Handling constraints in optimal control with saturation functions and system extension

Knut Graichen [a,*], Andreas Kugi [b], Nicolas Petit [c], Francois Chaplais [c]

[a] *Institute of Measurement, Control and Microtechnology, Universität Ulm, Germany*
[b] *Automation and Control Institute, Vienna University of Technology, Austria*
[c] *Centre Automatique et Systèmes, MINES ParisTech, France*

ARTICLE INFO

ABSTRACT

A method is presented to systematically transform a general inequality-constrained optimal control problem (OCP) into a new equality-constrained OCP by means of saturation functions. The transformed OCP can be treated more conveniently within the standard calculus of variations compared to the original constrained OCP. In detail, state constraints are substituted by saturation functions and successively constructed dynamical subsystems, which constitute a (dynamical) system extension. The dimension of the subsystems corresponds to the relative degree (or order) of the respective state constraints. These dynamical subsystems are linked to the original dynamics via algebraic coupling equations. The approach results in a new equality-constrained OCP with extended state and input vectors. An additional regularization term is used in the cost to regularize the new OCP with respect to the new inputs. The regularization term has to be successively reduced to approach the original constrained solution. The new OCP can be solved in a convenient manner, since the stationarity conditions are easily determined and exploited. An important aspect of the saturation function formulation is that the constraints cannot be violated during the numerical solution. The approach is illustrated for an extended version of the well-known Goddard problem with thrust and dynamic pressure constraints and using a collocation method for its numerical solution.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Numerical methods for the solution of optimal control problems (OCPs) can roughly be divided in two different classes. In *direct methods*, the OCP is discretized to obtain a finite-dimensional parameter optimization problem, see, e.g., [1–6]. Well-known advantages of the direct approach are the good domain of convergence as well as the efficient handling of constraints. On the other hand, *indirect approaches* are based on the calculus of variations and require the solution of a two-point boundary value problem (BVP), see, e.g., [7]. Indirect methods are known to show a fast numerical convergence in the neighborhood of the optimal solution and to deliver highly accurate solutions, which makes them particularly attractive for aerospace applications [8–12]. However, the handling of inequality constraints via Pontryagin's maximum principle [13] is in general non-trivial, since the overall structure of the BVP depends on the sequence between singular/nonsingular and

unconstrained/constrained arcs (if the respective constraint is active or not) and requires a-priori knowledge of the optimal solution structure.

In order to avoid these problems in handling constraints, a method has been presented in [14] to systematically incorporate a class of state and input constraints in a new unconstrained OCP formulation, which can be solved with standard unconstrained numerics from indirect optimal control. By using the state constraints as linearizing outputs, a normal form representation of the considered nonlinear system is derived. The constraint dynamics are then substituted by means of saturation functions and successive differentiation along the normal form cascades. This concept follows an approach originally presented in the context of feedforward control design [15,16]. The procedure results in a new unconstrained system representation having the *same* system dimension but new state and input variables. However, the specific transformation and replacement technique as presented in [14] is limited to a class of state constraints with well-defined relative degree. This means, for instance, that in case of a single-input system the approach is restricted to a single state constraint.

The intention of this paper therefore is to extend the saturation function approach to a more general class of constrained OCPs. The state constraints are represented by smooth saturation functions

* Corresponding author.
E-mail addresses: knut.graichen@uni-ulm.de (K. Graichen),
kugi@acin.tuwien.ac.at (A. Kugi), nicolas.petit@mines-paristech.fr (N. Petit),
francois.chaplais@mines-paristech.fr (F. Chaplais).

whose arguments satisfy a differential equation determined by successively differentiating the state constraint function up to its relative degree, i.e. until the input appears. In this way, dynamical subsystems in new coordinates are constructed for each state constraint. These subsystems are coupled to the original dynamics via equality constraints that relate their inputs to the variables of the original system. In addition, input constraints (or mixed state-input constraints) can also be considered by direct usage of saturation functions.

The resulting system of differential-algebraic equations (DAE) with *extended* state and input vectors is used to define a new OCP with equality constraints that can be handled conveniently in the calculus of variations (or alternatively with direct methods). The necessary optimality conditions define a two-point boundary value problem, which can be solved with unconstrained numerical methods. An additional regularization term is added to the cost in order to achieve regularity of the new OCP. The corresponding regularization parameter has to be successively reduced during the numerical solution of the new OCP to approach the optimal solution of the original constrained OCP.

An intrinsic property of the saturation function approach is that the constraints cannot be violated during the numerical solution due to their inherent incorporation in the new OCP [17,14]. This is particularly advantageous for the numerical initialization.

The paper is outlined as follows: Section 2 introduces the considered class of constrained OCPs. Section 3 is devoted to the transformation of the original constrained OCP into a new equality-constrained (and regularized) OCP with extended state and input vectors. The convergence properties of the new OCP for a successively reduced regularization parameter are investigated in Section 4. Section 5 is concerned with the solution of the new OCP by deriving the optimality conditions from the calculus of variations. In addition, a collocation method is shortly introduced to numerically solve the two-point BVP stemming from the optimality conditions. Section 6 applies the method to the well-known Goddard problem with thrust and dynamic pressure constraints and discusses the numerical results. Section 7 concludes the paper and gives an outlook on potential future research activities in this field.

2. Problem statement

The following general inequality-constrained optimal control problem, called OCP_u, is considered:

minimize

$$J(u) := \varphi(x(T), T) + \int_0^T L(x, u, t) \mathrm{d}t \tag{1}$$

subject to

$$\dot{x} = f(x, u), \quad x(0) = x_0, \tag{2}$$

$$\chi(x(T), T) = 0, \tag{3}$$

$$c_i(x) \in \left[c_i^-, c_i^+ \right], \quad i = 1, \dots, p, \tag{4}$$

$$d_i(x, u) \in \left[d_i^-, d_i^+ \right], \quad i = 1, \dots, q. \tag{5}$$

It is assumed that the nonlinear system (2) with state $x \in \mathbb{R}^n$, input $u \in \mathbb{R}^m$, and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ possesses a unique state trajectory $x(t)$ for each input trajectory $u(t)$, such that the cost (1) can be regarded as the functional $J(u)$. At the end of the time interval $t \in [0, T]$, the terminal conditions $\chi : \mathbb{R}^n \times \mathbb{R}_+ \to \mathbb{R}^l$ are imposed on the state x. The terminal time T in (1) and (3) may be fixed or unspecified. The constraints (4), (5) represent state constraints and mixed state-input constraints with interval bounds. The functions φ, L, f, χ, c_i, and d_i are assumed to be sufficiently smooth. Note that the two-sided constraints (4) and (5) are considered for the sake of generality. In practice, the constraints may also describe one-sided bounds, e.g. $c_i(x) \leq c_i^+$.

The order [7] or relative degree [18] of each state constraint function $c_i(x)$ in (4) is defined by

$$\frac{\partial c_i^{(j)}}{\partial u} = 0, \quad j = 1, \dots, r_i - 1, \qquad \frac{\partial c_i^{(r_i)}}{\partial u} \neq 0, \tag{6}$$

with

$$c_i^{(j)}(x) := L_f^j c_i(x), \quad j = 1, \dots, r_i - 1,$$
$$c_i^{(r_i)}(x, u) := L_f^{r_i} c_i(x), \quad i = 1, \dots, p. \tag{7}$$

The operator L_f denotes the Lie derivative defined by $L_f c_i(x) = \frac{\partial c_i}{\partial x} f(x, u)$ and $L_f^j c_i(x) = L_f L_f^{j-1} c_i(x)$. Literally, the order r_i corresponds to the number of times the state constraint function $c_i(x)$ has to be differentiated until at least one element of the input vector $u = (u_1, \dots, u_m)^\mathsf{T}$ appears explicitly.[1] In addition, the mixed state-input constraints (5) are assumed to be well-defined with respect to u, i.e.

$$\frac{\partial d_i}{\partial u} \neq 0, \quad i = 1, \dots, q. \tag{8}$$

The considered OCP_u covers a large class of optimal control problems. Note in particular that the interval bounds of the constraints (4), (5) actually represent two constraints for each function $c_i(x)$ and $d_i(x, u)$ in the standard notation of optimal control [13,7].

