}{(1+y)^3(1+\tilde{y})^2}(\tilde{y}-y)^2 & \text{on } \Omega, \\ z = 0 & \text{on } \partial\Omega, \end{cases}$$

to estimate

$$\begin{aligned} \|\Delta(\tilde{y}-y-w)\|_{L^2(\Omega)} \leqslant{}& \frac{2|\beta|}{(1-\overline{m})^3} \|u\|_{L^2(\Omega)} \|\tilde{y}-y-w\|_{C(\Omega)} \\ &+ \frac{2|\beta|}{(1-\overline{m})^3} \|\tilde{u}-u\|_{L^2(\Omega)} \|\tilde{y}-y\|_{C(\Omega)} \\ &+ \frac{|\beta|(3+2\overline{m})}{(1-\overline{m})^5} \|\tilde{u}\|_{L^2(\Omega)} \|\tilde{y}-y\|^2_{C(\Omega)} . \end{aligned}$$

Hence

$$\|\Delta(\tilde{y}-y-w)\|_{L^2(\Omega)} \leqslant K(\overline{m}, M_u) \|\tilde{u}-u\|^2_{L^2(\Omega)} ,$$

with

$$\begin{aligned} K(\overline{m}, M_u) := \left\{1 - \frac{2C_L|\beta|}{(1-\overline{m})^3} M_u\right\}^{-1} \\ \cdot \frac{C_L L(\overline{m}, M_u)|\beta|\Big((3+2\overline{m})M_u C_L L(\overline{m}, M_u) + 2(1-\overline{m})^2\Big)}{(1-\overline{m})^5} , \end{aligned}$$

the expression in braces being positive again due to the choice (2.3) of M_u. □

2.2 CONTROL PROBLEM

Since the state equation is well-posed for every $u \in \mathcal{U}_M$, we can use the control-to-state mapping $S : \mathcal{U}_M \to \mathcal{Y}$ to obtain the reduced problem

$$(\mathcal{P}'_{cc}) \qquad \min_{u \in \mathcal{U}_M} \frac{1}{2} \|S(u) - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2} \|u\|^2_{L^2(\Omega)}.$$

Since $\mathcal{U}_M$ is nonempty and weakly sequentially compact, and S is weakly continuous for all $u \in \mathcal{U}_M$ by Theorem 2.3, we obtain the existence of a minimizer $u^* \in \mathcal{U}_M$ by standard arguments (see, e.g. [17]).

Similarly we obtain first order necessary optimality conditions. For a local minimizer u^* of ($\mathcal{P}'_{cc}$) and $y^* := S(u^*) \in \mathcal{Y}$, we can introduce the adjoint state $p^* \in \mathcal{Y}$ solving

$$(2.9) \qquad \begin{cases} -\Delta p^* - \dfrac{2\beta u^*}{(1+y^*)^3} p^* = -(y^* - y_d) & \text{on } \Omega, \\ p^* = 0 & \text{on } \partial\Omega, \end{cases}$$

since the linearized state equation is self-adjoint and is surjective by Lemma 2.4. Furthermore, a Slater condition is trivially satisfied for the inequality constraint $\|u^*\|^2_{L^2(\Omega)} \leqslant M_u^2$ (take $u = 0$). From, e.g., Proposition 3.2 in [3], we thus deduce the existence of a corresponding Lagrange multiplier $\lambda^* \in \mathbb{R}$ and hence the following optimality system.

