Comput Optim Appl (2016) 64:295–326
DOI 10.1007/s10589-015-9800-2

Discretization of semilinear bang-singular-bang control problems

Ursula Felgenhauer[1]

Received: 25 September 2014 / Published online: 29 October 2015

Abstract Bang-singular controls may appear in optimal control problems where the control enters the system linearly. We analyze a discretization of the first-order system of necessary optimality conditions written in terms of a variational inequality (or: inclusion) under appropriate assumptions including second-order optimality conditions. For the so-called semilinear case, it is proved that the discrete control has the same principal bang-singular-bang structure as the reference control and, in L_1 topology, the convergence is of order one w.r.t. the stepsize.

Keywords Bang-singular control structure · Approximation of extremals · Euler method · L_1 error estimate

Mathematics Subject Classification 49M25 · 49M05 · 49J30

1 Introduction

The numerical solution of optimal control problems where the control function is of bang-bang or bang-singular type, often is a challenging problem. In the literature, one can find numerous examples of successful treatment for various applications, e.g. in [30,31,33,47,48]. More references are provided in [6,14].

The approaches include direct discretizations as well as indirect methods like shooting approaches, partly in combination with so-called arc–parameterization methods; cf. [6,28,34] for examples. The latter usually require an initial hypothesis on the prin-

✉ Ursula Felgenhauer
felgenh@b-tu.de

1 Institut für Angewandte Mathematik und Wissenschaftliches Rechnen, Brandenburgische Technische Universität Cottbus – Senftenberg, PF 101344, 03013 Cottbus, Germany

cipal control structure, i.e., the number and kind of subsequent bang-bang or singular control arcs. It was this fact which has aroused re-newed interest in classical discretization methods like Euler or Runge–Kutta schemes and their use for detecting control structures.

For optimal control problems with continuous control regimes, convergence results are found e.g. in [12,13,32]. In case of bang-bang control behavior, first convergence results have been obtained for Runge–Kutta methods in [45,46]. Recently, convergence of the Euler method has been obtained under appropriate bang-bang regularity and second-order optimality conditions in [1–4]. An extension to bang-bang switches of weaker regularity is addressed in [24].

Concerning the theoretical analysis of numerical methods in bang-singular control case, important advances have been achieved in deriving optimality conditions [5, 9,20,21,37–39,44], but also in stability investigation of solutions [14,15,40]. For the latter, the shooting method has been utilized e.g. in [47], and its convergence in combination with an arc parameterization was proven in [6]. An example for using Euler's discretization can be found in [1] but the error analysis therein was mainly taylored for bang-bang type optimal controls and remained fragmentary for the bang-singular case. With the present paper, we make an attempt to show that a particular Euler discretization of the first-order system of necessary optimality conditions (cf. [15]) converges, and enables us to detect the control structure with an accuracy of order one of mesh size. The analysis is so far restricted to scalar-valued controls, fixed initial states, and to semilinear state systems (where the control input enters the system linearly with constant coefficients).

1.1 Plan of the paper

After the Introduction, the problem and main assumptions are stated in Sect. 2. The conditions include growth restrictions on problem data, a structural assumption, and certain second-order optimality condition. Section 3 describes the discretized problem in variational inequality form including a right-hand side (*rhs*) perturbation, and analyzes the existence and uniqueness of solutions. Using an appropriate topology in the discrete function space, Robinson's concept of strongly regular generalized equations [41–43] is applied in Sect. 4. The main Lipschitz stability result (Theorem 1) is obtained by adapting the fixpoint iteration from [42] to the problem class. In Sect. 5, the properties of the controls are reconsidered in case when the *rhs* term vanishes. For sufficiently small stepsize parameters h, the discrete control is proved to have the same principal bang-singular-bang structure as the exact solution of the continuous problem. In $L_\infty \times L_1$ topology, the discrete state-control pairs converge to the reference solution as $\mathrm{O}(h)$ (see Theorem 2). Finally, a numerical example illustrates the applicability of the approach.

1.2 Notation

Let be given $N \gg 1$. With $h = 1/N$, find the set of equidistant nodes $t_i = ih$, $i = 0, \ldots, N$, on $[0, 1]$. A (multi-)vector $v = (v_0, \ldots, v_{N-1}) \in \mathbb{R}^{kN}$, $v_i \in R^k$, can be given the interpretation of a piecewise continuous function with constant values v_i on the interval $\omega_i = (t_i, t_{i+1})$ whereas $y = (y_0, \dots, y_N) \in \mathbb{R}^{k(N+1)}$ corresponds, e.g., to a piecewise linear function with nodes $y(t_i) = y_i$. Thus, the related function spaces denoted by $V^h \subset L_\infty$, $Y^h \subset W^1_\infty$ are finite-dimensional and their elements will be identified with the coefficient vectors $v \in \mathbb{R}^{kN}$, $y \in \mathbb{R}^{k(N+1)}$. In $Y^h \subset L_\infty \subset L_2$, introduce the following norms

$$\|y\|_\infty = \max_{0\le i\le N} |y_i|, \qquad \|y\|_{(2)} = \left(|y_N|^2 + h \sum_{i=0}^{N-1} |y_i|^2 \right)^{1/2}. \tag{1}$$

Analogously, set $\|v\|_\infty = \max_{0\le i\le N-1} |v_i|$, $\|v\|_p = \left(h \sum_{i=0}^{N-1} |v_i|^p\right)^{1/p}$ for $v \in V^h \subset L_\infty \subset L_p$, $1 \le p < \infty$. In the above definitions, $|\cdot|$ stands for the standard Euclidean vector norm in $\mathbb{R}^k$. If functions are restricted to a certain interval $I \subset [0, 1]$, the modified norms are written as $\|\cdot\|_{(p,I)}$.

The scalar product in V^h will be defined as the related L_2 scalar product by $(v, v') := h \sum_{i=0}^{N-1} v_i^T v_i'$ with the matrix notation $a^T b$ for the standard scalar product of two column vectors $a, b \in \mathbb{R}^k$. Finally, for $y \in \mathbb{R}^{N+1}$ define the finite difference operators $\Delta^1 y \in \mathbb{R}^N$ and $\Delta^2 y \in \mathbb{R}^{N-1}$,

$$(\Delta^1 y)_i = h^{-1}(y_{i+1} - y_i), \quad (\Delta^2 y)_i = h^{-2}(y_{i+1} - 2y_i + y_{i-1}), \quad i \le N-1.$$

2 Statement of the problem: assumptions

Consider semilinear control problems with scalar-valued control input and prescribed initial state,

(CP)

$$\text{minimize} \quad J(x, u) := k(x(1)) \tag{2}$$

subject to

$$\dot{x}(t) = f(x(t)) + B\,u(t) \quad \text{a.e. in } [0, 1], \tag{3}$$

$$x(0) = a, \tag{4}$$

$$\underline{\alpha} \le u(t) \le \overline{\alpha}, \quad \text{a.e. in } [0, 1], \tag{5}$$

$$x \in W^1_\infty(0, 1; \mathbb{R}^n), \quad u \in L_\infty(0, 1; \mathbb{R}). \tag{6}$$

For sake of simplicity, we assume $\underline{\alpha} = -1$, $\overline{\alpha} = 1$. Notice that, without changing the problem type, general control bounds transform easily to this case, and all results will be invariant w.r.t. possible control rescaling.

Assume that the data functions satisfy

(H0) (smoothness assumption)

The functions f, k depend smoothly on $x \in \mathbb{R}^n$: $f, k \in C^{3,1}$, and there exist constants c, L independent of x such that

$$|f(x)| \le c + L\,|x| \qquad \text{for all } x \in \mathbb{R}^n.$$

Moreover, $B \in \mathbb{R}^n$ is constant.

Define the Pontryagin function $H : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$ by

$$H(x, u, p) = p^T f(x) + p^T B\,u.$$

Then, for the given normal case, Pontryagin's Maximum Principle can be expressed in form of a variational inequality (or: inclusion)

(VI)

$$\dot{x} - f(x) - Bu = 0, \quad x(0) - a = 0, \tag{7}$$

$$\dot{p} + \nabla_x f(x)^T p = 0, \quad p(1) + \nabla k(x(1)) = 0, \tag{8}$$

$$B^T p - \mu_1 + \mu_2 = 0, \tag{9}$$

$$u - 1 \in N_+(\mu_1), \qquad -u - 1 \in N_+(\mu_2) \tag{10}$$

for almost every $t \in [0, 1]$. The set N_+ stands for the normal cone to $\mathbb{R}_+$. Conditions (9), (10) can be equivalently written in terms of the switching function $\sigma(\cdot) = B^T p(\cdot)$ and its positive resp. negative parts $[\sigma]_\pm$: for $\mu_1 = [\sigma]_+$, $\mu_2 = [\sigma]_-$ and $\sigma = [\sigma]_+ - [\sigma]_-$, the relations yield the complementarity system

$$[\sigma]_+,\, [\sigma]_- \ge 0, \quad -1 \le u \le 1, \quad (u-1)\,[\sigma]_+ = (u+1)\,[\sigma]_- = 0.$$

Assume that $(x^0, u^0) \in W^1_\infty \times L_\infty$ is a (strong) local minimizer of (CP), i.e., there exists an $\epsilon > 0$ such that

$$J(x, u) - J(x^0, u^0) \ge 0$$

for all admissible pairs (x, u) satisfying $\|x - x^0\|_\infty < \epsilon$.
Then, the adjoint and multiplier functions p^0, μ_1^0, μ_2^0 associated to (x^0, u^0) are uniquely determined by (VI) and belong to W^1_∞. Further, $\sigma^0 = B^T p^0 \in C^1$, and $\dot{\sigma}^0 = -B^T H_x^0 = -B^T \nabla f(x^0)^T p^0 \in W^1_\infty$. For $\ddot{\sigma}^0$ find

$$\ddot{\sigma}^0 = P + u^0 R, \quad P = B^T(\nabla f(x^0)^T H_x^0 - H_{xx}^0 f^0), \quad R = -B^T H_{xx}^0 B, \tag{11}$$

with functions P, R belonging to W^1_∞.

Optimality conditions for (CP) in case of a bang-singular-bang control switching structure have been derived in [5], see also [7–10,26,36,39,44]. Allover the paper, we use the following assumptions:

(H1) (strict structural assumption)

The function u^0 is of *strict* bang-singular-bang structure, i.e., there exist points τ_s, τ_s' with $0 < \tau_s < \tau_s' < 1$ and a positive constant m such that

$$u^0(t) = \begin{cases} u^1 & \text{if} \quad 0 \le t < \tau_s, \\ u_s^0(t) & \text{if} \quad \tau_s < t < \tau_s', \\ u^2 & \text{if} \quad \tau_s' < t \le 1, \end{cases}$$

with constants $u^1, u^2 \in \{-1, 1\}$, and $|u_s^0(t)| \le 1 - m$ for a.a. t on the singular arc $[\tau_s, \tau_s']$. Moreover, $\sigma^0 \neq 0$ on bang arcs $[0, \tau_s) \cup (\tau_s', 1]$.

For given $\beta > 0$, define the sets $J_\beta = \{t \in [0, 1] : |\sigma^0(t)| \ge \beta > 0\}$ and $I_\beta = (0, 1) \setminus J_\beta$. Then the singular arc $[\tau_s, \tau_s']$ is equal to $I_0 = \bigcap_{\beta > 0} I_\beta$.
(H2) (strong second-order optimality condition)
Let $\Omega = \Omega(v, y, b)$ denote the quadratic form

$$\Omega(v, y, b) = z(1)^T K\, z(1) + \int_0^1 z(t)^T Q(t) z(t)\, dt \tag{12}$$

with $K = \nabla^2 k(x^0(1))$, $Q = \nabla_{xx}^2 H(x^0, u^0, p^0)$, and $v \in L_2(0, 1; \mathbb{R})$, $y \in W_2^1(0, 1; \mathbb{R})$, $z \in W_2^1(0, 1; \mathbb{R}^n)$ satisfying

$$\begin{aligned} y(t) &= \int_0^t v(s)\, ds, \qquad y(1) =: b, \\ \dot z(t) &= A(t) z(t) + B v(t), z(0) = 0, \qquad A = \nabla f(x^0). \end{aligned} \tag{13}$$

There exist constants $\beta > 0$, $m > 0$ such that

$$\Omega(v, y, b) \ge m \left(\|y\|_2^2 + b^2 \right)$$

for all $v \in L_2$ with $v = 0$ on J_β, and $(y, b) \in W_2^1 \times \mathbb{R}$ given by (13).

(The formulation follows [15] but has been adapted to the special case of a semilinear state equation where both $\nabla_{uu}^2 H$ and $\nabla_{xu}^2 H$ vanish.) Without loss of generality, one can use the same constant m in both (H1), (H2) when shrinking it if necessary.