In the following, a method is demonstrated to compute solutions of the inequality-constrained OCP_u. Although there is no practical obstruction to using it on problems that cannot be guaranteed to possess a global solution, it simplifies the exposition to make this assumption. Therefore we formulate the following

Assumption 1. OCP_u has an optimal solution u^* (to which corresponds x^*) with the optimal cost $J(u^*) = J^*$.

3. Saturation function approach

Within the presented approach, the original inequality-constrained OCP_u is transformed into a new OCP by using saturation functions to systematically substitute the constraints (4), (5) by additional dynamical subsystems and algebraic coupling equations. The resulting DAE representation with extended state and input vectors leads to a new $OCP_{\bar{u}}^\varepsilon$ with equality constraints. An additional regularization term with parameter ε is used to regularize $OCP_{\bar{u}}^\varepsilon$ with respect to the introduced additional inputs.

3.1. Incorporation of state constraints

Consider in a first step the state constraints (4). The idea of the approach is to replace $c_i(x)$ by a saturation function

$$c_i(x) = \psi_i(\xi_{i,1}), \quad i = 1, \dots, p \tag{9}$$

with the new unconstrained variable $\xi_{i,1} \in \mathbb{R}$. To represent the constraint, the variable $\xi_{i,1}$ will satisfy some well-chosen differential equation whose construction is detailed below. The saturation functions $\psi_i : \mathbb{R} \to (c_i^-, c_i^+)$ are assumed to be smooth and strictly monotonically increasing, i.e. $\mathrm{d}\psi_i / \mathrm{d}\xi_{i,1} > 0 \,\forall \xi_{i,1} \in \mathbb{R}$. Hence, the limits $c_i^\pm$ are only reached asymptotically for $\xi_{i,1} \to \pm\infty$, see Fig. 1.

In order to calculate the derivative $c_i^{(r_i)}(x, u)$, where the input u appears explicitly, (9) is successively differentiated and new coordinates $\xi_{i,j+1}$ are introduced for the derivatives $\dot{\xi}_{i,j} =$

---

[1] For the sake of clarity, the dependance of $c_i^{(r_i)}$ on u has been made explicit in (7).
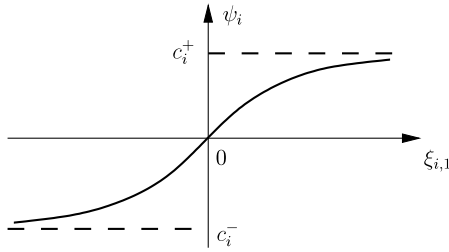
Fig. 1. Asymptotic saturation function (9) with saturation limits $c_i^{\pm}$ and coordinate $\xi_{i,1}$.

$\dot{\xi}_{i,j+1}, j = 1, \ldots, r_i - 1$. When $c_i^{(r_i)}(x, u)$ is reached, a new variable v_i serves as input to the final derivative $\dot{\xi}_{i,r_i} = v_i$. For instance, if the state constraint $c_i(x)$ is of order $r_i = 2$, the differentiation of (9) leads to

$$c_i^{(1)}(x) = \psi_i' \dot{\xi}_{i,1}, \quad \dot{\xi}_{i,1} = \xi_{i,2},$$
$$c_i^{(2)}(x, u) = \psi_i'' \xi_{i,2}^2 + \psi_i' \dot{\xi}_{i,2}, \quad \dot{\xi}_{i,2} = v_i \tag{10}$$

with the compact notation $\psi_i' := d\psi_i/d\xi_{i,1}$ and $\psi_i'' := d^2\psi_i/d\xi_{i,1}^2$. In the general case $r_i \geq 1$, the differentiations take the structure (obtained through chain differentiation)

$$c_i(x) = \psi_i(\xi_{i,1}) =: h_{i,1}(\xi_{i,1}), \quad i = 1, \ldots, p \tag{11a}$$

$$c_i^{(j)}(x) = \gamma_{i,j}(\xi_{i,1}, \ldots, \xi_{i,j}) + \psi_i' \xi_{i,j+1}$$
$$=: h_{i,j+1}(\xi_{i,1}, \ldots, \xi_{i,j+1}), \quad j = 1, \ldots, r_i - 1 \tag{11b}$$

$$c_i^{(r_i)}(x, u) = \gamma_{i,r_i}(\xi_i) + \psi_i' v_i$$
$$=: h_{i,r_i+1}(\xi_i, v_i), \tag{11c}$$

with the normal form cascades

$$\dot{\xi}_{i,j} = \xi_{i,j+1}, \quad j = 1, \ldots, r_i - 1,$$
$$\dot{\xi}_{i,r_i} = v_i, \quad i = 1, \ldots, p \tag{12}$$

and the new states

$$\xi_i = (\xi_{i,1}, \ldots, \xi_{i,r_i})^\mathsf{T}, \quad i = 1, \ldots, p. \tag{13}$$

The nonlinear terms $\gamma_{i,j}$ in the relations (11) follow from the previous functions $h_{i,j-1}$ with $\gamma_{i,1}(\xi_{i,1}) = 0$ and

$$\gamma_{i,j}\left(\xi_{i,1}, \ldots, \xi_{i,j}\right) = \sum_{k=1}^{j-1} \frac{\partial h_{i,j}}{\partial \xi_{i,k}} \xi_{i,k+1}, \quad j = 2, \ldots, r_i, \tag{14}$$

which are evaluated in a cascade (chain differentiation), see also (10). In summary, a new dynamic subsystem (12) with state $\xi_i = (\xi_{i,1}, \ldots, \xi_{i,r_i})^\mathsf{T}$ and new input v_i is generated for each state constraint (4), whereby the single normal form dynamics (12) are coupled to the original system dynamics (2) via (11c). In the following, these relations will be referred to using the DAE form

$$\dot{\xi}_i = g_i(\xi_i, v_i), \quad \xi_i(0) = h_i^{-1}(x_0), \tag{15}$$

$$0 = c_i^{(r_i)}(x, u) - h_{i,r_i+1}(\xi_i, v_i), \quad i = 1, \ldots, p, \tag{16}$$

with $g_i := (\xi_{i,2}, \ldots, \xi_{i,r_i}, v_i)^\mathsf{T}$ as the right-hand side of (12). To represent the constraints, the initial conditions $\xi_i(0)$ are chosen to be consistent with x_0 in (2). These initial conditions are determined by solving the relations (11a)–(11b) for $\xi_{i,1}, \ldots, \xi_{i,r_i}$:

$$\xi_{i,1} = \psi_i^{-1}(c_i(x)) =: h_{i,1}^{-1}(c_i(x))$$
$$\xi_{i,j+1} = \left[c_i^{(j)}(x) - \gamma_{i,j}(\xi_{i,1}, \ldots, \xi_{i,j})\right] / \psi_i'(\xi_{i,1}) \tag{17}$$
$$=: h_{i,j+1}^{-1}\left(\xi_{i,1}, \ldots, \xi_{i,j}, c_i^{(j)}(x)\right), \quad j = 1, \ldots, r_i - 1.$$

These relations have to be evaluated successively and can be written in the vector notation

$$\xi_i = h_i^{-1}(x) \tag{18}$$

as they appear in the initial conditions in (15). Naturally, the initial state $x(0) = x_0$ of the system (2) must strictly satisfy the constraints $c_i(x_0) \in (c_i^-, c_i^+)$, $i = 1, \ldots, p$ to ensure the existence of the inverse saturation function ψ_i^{-1} and to determine the initial conditions $\xi_i(0)$ in (15). The algebraic coupling equations (16) between (ξ_i, v_i) and (x, u) represent equality constraints, which couple the dynamics (15) with (2) and ensure that the relations (11) are satisfied.

To conclude Section 3.1, we summarize the results for the state constraints in the following

Proposition 1. *Let x with $x(0) = x_0$ satisfy $c_i(x) \in (c_i^-, c_i^+)$. Then, the variables ξ_i satisfying (12) with initial condition $\xi_i(0) = \psi^{-1}(x_0)$ and input v_i satisfying (16), are such that $c_i(x) = \psi_i(\xi_{i,1})$.*

3.2. Incorporation of input constraints

In contrast to the state constrained case, the mixed state-input constraints (5) can be directly substituted by further saturation functions (because its order is zero)

$$d_i(x, u) = \phi_i(w_i), \quad i = 1, \ldots, q \tag{19}$$

with the new unconstrained inputs w_i. As in the case of $\psi_i(v_i)$ in Fig. 1, the saturation functions $\phi_i : \mathbb{R} \rightarrow (d_i^-, d_i^+)$ are smooth and monotonically increasing.[2]

As a result, the constraints (4), (5) are replaced by the dynamic subsystems (15) and the algebraic equations (16), (19) with the new inputs $v = (v_1, \ldots, v_p)^\mathsf{T}$ and $w = (w_1, \ldots, w_q)^\mathsf{T}$, which extend the original dynamics (2) to a DAE system, as pictured in Fig. 2. The algebraic equations (16) and (19) represent equality constraints which can be treated much more conveniently in optimal control theory than the original inequality constraints (4), (5), e.g. by adjoining them in the calculus of variations, see [7] and Section 5.