Theorem 2.6. *Let* $u^* \in \mathcal{U}_M$ *be a local minimizer of* ($\mathcal{P}'_{cc}$), $y^* := S(u^*) \in \mathcal{Y}$, *and* $p^* \in \mathcal{Y}$ *satisfy* (2.9). *Then there exists* $\lambda^* \in \mathbb{R}$, $\lambda^* \geqslant 0$ *satisfying*

$$(OS_{cc}) \qquad \begin{cases} \alpha u^* - \dfrac{\beta}{(1+y^*)^2} p^* = 0, \\ \lambda^*(\|u^*\|^2_{L^2(\Omega)} - M_u^2) = 0. \end{cases}$$

3 STATE CONSTRAINTS

Consider now the state-constrained optimal control problem

$$(\mathcal{P}_{sc}) \qquad \begin{cases} \min\limits_{y \in \mathcal{Y}, u \in \mathcal{U}} \dfrac{1}{2} \|y - y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2} \|u\|^2_{L^2(\Omega)} \\ -\Delta y = -\dfrac{\beta u}{(1+y)^2}, \\ -y(x) \leqslant M_y \quad \text{for all } x \in \Omega, \end{cases}$$

for given $M_y < 1$. Since the state equation is not well-posed for every $u \in \mathcal{U}$, we cannot use the results of Section 2.1 but argue directly. We define

$$G : \mathcal{U} \times \tilde{\mathcal{Y}}_M \to L^2(\Omega), \quad G(u, y) = -\Delta y + \frac{\beta u}{(1+y)^2},$$

where $\tilde{\mathcal{Y}}_M := \{y \in \mathcal{Y} : -y(x) \leqslant \frac{1}{2}(M_y + 1), x \in \Omega\} \supset \mathcal{Y}_M$. The next section is concerned with continuity and differentiability of G.

3.1 STATE EQUATION

We first note that for all $y \in \mathcal{Y}_M$, the term $\frac{\beta}{(1+y)^2}$ is uniformly bounded by $\overline{m} := |\beta|(1-M_y)^{-2}$. Due to the linearity of G in u, it therefore immediately follows that G is continuously differentiable with respect to u for all $(\overline{u}, \overline{y}) \in \mathcal{U} \times \mathcal{Y}_M$, with Fréchet-derivative in direction $h \in \mathcal{U}$ given by

$$G_u(\overline{u}, \overline{y})h = \frac{\beta}{(1+\overline{y})^2}h.$$

To show differentiability with respect to y, we consider

$$\Psi : y(x) \mapsto \frac{\beta}{(1+y(x))^2}.$$

as a Nemytskii operator from $\mathcal{Y}_M \subset L^\infty(\Omega)$ to $L^\infty(\Omega)$ generated by the mapping

$$\psi : [-M_y, \infty) \times \Omega \mapsto \mathbb{R}, \quad \psi(t, x) := \frac{\beta}{(1+t)^2}.$$

Since ψ is measurable for fixed $t \geqslant -M_y > -1$ and uniformly continuous for all $x \in \Omega$, the Nemytskii operator Ψ is continuous from $\mathcal{Y}_M$ to $L^\infty(\Omega)$ [8, Thm. 5]. Similarly,

$$\psi_t : [-M_y, \infty) \times \Omega \mapsto \mathbb{R}, \quad \psi_t(t, x) := \frac{-2\beta}{(1+t)^3}$$

is uniformly continuous for all $x \in \Omega$ and hence generates a continuous Nemytskii operator from $\mathcal{Y}_M$ to $L^\infty(\Omega)$. By Theorem 7 of [8], Ψ is therefore continuously Fréchet-differentiable from $\mathcal{Y}_M$ to $L^\infty(\Omega)$ with derivative

$$\Psi' : y(x) \mapsto \frac{-2\beta}{(1+y(x))^3}.$$

Using this, it is straightforward to verify that G is continuously Fréchet-differentiable (and hence strictly differentiable) with respect to y for all $(\overline{u}, \overline{y}) \in \mathcal{U} \times \mathcal{Y}_M$, with Fréchet-derivative in direction $h \in \mathcal{Y}$ given by

$$G_y(\overline{u}, \overline{y})h = -\Delta h + \frac{-2\beta\overline{u}}{(1+\overline{y})^3}h.$$