In [5,9] it was proved that, under conditions (H1), (H2), the extremal (x^0, u^0) is a strict Pontryagin minimum [35] for (CP). Moreover, the following relations hold:

$$R \ge m > 0 \qquad \text{on} \quad I_\beta, \tag{14}$$

$$|\ddot\sigma(\tau_s - 0)| \ge m^2 > 0, \quad |\ddot\sigma(\tau_s' + 0)| \ge m^2 > 0. \tag{15}$$

Indeed, (14) had been proved in [14, Sect. 3.1]; cf. also [37]. While $\ddot\sigma(\tau_s + 0) = 0$, conclude $|\ddot\sigma(\tau_s - 0)| = |R \cdot (u^0(\tau_s + 0) - u^0(\tau_s + 0))| \ge m^2$ from (11), and from jump estimates on u^0 implied by (H1). Hence (15) is obtained, too.

It follows that the control u^0 has a singular arc of *order one* (cf. [29,49]) on $[\tau_s, \tau_s']$ and coincides there with $u_s^0 = -P/R$. If β is taken sufficiently small then, in addition, I_β is an open interval.

3 Variational inequality discretization

Let the interval [0, 1] be divided into subintervals $\omega_i = [t_i, t_{i+1}]$, $0 \le i < N$, of equal length $h = 1/N$, $N \gg 1$. Consider the following discretization of (VI):

(VI$^h_\delta$)

$$\delta_{1i} \in (\Delta^1 x)_i - f(x_i) - B\,u_i + \{0\}, \quad 0 = x_0 - a, \tag{16}$$

$$\delta_{2i} \in (\Delta^1 p)_i + \nabla f(x_{i+1})^T p_{i+1} + \{0\}, \delta_{2N} = p_N + \nabla k(x_N), \tag{17}$$

$$\delta_{3i} \in B^T p_i - \mu_{1i} + \mu_{2i} + \{0\}, \tag{18}$$

$$0 \in 1 - u_i + N_+(\mu_{1i}), \qquad 0 \in 1 + u_i + N_+(\mu_{2i}), \tag{19}$$

for $i = 0, \dots, N-1$. The formulation includes a right-hand side (or: *rhs*) perturbation $\delta = (\delta_1, \delta_2, \delta_3)$ of dimension $l = n(2N+1) + N$.
For $\xi = (p, x, u, \mu) \in C = \mathbb{R}^{n(N+1)} \times \mathbb{R}^{n(N+1)} \times \mathbb{R}^N \times \mathbb{R}^{2N}_+$ and given $\delta \in \mathbb{R}^l$, (VI$^h_\delta$) can be written in abstract form as

$$\bar{\delta} \in \psi(\xi) + N_C(\xi), \tag{20}$$

where $N_C(\xi)$ denotes the normal cone to the set C at point ξ, and δ is completed to $\bar{\delta}$ by zero entries $\delta_{1N} \in R^n$ and $\delta_4 \in R^{2N}$. The dimension of the vectors ξ and $\bar{\delta}$ is $d = 2n(N+1) + 3N$. For each $h = 1/N$, $N > 1$, it follows from (H0) that $\psi : \mathbb{R}^d \to \mathbb{R}^d$ is at least twice continuously differentiable.

Lemma 1 *For each $N > 1$, $h = 1/N$, and arbitrary $\delta \in R^l$, (VI$^h_\delta$) has a solution $\xi \in C$. There exists a constant $\bar{r} > 0$ independent of h and δ, such that ξ is bounded by $\|\xi\|_\infty \le \bar{r}(1 + \|\delta\|_\infty)$. Moreover, $\|u\|_\infty + \|x\|_\infty + \|p\|_\infty \le \bar{r}(1 + \|\delta_1\|_2 + \|\delta_2\|_{(2)})$ in the norms given by* (1).

Proof Any prospective solution of the variational inequality (20) satisfies $\|u\|_\infty \le 1$. Under Assumption (H0), for $i \ge 0$

$$|x_{i+1}| \le |x_i| + h\,|f(x_i) + Bu_i + \delta_{1i}| \le (1 + Lh)|x_i| + h\,(c + |B| + |\delta_{1i}|)$$

so that, in analogy to Gronwall's Lemma [23] (or Lemma 14, Appendix), the estimate

$$|x_i| \le e^L\,(|a| + c + |B|) + e^L \|\delta_1\|_2 \le \bar{r}_x(1 + \|\delta_1\|_2)$$

follows. All constants herein are independent of h. Further, from (17) we backwards obtain similar bounds for p. In order to estimate μ, it should be noticed that (19) yields $\mu_1, \mu_2 \ge 0$ and $\mu_{1i} \cdot \mu_{2i} = 0$ for all i. Consequently, $\mu_{1i} = \left[B^T p_i - \delta_{3i}\right]_+$, $\mu_{2i} = \left[B^T p_i - \delta_{3i}\right]_-$ are bounded, e.g., by $\bar{r}_\mu(1 + \|\delta\|_\infty)$ with constant r_μ independent of h, and the desired estimate for ξ follows.
Using the above information, for given h, δ one can restrict (VI$^h_\delta$) or (20) to $\xi \in C_\delta = C \cap \{\xi : \|\xi\|_\infty \le \bar{r}(1 + \|\delta\|_\infty)\} \subset \mathbb{R}^d$. As far as C_δ is non-empty, convex and compact, and ψ as an operator from C_δ to $\mathbb{R}^d$ is continuous, one can apply the existence result from Theorem 3.1 [27, Chap. 1] for variational inequalities in finite

dimensions: for compact sets, the assertion follows directly by Brouwer's Fixed-Point Theorem. □

Using the solution (x^0, u^0, p^0, μ^0) of the continuous problem (VI), one can find a reference solution for (VI^h_δ) with appropriately chosen *rhs* term $\delta = \tilde{\delta}^h$. To this aim, consider piecewise linear interpolations $\tilde{x}^h$, $\tilde{p}^h$ with node values $\tilde{x}^h(t_i) = \tilde{x}^h_i = x^0(t_i)$, $\tilde{p}^h(t_i) = p^0(t_i)$ and constant $\dot{\tilde{x}}^h$, $\dot{\tilde{p}}^h$ on each interval $\omega_i,\ i = 0, \ldots, N-1$. Further, define $\tilde{u}^h$ as as a piecewise continuous function with constant values $\tilde{u}^h_i$ on $\omega_i,\ i = 0, \ldots, N-1$, by

$$\tilde{u}^h\Big|_{\omega_i} \equiv \tilde{u}^h_i = h^{-1}\int_{\omega_i} u^0(t)\,dt. \tag{21}$$

If we denote by k, k' indices such that $t_k \le \tau_s < t_{k+1},\ t_{k'} < \tau'_s \le t_{k'+1}$ then $\tilde{u}^h_i$ takes constant extremal values u^1 (or u^2 resp.) for $i \le k-1$ (or $i \ge k'+1$), but has singular values $\tilde{u}^h_i \in (0, 1)$ for $k \le i \le k'$.
For $S_\tau := \{k, k+1, k', k'+1\}$, introduce

$$\|\cdot\|_{(\infty,\tau)} = \max\{|\cdot|_i : i \notin S_\tau\}. \tag{22}$$

Lemma 2 *Let $h = 1/N$ be sufficiently small. Then the functions $\tilde{x}^h, \tilde{u}^h$ and $\tilde{p}^h$ together with $\tilde{\mu}^h$ given by*

$$\tilde{\mu}^h_{1i} = \left[\tilde{\sigma}^h_i\right]_+, \quad \tilde{\mu}^h_{2i} = \left[\tilde{\sigma}^h_i\right]_-$$

with the definition

$$\tilde{\sigma}^h_i = \sigma^\tau(t_i), \qquad \sigma^\tau(t) = \begin{cases} \sigma^0(t + \tau_s - t_k) & \text{if} \quad t \le t_k, \\ 0 & \text{if} \quad t_k < t \le t_{k'}, \\ \sigma^0(t + \tau'_s - t_{k'}) & \text{if} \quad t > t_{k'}, \end{cases}$$

solve (VI^h_δ) for some $\delta = \tilde{\delta}^h$ satisfying $|\tilde{\delta}^h_{1i}| + |\tilde{\delta}^h_{2i}| + |\tilde{\delta}^h_{3i}| = \mathrm{O}(h)$, together with $\tilde{\delta}^h_{2N} = 0$. Moreover, $|(\Delta^1\tilde{\delta}^h_3)_i| = \mathrm{O}(h)$ uniformly for $i < N$, $\|\Delta^2\tilde{\delta}^h_3\|_2 = \mathrm{O}(\sqrt{h})$ and $\|\Delta^1\tilde{\delta}^h_2\|_{(\infty,\tau)} + \|\Delta^2\tilde{\delta}^h_3\|_{(\infty,\tau)} = \mathrm{O}(h)$.

The proof is given in Appendix.

Lemma 3 *Under Assumptions* (H1) *and* (H2)*, the discrete switching function $\tilde{\sigma}^h$ satisfies $\|\Delta^1\tilde{\sigma}^h\|_2 + \|\Delta^2\tilde{\sigma}^h\|_2 \le c$ with a constant c independent of h. Further,*

$$(\Delta^2\tilde{\sigma}^h)_i = \tilde{P}^h_i + R_i\tilde{u}^h_i$$

for some $\tilde{P}^h$ with the property $\|\tilde{P}^h - P\|_{(\infty,\tau)} = \mathrm{O}(h)$, and $(P_i, R_i) = (P(t_i), R(t_i))$.

Proof By direct calculation, for $\Delta^2\tilde{\sigma}^h$ obtain

$$(\Delta^2\tilde{\sigma}^h)_i = \begin{cases} (\Delta^2\sigma^\tau)_i & \text{if} \quad i \le k-1 \;\text{ or }\; i \ge k'+1, \\ \ddot{\sigma}^0(t_i) & \text{if} \quad k+1 \le i \le k'-1, \\ h^{-2}\sigma^0(\tau_s - h) & \text{if} \quad i = k, \\ h^{-2}\sigma^0(\tau_s' + h) & \text{if} \quad i = k'. \end{cases}$$

Therefore, for all $i \notin \{k, k'\}$, we can write $(\Delta^2\sigma^\tau)_i = \ddot{\sigma}^0(\theta_i)$ for some θ_i satisfying $|\theta_i - t_i| = \mathrm{O}(h)$ so that

$$(\Delta^2\sigma^\tau)_i = P(\theta_i) + R(\theta_i)u^0(\theta_i).$$

A closer look on the possible localization of θ_i yields

$$u^0(\theta_i) = \tilde{u}_i^h = u^1 \;\text{ for }\; i \le k-1, \qquad u^0(\theta_i) = \tilde{u}_i^h = u^2 \;\text{ for }\; i \ge k'+1,$$

and $u^0(\theta_i) = u_s^0(t_i) = -P_i/R_i,\; k+1 \le i \le k'-1$. For the latter, it follows from $x^0, p^0, P, R \in W_\infty^1$ and condition (14) that $|u_s^0(t_i) - \tilde{u}_i^h| = \mathrm{O}(h)$. Using further $|R(\theta_i) - R_i| = \mathrm{O}(h)$ and the boundedness of R, the desired representation and estimates are obtained. □

4 Solution uniqueness and estimates for $(\mathbf{VI}_\delta^h)$

4.1 Linearized VI problem

Let the functions $\tilde{x}^h$, $\tilde{p}^h$, $\tilde{u}^h$ and $\tilde{\sigma}^h$ be given as in Sect. 3, (21) ff. For $\xi = (p, x, u, \nu)$, the following linearized variational inequality related to (VI_δ^h) will be considered:

$$\bar{\delta} \in T(\xi) + N_C(\xi) \tag{23}$$

where $T : \mathbb{R}^d \to \mathbb{R}^d$ is the linearization of $(\psi - \bar{\delta})$ at $(\tilde{\xi}^h, \tilde{\delta}^h)$,

$$T(\xi) = \psi(\tilde{\xi}^h) + \psi'(\tilde{\xi}^h)(\xi - \tilde{\xi}^h) - \bar{\delta}^h, \tag{24}$$

and $\bar{\delta}^h \in \mathbb{R}^d$ is the extension by zeros related to $\tilde{\delta}^h$.
Denoting $z_i = x_i - \tilde{x}_i^h$, $q_i = p_i - \tilde{p}_i^h$ and $v_i = u_i - \tilde{u}_i^h$, in detail the relations read as follows

$(\mathbf{LVI}_\delta^h)$

$$\delta_{1i} \in (\Delta^1 z)_i - A_i z_i - B v_i + \{0\}, \quad z_0 = 0, \tag{25}$$

$$\delta_{2i} \in (\Delta^1 q)_i + A_{i+1}^T q_{i+1} + Q_{i+1} z_{i+1} + \{0\}, \delta_{2N} = q_N + K z_N, \tag{26}$$

$$\delta_{3i} \in B^T q_i + \tilde{\sigma}_i^h - \nu_{1i} + \nu_{2i} + \{0\}, \tag{27}$$

$$\tilde{u}_i^h + v_i - 1 \in N_+(\nu_{1i}), \qquad -\tilde{u}_i^h - v_i - 1 \in N_+(\nu_{2i}), \tag{28}$$

for $i = 0, \ldots, N-1$. The coefficient matrices are

$$A_i = \nabla f(x^0(t_i)), \qquad Q_i = \nabla^2_{xx} H^0[t_i], \qquad K = \nabla^2 k(x^0(1)).$$

Notice that, for $\delta = 0$, the system (LVI$^h_\delta$) has the solution $(q, z, v, \nu) = (0, 0, 0, \tilde{\mu}^h)$. In general, every solution (z, q) of (25), (26) can be represented as

$$z = S_1 v + z^h_\delta, \quad q = S_2 v + q^h_\delta,$$

where $(S_1 v, S_2 v)$ are the solutions for $\delta = 0$ written in terms of linear solution maps $S_{1,2}$, but (z^h_δ, q^h_δ) are related particular solutions for $v = 0$, and given $\delta_1 \in \mathbb{R}^{Nn}$, $\delta_2 \in \mathbb{R}^{(N+1)n}$, $\delta_3 \in \mathbb{R}^N$ and $q_N = \delta_{2N} - K z^h_{\delta,N}$. Inserting the expressions into (27) leads to

$$B^T q_i - \delta_{3i} =: -(C^h v)_i + r^h_i(\delta) \tag{29}$$

with $r^h_i(\delta) = B^T q^h_{\delta,i} - \delta_{3i}$ and $(C^h v)_i = -B^T (S_2 v)_i$.