3.3. New regularized optimal control problem

The new subsystems (15) with the equality constraints (16) and (19) are used to define a new OCP with the extended input and state vectors

$$\bar{u}^\mathsf{T} = \left(u^\mathsf{T}, v^\mathsf{T}, w^\mathsf{T}\right) \in \mathbb{R}^{m+p+q},$$
$$\bar{x}^\mathsf{T} = \left(x^\mathsf{T}, \xi_1^\mathsf{T}, \ldots, \xi_p^\mathsf{T}\right) \in \mathbb{R}^{n+r}, \quad r = \sum_{i=1}^{p} r_i. \tag{20}$$

Since the original cost functional (1) is independent of v and w, an additional regularization term[3] $\varepsilon(\|v\|^2 + \|w\|^2)$ with parameter ε is added to the cost $J(u)$, which leads to the new regularized and equality-constrained optimal control problem $\mathrm{OCP}_{\bar{u}}^{\varepsilon}$

minimize

$$\bar{J}(\bar{u}, \varepsilon) := J(u) + \varepsilon \int_0^T \|v\|^2 + \|w\|^2 \mathrm{d}t \tag{21}$$

subject to

---

[2] In this article, the saturation functions $\psi_i(\xi_{i,1})$ and $\phi_i(w_i)$ are constructed with the same formula given in (A.1) and (A.2) of the Appendix. In addition, (A.4) and (A.5) represent one-sided saturation functions if one of the general constraints (4) and (5) is only upper or lower bounded.

[3] $\|v\|$ and $\|w\|$ denote the Euclidean norm of the vectors v and w.
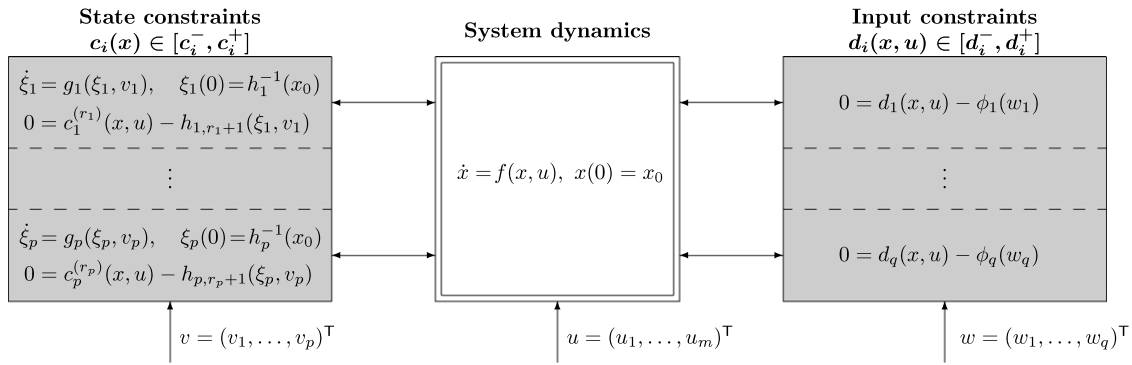
Fig. 2. Structure of the extended DAE system (2), (15)–(16), (19) with incorporated inequality constraints (4), (5).

$$\dot{x} = f(x, u), \quad x(0) = x_0, \tag{22}$$

$$\dot{\xi}_i = g_i(\xi_i, v_i), \qquad \xi_i(0) = h_i^{-1}(x_0), \quad i = 1, \ldots, p, \tag{23}$$

$$0 = \chi(x(T), T), \tag{24}$$

$$0 = c_i^{(r_i)}(x, u) - h_{i,r_i+1}(\xi_i, v_i), \quad i = 1, \ldots, p, \tag{25}$$

$$0 = d_i(x, u) - \phi_i(w_i), \quad i = 1, \ldots, q. \tag{26}$$

The effect of the regularization term with parameter ε will be explained in more detail in Section 5.2. In particular, it yields an easily exploitable stationarity condition with respect to v and w. The regularization term in (21) can also be interpreted as a penalty term to account for unboundedness of the new inputs v_i (as the r_i-th time derivative of $\xi_{i,1}$) and w_i if the corresponding constraint in (4) or (5) is touched.

In order to proceed, the following assumption is made:

Assumption 2. $\text{OCP}_{\bar{u}}^{\varepsilon}$ has a bounded solution $(\bar{u}^{\varepsilon}, \bar{x}^{\varepsilon})$ for each $\varepsilon > 0$.

Assumption 2 is reasonable from a practical point of view since the penalty is also introduced to penalize the "reaching" of the constraints. Moreover, the boundedness of the variables $(\xi_1, \ldots, \xi_p)$, v, and w implies that the constraints (4), (5) strictly remain inside their bounds.

In practice, $\text{OCP}_{\bar{u}}^{\varepsilon}$ has to be successively solved with decreasing values of the parameter ε. For $\varepsilon \to 0$, one intuitively expects that the regularized cost $\hat{J}(\bar{u}^{\varepsilon}, \varepsilon)$ converges to the optimal value J^* and that the original variables $(u^{\varepsilon}, x^{\varepsilon})$ converge to the optimal solution (u^*, x^*) of OCP_u. This point will be elaborated in the next section.

Remark 1. In certain cases, the boundedness in Assumption 2 can be explicitly verified. For the state-constrained case, this property is considered more deeply in [17,14] under an additional analycity assumption on the state $x^{\varepsilon}(t)$. In the input-constrained case, it is shown in [19] that the saturation function approach can be related to interior penalty methods, for which results are available concerning the strict satisfaction of input constraints [20].

4. Investigation of convergence

This section discusses the convergence of the cost (21) of $\text{OCP}_{\bar{u}}^{\varepsilon}$ for $\varepsilon \to 0$. In addition, the convergence of the original variables $(u^{\varepsilon}, x^{\varepsilon})$ to the optimal solution (x^*, u^*) is investigated under a quadratic growth assumption. In the following considerations the terminal time T is assumed to be fixed for the sake of clarity and simplicity.

4.1. Abstract formulation

Several norms are used in the following. Besides the Euclidian norm mentioned before, the standard norms $L^i(0, T; \mathbb{R}^b)$, $i = 1$,

2, ∞ are used over the set of continuous functions $a : [0, T] \to \mathbb{R}^b$ and are denoted by $\|a\|_1$, $\|a\|_2$, and $\|a\|_\infty$.

Some further definitions and assumptions are required. Define the set S of admissible control inputs $u \in L^\infty(0, T; \mathbb{R}^m)$ which together with their (unique) states x satisfy the dynamics and initial conditions (2), the terminal conditions (3), and the constraints (4), (5) for all times $t \in [0, T]$. In terms of S, OCP_u can alternatively be formulated in the form

$$\min_{u \in S} J(u) \tag{27}$$

with the optimal solution $J(u^*) = J^*$ (see Assumption 1). In order to investigate the convergence of the regularized $\text{OCP}_{\bar{u}}^{\varepsilon}$, consider the following subset of admissible inputs u for which the constraints (4), (5) are strictly satisfied on the open intervals:

$$S^0 = \big\{ u \in S : c_i(x) \in (c_i^-, c_i^+), i = 1, \ldots, p$$
$$d_i(x, u) \in (d_i^-, d_i^+), i = 1, \ldots, q \big\}. \tag{28}$$

For each admissible input $u \in S^0$, the new variables ξ_i and v_i of $\text{OCP}_{\bar{u}}^{\varepsilon}$ corresponding to the state constraints (4) are *uniquely* determined by (18), (9), and the inverse relation to (11c)

$$v_i = \frac{c_i^{(r_i)}(x, u) - \gamma_{i,r_i}(h_i^{-1}(x))}{\psi_i'(\psi_i^{-1}(c_i(x)))} =: h_{v,i}(x, u). \tag{29}$$