We next argue surjectivity of G_y on $L^2(\Omega)$, i.e., well-posedness of the linearized state equation (2.5) at $\overline{u} \in \mathcal{U}, \overline{y} \in \mathcal{Y}_M$. We assume that 1 is not an eigenvalue of the operator $A : L^2(\Omega) \to L^2(\Omega)$ defined by $Az = (-\Delta + cI)^{-1}bz$, where $-\Delta + cI$ is to be understood with homogeneous Dirichlet boundary conditions and $a = \frac{2\beta\overline{u}}{(1+\overline{y})^3}$, $b = \max\{a, 0\}$, $c = \max\{-a, 0\} = -a + b$. Note that for $n \leqslant 3$, the operator A is compact due to the estimate

$$\|Az\|_{H^\epsilon(\Omega)} \leqslant C\,\|bz\|_{H^{-(2-\epsilon)}(\Omega)} \leqslant C\,\|b\|_{L^2(\Omega)}\,\|z\|_{L^2(\Omega)}$$

for some $0 < \epsilon < 2 - \frac{n}{2}$. This condition is equivalent to

$$\left.\begin{cases} -\Delta z - \dfrac{2\beta\bar{u}}{(1+\bar{y})^3} z = 0 & \text{on } \Omega, \\ z = 0 & \text{on } \partial\Omega, \end{cases}\right\} \Rightarrow z = 0. \tag{3.1}$$

Lemma 3.1. *Let (3.1) hold. Then for all* $\bar{u} \in \mathcal{U}$, $\bar{y} \in \mathcal{Y}_M$ *and* $f \in L^2(\Omega)$, *the linearized state equation (2.5) has a unique solution* $y \in \mathcal{Y}$.

Proof. Consider the compact operator $\tilde{T} : C_0(\Omega) \to C_0(\Omega)$ defined by $\tilde{T}v = (-\Delta)^{-1}(av)$, where $-\Delta$ is to be understood with homogeneous Dirichlet boundary conditions and $a = \frac{2\bar{u}}{(1+\bar{y})^3}$. By (3.1), the operator $I-\tilde{T}$ has Riesz index one. For any $f \in L^2(\Omega)$, we have $(-\Delta)^{-1}f \in C_0(\Omega)$. Hence, (2.5) is equivalent to $(I-\tilde{T})y = (-\Delta)^{-1}f$, which implies that (2.5) is uniquely solvable. □

Note that condition (3.1) does not imply smallness of $\bar{u}$ (as is required for Lemma 2.4), and can be expected to be satisfied except in pathological situations.

For the derivation of the optimality system, we also require the well-posedness of the adjoint problem with measure-valued right hand side. We follow the standard approach [16] (see also [10]).

Lemma 3.2. *Let* $\Omega \subseteq \mathbb{R}^n$, $n \in \{1,2,3\}$ *be bounded, and let (3.1) hold. Then there exists* $q \in [1, \frac{n}{n-1})$ *such that for all* $f \in W^{-1,q}(\Omega)$, *the linearized state equation (2.5) has a unique solution* $y \in W^{1,q}(\Omega)$, *and there exist constants* $c, C > 0$ *such that for all* $f \in W^{-1,q}(\Omega)$

$$c\,\|f\|_{W^{-1,q}(\Omega)} \leqslant \|y\|_{W^{1,q}(\Omega)} \leqslant C\,\|f\|_{W^{-1,q}(\Omega)}\,. \tag{3.2}$$