Lemma 4 *Every solution (q, z, v, ν) of (LVI$^h_\delta$) solves*

(LVI') *find* $v \in W^h : \ \Big((C^h v)_i - \tilde{\sigma}^h_i - r^h_i(\delta),\ v'_i - v_i\Big) \geq 0 \ \ \forall\, v' \in W^h$

for $i = 0, \ldots, N-1$ *on* $W^h = \{v' \in \mathbb{R}^N : -1 \leq v'_i + \tilde{u}^h_i \leq 1,\ i = 0, \ldots, N-1\}$. *If, in addition,* $|\nu_{1i} - \tilde{\mu}^h_{1i}| + |\nu_{2i} - \tilde{\mu}^h_{2i}| < \beta/2$ *on* J_β*, and h is sufficiently small, then* v *solves* (LVI') *on* $W^h_\beta = \{v' \in W^h : \ v'_i = 0 \text{ for } t_i \in J_\beta\}$ *where* $J_\beta = \{t \in [0, 1] : \sigma^0(t) \geq \beta\}$*, and* β *is the constant from* (H2).

The proof is a direct consequence of the complementarity relations (28).

4.2 Discrete Goh transformation

Define the vector $(y, \zeta) \in \mathbb{R}^{N+1} \times \mathbb{R}^{n(N+1)}$ by

$$(\Delta^1 y)_i = v_i = h^{-1} \int_{\omega_i} v(t)\, dt, \quad y_0 = 0, \tag{30}$$

$$\zeta_i = z_i - B y_i. \tag{31}$$

Inserting the expressions into (25), (26) yields

$$\begin{aligned}
(\Delta^1 \zeta)_i &= (\Delta^1 z)_i - B(\Delta^1 y)_i \\
&= A_i z_i + \delta_{1i} \\
&= A_i \zeta_i + B^1_i y_i + \delta_{1i} \qquad \text{with } B^1_i = A_i B;
\end{aligned} \tag{32}$$

$$\begin{aligned}
(\Delta^1 q)_i &= -A^T_{i+1} q_{i+1} - Q_{i+1} z_{i+1} + \delta_{2i} \\
&= -A^T_{i+1} q_{i+1} - Q_{i+1} \zeta_{i+1} - M^T_{i+1} y_{i+1} + \delta_{2i} \\
&\qquad \text{with } \ M^T_i = Q_i B.
\end{aligned} \tag{33}$$

The boundary conditions transform into

$$\zeta_0 = 0, \quad y_N =: b, \quad q_N = -K\,(\zeta_N + B\,b) + \delta_{2N}. \tag{34}$$

The solutions of the transformed linearized state-adjoint system (32), (33) for $\delta = 0$ can be expressed by $\zeta = Sy,\ q = \tilde{S}y + \tilde{W}b$ where $S,\ \tilde{S}$ are linear solution operators to (32), (33). Thus, (27) turns into

$$-\,(\hat{C}^h y)_i + \tilde{\sigma}_i^h - \nu_{1i} + \nu_{2i} + r_i^h(\delta) = 0 \tag{35}$$

with the abbreviation $(\hat{C}^h y)_i = -B^T(\tilde{S}y + \tilde{W}b)_i$, and the function r^h given by (29).

Lemma 5 *Let the Assumptions* (H0)–(H2) *hold. If* $h < \bar{h}$ *with* $\bar{h}$ *sufficiently small then* $\Omega^h(y) := \big(C^h v,\ v\big)$ *satisfies*

$$\Omega^h(y) \geq \frac{m}{2}\left(\|y\|_2^2 + |y_N|^2\right) \tag{36}$$

for all $v \in \mathbb{R}^N,\ y \in \mathbb{R}^{N+1}$ *such that* (30) *holds together with* $v_i = 0$ *for* $t_i \in J_\beta$.

(As far as the constant m is taken from (H2) it is independent of N resp. h.)

Proof Denote by $\Psi,\ \Phi$ the discrete fundamental solutions of (32), (33),

$$\Psi_{i+1} - \Psi_i = hA_i\Psi_i, \quad \Phi_{i+1} - \Phi_i = -hA_{i+1}^T\Phi_{i+1}, \quad \Psi_0 = \Phi_{-1} = I.$$

Notice that, for all i, $\Phi_{i-1}^T\Psi_i = \Psi_i^T\Phi_{i-1} = I$. Then, for the discrete data $\zeta,\ q$ explicit formulas can be given:

$$\begin{aligned}
\zeta_i &= (Sy)_i = h\,\Psi_i \sum_{k=0}^{i-1} \Phi_k^T B_k^1 y_k, \\
q_i &= (\tilde{S}y)_i + \tilde{W}_i b \\
&\quad = -\Phi_i\Psi_{N+1}^T K(\zeta_N + Bb) + h\,\Phi_i \sum_{k=i}^{N-1} \Psi_{k+1}^T(Q_{k+1}\zeta_{k+1} + M_{k+1}^T y_{k+1}).
\end{aligned}$$

Consequently, for some positive c independent of h,

$$\|\zeta\|_\infty + \|q\|_\infty \leq c\,(\|y\|_2 + |b|)\,.$$

In order to prove the assertion of the lemma, start with $(\hat{C}^h y)_i = (C^h v)_i = -B^T q_i$; cf. (35). Denoting $\tilde{y}_i = h\sum_{k=0}^{i-1}\tilde{v}_k,\ i = 0, \ldots, N$, the quadratic form transforms as follows:

$$\begin{aligned}\left(C^h v, \tilde{v}\right) &= -h \sum_{i=0}^{N-1} B^T q_i (\Delta^1 \tilde{y})_i = -\tilde{b} B^T q_N + h \sum_{i=1}^{N} \tilde{y}_i B^T (\Delta^1 q)_{i-1} \\ &= \tilde{b} B^T K(\zeta_N + Bb) - h \sum_{i=1}^{N} \tilde{y}_i B^T (A_i^T q_i + Q_i \zeta_i + M_i^T y_i) \\ &= \tilde{b} B^T K(\zeta_N + Bb) + h \sum_{i=1}^{N} \tilde{y}_i (R_i y_i - M_i \zeta_i - (B_i^1)^T q_i).\end{aligned}$$

In the last sum on the right, one can substitute the explicit representation for q_i and obtain

$$\begin{aligned}-h \sum_{i=1}^{N} \tilde{y}_i (B_i^1)^T q_i &= h \sum_{i=1}^{N} \tilde{y}_i (B_i^1)^T \Phi_i \Psi_{N+1}^T K(\zeta_N + Bb) \\ &\quad - h^2 \sum_{i=0}^{N-1} \sum_{k=i}^{N-1} \tilde{y}_i (B_i^1)^T \Phi_i \Psi_{k+1}^T (Q_{k+1} \zeta_{k+1} + M_{k+1}^T y_{k+1}) \\ &= \tilde{\zeta}_{N+1}^T K(\zeta_N + Bb) - h \sum_{k=0}^{N-1} \tilde{\zeta}_{k+1}^T (Q_{k+1} \zeta_{k+1} + M_{k+1}^T y_{k+1}) \\ &= \tilde{\zeta}_N^T K(\zeta_N + Bb) + \pi_N(y, \tilde{y}) - h \sum_{i=1}^{N} \tilde{\zeta}_i^T (Q_i \zeta_i + M_i^T y_i).\end{aligned}$$

For $\tilde{y} = y$, the error term $\pi_N = (\zeta_{N+1} - \zeta_N)^T K(\zeta_N + Bb)$ is bounded by

$$|\pi_N| \le h\, c_\Omega \left(\|y\|_2^2 + |b|^2\right)$$

with some positive constant $c = c_\Omega$ independent of h.
The sum over the remaining expressions corresponds to a standard integral sum for the discretization of $\Omega(y, b) = (Cv, v)$ from (H2). Therefore, the estimate (36) follows directly from this second-order condition if only h is taken sufficiently small. □

For our further analysis it will be useful to restrict the *rhs* terms δ to a certain neighborhood of $\delta_0 = 0$. To this aim, for $\delta = (\delta_1, \delta_2, \delta_3)$ and given $\rho > 0$ define

$$\|\delta\|_D^2 := \|\delta_1\|_2^2 + \|\delta_2\|_{(2)}^2 + \|\delta_3\|_2^2 + \|\Delta^1 \delta_3\|_2^2, \tag{37}$$

$$D_\rho = \{\delta \in \mathbb{R}^l : \|\delta\|_D < \rho\}. \tag{38}$$

Lemma 6 *Under assumptions* (H0)–(H2)*, there exist constants* $\rho > 0$ *and* $h' < \bar{h}$ *such that, for each* $\delta \in D_\rho$ *and* $h \le h'$*, the variational inequality* (LVI') *restricted* *to* W_β^h *has a unique solution* $v^\delta \in R^N$. *On* D_ρ, *the elements* y^δ *defined by* (30) *with* $v = v^\delta$ *satisfy*

$$\|y^{\delta'} - y^\delta\|_2 + |y_N^{\delta'} - y_N^\delta| \le c\|\delta' - \delta\|_D, \tag{39}$$

and the constants c, ρ *are independent of* h. *Moreover, there exist unique multipliers* $\nu_{1,2}^\delta$ *such that* $\xi^\delta = (z^\delta, q^\delta, v^\delta, \nu^\delta)$ *(with* (z^δ, q^δ) *defined by* (25), (26) *for* $v = v^\delta$*) is a solution of* (LVI$_h^\delta$).

Proof The existence and uniqueness of v^δ is a consequence of the continuity and strict monotonicity of C^h together with the boundedness of the closed convex set W_β^h in $\mathbb{R}^N$, resp. the finite-dimensional subspace V^h; cf. [27].