Correspondingly, the new inputs w_i follow from solving the relations (19) for w_i:

$$w_i = \phi_i^{-1}(d_i(x, u)) := h_{w,i}(x, u). \tag{30}$$

Since $v = (v_1, \ldots, v_p)^\mathsf{T}$ and $w = (w_1, \ldots, w_q)^\mathsf{T}$ are *uniquely* determined from the upper relations for all $u \in S^0$, the cost functional (21) can be stated as

$$\hat{J}(u, \varepsilon) := J(u) + \varepsilon p(u) \tag{31}$$

with

$$p(u) := \int_0^T \|h_v(x, u)\|^2 + \|h_w(x, u)\|^2 \mathrm{d}t \tag{32}$$

and $h_v = (h_{v,1}, \ldots, h_{v,p})^\mathsf{T}$, $h_w = (h_{w,1}, \ldots, h_{w,q})^\mathsf{T}$. In summary, the regularized $\text{OCP}_{\bar{u}}^{\varepsilon}$ in (21)–(26) is equivalent to

$$\min_{u \in S^0} \hat{J}(u, \varepsilon), \tag{33}$$

whereby Assumption 2 implies the non-emptiness of S^0. Finally, some additional assumptions are made:

Assumption 3. (a) The functional J is continuous in u for all $u \in S$.
(b) The optimal control u^* of OCP_u lies in the closure of the set S^0.

(c) The function f (being sufficiently smooth) is Lipschitz in x and u.

A direct consequence of the above Lipschitz assumption is that two solutions x and x' for associated inputs u and u' satisfy

$$\|x - x'\|_\infty \le C_1 \|u - u'\|_1 \quad \forall u, u' \in S \tag{34}$$

for some constant $C_1 > 0$. This relation can be proved with Gronwall's lemma. The additional assumption that the optimal control $u^* \in S$ lies in the closure of S^0 is necessary to ensure that u^* can be approached from within S^0, see, e.g., [21] for a similar assumption in the context of interior point methods.

4.2. Convergence results for $\varepsilon \to 0$

Since OCP (33) has to be successively solved for decreasing $\varepsilon^{k+1} < \varepsilon^k$, the following lemma is of importance concerning the non-growth of the cost (31).

Lemma 1. *Let u^{k+1} and u^k be the optimal control inputs of OCP (33) for $0 < \varepsilon^{k+1} < \varepsilon^k$. Then, the following inequalities hold for the cost functional (31):*

$$J(u^{k+1}) \le J(u^k), \qquad p(u^{k+1}) \ge p(u^k) \tag{35a}$$

and

$$\hat{J}(u^{k+1}, \varepsilon^{k+1}) \le \hat{J}(u^k, \varepsilon^k). \tag{35b}$$

The proof can be found in [21] and is omitted here due to the lack of space. The following theorem concerns the convergence of the cost $\hat{J}(u^k, \varepsilon^k)$ using the results of Lemma 1.

Theorem 1. *With the notations of Lemma 1, let $\{\varepsilon^k\}$ be a decreasing sequence of positive regularization parameters with $\lim_{k\to\infty} \varepsilon^k = 0$. Then, $\hat{J}(u^k, \varepsilon^k)$ converges to the optimal cost*

$$\lim_{k\to\infty} \hat{J}(u^k, \varepsilon^k) = J^* \tag{36a}$$

with

$$\lim_{k\to\infty} J(u^k) = J^*, \qquad \lim_{k\to\infty} \varepsilon^k p(u^k) = 0. \tag{36b}$$

Proof. The proof of the theorem is adapted from [22]. Since $J(u)$ is continuous over S and $u^* \in S$ lies in the closure of S^0 (see Assumption 3), it follows that, for any parameter $\delta > 0$, one can always find an admissible input $u^\delta \in S^0$ with associated state x^δ such that

$$J(u^\delta) < J^* + \delta/2. \tag{37a}$$

Select ε^l with $\varepsilon^l p(u^\delta) < \delta/2$. Then, for any $k > l$ with $\varepsilon^k < \varepsilon^l$ and using Lemma 1,

$$\hat{J}(u^k, \varepsilon^k) \le \hat{J}(u^l, \varepsilon^l) \le \hat{J}(u^\delta, \varepsilon^l), \tag{37b}$$

where u^k and u^l are the optimal solutions for ε^k and ε^l, respectively. With (37a) and $\varepsilon^l p(u^\delta) < \delta/2$, there exists an upper estimate on $\hat{J}(u^\delta, \varepsilon^l)$ with

$$\hat{J}(u^\delta, \varepsilon^l) < J^* + \delta/2 + \delta/2 = J^* + \delta. \tag{37c}$$

Finally, using $\hat{J}(u^k, \varepsilon^k) > J(u^k) \ge J^*$, (37b) and (37c) lead to the conclusion that $\forall \delta > 0$, $\exists l$ such that $\forall k > l$, $|\hat{J}(u^k, \varepsilon^k) - J^*| < \delta$. This proves (36a) and additionally (36b) in view of (31). $\quad\square$

In order to prove convergence of the input u^k and its associated states x^k, we require an additional quadratic growth condition as given in the next theorem:

Theorem 2. *Assume that (in addition to Assumptions 1–3) the cost functional $J(u)$ satisfies the quadratic growth property*

$$C_2 \|u - u^*\|_2^2 \le J(u) - J(u^*) \quad \forall u \in S \tag{38}$$

for some constant $C_2 > 0$. Then, the solution (u^k, x^k) of OCP (33) for the decreasing sequence $\{\varepsilon^k\}$ with $\lim_{k\to\infty} \varepsilon^k = 0$ converges to (u^*, x^*) according to

$$\lim_{k\to\infty} \|u^k - u^*\|_2 = 0, \qquad \lim_{k\to\infty} \|x^k - x^*\|_\infty = 0. \tag{39}$$

Proof. Due to the results of Theorem 1, it follows from (38) that u^k converges to u^* in L^2. The convergence of x^k relies on Assumption 3(c) which leads to (34). The L^1-norm in (34) can be related to L^2 in (38) by means of Hölder's inequality, i.e. $\|u - u^*\|_1 \le \sqrt{T} \|u - u^*\|_2$. Hence, (34) can be written as $\|x - x^*\|_\infty \le C_1 \|u - u^*\|_1 \le C_1 \sqrt{T} \|u - u^*\|_2$, which shows the convergence of x^k in L^∞. $\quad\square$

Remark 2. It can be shown that the quadratic growth property (38) holds for linear system dynamics and quadratic cost functionals. Moreover, (38) can be seen as a strong smoothness assumption which however is clearly weaker than assuming strong convexity (it is well known that strong convexity on a compact set implies quadratic growth, see e.g. [23]). A simple finite-dimensional example is the function $f(x) = x^2 + 10 \sin^2 x$, where x is restricted to the interval $x \in [-5, 5]$. This function has a global minimum $f(x^*) = 0$ at $x^* = 0$. Clearly, the quadratic growth property $\frac{1}{2}|x - x^*| \le f(x) - f(x^*)$ is satisfied for all $x \in [-5, 5]$ although $f(x)$ is not convex on this interval.

Note that no statement is made in Theorem 2 about the limits of the new variables ξ^k, v^k, and w^k of OCP$_{\bar{u}}^\varepsilon$, since they become unbounded if the trajectories x^k and u^k touch the corresponding constraint in (4) and (5) for $k \to \infty$. Nevertheless, the original state x^k and input u^k converge to the optimal trajectories u^* and x^* as stated in Theorem 2. In practice, the sequence $\{\varepsilon^k\}$ is stopped after a convergence criterion is satisfied, which prevents the variables (ξ^k, v^k, w^k) from becoming unbounded.