Proof. Analogously to Lemma 3.1, we obtain unique solvability of (2.5) in $W^{1,q}(\Omega)$ for any $f \in W^{-1,q}(\Omega)$ by considering the compact operator $\tilde{T} : W^{1,q}(\Omega) \to W^{1,q}(\Omega)$ defined as above by $\tilde{T}v = (-\Delta)^{-1}(av)$. Furthermore we can estimate

$$\begin{aligned} \left\|\tilde{T}v\right\|_{W^{s,q}(\Omega)} &\leqslant C\,\|av\|_{W^{-(2-s),q}(\Omega)} \\ &= C \sup\left\{\int_\Omega \varphi a v \,dx \;:\; \varphi \in C_0^\infty(\Omega),\ \|\varphi\|_{W^{2-s,q'}(\Omega)} = 1\right\} \\ &\leqslant C\,\|a\|_{L^2(\Omega)}\,\|v\|_{W^{1,q}(\Omega)}\,, \end{aligned}$$

where $q' = \frac{q}{q-1}$, $C > 0$ denotes a generic constant, and we have used the Sobolev embeddings $W^{2-s,q'}(\Omega) \hookrightarrow L^{2p}(\Omega)$ and $W^{1,q}(\Omega) \hookrightarrow L^{2p/(p-1)}(\Omega)$ with appropriately chosen $p \in [1,\infty]$ and $s \in (1,2]$. Indeed, the inequalities required for continuity of these embeddings,

$$2 - s + \frac{n}{q} \geqslant \frac{n}{2} + \frac{n}{2p}\,, \qquad 1 - \frac{n}{q} \geqslant -\frac{n}{2p}\,, \tag{3.3}$$

(with strict inequality in case $p = 1$ and $p = \infty$, respectively) can be satisfied by the following choices:

$$\begin{aligned}
n = 1: \quad & q \in [1, \infty]: && p = 2,\ s = \tfrac{5}{4},\\
n = 2: \quad & q = 1: && p = 1,\ s = 2,\\
& q \in (1, 2): && p = \tfrac{q}{2-q},\ s = 2,\\
& q = 2: && p = 2,\ s = \tfrac{3}{2},\\
& q \in (2, \infty): && p = \tfrac{q+1}{2},\ s = 1 + \tfrac{2}{q(q+1)},\\
n = 3: \quad & q = \tfrac{6}{5}: && p = 1,\ s = \tfrac{5}{4},\\
& q \in (\tfrac{6}{5}, 3): && p = \tfrac{3q}{6-2q},\ s = \tfrac{3}{2},\\
& q \in [3, 6): && p \in (\tfrac{3q}{6-q}, \infty),\ s = \tfrac{1}{2} + \tfrac{3}{q} - \tfrac{3}{2p},
\end{aligned}$$

hence for

$$q \in [1, \infty] \quad \text{if} \quad n = 1, \qquad q \in [1, \infty) \quad \text{if} \quad n = 2, \qquad q \in [\tfrac{6}{5}, 6) \quad \text{if} \quad n = 3.$$

(Note that by adding the two inequalities in (3.3), it becomes obvious that they cannot be satisfied for any $s > 1$ as soon as $n \geqslant 4$.) Hence, $\tilde{T}$ is bounded as an operator from $W^{1,q}(\Omega)$ to $W^{s,q}(\Omega)$, and by the Rellich–Kondrachov Theorem and boundedness of Ω, compact as an operator from $W^{1,q}(\Omega)$ to $W^{1,q}(\Omega)$. By (3.1), the operator $I - \tilde{T}$ has Riesz index one and is therefore bijective by the third Riesz Theorem (e.g., Theorem 3.3 in [9]). Moreover, by the second Riesz Theorem (e.g., Theorem 3.2 in [9]) it has closed range, hence by the Open Mapping Theorem (e.g., Theorem 2.11 in [14]) its inverse $(I - \tilde{T})^{-1} : W^{1,q}(\Omega) \to W^{1,q}(\Omega)$ is bounded. Using the fact that (2.5) is equivalent to $(I - \tilde{T})y = (-\Delta)^{-1} f$, we conclude estimate (3.2). □

Due to the dense embedding $W^{1,q'} \hookrightarrow C_0(\Omega)$ for $q' > n$, we have $\mathcal{M}(\Omega) \hookrightarrow W^{-1,q}(\Omega)$, $q \in [1, \frac{n}{n-1})$, and we deduce well-posedness of the adjoint equation.