It remains to prove the Lipschitz continuity of $y = y^\delta$ w.r.t. δ: First notice that δ_3 can be interpreted as a piecewise linear function belonging to $W_2^1(0, 1; \mathbb{R})$. By the Theorem of Morrey [11, Theorem 6.25], on intervals the embedding $W_2^1 \to L_\infty$ is continuous. Consequently, for r^h the estimate

$$\|\Delta^1 r^h(\delta') - \Delta^1 r^h(\delta)\|_2 + \|r^h(\delta') - r^h(\delta)\|_\infty \le c'\,\|\delta' - \delta\|_D$$

(with some $c' > 0$ independent of h) follows directly from (29) and the definition of (z_δ^h, q_δ^h). With the abbreviations $v = v^\delta$, $v' = v^{\delta'}$ etc., from (LVI') further obtain

$$\begin{aligned}
(C^h(v' - v), v' - v) &\le \Big(r^h(\delta') - r^h(\delta), \Delta^1(y' - y)\Big) \\
&= (r_{N-1}^h(\delta') - r_{N-1}^h(\delta))(y_N' - y_N) \\
&\qquad - h\sum_{i=1}^{N-1}\Big(\Delta^1(r^h(\delta') - r^h(\delta))\Big)_{i-1}(y_i' - y_i) \\
&\le c\|\delta' - \delta\|_D\left(|y_N' - y_N| + \|y' - y\|_2\right).
\end{aligned}$$

Together with (36), the desired Lipschitz property (39) of y^δ follows. In particular,

$$\|y^\delta\|_2 + |y_N^\delta| \le c\,\|\delta\|_D. \tag{40}$$

Denote $\phi := -C^h v^\delta + \tilde{\sigma}^h + r^h(\delta) \in \mathbb{R}^N$. Without loss of generality, one can write $\phi = \phi^1 - \phi^2$ for some $\phi^1, \phi^2 \ge 0$. From (LVI') the following relation is obtained:

$$(\phi, v - v^\delta) = (\phi^1, w - w^\delta) - (\phi^2, w - w^\delta) \le 0 \tag{41}$$

for $w^\delta = v^\delta + \tilde{u}^h$ and all $w \in \tilde{u}^h + W_\beta^h$. Using $\phi^{1,2}$ as candidates for $\nu_{1,2}^\delta$, it follows directly from (28) that the only possible choice is given by

$$\phi^1 = [\phi]_+, \qquad \phi^2 = [\phi]_-.$$

Then, by (41) we have $(\phi^1, w - w^\delta) \le 0$, $(\phi^2, w - w^\delta) \ge 0$ for all $w \in \tilde{u}^h + W^h_\beta$. Consequently, for all i with $t_i \notin J_\beta$ obtain

$$\phi_i = [\phi_i]_+ > 0 \;\Rightarrow\; w_i^\delta = 1, \qquad \phi_i = [\phi_i]_- > 0 \;\Rightarrow\; w_i^\delta = -1.$$

In case $t_i \in J_\beta$ remember

$$\phi_i \;=\; B^T(\tilde{S}y + \tilde{W}b)_i + \tilde{\sigma}_i^h + r_i^h(\delta).$$

Inserting $y = y^\delta$, $b = y_N^\delta$, it follows that

$$|\phi_i - \tilde{\sigma}_i^h| \;\le\; c(\|y^\delta\|_2 + |y_N^\delta| + \|r^h(\delta)\|_\infty) = \mathrm{O}(\|\delta\|_D);$$

cf. (40). Choosing ρ and h' sufficiently small one can guarantee $|\phi_i| > \beta/2 > 0$ uniformly for $\delta \in D_\rho$, $h < h'$ and $t_i \in J_\beta$ so that ϕ has the same sign as σ^0. Hence, (27), (28) are fulfilled if and only if $\nu_1^\delta = [\phi]_+$, $\nu_2^\delta = [\phi]_-$. □

Lemma 7 *Under Assumptions* (H0)–(H2)*, there exist constants h', ρ and ρ' such that, for all $h < h'$ and $\delta \in D_{\rho'}$, the following statements hold:*

(i) *On the set* $\Xi_\rho = \{\xi = (z, q, v, \nu) : v = \Delta^1 y,\ y_0 = 0,\ \|y\|_2^2 + |y_N|^2 < \rho^2\}$, *the variational inequality (LVI$^h_\delta$) has a unique solution $\xi^h(\delta)$,*

(ii) *the components $(z^h(\delta), q^h(\delta), y^h(\delta), \nu^h(\delta))$ depend Lipschitz continuously on δ in the sense*

$$\max\{\|z^\delta - z^{\delta'}\|_2,\ \|q^\delta - q^{\delta'}\|_2,\ \|y^\delta - y^{\delta'}\|_2,\ \|\nu^\delta - \nu^{\delta'}\|_\infty\} \le c\,\|\delta - \delta'\|_D,$$
$$|\pi z^\delta - \pi z^{\delta'}| + |\pi q^\delta - \pi q^{\delta'}| \;\le\; c\,\|\delta - \delta'\|_D$$

where $\pi : \mathbb{R}^{n(N+1)} \to \mathbb{R}^{2n}$ denotes the boundary trace operator $\pi\phi = (\phi_0, \phi_N)$. The constant c herein does not depend on h or $\delta, \delta' \in D_{\rho'}$.

Proof For $h < h'$ and $\|\delta\|_D < \rho'$ with appropriately chosen ρ', it follows from Lemma 6 that (LVI$^\delta_h$) has a solution $\xi = \xi^\delta$. Due to (40), one can always take ρ' small enough to ensure $\xi \in \Xi_\rho$ for given $\rho > 0$. Thus, part (i) of the proof reduces to verifying the solution uniqueness on the set Ξ_ρ:
Assume for the moment, that (LVI$^\delta_h$) has a second solution $\hat{\xi}^\delta \neq \xi^\delta$, and consider

$$\hat{\phi} \;=\; -\hat{C}^h \hat{y}^\delta + \tilde{\sigma}^h + r^h(\delta).$$

By (35) and standard estimates for (32), (33), obtain

$$\|\hat{C}^h \hat{y}^\delta\|_\infty \le c\left(\|\hat{y}^\delta\|_2 + |\hat{y}_N^\delta|\right) \;=\; \mathrm{O}(\rho).$$

On the other hand, (29) yields

$$\|r^h(\delta)\|_\infty \le c\|\delta\|_D \;=\; \mathrm{O}(\rho').$$

If now ρ, ρ' and h' are sufficiently small then, in particular, $\|\hat{\phi} - \tilde{\sigma}^h\|_\infty < \beta/4$ and thus, for $\hat{\nu}_1^\delta = [\hat{\phi}]_+$, $\hat{\nu}_2^\delta = [\hat{\phi}]_-$ deduce

$$\|\hat{\nu}_1^\delta - \tilde{\mu}_1^h\|_\infty + \|\hat{\nu}_2^\delta - \tilde{\mu}_2^h\|_\infty < \beta/2.$$

Lemma 4 therefore shows that $\hat{\upsilon}^\delta$ solves (LVI') on W_β^h and hence, coincides with υ^δ. By construction of $\hat{z}^\delta$, $\hat{q}^\delta$ and $\hat{\nu}^\delta$ according to (LVI_h^δ) it follows that $\hat{\xi}^\delta = \xi^\delta$, i.e., for $\delta \in D_{\rho'}$ the solution is unique on Ξ_ρ.

Part (ii): The estimates are direct consequences of (39), Lemma 6, together with the construction

$$\begin{aligned} \zeta = Sy, \qquad z &= \zeta + By, \qquad q = \tilde{S}y + \tilde{W}y_N, \\ \nu_1 = [\hat{C}^h y + r^h(\delta)]_+, &\quad \nu_2 = [\hat{C}^h y + r^h(\delta)]_-, \end{aligned}$$

and formula (29) for $r^h(\delta)$. □

4.3 Strong regularity

For variational inequalities of type (20), a fundamental existence and Lipschitz stability result was given by Robinson in [42]. The crucial characterizing property concerns in the strong regularity which, for continuously Fréchet differentiable mappings ψ from certain Banach space Ξ to its dual space, allows to prove existence and local uniqueness of solution, and further its Lipschitz continuous dependence on problem parameters [42, Theorem 2.1]:

Definition 1 [42, *Definition 1*] Let Ξ be a normed linear space, and let Ω be an open subset of Ξ containing a point $\bar{\xi}$. Let C be a closed convex set in Ξ, and let $\psi : \Omega \to \Xi^*$ (where Ξ^* is the topological dual of Ξ) be Fréchet differentiable at $\bar{\xi}$. Suppose that

$$0 \in \psi(\xi) + N_C(\xi) \tag{42}$$

has $\bar{\xi}$ as a solution, and define, for $\xi \in \Xi$,

$$T_C(\xi) := \psi(\bar{\xi}) + \psi'(\bar{\xi})(\xi - \bar{\xi}) + N_C(\xi).$$

We say that (42) is *strongly regular* at $\bar{\xi}$ with associated Lipschitz constant λ if there exist neighborhoods U of the origin in Ξ^*, V of $\bar{\xi}$ in Ξ such that the restriction to U of $T_C^{-1} \cap V$ is a single-valued function from U to V which is Lipschitzian on U with modulus λ.

Remark 1 A closer look on the main steps of the proof in [42] shows that, without loss of generality, the neighborhood U of *rhs* perturbations $\bar{\delta}$ for the linearized problem (23) can be restricted to the subspace $\Xi^0 \subset \Xi^*$ where all $\bar{\delta}$ components corresponding to *linear* components of ψ vanish. In the forthcoming, we will call (42) also strongly regular (in the sense of Robinson) if the definition holds in this slightly relaxed form.

In order to make usage of the strong regularity approach for the variational inequalities (VI_δ^h), the space $\mathbb{R}^d$ will be equipped with a norm reflecting the use of Goh's transformation in deriving (39) and the estimates in Lemma 7: Define $\Xi = (\mathbb{R}^d, |||\cdot|||)$ with the norm of an element $\xi = (p, x, u, \mu)$ given by

$$|||\xi||| := \left[\|p\|_{(2)}^2 + \|x\|_{(2)}^2 + \|u\|_{(-1)}^2 + \|\mu\|_2^2\right]^{1/2}, \tag{43}$$

where

$$\|u\|_{(-1)} := \left(y_N^2 + h\sum_{i=1}^{N} y_i^2\right)^{1/2}, \qquad y_i = h\sum_{k=0}^{i-1} u_k.$$

Then Ξ is a Banach space and its dual Ξ^* can be described as $(\mathbb{R}^d, |||\cdot|||_*)$ with

$$|||\Theta|||_* := \left[\|\phi\|_{(2)}^2 + \|\rho\|_{(2)}^2 + \|\Delta^1\vartheta\|_2^2 + \vartheta_N^2 + \|\omega\|_2^2\right]^{1/2} \tag{44}$$

for $\Theta = (\phi, \rho, \vartheta, \omega) \in \mathbb{R}^d$. Indeed, one can transform (ϑ, u) as follows

$$\begin{aligned}(\vartheta, u) = h\sum_{i=0}^{N-1} \vartheta_i u_i \ &= \ h\sum_{i=0}^{N-1} \vartheta_i (\Delta^1 y)_i \\ &= -h\sum_{i=1}^{N} y_i (\Delta^1\vartheta)_{i-1} \ + \ \vartheta_N y_N,\end{aligned}$$

and consequently, $|(\vartheta, u)| \leq \|u\|_{(-1)}(\|\Delta^1\vartheta\|_2^2 + |\vartheta_N|^2)^{1/2}$ holds true. Notice that the last estimate is sharp whenever ϑ solves

$$(\Delta^1\vartheta)_{i-1} = y_i, \ \ i = 1, \dots, N, \quad \vartheta_N = y_N.$$

Moreover, on the subspace Ξ^0 of Ξ^* with zero Θ-entries $\phi_N = 0$ and $\omega = 0$ corresponding to linear parts of ψ, the norm $|||\bar{\delta}|||_*$ is equivalent to $\|\delta\|_D$ from (37), and the constants in the related estimates do not depend on h.

Remark 2 It is well known that all norms in finite-dimensional space are equivalent. In particular, for the construction (43) the following estimates hold:

$$\|u\|_{(-1)}^2 = y_N^2 + h\sum_{i=1}^{N} y_i^2 \leq 2\,\|y\|_\infty^2 \leq 2\,\|u\|_2^2$$

(cf. Lemma 14, Appendix); and

$$\|u\|_2^2 = h^{-1}\sum_{i=0}^{N-1}(y_{i+1} - y_i)^2 \leq 4h^{-1}\sum_{i=1}^{N} y_i^2 \leq 4h^{-2}\|u\|_{(-1)}^2.$$

Thus, for each *fixed* $h > 0$, the norm $|||\cdot|||$ is equivalent to the Euclidean norm in $\mathbb{R}^d$ but the related constants are not independent of h.

Lemma 8 *Under assumptions* (H0)–(H2)*, there exists a constant h' such that each variational inequality* (VI^h_δ)*, $h < h'$, is strongly regular (in the sense of Robinson) for $\delta = \tilde{\delta}^h$ at the solution $\tilde{\xi}^h$, and the related neighborhoods $U = \{\bar{\delta} \in \varXi^0 : \|\delta - \tilde{\delta}^h\|_D \le \rho'\}$, $V = \{\xi \in \varXi : |||\xi - \tilde{\xi}^h||| < \bar{\rho}\}$ and the Lipschitz modulus λ can be chosen independently of h.*

Proof With the definitions of $\varXi$, $\varXi^*$ and related norms $|||\cdot|||$ and $|||\cdot|||_*$ from (43), (44), the assertion is deduced for $(\psi - \delta)$ from Lemmas 6, 7 and the estimates for $\|\tilde{\delta}^h\|_D$ from Lemma 2: with the constants ρ, ρ', c from Lemma 7, sufficiently small h' and appropriately chosen $\bar{\rho} = \mathrm{O}(\rho)$, the Lipschitz continuity follows with a modulus $\lambda = \mathrm{O}(c)$. □

In order to verify local uniqueness and Lipschitz behavior of solutions in (VI^h_δ), first refer to [42, Theorem2.1]: indeed, assumption (H0) together with Lemma 1 guarantees solution existence for each of the problems (VI^h_δ). Further, for arbitrary $\rho_0 > 0$ and $\bar{D}_0 = \{\bar{\delta} \in \varXi^0 : \|\delta\|_D \le \rho_0\}$, $\psi(\bar{\delta}, \xi) := \psi(\xi) - \bar{\delta}$ as a mapping from $\bar{D}_0 \times \varXi$ to $\varXi^*$ is twice continuously differentiable so that, for fixed h, the assumptions of Robinson's Inverse Function Theorem are fulfilled. However, for obtaining uniform w.r.t. h estimates of related neighborhoods and Lipschitz moduli, the fixpoint approach used in [42] has to reconsidered.