4.3. Summary of results

The following paragraph summarizes the results obtained so far in a condensed manner before proceeding to Section 5:

Consider the inequality-constrained OCP$_u$ (1)–(5) with the optimal control u^* lying in the closure of the set S^0 defined in (28) (also see Assumptions 1 and 3). Moreover, consider the penalized OCP$_{\bar{u}}^\varepsilon$ (21)–(26) defined by the introduction of saturation functions and dynamic extension to integrate the original inequality constraints of OCP$_u$. Under Assumption 2, OCP$_{\bar{u}}^\varepsilon$ possesses a bounded solution in the new extended coordinates $\bar{u}^\mathsf{T} = (u^\mathsf{T}, v^\mathsf{T}, w^\mathsf{T})$ and $\bar{x}^\mathsf{T} = (x^\mathsf{T}, \xi_1^\mathsf{T}, \ldots, \xi_p^\mathsf{T})$ defined in (20). This equality-constrained OCP can be readily be handled in the classical calculus of variations (see Section 5). Moreover, OCP$_{\bar{u}}^\varepsilon$ is equivalent to

$$\min_{u \in S^0} J(u) + \varepsilon p(u)$$

defined in (31)–(33). When $\varepsilon \to 0$, convergence of the optimal cost of OCP$_{\bar{u}}^\varepsilon$ towards $J(u^*)$ is guaranteed by Theorem 1. If the quadratic growth property (38) holds, Theorem 2 additionally guarantees the convergence of the trajectories towards (x^*, u^*).

5. Solution of OCP$_{\bar{u}}^\varepsilon$

This section focuses on the numerical solution of OCP$_{\bar{u}}^\varepsilon$. In the first step, the first-order optimality conditions are derived, which lead to a two-point boundary value problem (BVP). This BVP can be solved numerically, for instance with the collocation method.

5.1. Necessary optimality conditions

The necessary optimality conditions for $\text{OCP}_{\bar{u}}^{\varepsilon}$ follow from the calculus of variations. To this end, we define the Hamiltonian

$$H(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}, t) = L(x, u, t) + \varepsilon \left(\|v\|^2 + \|w\|^2 \right)$$

$$+ \lambda^{\mathsf{T}} f(x, u) + \sum_{i=1}^{p} \eta_i^{\mathsf{T}} g_i(\xi_i, v_i)$$

$$+ \sum_{i=1}^{p} \mu_i \left(c_i^{(r_i)}(x, u) - h_{i,r_i+1}(\xi_i, v_i) \right)$$

$$+ \sum_{i=1}^{q} \nu_i \left(d_i(x, u) - \phi_i(w_i) \right) \tag{40}$$

with the adjoint states $\bar{\lambda}^{\mathsf{T}} = (\lambda^{\mathsf{T}}, \eta_1^{\mathsf{T}}, \ldots, \eta_p^{\mathsf{T}})$, where $\lambda \in \mathbb{R}^n$ and $\eta_i \in \mathbb{R}^{r_i}$ are related to the dynamical subsystems (22) and (23). The additional multipliers $\bar{\mu}^{\mathsf{T}} = (\mu^{\mathsf{T}}, \nu^{\mathsf{T}})$ with $\nu \in \mathbb{R}^p$ and $\mu \in \mathbb{R}^q$ are used to adjoin the equality constraints (25) and (26) to the Hamiltonian (40).

An optimal solution of $\text{OCP}_{\bar{u}}^{\varepsilon}$ has to satisfy the stationarity conditions

$$\frac{\partial H}{\partial u} = \frac{\partial L}{\partial u} + \lambda^{\mathsf{T}} \frac{\partial f}{\partial u} + \sum_{i=1}^{p} \mu_i \frac{\partial c_i^{(r_i)}}{\partial u} + \sum_{i=1}^{q} \nu_i \frac{\partial d_i}{\partial u} = 0 \tag{41a}$$

$$\frac{\partial H}{\partial v_i} = 2\varepsilon v_i + \eta_{i,r_i} - \mu_i \psi_i'(\xi_{i,1}) = 0, \quad i = 1, \ldots, p \tag{41b}$$

$$\frac{\partial H}{\partial w_i} = 2\varepsilon w_i - \nu_i \phi_i'(w_i) = 0, \quad i = 1, \ldots, q, \tag{41c}$$

which, together with the equality constraints (25)–(26), determine the inputs $\bar{u}^{\mathsf{T}} = (u^{\mathsf{T}}, v^{\mathsf{T}}, w^{\mathsf{T}})$ and the multipliers (ν, μ).

The adjoint states $\bar{\lambda}^{\mathsf{T}} = (\lambda^{\mathsf{T}}, \eta_1^{\mathsf{T}}, \ldots, \eta_p^{\mathsf{T}})$ follow from the dynamics $\dot{\bar{\lambda}}^{\mathsf{T}} = -\partial H / \partial x$ and $\dot{\eta}_i^{\mathsf{T}} = -\partial H / \partial \xi_i$, which can be written in more detail as

$$\dot{\lambda}^{\mathsf{T}} = -\frac{\partial L}{\partial x} - \lambda^{\mathsf{T}} \frac{\partial f}{\partial x} - \sum_{i=1}^{p} \mu_i \frac{\partial c_i^{(r_i)}}{\partial x} - \sum_{i=1}^{q} \nu_i \frac{\partial d_i}{\partial x} \tag{42a}$$

$$\dot{\eta}_{i,1} = \mu_i \frac{\partial h_{i,r_i+1}}{\partial \xi_{i,1}}, \quad i = 1, \ldots, p \tag{42b}$$

$$\dot{\eta}_{i,j} = -\eta_{i,j-1} + \mu_i \frac{\partial h_{i,r_i+1}}{\partial \xi_{i,j}}, \quad j = 2, \ldots, r_i. \tag{42c}$$

The terminal conditions for λ and η_i, $i = 1, \ldots, p$ are given by

$$\lambda^{\mathsf{T}}(T) = \frac{\partial \varphi}{\partial x}\bigg|_T + \kappa^{\mathsf{T}} \frac{\partial \chi}{\partial x}\bigg|_T, \qquad \eta_i(T) = 0, \quad i = 1, \ldots, p \tag{43}$$

with the additional (constant) multipliers $\kappa \in \mathbb{R}^l$. The final conditions for η_i are zero, since no terminal conditions are imposed on the states $(\xi_1, \ldots, \xi_p)$ of $\text{OCP}_{\bar{u}}^{\varepsilon}$. If the terminal time T is unspecified, the additional transversality condition

$$H(\bar{x}, \bar{u}, \bar{\lambda}, \bar{\mu}, t)\big|_T = -\frac{\partial \varphi}{\partial t}\bigg|_T - \kappa^{\mathsf{T}} \frac{\partial \chi}{\partial t}\bigg|_T \tag{44}$$

must be satisfied. The differential equations and boundary conditions (22)–(24), (42)–(44) together with the algebraic equations (25), (26), (41) define a two-point BVP. Its solution yields the trajectories of the (extended) states $\bar{x}^{\varepsilon}(t)$, $\bar{\lambda}^{\varepsilon}(t)$ and input $\bar{u}^{\varepsilon}(t)$, the multipliers $\bar{\mu}^{\varepsilon}(t)$, and κ^{ε} as well as the optimal end time T^{ε} (if initially unspecified).

5.2. Jacobian of the algebraic equations

Some interesting properties of the saturation function formulation and dynamic extension can be revealed from the Jacobian of the algebraic equations (41), (25), (26) with respect to (u, v, w, μ, ν):

$$F = \begin{bmatrix} \frac{\partial^2 H}{\partial u^2} & 0 & 0 & \left(\frac{\partial c^{(r)}}{\partial u}\right)^{\mathsf{T}} & \left(\frac{\partial d}{\partial u}\right)^{\mathsf{T}} \\ 0 & 2\varepsilon I_p & 0 & -\Psi'(\xi) & 0 \\ 0 & 0 & 2\varepsilon I_q - \nu^{\mathsf{T}} \Phi''(w) & 0 & -\Phi'(w) \\ \frac{\partial c^{(r)}}{\partial u} & -\Psi'(\xi) & 0 & 0 & 0 \\ \frac{\partial d}{\partial u} & 0 & -\Phi'(w) & 0 & 0 \end{bmatrix}, \tag{45}$$

with the $p \times p$ and $q \times q$ unit matrices I_p and I_q, the abbreviations $c^{(r)} = (c_1^{(r_1)}, \ldots, c_p^{(r_p)})^{\mathsf{T}}$, $d = (d_1, \ldots, d_q)^{\mathsf{T}}$, and the matrices $\Psi'(\xi)$, $\Phi'(w)$ (and $\Phi''(w)$ accordingly) being defined by

$$\Psi'(\xi) = \begin{bmatrix} \psi_1' & & 0 \\ & \ddots & \\ 0 & & \psi_p' \end{bmatrix}, \qquad \Phi'(w) = \begin{bmatrix} \phi_1' & & 0 \\ & \ddots & \\ 0 & & \phi_q' \end{bmatrix}.$$