Corollary 3.3. *For any* $\overline{u} \in \mathcal{U}$, $\overline{y} \in \mathcal{Y}_M$, $\mu \in \mathcal{M}(\Omega)$, *and* $q \in [\max\{1, \frac{2n}{n+2}\}, \frac{n}{n-1})$, *there exists a unique* $p \in W^{1,q}(\Omega)$, *solving*

$$\begin{cases} -\Delta p - \dfrac{2\beta\overline{u}}{(1+\overline{y})^3} p = \mu & \textit{on } \Omega, \\ \hfill p = 0 & \textit{on } \partial\Omega. \end{cases}$$

Furthermore, there exists a $C > 0$ *such that for all* $\mu \in \mathcal{M}(\Omega)$, *there holds*

$$\|p\|_{W^{1,q}(\Omega)} \leqslant C \, \|\mu\|_{\mathcal{M}(\Omega)}.$$

3.2 CONTROL PROBLEM

The existence of minimizers follows again from standard arguments. For the sake of completeness, we nonetheless state the proof.

Theorem 3.4. *There exists a minimizer* $(u^*, y^*) \in \mathcal{U} \times \mathcal{Y}_M$ *of* ($\mathcal{P}_{sc}$).

Proof. The set of feasible (u, y) satisfying the equality and inequality constraints is non-empty (take $(u, y) = (0, 0)$). By the boundedness of $\mathcal{Y}_M$ and the coercivity of the functional in u, we obtain the existence of a minimizing sequence that is bounded in $\mathcal{U} \times \mathcal{Y}_M$. Hence, there exists a subsequence, denoted by $\{(u_n, y_n)\}$, that weakly converges to $(u^*, y^*) \in \mathcal{U} \times \mathcal{Y}_M$. Due to the compact embedding of $\mathcal{Y}$ in $C_0(\Omega)$ and the continuity of Ψ, we have that a subsequence of $\frac{\beta}{(1+y_n)^2}$ converges strongly to $\frac{\beta}{(1+y^*)^2}$. Thus, we can pass to the limit in (the weak formulation of) $G(u_n, y_n) = 0$ to obtain $G(u^*, y^*) = 0$. □

For the derivation of optimality conditions, we follow [1]. Note that – as can be seen by inspection of the proofs in [1] – it suffices to consider G on the open set $\tilde{\mathcal{Y}}_M$ containing the admissible set, which is possible due to $M_y < 1$.

Theorem 3.5. *Let* $(u^*, y^*) \in \mathcal{U} \times \mathcal{Y}_M$ *be a local minimizer of* ($\mathcal{P}_{sc}$). *Then there exist* $\mu^* \in \mathcal{M}(\Omega)$ *and* $p^* \in W_0^{1,q}(\Omega)$ *satisfying*

$$\begin{cases} -\Delta p^* - \dfrac{2\beta u^*}{(1+y^*)^3} p^* = -(y^* - y_d) - \mu^*, \\ \alpha u^* - \dfrac{\beta}{(1+y^*)^2} p^* = 0, \\ \qquad \langle \mu^*, y - y^* \rangle_{\mathcal{M},C} \leqslant 0 \quad \textit{for all } y \in \mathcal{Y}_M. \end{cases}$$

Proof. We recall that G is strictly differentiable at (u^*, y^*). Furthermore, by Lemma 3.1, the linearized constraint $G_y(u^*, y^*)$ is surjective on $L^2(\Omega)$. Finally, a regular point condition is satisfied: There exists $(u_0, y_0) \in \mathcal{U} \times \mathcal{Y}_M$ such that $(y^* + y_0) \in \operatorname{int} \mathcal{Y}_M$ and

$$G_y(u^*, y^*)y_0 + G_u(u^*, y^*)(u_0 - u^*) = 0.$$

Specifically, if $y^* \in \operatorname{int} \mathcal{Y}_M$, take $(u_0, y_0) = (u^*, 0)$, otherwise take

$$y_0 = -\frac{M_y}{2 \sup_{x \in \Omega} y^*(x)} y^* \in \mathcal{Y}_M$$

and

$$u_0 = u^* - \frac{M_y}{2 \sup_{x \in \Omega} y^*(x)} \left(\frac{(1+y^*)^2}{\beta} \Delta y^* + \frac{2u^* y^*}{1+y^*} \right) \in \mathcal{U}.$$