4.4 Lipschitz stability result

The inclusion (VI^h_δ) can be transformed to a fixpoint problem as follows: for given $(\delta, u) \in \mathbb{R}^l \times \mathbb{R}^N$ with $\|u\|_\infty \le 1$, denote by $(x, p) = (x(\delta, u), p(\delta, u))$ the solution of (16), (17). Further, define $\mu = \mu(\delta, u)$ by $\sigma = B^T p(\delta, u) - \delta_3$ and $\mu_1 = [\sigma]_+$, $\mu_2 = [\sigma]_-$, and introduce

$$r(\delta, u) := T(\xi(\delta, u)) - \psi(\xi(\delta, u)) + \bar{\delta} \tag{45}$$

with T given by (24). Obviously, $\xi = \xi(\delta, u)$ solves (VI^h_δ) if and only if

$$r(\delta, u) \in T(\xi(\delta, u)) + N_C(\xi(\delta, u)) =: T_C(\xi(\delta, u)).$$

By Lemma 6, for $\delta \in D_\rho$ the set $\varLambda(\delta) := T_C^{-1}(\bar{\delta})$ is nonempty. If $\varLambda_u(\delta)$ denotes the set of u-components related to $\xi \in \varLambda(\delta)$, and

$$\varPhi_\delta(u) := \varLambda_u(r(\delta, u)), \tag{46}$$

then the element $\xi = \xi(\delta, u)$ solves (VI^h_δ) if and only if $u \in \varPhi_\delta(u)$ (see also [15, Lemma 2.1]). This fixpoint characterization allows to prove

Theorem 1 *Under assumptions* (H0)–(H2)*, there exist constants h', $\bar{\rho}'$ and ϵ such that, for all $h < h'$ and $\delta \in D' = \{\delta \in \mathbb{R}^l : \|\delta - \tilde{\delta}^h\|_D < \bar{\rho}'\}$, problem* (VI^h_δ) *has* *a unique solution* ξ_δ^h *on the set* $V_\epsilon = \{\xi \in \mathbb{R}^d : \|u - \tilde{u}^h\|_{(-1)} \le \epsilon\}$ *and, as an element of* Ξ, ξ_δ^h *depends Lipschitz continuously on* $\bar{\delta} \in \Xi^*$. *The associated Lipschitz modulus as well as the radii* ϵ, $\bar{\rho}'$ *are independent of* h.

Before proving the theorem, some auxiliary estimates will be provided.

First notice that, under assumption (H0), the solutions $(x(u, \delta), p(u, \delta))$ of (16), (17) are uniformly bounded, cf. Lemma 1: if $\|\delta\|_D < \rho_0$ then

$$\|x\|_\infty + \|\Delta^1 x\|_2 + \|p\|_\infty + \|\Delta^1 p\|_2 \le M_0 \tag{47}$$

for some constant M_0 independent of u, δ and h. Shrinking h' if necessary, similar bounds are obtained for $(\tilde{x}^h, \tilde{p}^h)$ by definition of these functions and the estimates for $\tilde{\delta}^h$ from Lemma 2. Therefore,

$$\|z'\|_\infty + \|\Delta^1 z'\|_2 + \|q'\|_\infty + \|\Delta^1 q'\|_2 \le M_1 \tag{48}$$

holds true for $z' = x - \tilde{x}^h, q' = p - \tilde{p}^h$ and $M_1 := 2M_0$. As long as $h < h'$ and $\|\delta\|_D < \rho_0$, Lemma 13 from Appendix further yields

$$\|z'\|_\infty^2 + \|q'\|_\infty^2 \le M_2 \left(h' + \|z'\|_{(2)} + \|q'\|_{(2)}\right) \tag{49}$$

with a constant M_2 independent of u, δ and h.

Lemma 9 *Let* $\rho > 0$ *be arbitrarily given. If* $\bar{\rho}'$ *and* ϵ *are suffiently small,* $h < h'$ *and* $\delta \in D'$, *then every solution of* (VI_δ^h) *in* V_ϵ *satisfies* $\|\!|\xi - \tilde{\xi}^h\|\!| < \rho$.

Proof If $\xi = (p, x, u, \mu)$ is a solution of (VI_δ^h) then $\xi = \xi(\delta, u)$. By $y, \tilde{y}^h$ denote the integrated controls $y_i = h\sum_{k=0}^{i-1} v_i$ related to $v = u$ or $v = \tilde{u}^h$ resp., and set $\eta := x - By$, $\tilde{\eta}^h = \tilde{x}^h - B\tilde{y}^h$. For $z := x - \tilde{x}^h$, $q := p - \tilde{p}^h$, $\zeta := \eta - \tilde{\eta}^h$ and $\omega := y - \tilde{y}^h$, under (H0) we conclude

$$\begin{aligned}
(\Delta^1\zeta)_i &= f(\eta_i + By_i) - f(\tilde{\eta}_i^h + B\tilde{y}_i^h) + \delta_{1i} - \tilde{\delta}_{1i}^h,\\
|(\Delta^1\zeta)_i| &\le L'(|\zeta_i| + |B|\,|\omega_i|) + |\delta_{1i} - \tilde{\delta}_{1i}^h|,\\
|(\Delta^1 q)_i| &\le L'(|q_{i+1}| + |z_{i+1}|) + |\delta_{2i} - \tilde{\delta}_{2i}^h|,
\end{aligned}$$

where L' is the common Lipschitz modulus of f and H_x on the set $\{(x, p) : \|x\|_\infty + \|p\|_\infty \le M_0\}$; cf. (47). Applying Lemma 14 from Appendix to the equation for ζ yields

$$\|\zeta\|_\infty \le c(\|\delta - \tilde{\delta}^h\|_D + \|\omega\|_2)$$

with some constant c independent of h, δ and u.

Consequently, $\|z\|_2 \le (c + |B|)(\|\delta - \tilde{\delta}^h\|_D + \|\omega\|_2)$. Taking into account this relation and applying Lemma 14 to q now, an analogous estimate for $\|q\|_\infty$ is obtained. Summing up, the inequality

$$\|x - \tilde{x}^h\|_2 + \|p - \tilde{p}^h\|_\infty + \|\mu - \tilde{\mu}^h\|_\infty \le \mathrm{O}(\|\delta - \tilde{\delta}^h\|_D + \|u - \tilde{u}^h\|_{(-1)}) \tag{50}$$

follows and hence the assertion. □

The preparation for proving Theorem 1 will be continued with an analysis of $r = r(\delta, u)$: by construction, $r(\delta, u) = (r_1, r_2, r_3, r_4) \in \Xi^0$ (i.e. r_{1N}, r_4 vanish), and $r_3 = \delta_3 - \tilde{\delta}_3^h$. Further,

$$\begin{aligned} r_{1i} &= f(x_i) - f(x_i^0) - A_i(x_i - x_i^0) + \delta_{1i} - \tilde{\delta}_{1i}^h, \\ r_{2i} &= (\nabla f(x_{i+1}^0) - \nabla f(x_{i+1}))^T p_{i+1} - \nabla_x^2(\nabla f(x^0)^T p^0)\Big|_{t_{i+1}}(x_{i+1} - x_{i+1}^0) \\ &\quad + \delta_{2i} - \tilde{\delta}_{2i}^h, \\ r_{2N} &= \nabla k(x_N^0) - \nabla k(x_N) + K(x_N - x_N^0) + \delta_{2N} - \tilde{\delta}_{2N}^h. \end{aligned}$$

Consequently, from (48) and

$$\begin{aligned} &|r_{1i}| \le c\|z'\|_\infty |x_i - x_i^0| + |\delta_{1i} - \tilde{\delta}_{1i}^h| \le cM_1|x_i - x_i^0| + |\delta_{1i} - \tilde{\delta}_{1i}^h|, \quad i \le N-1, \\ &|r_{2i}| \le c(\|z'\|_\infty + \|q'\|_\infty) \cdot (|x_{i+1} - x_{i+1}^0| + |p_{i+1} - p_{i+1}^0|) + |\delta_{2i} - \tilde{\delta}_{2i}^h|, \quad i \le N, \end{aligned}$$

it is deduced that, for some constant c_r independent of u, h, δ,

$$\begin{aligned} |||r|||_* &\le c\left(\|z'\|_\infty + \|q'\|_\infty\right) \cdot (\|x - \tilde{x}^h\|_{(2)} + \|p - \tilde{p}^h\|_{(2)}) + \|\delta - \tilde{\delta}^h\|_D \quad (51) \\ &\le c_r\rho + \bar{\rho}' \quad (52) \end{aligned}$$

if $h < h'$, $\delta \in D'$, $|||\xi - \tilde{\xi}^h||| \le \rho$, and $\bar{\rho}' < \rho_0$.
Finally consider $r^1 = r(\delta, u^{(1)})$ and $r^2 = r(\delta, u^{(2)})$ for given $u^{(1)}, u^{(2)}$ and fixed $\delta \in D'$. Denoting by $(x^{(j)}, p^{(j)})$ the solution of (16), (17), and setting $z^j = x^{(j)} - \tilde{x}^h$, $q^j = p^{(j)} - \tilde{p}^h$ for $j = 1, 2$, in analogy to (51) obtain

$$|||r^1 - r^2|||_* \le c \max_{j=1,2}\{\|z^j\|_\infty + \|q^j\|_\infty\} \cdot (\|x^{(1)} - x^{(2)}\|_{(2)} + \|p^{(1)} - p^{(2)}\|_{(2)}). \quad (53)$$

For the elements z^j, q^j, $j = 1, 2$, (49) can be applied so that, with a constant ρ such that $|||\xi^{(j)} - \tilde{\xi}^h||| < \rho$ for $j = 1, 2$, the auxiliary result

$$|||r^1 - r^2|||_* \le \mathrm{O}(\sqrt{h'} + \sqrt{\rho}) \cdot (\|x^{(1)} - x^{(2)}\|_{(2)} + \|p^{(1)} - p^{(2)}\|_{(2)}).$$

is obtained. If we repeat the arguments from the proof of (50), Lemma 9, we deduce $\|x^{(1)} - x^{(2)}\|_{(2)} + \|p^{(1)} - p^{(2)}\|_{(2)} \le c\|u^{(1)} - u^{(2)}\|_{(-1)}$ and thus, end up with

$$|||r^1 - r^2|||_* \le c_r'(\sqrt{h'} + \sqrt{\rho}) \cdot \|u^{(1)} - u^{(2)}\|_{(-1)} \quad (54)$$

where the constant c_r' is independent of $h, \delta, u^{(1)}$ and $u^{(2)}$ as long as $\xi^{(1)}, \xi^{(2)}$ satisfy $\max\{|||\xi^{(1)} - \tilde{\xi}^h|||, |||\xi^{(2)} - \tilde{\xi}^h|||\} < \rho$.

Proof of Theorem 1 Define $W_\epsilon = \{w \in \mathbb{R}^N : \|w - \tilde{u}^h\|_{(-1)} \le \epsilon\}$. According to Lemma 9, one can choose ϵ and $\bar{\rho}'$ small enough to ensure

$$|||\xi(\delta, w) - \tilde{\xi}^h||| \le \rho \le \bar{\rho}$$

with the constant $\bar{\rho}$ from Lemma 8 for all $w \in W_\epsilon$ and arbitrary $h < h'$. Shrinking ϵ, h' and $\bar{\rho}'$ if necessary and taking into account (52) as well as Lemma 2, the estimate

$$|||r(\delta, w) - \bar{\delta}^h|||_* \le c_r\rho + \bar{\rho}' + \|\tilde{\delta}^h\|_D < \rho'$$

with constant ρ' from Lemma 8 will hold uniformly for $w \in W_\epsilon$. Thus, for arbitrary $\delta \in D'$ and $h < h'$, the mapping $\Phi_\delta(w)$ from (46) is well-defined and Lipschitz continuous on W_ϵ: it is sufficient to notice that the underlying map Λ has the Lipschitz modulus λ.