For the non-singularity of the Jacobian F that is required for the numerical solution of the algebraic Eqs. (41), (25) and (26), the Hamiltonian $H(\bar{x}(t), \bar{u}(t), \bar{\lambda}(t), \bar{\mu}(t), t)$ is assumed to have a positive definite Hessian

$$\frac{\partial^2 H}{\partial \bar{u}^2} = \begin{bmatrix} \frac{\partial^2 H}{\partial u^2} & 0 & 0 \\ 0 & 2\varepsilon I_p & 0 \\ 0 & 0 & 2\varepsilon I_q - \nu^{\mathsf{T}} \Phi''(w) \end{bmatrix} > 0$$

$$\forall t \in [0, T]. \tag{46}$$

The positive definiteness of $\partial^2 H / \partial \bar{u}^2$ represents a sufficient second-order optimality condition (strengthened Legendre–Clebsch condition) and is often implicitly assumed in optimal control when the first-order optimality conditions are solved numerically.[4]

The non-singularity of the Jacobian (45) can easily be shown by means of (46) and $\Psi'(\xi)$ and $\Phi'(w)$ being positive definite. Moreover, it is obvious that the regularization term with parameter ε is necessary for $\partial^2 H / \partial \bar{u}^2 > 0$ in (46) and that the ε-entries in (45) avoid a rank deficiency of the second and third lines of F for an "almost active" constraint $c_i(x) \to c_i^{\pm}$ or $d_i(x, u) \to d_i^{\pm}$ with $\psi_i' \to 0$ or $\phi_i', \phi_i'' \to 0$, respectively.

Remark 3. A theoretically interesting option of the saturation function approach is to use "exact" saturation functions for the state constraints (4), which lead to a tangential entry and exit of the constraints $c_i^{\pm} = \psi_i(\xi_{i,1}^{\pm})$ at certain saturation values $\xi_{i,1}^{-}$ and $\xi_{i,1}^{+}$. Then, under the standard state constraint qualification (i.e., the gradient vectors $\partial c_i^{(r_i)} / \partial u$ of all active state constraints $c_i(x)$ being linearly independent) the Jacobian (45) remains non-singular, although $\psi_i'(\xi_{i,1}) \equiv 0$ holds for the active state constraints. The properties of exact saturation functions and the influence of their order of continuity on the boundedness of the ξ_i-variables are analytically investigated in [17] for the special case of a single state constraint.

---

[4] Note that the Hessian with respect to the original control, $\partial^2 H / \partial u^2$, is always positive definite if the system (2) and the constraints (5) are input-affine and the cost (1) satisfies $\partial^2 L / \partial u^2 > 0$, e.g. for $L(u) = \|u\|^2$.

5.3. Numerical solution with collocation

An efficient way to numerically solve two-point BVPs is the collocation method, see, e.g., [24]. In order to solve DAE-BVPs as they arise from the optimality conditions, a collocation solver under MATLAB is used, which uses modifications of the standard MATLAB function bvp4c [25]. This solver is applicable to general BVPs of (index 1) differential-algebraic equations (DAE)

$$\dot{x}_d = f_d(x_d, x_a, t, p), \tag{47a}$$
$$0 = f_a(x_d, x_a, t, p), \tag{47b}$$
$$0 = f_{bc}(x_d(t_0), x_d(t_f), x_a(t_0), x_a(t_f), p) \tag{47c}$$

with the differential and algebraic equations (47a), (47b) for the dynamic and algebraic states $x_d(t)$ and $x_a(t)$ on the time interval $t \in [t_0, t_f]$ and the boundary conditions (47c). Unknown parameters p can additionally be considered in the DAE formulation (47).

The collocation scheme discretizes the differential equations (47a) along a time mesh $t_i \in [t_0, t_f]$, $i = 1, \ldots, N$. The resulting discretized equations (47a), (47b) together with the boundary conditions (47c) result in a set of nonlinear algebraic equations for the variables $x_d(t_i)$ and $x_a(t_i)$, $i = 1, \ldots, N$, which is solved with a damped Newton iteration scheme. If the initial guess provided by the user is far away from the actual solution, a line search algorithm is used to bring the numerical iterates closer to the domain of convergence of the Newton iteration scheme. In addition, a mesh refinement strategy is used to adapt the time mesh $t_i \in [t_0, t_f]$, $i = 1, \ldots, N$ and the number of grid points N in each Newton step based on the residual along the discretized ODEs (47a).

In order to use the collocation method to solve the optimal control problem OCP$_{\bar{u}}^{\varepsilon}$, the BVP (22)–(26), (42)–(44) has to be adapted to the DAE form (47). The ODEs (47a) are given by the system and adjoint equations in (22), (23), (42) for the dynamic states $x_d^T = (\bar{x}^T, \bar{\lambda}^T)$. The extended input $\bar{u}^T = (u^T, v^T, w^T)$ and the multipliers $\bar{\mu}^T = (v^T, \mu^T)$ refer to the algebraic variables x_a with (25), (26), (41) corresponding to (47b). The boundary conditions for $\bar{x}$ and $\bar{\lambda}$ according to (22)–(24), (43) together with the transversality condition (44) if T is unspecified constitute (47c). The multipliers κ in the terminal conditions (43) and (44) can be treated as unknown parameters $p = \kappa$.

6. Example — Goddard problem

A well-known benchmark problem in trajectory optimization suggested by Goddard in 1919 [26] is to maximize the final altitude of a vertically ascending rocket under the influence of atmospheric drag and gravitation. The given constraints on the thrust (input constraint) and dynamic pressure (state constraint) as well as the characteristic singular arc behavior of the Goddard problem make it an ideal example to illustrate the presented approach.

6.1. Optimal control problem

The optimal control problem for the Goddard rocket subject to thrust and dynamic pressure constraints can be stated as [27,28, 10,29]

minimize

$$J(u) := -h(T) \tag{48}$$

subject to

$$\dot{h} = V, \quad \dot{V} = \frac{u - D(h, V)}{m} - \frac{1}{h^2}, \quad \dot{m} = -\frac{u}{c}, \tag{49}$$

$$h(0) = 1, \quad v(0) = 0, \quad m(0) = 1, \quad m(T) = 0.6, \tag{50}$$

$$Aq(h, V) < 10, \quad u \in [0, 3.5] \tag{51}$$

with the altitude h from the center of Earth, the velocity V, and the mass m of the rocket. The states (h, V, m), the thrust u as the control input, and the time t are normalized and dimension-free. The drag function $D(h, V) = D_0 q(h, V)$ in (49) is proportional to the dynamic pressure

$$q(h, V) = \frac{1}{2} \rho_0 V^2 \exp[\beta(1 - h)], \tag{52}$$

that depends on the altitude h and velocity V.[5]

As described by the boundary conditions (50), the rocket starts from the Earth's surface and is initially at rest. The terminal condition at the free end time T is imposed on the mass to account for the consumed fuel of the rocket with respect to its initial mass. The dynamic pressure constraint (with the reference area A) in addition to the thrust constraint in (51) is an extension of the original Goddard problem. It was introduced by Seywald and Cliff [28].

6.2. Incorporation of constraints with saturation functions

The dynamic pressure constraint in (51) can be reordered and written in the form (4)

$$c(h, V) := V^2 \exp[\beta(1 - h)] \leq \frac{20}{A\rho_0} = c^+. \tag{53}$$

The state inequality constraint (53) is of order one (i.e. $r = 1$), since the input u appears in the first time derivative

$$c^{(1)}(h, V, m, u) := 2V \exp[\beta(1 - h)]$$
$$\times \left(\frac{u - D(h, V)}{m} - \frac{1}{h^2} - \frac{1}{2}\beta V^2 \right). \tag{54}$$

Following Section 3.1, the inequality constraint (53) is represented by a saturation function $\psi(\xi)$, cf. (9), and is differentiated in order to introduce a new input v for $\dot{\xi}$:

$$c(h, V) = \psi(\xi) \in (c^-, c^+) \tag{55}$$

$$c^{(1)}(h, V, m, u) = \psi'\dot{\xi} \quad \text{with } \dot{\xi} = v. \tag{56}$$

Since the reformulated pressure constraint (53) is only bounded from above, the saturation limits for $\psi : \mathbb{R} \to (c^-, c^+)$ are chosen symmetrically with $c^- = -c^+$.