Using Corollary 3.3, we therefore obtain from Theorem 2.1 in [1] the existence of unique Lagrange multipliers $p^* \in W_0^{1,q}(\Omega)$ and $\mu^* \in \mathcal{M}(\Omega)$ corresponding to the equality and inequality constraint, respectively. □

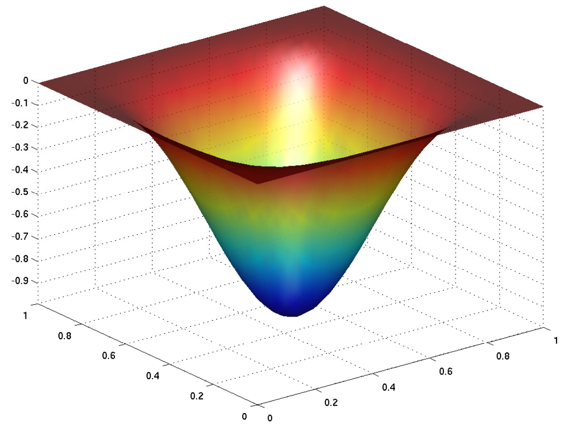


Figure 1: Target y_d

4 NUMERICAL EXAMPLES

We contrast the two approaches for a simple two-dimensional example, where we set $\Omega = (0,1)^2$ and $\beta = 1$. For the case of control constraints we set $M_u = 0.14815$ according to (2.4); for the state-constrained problem we used $M_y = 0.99$. The common target y_d (shown in Figure 1) is chosen to attain the bound $\min_{x\in\Omega} y_d(x) = -M_y$ by setting

$$y_d(x_1, x_2) = M_y \frac{(1 + \cos(2\pi x_1 + \pi))(1 + \cos(2\pi x_1 + \pi))}{4}.$$

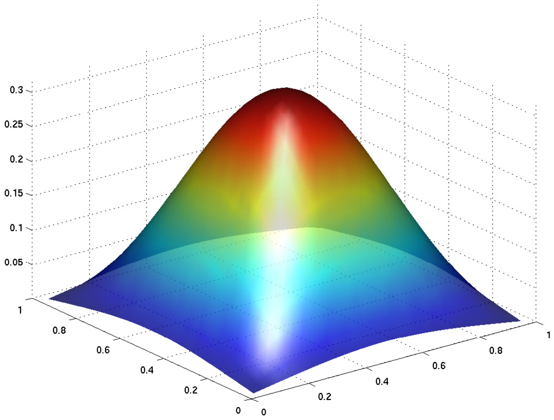
The penalty parameter was set to $\alpha = 10^{-6}$ in both cases. The differential operators were discretized using second order finite differences on a mesh with 64×64 grid points.

To ensure feasibility of the iterates, the minimizers of ($\mathcal{P}'_{cc}$) and ($\mathcal{P}_{sc}$) were computed using an interior point method (`fmincon` in MATLAB). In both cases, the final iterates attain the constraints and satisfy the optimality conditions to within a specified tolerance of 10^{-6}.