Let w^1, w^2 be any two points from W_ϵ. Using once more Lemma 8 we have

$$\begin{aligned}\left\|\Phi_\delta(w^1) - \Phi_\delta(w^2)\right\|_{(-1)} &\le |||\Lambda(r(\delta, w^1)) - \Lambda(r(\delta, w^2))|||\\ &\le \lambda\, |||r(\delta, w^1) - r(\delta, w^2)|||_*\\ &\le \lambda c_r'(\sqrt{h'} + \sqrt{\rho}) \cdot \|w^1 - w^2\|_{(-1)}\end{aligned}$$

and $\rho = \mathrm{O}(\epsilon + \bar{\rho}')$; see (53), (50) and Lemma 13. Shrinking h', ϵ and $\bar{\rho}'$ if necessary, one can make the last term satisfy

$$\left\|\Phi_\delta(w^1) - \Phi_\delta(w^2)\right\|_{(-1)} \le \kappa\, \|w^1 - w^2\|_{(-1)} \tag{55}$$

with a constant $\kappa < 1$. Hence Φ_δ is strictly contractive.

Next we will show that Φ_δ is a self-map of W_ϵ:
First, consider $w^0 = \tilde{w}^h$, $\delta_0 = \tilde{\delta}^h$: obviously, $w^0 = \Phi_{\delta_0}(w^0)$ and

$$\begin{aligned}\left\|\Phi_\delta(w^0) - w^0\right\|_{(-1)} &\le |||\Lambda(r(\delta, w^0)) - \Lambda(r(\delta_0, w^0))|||\\ &\le \lambda\, |||r(\delta, w^0) - r(\delta_0, w^0)|||_*\\ &\le \lambda\, \|\delta - \delta_0\|_D \le \lambda\, \bar{\rho}'\end{aligned}$$

follows by the definition of r, and from (50), Lemma 9. Further, for arbitrary $w \in W_\epsilon$, from the contractivity of Φ_δ conclude

$$\begin{aligned}\left\|\Phi_\delta(w) - w^0\right\|_{(-1)} &\le \left\|\Phi_\delta(w) - \Phi_\delta(w^0)\right\|_{(-1)} + \left\|\Phi_\delta(w^0) - w^0\right\|_{(-1)}\\ &\le \kappa\, \epsilon + \lambda\, \bar{\rho}'.\end{aligned}$$

The latter can be made smaller than ϵ : indeed, the above construction shows $\kappa = \mathrm{O}(\sqrt{h'} + \sqrt{\epsilon} + \sqrt{\bar{\rho}'})$ so that $\lambda\bar{\rho}'$ can be made smaller than $(1-\kappa)\epsilon$ by shrinking $\bar{\rho}'$ if necessary.

The properties of $\Phi_\delta: W_\epsilon \to W_\epsilon$ guarantee the existence of a fixpoint $w = u_\delta^h$ in W_ϵ corresponding to the unique solution $\xi_\delta^h = \xi(\delta, u_\delta^h)$ of (VI_δ^h) in V_ϵ. The Lipschitz continuity of u_δ^h w.r.t. δ is obtained similarly to the proof in [42, Theorem 2.1] with Lipschitz modulus $\lambda' = \lambda(1-\kappa)^{-1}$. Together with condition (50), this fact yields the Lipschitz continuous dependence of ξ_δ^h on δ with a modulus independent of h and thus, completes the proof. □

5 Control properties

In this section, consider (VI_δ^h) in the special case that $\delta = 0$. As it was shown in the previous section, there exists a unique solution $\xi_0^h = \hat{\xi} = (\hat{p}, \hat{x}, \hat{u}, \hat{\mu})$ of the problem close to $\tilde{\xi}^h$ provided h is sufficiently small. Due to (18), (19), $\hat{\sigma} := B^T\hat{p}$ is the discrete switching function for $\hat{u}$: with $\hat{\mu}_1 = [\hat{\sigma}]_+$, $\hat{\mu}_2 = [\hat{\sigma}]_-$ we have

$$-1 \le \hat{u}_i \le 1, \qquad (\hat{u}_i - 1)[\hat{\sigma}_i]_+ = (\hat{u}_i + 1)[\hat{\sigma}_i]_- = 0$$

for $i = 0, \dots, N-1$.

In [14], for continuous problems of type (VI) depending on a real parameter it was shown that (H1) together with the assumptions

$$\begin{aligned} R \;=\; -p^T\left[B, [f^0, B]\right] = -B^T H_{xx}^0 B \;\ge\; \bar{m} \;>\; 0 \quad &\text{on } I_\beta, \\ \min\{|\ddot{\sigma}^0(\tau_s - 0)|, |\ddot{\sigma}^0(\tau_s' + 0)|\} \;\ge\; \bar{m} \;>\; 0 & \end{aligned}$$

(see [14, Assumptions 2.2-3]), ensures the bang-singular-bang control structure to be stable under parameter perturbation. Similar results will be proved now for the discretized problems (VI_0^h) for sufficiently small step size parameter h. To this aim, the discrete solution $(\hat{u}, \hat{\sigma})$ will be compared to the reference data (u^0, σ^0) from (VI). For simplicity, the discrete functions $(\tilde{u}, \tilde{\sigma}) \in \mathbb{R}^N \times \mathbb{R}^N$ defined by $\tilde{u}_i = u^0(t_i + 0)$, $\tilde{\sigma}_i = \sigma^0(t_i)$ will again be denoted by (u^0, σ^0).

Lemma 10 *For the switching function $\hat{\sigma}$, the following estimate holds:*

$$\|\hat{\sigma} - \sigma^0\|_\infty + \|\Delta^1(\hat{\sigma} - \sigma^0)\|_2 = O(h).$$

Moreover, $\hat{\sigma}_i = \hat{P}_i + \hat{R}_i\,\hat{u}_i$ where $\hat{P}$, $\hat{R}$ satisfy

$$\begin{aligned} &\|\hat{P} - P\|_2 + \|\hat{R} - R\|_2 = O(h), \\ &\|\Delta^1(\hat{\sigma} - \sigma^0)\|_\infty + \|\hat{P} - P\|_\infty + \|\hat{R} - R\|_\infty = O(\|\hat{x} - x^0\|_\infty + h) \end{aligned}$$

(The proof is given in Appendix.)

The last lemma in particular yields the following property: there exist $\epsilon_0 > 0$, $\bar{h} > 0$ and a constant $\hat{m} > 0$ independent of h such that

$$\hat{P}_i - \hat{R}_i \le -\hat{m}, \quad \hat{P}_i + \hat{R}_i \ge \hat{m} \qquad \text{for } t_i \in I_\beta \tag{56}$$

whenever $h < \bar{h}$ and $\|\hat{x} - x^0\|_\infty < \epsilon_0$; cf. (14), (15).

Lemma 11 *Under assumptions* (H0)–(H2)*, there exist indices* $\hat{k}, \hat{k}'$ *such that, with the constant values* u^1, u^2 *from Assumption* (H1)*,*

$$\hat{u}_i \equiv u^1, \quad i = 0, \ldots, \hat{k}, \qquad \hat{u}_i \equiv u^2, \quad i = \hat{k}', \ldots, N-1,$$

and $|t_{\hat{k}} - \tau_s| + |t_{\hat{k}'} - \tau_s'| = \mathrm{O}(h^{1/2})$.

Proof Without loss of generality, consider the case $u^1 = -1$ only:
At the bang-singular junction point τ_s, both the function σ^0 and its time derivative vanish. Thus, for all j such that $t_j < \tau_s$, it follows from (15) that

$$\sigma^0(t_j) = \int_{t_j}^{\tau_s} (\theta - t_j)\ddot{\sigma}^0(\theta)\, d\theta \;\le\; -\frac{m^2}{2}(\tau_s - t_j)^2.$$

Consider $\hat{\sigma}$: if $\hat{\sigma}_j < 0$ for all $t_j < \tau_s$ then we can choose $\hat{k} = k$ (see Lemma 2 and related notation). Otherwise, let $\hat{k}$ be the first index such that $\hat{\sigma}_{\hat{k}} \ge 0$, $t_{\hat{k}} < \tau_s$: from the previous lemma conclude

$$0 \le \hat{\sigma}_{\hat{k}} \le ch - \frac{m^2}{2}(\tau_s - t_{\hat{k}})^2$$

and thus, the desired estimate for the left bang arc follows.
Analogously, the right bang arc is confirmed and estimated. □

In the next step of analyzing the control structure it will be shown that, in a certain neighborhood of the state trajectory, the discrete controls $\hat{u}$ consist of two bang arcs at right and left ends of the interval, and a singular arc located in the interior of I_β. Following the ideas from [14], the proof will make use of auxiliary state-adjoint functions $\hat{x}^{\pm}$, $\hat{p}^{\pm}$ which are obtained from solving (16), (17) with $\delta = 0$, constant $u = u^+ \equiv 1$ (or $u = u^- \equiv -1$ respectively) and initial value $(\hat{x}_i, \hat{p}_i)$ at given $t = t_i$. Similarly, $x^{\pm}$, $p^{\pm}$ solve (7), (8) with constant $u \in \{-1, 1\}$ and initial value (x_i^0, p_i^0) at $t = t_i$. Notice that, for sufficiently small h, the assumptions on the input functions guarantee the existence of solutions at least for $|t_j - t_i| \le \Delta_0$ (or $|t - t_i| \le \Delta_0$ for the continuous version) for some $\Delta_0 > 0$ independent of h.

Lemma 12 *Let* $\bar{h} > 0$, $\epsilon_0 > 0$ *be the constants used for* (56)*. Further, for given* $h < \bar{h}$*, denote by* $\xi_0^h = \hat{\xi} = (\hat{p}, \hat{x}, \hat{u}, \hat{\mu})$ *the solution of* (VI_0^h) *from Theorem* 1*. Then there exists a positive* $\epsilon_1 < \epsilon_0$ *with the following property: Whenever the state component* $\hat{x}$ *satisfies* $\|\hat{x} - x^0\|_\infty < \epsilon_1$*, the control* $\hat{u}$ *has bang-singular-bang structure. More precisely, there exist indices* $\hat{k}$ *and* $\hat{k}'$ *such that*

$$\hat{u}_i = \begin{cases} u^1 & \text{if} \quad 0 \le i \le \hat{k}, \\ \hat{u}_{s,i} & \text{if} \quad \hat{k} < i < \hat{k}', \\ u^2 & \text{if} \quad \hat{k}' \le i \le N-1. \end{cases} \tag{57}$$

In addition, $|t_{\hat{k}} - \tau_s| + |t_{\hat{k}'} - \tau_s'| + \|\hat{u}_s - u_s^0\|_{(\infty, I_\beta)} = \mathrm{O}(\epsilon_1 + h)$. *The singular values are given by* $\hat{u}_{s,i} = -\hat{P}_i/\hat{R}_i$ *and belong to* $[m_s - 1, 1 - m_s] \subset (-1, 1)$ *if* ϵ_1 *and* h *are taken sufficiently small. The constant* m_s *is independent of* h.

Proof The proof is similar to that of [14, Theorem 3.1]; see Sect. 4 therein. For clarity reasons, the main arguments are repeated in the needed discrete formulation assuming (without loss of generality) that $u^1 = -1$.

Lemma 11 says that $\hat{u}$ has a bang arc in the left part of the interval [0, 1] where $\hat{u} = -1$ and thus, $\hat{\sigma} = \hat{\sigma}^-$ (continued from $t_0 = 0$, e.g.) as long as $\hat{\sigma}^- < 0$. If h is sufficiently small, the right end of the arc lies inside I_β.
Now, from Lemma 10 we have

$$\|\Delta^1(\hat{\sigma}^- - \sigma^-)\|_\infty = \mathrm{O}(\epsilon_1 + h).$$

and $\Delta^1\hat{\sigma}^-$ (as well as the finite difference approximation of the time derivative of σ^-) changes its sign, i.e., $(\Delta^1\hat{\sigma}^-)_i \le 0$ for $i \ge j$ and some node t_j near τ_s: Due to (15) and (56), $\Delta^2\hat{\sigma}^- \le -\hat{m} < 0$ so that $|t_j - \tau_s| = \mathrm{O}(\epsilon_1 + h)$ follows.

Let j be the first index where $(\Delta^1\hat{\sigma}^-)_i \le 0$. We will distinguish between the following cases:

Case 1 $\hat{\sigma}_j^- < 0$.