The thrust constraint in (51) is taken into account by a further saturation function according to (19):

$$u = \phi(w). \tag{57}$$

To summarize, the inequality constraints (51) are replaced by the new dynamics $\dot{\xi} = v$ and the equality constraints in (56), (57). This leads to the new regularized OCP$_{\bar{u}}^{\varepsilon}$ as defined in (21)–(26) with the extended state $\bar{x} = (h, V, m, \xi)^T$ and input $\bar{u} = (u, v, w)^T$:

minimize

$$\bar{J}(\bar{u}, \varepsilon) := -h(T) + \varepsilon \int_0^T v^2 + w^2 \, dt \tag{58}$$

subject to

$$\dot{h} = V, \quad \dot{V} = \frac{u - D(h, V)}{m} - \frac{1}{h^2}, \quad \dot{m} = -\frac{u}{c}, \tag{59}$$

$$h(0) = 1, \quad V(0) = 0, \quad m(0) = 1, \quad m(T) = 0.6, \tag{60}$$

$$\dot{\xi} = v, \quad \xi(0) = \psi^{-1}(0) = 0, \tag{61}$$

$$0 = c^{(1)}(h, V, m, u) - \psi'(\xi)v, \tag{62}$$

$$0 = u - \phi(w). \tag{63}$$

---

[5] The following (normalized) values are taken from [28]: $\beta = 500$, $c = 0.5$, $D_0\rho_0 = 620$, and $A\rho_0 = 12\,400$.
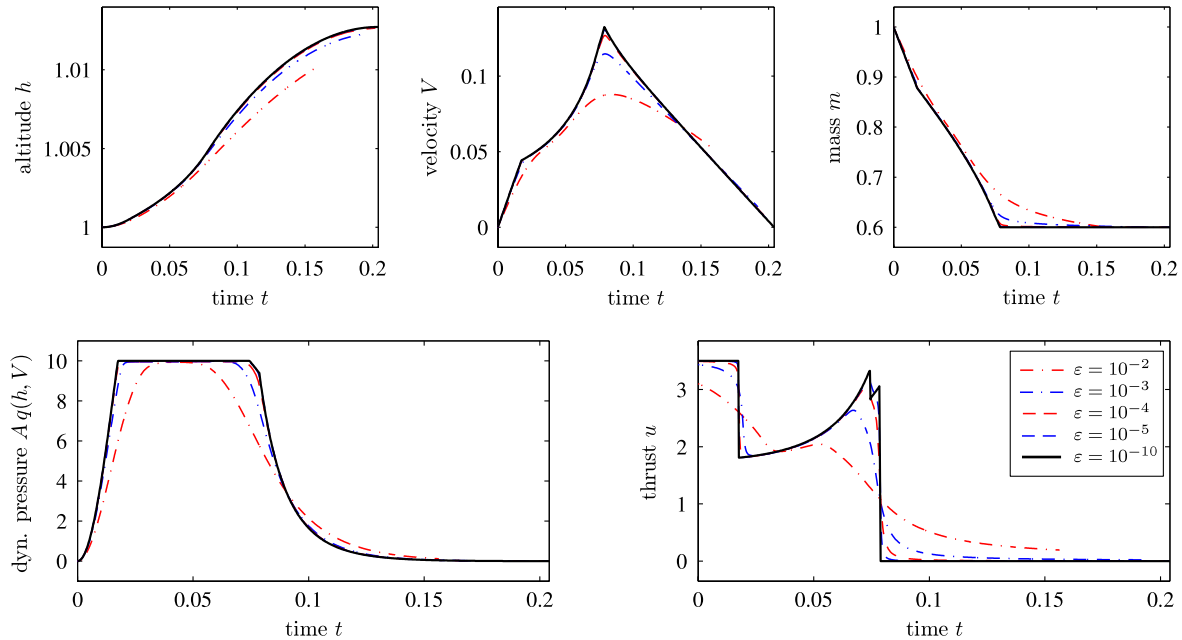
Fig. 3. Optimal trajectories of the Goddard problem for decreasing regularization parameters ε.

Note that the initial condition for ξ is zero due to the symmetric bounds $c^{\pm}$ for the saturation function in (55) and $c(h(0), V(0)) = 0$ with $V(0) = 0$.

6.3. Numerical results

The necessary optimality conditions (41)–(44) for the Goddard problem are calculated with the computer algebra system MATHEMATICA and are stored as MATLAB functions.[6] The Goddard problem is solved with the collocation method as described in Section 5.3, whereby the free end time T is taken into account by means of the time transformation

$$t = \delta\tau, \qquad T = \delta, \qquad \frac{\mathrm{d}}{\mathrm{d}t} = \frac{1}{\delta}\frac{\mathrm{d}}{\mathrm{d}\tau} \tag{64}$$

with the normalized time coordinate $\tau \in [0, 1]$. The scaling factor δ is treated as a free parameter $p = \delta$ in the DAE system (47) and the new time coordinate τ replaces $t \in [t_0, t_f]$ with the normalized interval boundaries $t_0 = 0$ and $t_f = 1$.

The initial guess for the states (h, V, m) is a linear interpolation between the boundary conditions in (60) and $h(T) = 1$, $V(T) = 0$ for the unspecified final states. The initial trajectories for ξ, λ, and $\bar{u}$ are zero. The free parameter δ in the time transformation (64) is chosen to $\delta = 0.1$, which corresponds to an initial final time $T = 0.1$.

Fig. 3 shows the optimal trajectories of the states (h, V, m) and of the dynamic pressure constraint in (51) as well as the original control input $u(t)$ for the Goddard problem with different regularization parameters ε by solving the Goddard problem with the collocation method and successively decreasing ε from 10^0 to 10^{-10}. After the initial run with $\varepsilon = 10^0$, the subsequent runs use the previous solutions as initialization. The maximum altitude $h(T)$ and the terminal time T obtained for $\varepsilon = 10^{-10}$ are (both normalized values)

$$h(T) = 1.0127176, \quad T = 0.20404872. \tag{65}$$

---

[6] In addition to the boundary conditions in (60), (61), the terminal conditions (43) for the adjoint states and the transversality condition (44) simplify to $\lambda_h(T) = -1$, $\lambda_V(T) = 0$, $\eta(T) = 0$, and $H(\cdot)|_T = 0$.

The constraints (51) of the dynamic pressure and the thrust are clearly satisfied. Particularly interesting is the singular arc of the thrust $u(t)$ appearing at $t \approx 0.075$, which directly follows the constrained arc of the dynamic pressure constraint. This behavior is explained in more detail in [28]. Although this singular arc barely influences the final cost, it shows the applicability of the saturation function approach and the accuracy of the collocation method for solving the Goddard problem.

7. Conclusions

An approach has been presented to systematically transform a given constrained optimal control problem (OCP) into a new equality-constrained OCP by means of saturation functions. The method accounts for the relative degree (or order) of state constraints by constructing dynamical subsystems which are coupled to the original dynamics. In contrast to the original constrained OCP, the derived equality-constrained OCP with extended state and input vectors can be treated conveniently in the calculus of variations. An additional regularization term is introduced in the cost to regularize the new OCP with respect to the new input variables. After deriving the optimality conditions, the resulting two-point BVP can be solved numerically (e.g. with the collocation method as used in this paper), whereby the regularization parameter has to be successively reduced to approach the constrained optimal solution. Besides the Goddard problem used in this paper, the applicability of the approach and the potential of the presented collocation solver has been demonstrated for the space shuttle reentry problem [19], which is a benchmark problem in optimization due to a high numerical sensitivity and the presence of input and heating constraints.

An important aspect is that the presented method is independent of the methodology, which is used to solve the new OCP. Although the indirect method is used in this paper because of its well-known accuracy and to exploit the structure of the optimality conditions, direct optimization methods can equally be applied to solve the resulting equality-constrained OCP. A particular advantage in this context is that the constraints cannot be violated in the new OCP formulation due to their inherent consideration by the saturation functions. Hence, an alternative to the indirect solution

in this paper are standard numerics from unconstrained optimization.

A further point of interest is to investigate the impact of the system extension on the overall convergence behavior, since the saturation function approach naturally increases the dimension of the overall problem. On the other hand, the additional subsystems have a characteristic structure that can be accounted for in the numerical solution of the extended OCP. Our own numerical experiences as well as the numerical results in the paper show that the increased dimensionality behaves well regarding numerical robustness and convergence. A more rigorous investigation of this point may be subject of future research.