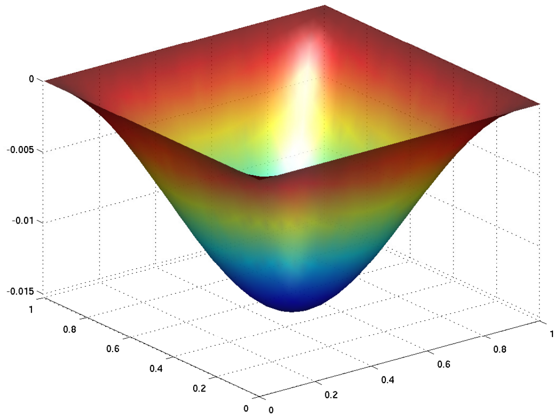
The resulting optimal controls and states are given in Figures 2 and 3, while Figure 4 shows a cross-section at $x_1 = 0.5$ of y_d, y_{cc} and y_{sc}. It can be observed that the control constraint leads to a very smooth optimal control, and that the smallness condition on the control is very restrictive and prevents attainment of the target. On the other hand, the state constraint allows much richer behavior of the optimal control and hence a significantly closer attainment.

5 CONCLUSION

Blow-up in optimal control of singular partial differential equations can be avoided by imposing suitable constraints on either the admissible controls or the admissible states. While the former allows applying standard techniques (via the introduction of the control-to-state mapping) to obtain existence and optimality conditions, the numerical examples for our

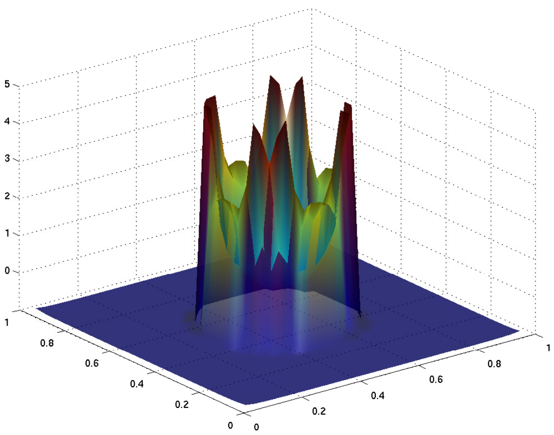


(a) optimal control u_{cc}

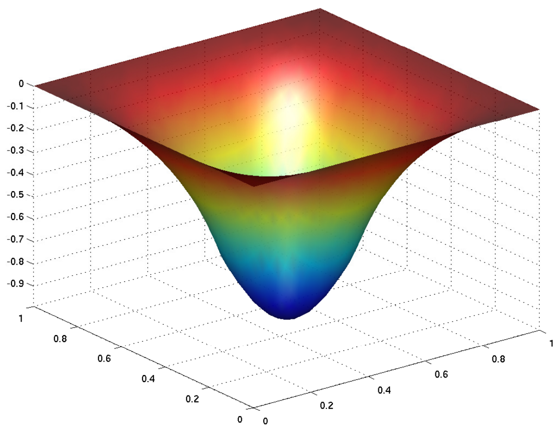


(b) optimal state y_{cc}

Figure 2: Solutions to control-constrained problem ($\mathcal{P}'_{cc}$)



(a) optimal control u_{sc}



(b) optimal state y_{sc}

Figure 3: Solutions to state-constrained problem ($\mathcal{P}_{sc}$)

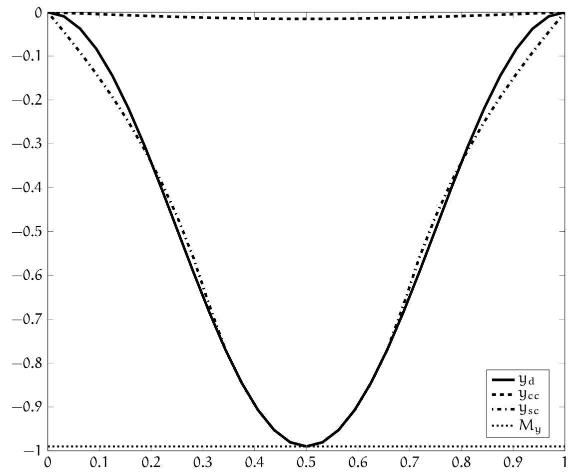


Figure 4: Cross sections of states for control-constrained (y_{cc}) and state-constrained (y_{sc}) problems, as well as target y_d and bound M_y