In this case, $\hat{u}_j = -1$, $(\Delta^2\hat{\sigma})_j \le -\hat{m} < 0$ and $\hat{\sigma}^-$ is concavely decreasing for $t_i > t_j$. Therefore, $\hat{\sigma}_i = \hat{\sigma}_i^-$ remains to be valid at least for all i with $t_i - t_j \le \Delta_0$. We obtain:

$$\begin{aligned}
(\Delta^1\hat{\sigma})_{j+\nu} &\le -\hat{m}\cdot\nu h,\\
\hat{\sigma}_{j+\nu} &< h\sum_{k=1}^{\nu-1}(\Delta^1\hat{\sigma})_{j+k} \le -\frac{\hat{m}}{2}(\nu-1)^2h^2.
\end{aligned}$$

If t_i is the last node on the interval $[t_j, t_j+\Delta_0]$ then $\hat{\sigma}_i \le -\hat{m}\,(\Delta_0-2h)^2/2$ contradicting the conditions $\sigma^0(t) = 0$ for $t \ge \tau_s$, $|t_j - \tau_s| = \mathrm{O}(\epsilon_1+h)$, and $\|\hat{\sigma} - \sigma_0\|_\infty = \mathrm{O}(h)$ from Lemma 10.

Case 2 $\hat{\sigma}_j^- > 0$.

In this case, $\hat{\sigma} \ge 0$ has appeared first for some node $t_{j'} \le t_j$: if $\hat{\sigma}_{j'} > 0$, the function coincides with the convexly increasing function $\hat{\sigma}^+$ for $t_{j'} \le t_i \le t_{j'} + \Delta_0$. Estimating $\hat{\sigma}^+_{j'+\nu}$ in analogy to Case 1, we end up with a contradiction. If $\hat{\sigma}^-_{j'} = 0$ then proceed as in Case 3 below.

Case 3 $\hat{\sigma}_j^- = 0$.

It is sufficient to consider $(\Delta^1\hat{\sigma})_j = 0$ since $(\Delta^1\hat{\sigma})_j \ne 0$ yields $\hat{\sigma}_{j+1} \ne 0$ and thus, a situation as in Case 1 or Case 2. (Similarly, one can exclude $(\Delta^1\hat{\sigma})_{j+1} \ne 0$ by contradiction).

For $(\Delta^1\hat{\sigma})_j = (\Delta^1\hat{\sigma})_{j+1} = 0$, obviously $(\Delta^2\hat{\sigma})_{j+1} = 0$ so that $\hat{u}_{j+1} = \hat{u}_{s,j+1}$. Whenever $t_i \le 0.5(\tau_s + \tau_s') + \bar{h}$, a change to positive resp. negative values of $\hat{\sigma}$ at t_i can be excluded by repeating the arguments above. Thus, the left bang arc is concatenated to a central singular arc.

The analysis can be carried out similarly from the right end, and the statement on the control structure together with

$$|t_j - \tau_s| + |t_{j'} - \tau_s'| = \mathrm{O}(\epsilon_1 + h)$$

follows. Finally, by use of Lemma 10, the estimate for $\hat{u}_s$ is obtained. □

The structural information allows to strengthen the general result from Theorem 1 and formulate a convergence result for the solutions of (VI$_0^h$) w.r.t. (discrete) $L_\infty \times L_\infty \times L_1 \times L_\infty$ topology:

Theorem 2 *Let the assumptions* (H0)–(H2) *hold for the solution* $\xi^0 = (p^0, x^0, u^0, \mu^0)$ *of* (VI). *Further, suppose* h', ϵ, $\bar{\rho}'$ *and the set* V_ϵ *be given as in Theorem 1. If* $h < h'$ *is small enough to ensure* $\|\tilde{\delta}^h\|_D < \bar{\rho}'$ *then the following statements are true:*

(i) *On* V_ϵ*, the discrete problem* (VI$_0^h$) *has a unique solution* $\xi_0^h = \hat{\xi} = (\hat{p}, \hat{x}, \hat{u}, \hat{\mu})$ *estimated by*

$$\|\hat{x} - x^0\|_2 + \|\hat{p} - p^0\|_\infty + \|\hat{\mu} - \mu^0\|_\infty = O(h),$$

and $\|\hat{x} - x^0\|_\infty = \mathrm{O}(h^{1/2})$.

(ii) *There exists a constant* $h_1 < h'$ *such that, for each* $h < h_1$*, the control* $\hat{u}$ *has bang-singular-bang structure in the sense of* (57). *The state-control pair and associated bang-singular junction points* $\hat{\tau}$, $\hat{\tau}'$ *satisfy*

$$\|\hat{x} - x^0\|_\infty + \|\hat{u} - u^0\|_1 + |\hat{\tau} - \tau_s| + |\hat{\tau}' - \tau_s'| = O(h).$$

Proof The first estimates from part (i) are a direct consequence of Theorem 1 applied to the choice $\delta = 0$, and of Lemma 10 (see also the proof of the latter in Appendix). In order to obtain the estimate for $\|\hat{x} - x^0\|_\infty$ we observe that, under assumption (H0) on the data, there exists a constant M_1 independent of h such that

$$\|\hat{x} - x^0\|_\infty + \|\Delta^1(\hat{x} - x^0)\|_2 \leq M_1.$$

Applying Lemma 13 from Appendix, the desired estimate follows.

For proving part (ii), the technique from [15, Lemma 4.1] will be adapted. To this aim, define

$$y_i^0 = h \sum_{j=0}^{i-1} u^0(t_j + 0), \qquad \hat{y}_i = h \sum_{j=0}^{i-1} \hat{u}_j.$$

Notice that y^0 slightly differs from $\tilde{y}^h$ defined via Goh transformation with the input $\tilde{u}^h$, (21). Assumptions (H1) and (H2), however, guarantee that the pointwise differences do not exceed $\mathrm{O}(h)$. By Theorem 1 and Lemma 2, we therefore have

$$\|\hat{y} - y^0\|_1 \leq \|\hat{y} - \tilde{y}^h\|_2 + \|\tilde{y}^h - y^0\|_\infty = \mathrm{O}(h). \tag{58}$$

Similarly obtain $|\hat{y}_N - y^0_N| = \mathrm{O}(h)$.

In part (i) it was particularly proved that $\|\hat{x} - x^0\|_\infty \to 0$ as $h \to 0$. Hence, there exists a positive $h_1 < \bar{h}$ such that $\|\hat{x} - x^0\|_\infty < \epsilon_1$ holds with the constant ϵ_1 from Lemma 12 for $h < h_1$, and $\hat{u}$ has the structure (57). Therefore, the y–differences given at nodes t_i by

$$\hat{y}_i - y^0_i = h \sum_{j=0}^{i-1} (\hat{u}_j - u^0_j) = (\hat{y}_N - y^0_N) - h \sum_{j=i}^{N-1} (\hat{u}_j - u^0_j)$$

can be estimated, in particular, for points t_i from the intersection $(\tau_s, \tau_s') \cap (\hat{\tau}, \hat{\tau}')$ as follows: in case $\tau_s < \hat{\tau}$ the above formula yields

$$\begin{aligned} |\hat{y}_i - y^0_i| &= h \left| \sum_{j:\, \tau_s < t_j < \hat{\tau}} (u^1 - u^0_{s,j}) + \sum_{j: \hat{\tau} \le t_j < t_i} (\hat{u}_{s,j} - u^0_{s,j}) \right| \\ &\ge \min\{m, \hat{m}\}\, |\hat{\tau} - \tau_s| - \|\hat{u}_s - u^0_s\|_{(1, I_\beta)} \end{aligned} \tag{59}$$

(and the same relation results in case $\tau_s \ge \hat{\tau}$, too). Analogously obtain

$$|\hat{y}_i - y^0_i| \ \ge\ \min\{m, \hat{m}\}\, |\hat{\tau}' - \tau_s'| - \|\hat{u}_s - u^0_s\|_{(1, I_\beta)} - |\hat{y}_N - y^0_N|. \tag{60}$$

For $h < h_1$, the singular control $\hat{u}_s$ restricted to nodes $t_i \in I_\beta$ is well defined. Further, from (14), (56) obtain $R_i \ge m > 0,\ \hat{R}_i \ge \hat{m} > 0,\ t_i \in I_\beta$, so that

$$\|\hat{u}_s - u^0_s\|_{(1, I_\beta)} \le \hat{m}^{-1} \|\hat{P} - P\|_1 + (m\hat{m})^{-1} \|P\|_\infty \|\hat{R} - R\|_1 = \mathrm{O}(h)$$

follows from (H0) and Lemma 10.

Combining the estimates (59), (60), we come to the following conclusion: there exists a constant $c > 0$ independent of h such that

$$|\hat{\tau} - \tau_s| + |\hat{\tau}' - \tau_s'| \le c \left(\|\hat{y} - y^0\|_1 + \|\hat{u}_s - u^0_s\|_{(1, I_\beta)} + |\hat{y}_N - y^0_N| \right) = \mathrm{O}(h).$$

Finally, by $0 \le \hat{u}_i, u^0_i \le 1$ for $i = 0, \ldots, N-1$,

$$\|\hat{u} - u^0\|_1 \le \|\hat{u}_s - u^0_s\|_{(1, I_\beta)} + |\hat{\tau} - \tau_s| + |\hat{\tau}' - \tau_s'| = \mathrm{O}(h)$$

follows. Remembering (16), for $\delta = 0$ this last estimate also yields $\|\hat{x} - x^0\|_\infty = \mathrm{O}(h)$ and hence, the theorem. □

6 Numerical test

The discretization approach will be illustrated by approximating the solution of the following problem,

$$
\begin{aligned}
&(\mathbf{P}_\delta) \qquad \min \quad J'(x,u) := 0.5\left(x_1^2(T)+x_2^2(T)\right)\\
&\text{w.r.t.} \qquad \begin{array}{ll}
\dot{x}_1 = \cos x_3 - \delta u, & x_1(0) = a,\\
\dot{x}_2 = \sin x_3, & x_2(0) = 0,\\
\dot{x}_3 = u, & x_3(0) = \pi/2,\\
|u| \le 1, & (T > 0,\ a > 2,\ 0 \le \delta \ll 1).
\end{array}
\end{aligned}
\tag{61}
$$

As it was shown in [14], for small δ the solution is of bang-singular-bang structure, and the control vanishes along the singular arc. For data $\delta = 0.15$, $a = 3$, the optimal termination time for (61) is found from a synthesis approach as $T^* = 3.5117$. The terminal time T for (P_δ) is chosen as $T = 3.2 < T^*$. Assuming the same principal control structure as in time-optimal termination case, one can find piecewise analytic expressions for x, p and determine the bang-singular junction points $\tau_s = 2.1482$, $\tau_s' = 2.6932$ for (P_δ): this defines the reference solution u^0.

In the numerical tests, we replace (VI_0^h) by an equivalent system of equations: for $i = 1, \ldots, N$, the complementarity relations (18), (19) are expressed by

$$\phi_\gamma(\mu_{1i}, 1 - u_i) = 0, \quad \phi_\gamma(\mu_{2i}, 1 + u_i) = 0,$$

where $\phi_\gamma(w, \nu) = w + \nu - \sqrt{w^2 + \nu^2 + \gamma}$ with $\gamma = 0$ stands for the Fischer–Burmeister function [16,17]. Notice that the original function ϕ_0 is non-differentiable at the origin. In order to ensure robustness of the iterative solution when smooth solvers like the MATLAB[1] routine `fsolve` are used, the additional regularization parameter $\gamma > 0$ was introduced (cf. [25] and the survey [18]). Further, the system (VI_0^h) was reduced to a problem of finding (u, μ_1, μ_2). To this aim, the system (16), (17) is solved explicitely and the resulting vector $p = p(u)$ is inserted into (18). The final problem consists in solving $\bar{F}^\gamma(u, \mu_1, \mu_2) = 0$ where

$$\bar{F}^\gamma_{1i} = B^T p_i(u) - \mu_{1i} + \mu_{2i}, \quad \bar{F}^\gamma_{2i} = \phi_\gamma(\mu_{1i}, 1 - u_i), \quad \bar{F}^\gamma_{3i} = \phi_\gamma(\mu_{2i}, 1 + u_i).$$

For the initial values used in `fsolve`, for given $\bar{u}$ find $(\bar{x}, \bar{p}) = (x(\bar{u}), p(\bar{u}))$ from the canonical system (16), (17), and $\bar{\mu}_{1,2} = \mu_{1,2}(\bar{u})$ as positive resp. negative parts of $\sigma = B^T p(\bar{u})$. Afterwards, the vector $(\bar{u}, \bar{\mu}_1, \bar{\mu}_2)$ is projected onto the closed set $\{[-1+\gamma, 1-\gamma]^N \times [\gamma, \infty)^{2N}\}$: this change turns $\bar{u}$ and $\bar{\mu}_{1,2}$ into strictly inner points of their respective domain of definition and again, improves the robustness of the solution process.

[1] MATLAB is a registered trademark of TheMathWorks, Inc. (see www.mathworks.com).