Moreover, current research is spent on the further development of the collocation solver and on the definition of a strategy to automatically update the penalty parameter. Potential future work concerns the combination of direct and indirect methods as studied in [30–33] in order to take advantage of the larger domain of convergence of direct methods to find a suitable initial guess for the more accurate indirect method.

Acknowledgements

Appendix. Choice of saturation functions

An appropriate choice for the smooth saturation function $\psi_i(\xi_{i,1}) \in (c_i^-, c_i^+)$ in (9) is given by

$$\psi_i(\xi_{i,1}) = c_i^+ - \frac{c_i^+ - c_i^-}{1 + \exp(s\xi_{i,1})}, \quad s = \frac{4}{c_i^+ - c_i^-}. \tag{A.1}$$

The parameter s adjusts the slope at the position $\xi_{i,1} = 0$ and is chosen in (A.1) to normalize the slope at $\xi_{i,1} = 0$ to $\frac{\partial \psi_i}{\partial \xi_{i,1}} = 1$. Note that the saturation limits $c_i^\pm$ are only reached asymptotically for $\xi_{i,1} \to \pm\infty$.

The saturation functions for the input constraints in (19) are constructed accordingly by

$$\phi_i(w_i) = d_i^+ - \frac{d_i^+ - d_i^-}{1 + \exp(sw_i)}, \quad s = \frac{4}{d_i^+ - d_i^-}. \tag{A.2}$$

Further choices for asymptotic saturation functions can be constructed by means of tanh-functions.

The saturation function approach can readily be applied to one-sided state or control constraints of the form

$$\begin{array}{ll} c_i(x) \le c_i^+ & \quad d_i(x, u) \le d_i^+ \\ \text{or} & \\ c_i(x) \ge c_i^- & \quad d_i(x, u) \ge d_i^-. \end{array} \tag{A.3}$$

A suitable choice of an upper or lower bounded saturation function $\psi_i^+ : \mathbb{R} \to (-\infty, c_i^+)$ or $\psi_i^- : \mathbb{R} \to (c_i^-, \infty)$, respectively, is

$$\psi_i^+(\xi_{i,1}) = c_i^+ - e^{-\xi_{i,1}}, \qquad \psi_i^-(\xi_{i,1}) = c_i^- + e^{\xi_{i,1}}. \tag{A.4}$$

Accordingly, the one-sided version of the saturation function (A.2) for an input constraint $d_i(x, u) \le d_i^+$ or $d_i(x, u) \ge d_i^-$ becomes

$$\phi_i^+(w_i) = d_i^+ - e^{-w_i}, \qquad \phi_i^-(w_i) = d_i^- + e^{w_i}. \tag{A.5}$$

Similar to the two-sided saturation functions in (A.1) and (A.2), the functions $\psi_i^\pm$ and $\phi_i^\pm$ have the slope $\frac{d\psi_i^\pm}{d\xi_{i,1}} = 1$ and $\frac{d\phi_i^\pm}{dw_i} = 1$ at the positions $\xi_{i,1} = 0$ and $w_i = 0$.

References

[1] C. Hargraves, S. Paris, Direct trajectory optimization using nonlinear programming and collocation, AIAA Journal of Guidance, Control, and Dynamics 10 (1987) 338–342.

[2] O. von Stryk, Numerical solution of optimal control problems by direct collocation, International Series of Numerical Mathematics 111 (1993) 129–143.

[3] H. Seywald, Trajectory optimization based on differential inclusion, AIAA Journal of Guidance, Control, and Dynamics 17 (1994) 480–487.

[4] J. Betts, Survey of numerical methods for trajectory optimization, AIAA Journal of Guidance, Control, and Dynamics 21 (1998) 193–207.

[5] J. Betts, Practical Methods for Optimal Control Using Nonlinear Programming, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[6] J. Nocedal, S. Wright, Numerical Optimization, Springer, New York, 2006.

[7] A. Bryson, Y.-C. Ho, Applied Optimal Control, Wiley & Sons, New York, 1975.

[8] H. Pesch, A practical guide to the solution of real-life optimal control problems, Control and Cybernetics 23 (1994) 7–60.

[9] H. Kreim, B. Kugelmann, H. Pesch, M. Breitner, Minimizing the maximum heating of a reentry space shuttle: an optimal control problem with multiple control constraints, Optimal Control Applications and Methods 17 (1996) 45–69.

[10] A. Bryson, Dynamic Optimization, Addison–Wesley, Menlo Park, CA, 1999.

[11] B. Bonnard, L. Faubourg, G. Launay, E. Trélat, Optimal control with state constraints and the space shuttle re-entry problem, Journal of Dynamical and Control Systems 9 (2003) 155–199.

[12] F. Chaplais, N. Petit, Inversion in indirect optimal control of multivariable systems, ESAIM: Control, Optimisation and Calculus of Variations 14 (2) (2008) 294–317.

[13] L. Pontryagin, V. Boltyansky, V. Gamkrelidze, E. Mischenko, Mathematical Theory of Optimal Processes, Wiley-Interscience, New York, 1962.

[14] K. Graichen, N. Petit, Incorporating a class of constraints into the dynamics of optimal control problems, Optimal Control Applications and Methods 30 (2009) 537–561.

[15] K. Graichen, Feedforward control design for finite-time transition problems of nonlinear systems with input and output constraints, Doctoral Thesis, Universität Stuttgart, Shaker Verlag, Aachen (Germany), 2006. Available at: http://elib.uni-stuttgart.de/opus/volltexte/2007/3004.

[16] K. Graichen, M. Zeitz, Feedforward control design for finite-time transition problems of nonlinear systems with input and output constraints, IEEE Transactions on Automatic Control 53 (2008) 1273–1278.

[17] K. Graichen, N. Petit, A. Kugi, Transformation of optimal control problems with a state constraint avoiding interior boundary conditions, in: Proc. 47th IEEE Conf. on Decision and Control, CDC, Cancun, Mexico, 2008, pp. 913–920.

[18] A. Isidori, Nonlinear Control Systems, Springer, London, 1995.

[19] K. Graichen, N. Petit, Constructive methods for initialization and handling mixed state-input constraints in optimal control, AIAA Journal of Guidance, Control, and Dynamics 31 (2008) 1334–1343.

[20] J. Bonnans, T. Guilbaud, Using logarithmic penalties in the shooting algorithm for optimal control problems, Optimal Control Applications and Methods 24 (2003) 257–278.

[21] M. Wright, Interior methods for constrained optimization, Acta Numerica 1 (1992) 341–407.

[22] L. Lasdon, A. Waren, R. Rice, An interior penalty method for inequality constrained optimal control problems, IEEE Transactions on Automatic Control 12 (1967) 388–395.

[23] G. Allaire, Numerical Analysis and Optimization, Oxford University Press, New York, 2007.

[24] U. Ascher, R. Mattheij, R. Russell, Numerical Solution of Boundary Value Problems of Ordinary Differential Equations, Prentice Hall, 1988.

[25] J. Kierzenka, L. Shampine, A BVP solver based on residual control and the MATLAB PSE, ACM Transactions on Mathematical Software 27 (2001) 299–316.

[26] R. Goddard, A method for reaching extreme altitudes, Smithsonian Miscellaneous Collections 71.

[27] P. Tsiotras, H. Kelley, Drag-law effects in the Goddard problem, Automatica 27 (1991) 481–490.

[28] H. Seywald, E. Cliff, Goddard problem in presence of a dynamic pressure limit, AIAA Journal of Guidance, Control, and Dynamics 16 (1992) 776–781.

[29] R. Rugescu, Goddard's 85 years optimal ascent problem finally solved, in: 57th International Astronautical Congress, IAC-06-E4.3.05, Valencia, Spain, 2006.

[30] O. von Stryk, R. Bulirsch, Direct and indirect methods for trajectory optimization, Annals of Operational Research 37 (1992) 357–373.

[31] C. Martell, J. Lawton, Adjoint variable solutions via an auxiliary optimization problem, AIAA Journal of Guidance, Control, and Dynamics 18 (1995) 1267–1271.

[32] H. Seywald, R. Kumar, Method for automatic costate calculation, AIAA Journal of Guidance, Control, and Dynamics 19 (1996) 1252–1261.

[33] H. Shena, P. Tsiotras, Time-optimal control of axi-symmetric rigid spacecraft with two controls, AIAA Journal of Guidance, Control, and Dynamics 22 (1999) 682–694.