model problem indicate that the required norm bounds can be overly pessimistic and prohibit even approximate attainment of the target. This indicates that state constraints are the more appropriate choice, even though they are computationally more involved and the constraint qualifications are easier to verify in the control-constrained case. In addition, the key requirements for deriving optimality conditions (via surjectivity of the linearized constraint) are less restrictive for state constraints. The suggested approach is expected to be similarly advantageous in many practically relevant applications involving the control of potentially singular or degenerate equations.

ACKNOWLEDGMENTS

The work of the first-named author was supported in part by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences"). The second-named author would like to thank the German Science Foundation (DFG) for financial support within the grant KA 1778/5-1 ("Adaptive Discretization Methods for the Regularization of Inverse Problems").

REFERENCES

[1] J.-J. Alibert and J.-P. Raymond, *A Lagrange multiplier theorem for control problems with state constraints*, Numer. Funct. Anal. Optim., 19 (1998), pp. 697–704.

[2] M. Bergounioux and K. Kunisch, *Augmented Lagrangian techniques for elliptic state constrained optimal control problems*, SIAM J. Control Optim., 35 (1997), pp. 1524–1543.

[3] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer-Verlag, New York, 2000.

[4] E. Casas, *Control of an elliptic problem with pointwise state constraints*, SIAM J. Control Optim., 24 (1986), pp. 1309–1318.

[5] G.-Q. G. Chen, *On degenerate partial differential equations*, in Nonlinear partial differential equations and hyperbolic wave phenomena, Providence, RI, 2010, Amer. Math. Soc., pp. 53–90.

[6] C. Clason, B. Kaltenbacher, and S. Veljovic, *Boundary optimal control of the Westervelt and the Kuznetsov equation*, Journal of Mathematical Analysis and Applications, 356 (2009), pp. 738–751.

[7] F. Colonius and K. Kunisch, *Stability for parameter estimation in two boundary value problems*, Journal für Reine und Angewandte Mathematik, 370 (1986), pp. 1–29.

[8] H. Goldberg, W. Kampowsky, and F. Tröltzsch, *On Nemytskij operators in* L_p*-spaces of abstract functions*, Math. Nachr., 155 (1992), pp. 127–140.

[9] R. Kress, *Linear integral equations*, Springer, 1999.

[10] C. Meyer, L. Panizzi, and A. Schiela, *Uniqueness criteria for the adjoint equation in state-constrained elliptic optimal control*, Numer. Funct. Anal. Optim., 32 (2011), pp. 983–1007.

[11] J. A. Pelesko, *Mathematical modeling of electrostatic MEMS with tailored dielectric properties*, SIAM J. Appl. Math., 62 (2001/02), pp. 888–908.

[12] J. A. Pelesko and D. H. Bernstein, *Modeling MEMS and NEMS*, Chapman & Hall/CRC, Boca Raton, FL, 2003.

[13] I. Pop and B. Schweizer, *Regularization schemes for degenerate Richards equations and outflow conditions*, Mathematical Models and Methods in the Applied Sciences, 21 (2011), pp. 1685–1712.

[14] W. Rudin, *Functional analysis*, McGraw-Hill Inc., New York, second ed., 1991.

[15] A. Schiela, *State constrained optimal control problems with states of low regularity*, SIAM J. Control Optim., 48 (2009), pp. 2407–2432.

[16] G. Stampacchia, *Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus*, Ann. Inst. Fourier (Grenoble), 15 (1965), pp. 189–258.

[17] F. Tröltzsch, *Optimal control of partial differential equations*, vol. 112, American Mathematical Society, Providence, RI, 2010. Theory, methods and applications, Translated from the 2005 German original by Jürgen Sprekels.