Table 1 Approximation errors for control function u and junction times $\hat{\tau}$, $\hat{\tau}'$

γ	$(N=50)$		$(N=100)$		$(N=200)$	
	$\|u-u^0\|_1$	$\Delta\tau$	$\|u-u^0\|_1$	$\Delta\tau$	$\|u-u^0\|_1$	$\Delta\tau$
5×10^{-5}	0.0528	0.1868	0.0399	0.2188	0.0402	0.2028
2×10^{-5}	0.0434	0.1332	0.0265	0.1642	0.0252	0.1708
1×10^{-5}	0.0398	0.1228	0.0203	0.1548	0.0176	0.1388
5×10^{-6}	0.0378	0.1228	0.0166	0.1228	0.0123	0.1228
0	0.0358	0.0918	0.0120	0.0372	0.0012	0.0212

Table 1 summarizes selected results. The initial control data are $\bar{u} \equiv 1$. (It should be mentioned that tests with $\bar{u} \equiv 0.5$ as well as for some prescribed bang-singular-bang regimes, showed essentially the same behavior.) One can see that the solution process gives fairly good approximations of the reference control. Let us mention that, in all reported cases, the discrete controls have clearly visible bang arcs at left and right ends of the time interval and an inner singular arc with $u = 0$ (up to marginal rounding errors perhaps). Due to the regularization by $\gamma > 0$, but also by rounding effects, the bang-singular junction appears as an intermediate monotonicity interval rather then a "sharp" jump but is restricted to few nodes: with $m_\epsilon = 0.1\,\|u-u^0\|_1$, we determined two intervals I_s, I_s' where $1-m_\epsilon \geq u_i \geq m_\epsilon$. The approximation error $\Delta\tau$ for τ_s, τ_s' is defined as the maximum of distances from τ_s to interval ends of I_s, and from τ_s' to interval ends of I_s'.

Finally notice that, in the given example, the solution could be also found for $\gamma = 0$: in all tests the junction points are approximated with accuracy $\Delta\tau < 1.5\,h$, i.e., the "1–0–1" switching structure of the discrete control becomes evident. For $N = 200$, the control is approximated with a remaining tolerance bound of $\|u-u^0\|_1 = 0.0012$.

Acknowledgments The author is grateful to the anonymous referees for their instructive comments which, in particular, helped to close a gap in one of the main proofs.

Appendix

In this section, we first derive several norm estimates for functions from W_2^1 resp. their discrete analoga from Y^h. Afterwards, the proofs of Lemmas 2 and 10 are given.

Lemma 13 *Let $w \in \mathbb{R}^{k(N+1)}$ with $\Delta^1 w \in \mathbb{R}^{kN}$ be such that*

$$\|w\|_\infty + \|\Delta^1 w\|_2 \leq M_0 < \infty.$$

Then, $\|w\|_\infty^2 \leq M\,(h + \|w\|_{(2)})$ for some $M > 0$ independent of h, w.

Proof Let j be an index where $|w_j| = \|w\|_\infty$, and c a constant greater than M_0. Using

$$w_{j+k} = w_j + \sum_{l=0}^{k-1} h(\Delta^1 w)_{j+l}$$

and analogous formulas for w_{j-k}, it follows from the assumptions on w that

$$|w_i| \geq |w_j| - c\sqrt{h\,|i-j|} \geq |w_j|\left(1 - \frac{1}{\sqrt{2}}\right)$$

for all i such that $|i - j| \leq m := \lfloor |w_j|^2 \cdot N/(2c^2) \rfloor$ and $t_i \in [0, 1]$. Taking into account $|w_j|^2 \cdot N/(2c^2) \leq M_0^2\, N/(2c^2) < N/2$, it is easy to see that the conditions are fulfilled at least for $j < i \leq j + m$ in case $t_j \leq 1/2$ (or $j - m \leq i < j$ for $t_j > 1/2$) , i.e. on some index set $I(w)$ containing at least m knots on [0, 1]. Consequently,

$$\begin{aligned}\|w\|_{(2)}^2 &\geq h \sum_{i \in I(w)} |w_i|^2 \geq \frac{mh}{2}(1-\sqrt{2})^2\, \|w\|_\infty^2 \\ &\geq \frac{h}{2}(3 - 2\sqrt{2})\, \|w\|_\infty^2 \left(\frac{|w_j|^2 \cdot N}{2c^2} - 1\right) = 2\, M_1 \|w\|_\infty^2 \left(\|w\|_\infty^2 - 2c^2 h\right)\end{aligned}$$

(with $M_1 = (3 - 2\sqrt{2})/(8c^2)$) due to $h \cdot N = 1$.
If now $\|w\|_\infty^2 \geq 4hc^2$ then $\|w\|_{(2)}^2 \geq M_1 \|w\|_\infty^4$. Otherwise, $\|w\|_\infty^2 < 4hc^2$ and hence the lemma. □

Remark For continuous functions $w \in W_2^1(0, 1; \mathbb{R}^k)$ with $\|w\|_\infty + \|\dot{w}\|_2 \leq M_0$, similarly get $\|w\|_\infty \leq M\, \|w\|_2^{1/2}$ where M depends only on M_0.
The next lemma is a special case of the discrete analogon of Gronwall's Lemma [22]. In order to emphasize the independence of related constants of the step size h, a short direct proof is provided.

Lemma 14 (Discrete Gronwall Lemma) *Suppose there are given an arbitrary $\phi \in \mathbb{R}^{kN}$ and some constant $L > 0$.*

(i) *If $\eta \in \mathbb{R}^{k(N+1)}$ satisfies $|(\Delta^1\eta)_i| \leq L\,|\eta_i| + |\phi_i|$ for $i = 0, \ldots, N-1$, then $\|\eta\|_\infty \leq e^L(|\eta_0| + \|\phi\|_2)$.*
(ii) *If $\eta \in \mathbb{R}^{k(N+1)}$ satisfies $|(\Delta^1\eta)_i| \leq L\,|\eta_{i+1}| + |\phi_i|$ for $i = 0, \ldots, N-1$, then $\|\eta\|_\infty \leq e^L(|\eta_N| + \|\phi\|_2)$.*

Proof The proof starts with the observation

$$\big|\,|\eta_{i+1}| - |\eta_i|\,\big| \leq h\,|(\Delta^1\eta)_i|.$$

For part (i) we thus obtain by induction

$$|\eta_{i+1}| \le (1+Lh)|\eta_i| \,+\, h\,|\phi_i| \;\le\; (1+Lh)^{i+1}|\eta_0| \,+\, h\sum_{j=0}^{i}(1+Lh)^{i-j}|\phi_j|$$
$$\le (1+Lh)^N\left(|\eta_0| \,+\, h\sum_{j=0}^{N-1}|\phi_j|\right) \;\le\; e^L\,(|\eta_0| \,+\, \|\phi\|_2)\,.$$

Similarly, part (ii) follows from $|\eta_i| \le (1+Lh)|\eta_{i+1}| + h|\phi_i|,\; i \le N-1.$ □

Proof of Lemma 2 Consider first $x_i = \tilde{x}_i^h,\ i = 0,\ldots,N$, and $u_i = \tilde{u}_i^h$:

$$x_{i+1} - x_i = \int_{\omega_i} \dot{x}^0(t)\,dt \;=\; \int_{\omega_i}\Big[f(x^0(t)) \,+\, B\,u^0(t)\Big]\,dt$$
$$= \int_{\omega_i}\Big[f(x^0(t)) - f(x^0(t_i))\Big]\,dt \;+\, h\,[f(x_i) + B\,u_i]$$

due to the construction of $\tilde{u}^h$. Therefore,

$$\tilde{\delta}_{1i}^h \,=\, h^{-1}\int_{t_i}^{t_i+h}\Big[f(x^0(t)) - f(x^0(t_i))\Big]\,dt \;=\; \mathrm{O}(h).$$

Analogous estimates show that $\tilde{\delta}_{2i}^h = \mathrm{O}(h)$ uniformly for $i = 0,\ldots,N-1$.

The construction of $\tilde{\mu}_i^h$ ensures (19) to be valid for each $i \le N-1$ if only h is sufficiently small. In order to find $\tilde{\delta}_3^h$, insert $\tilde{x}_i^h$, $\tilde{p}_i^h$ and $\tilde{\mu}_i^h$ into (18): for $i = 0,\ldots,k$,

$$B^T p_i \,=\, \sigma^0(t_i) \;=\; \begin{cases} \tilde{\sigma}_i^h + (\sigma^0(t_i) - \sigma^\tau(t_i)) & \text{if} \quad i \le k, \\ \tilde{\sigma}_i^h & \text{if} \quad i > k. \end{cases}$$

where σ^τ abbreviates $\sigma^\tau(t) = \sigma(t + \tau_s - t_k)$. Consequently,

$$\left|\tilde{\delta}_{3i}^h\right| = |\sigma^0(t_i) - \sigma^\tau(t_i)| \;=\; \mathrm{O}(h) \qquad \text{for } i \le k,$$

and $\tilde{\delta}_{3i}^h = 0$ in case $i > k$. For $\Delta^1\tilde{\delta}_3^h,\ \Delta^2\tilde{\delta}_3^h$ we have

$$(\Delta^1\tilde{\delta}_3^h)_i \,=\, \begin{cases} (\Delta^1\sigma^0)_i \,-\, (\Delta^1\sigma^\tau)_i & \text{if} \quad i \le k-1, \\ -h^{-1}\sigma^0(t_k) & \text{if} \quad i = k, \\ 0 & \text{if} \quad i > k, \end{cases}$$
$$(\Delta^2\tilde{\delta}_3^h)_i \,=\, \begin{cases} (\Delta^2\sigma^0)_i \,-\, (\Delta^2\sigma^\tau)_i & \text{if} \quad i \le k-1, \\ h^{-2}(\sigma^0(t_{k-1}) - \sigma^\tau(t_{k-1}) - 2\sigma^0(t_k)) & \text{if} \quad i = k, \\ h^{-2}\sigma^0(t_k) & \text{if} \quad i = k+1, \\ 0 & \text{if} \quad i \ge k+2. \end{cases}$$

For $t \le \tau_s$,

$$\sigma^0(t) = \int_{\tau_s}^{t} \int_{\tau_s}^{s} \ddot{\sigma}^0(\theta)\, d\theta\, ds = \mathrm{O}((\tau_s - t)^2)$$

so that $|\sigma^0(t_k)| + |\sigma^0(t_{k-1})| = \mathrm{O}(h^2)$. Analogously obtain $|\sigma^\tau(t_{k-1})| = \mathrm{O}(h^2)$, too. Together with the estimates

$$|(\Delta^1 \sigma^\tau)_i - (\Delta^1 \sigma^0)_i| + |(\Delta^2 \sigma^\tau)_i - (\Delta^2 \sigma^0)_i| = \mathrm{O}(h),$$

the desired results for $\tilde{\delta}_3^h$ and its finite differences directly follow. □

Proof of Lemma 10 By definition, $\hat{\sigma} = B^T \hat{p}$ where $\hat{p}$ solves the backward initial value problem for the finite difference equation

$$(\Delta^1 \hat{p})_i = -\nabla f(\hat{x}_{i+1})^T \hat{p}_{i+1}, \qquad \hat{p}_N = -\nabla k(\hat{x}_N).$$

Thus,

$$\begin{aligned} (\Delta^1(\hat{p} - p^0))_i + A_{i+1}(\hat{p} - p^0)_{i+1} &= \left[\nabla f(x^0_{i+1}) - \nabla f(\hat{x}_{i+1})\right]^T \hat{p}_{i+1} - \tilde{\delta}^h_{2i}, \\ \hat{p}_N - p^0_N &= \nabla k(x^0_N) - \nabla k(\hat{x}_N). \end{aligned}$$

Taking into account assumption (H0), from Theorem 1 and Lemma 2 conclude

$$\|\hat{x} - x^0\|_2 + \|\hat{p} - p^0\|_2 = \mathrm{O}(\|\tilde{\delta}^h\|_D) = \mathrm{O}(h).$$

In analogy to Lemma 7, the boundary terms are equally estimated by $\mathrm{O}(h)$. Using Lemma 14 we see that

$$\|\hat{p} - p^0\|_\infty = \mathrm{O}(h) \tag{62}$$

and the estimate for $\|\hat{\sigma} - \sigma^0\|_\infty$ directly follows.

Consider next $\Delta^1(\hat{\sigma} - \sigma^0)$:

$$\begin{aligned} (\Delta^1(\hat{\sigma} - \sigma^0))_i &= B^T (\Delta^1(\hat{p} - p^0))_i \\ &= -B^T \left[(\nabla \hat{f}_{i+1})^T \hat{p}_{i+1} - (\nabla f^0_{i+1})^T p^0_{i+1} + \tilde{\delta}^h_{2i}\right] \\ &= \mathrm{O}(|\hat{x}_{i+1} - x^0_{i+1}| + |\hat{p}_{i+1} - p^0_{i+1}| + |\tilde{\delta}^h_{2i}|) \end{aligned}$$

so that $\|\Delta^1(\hat{\sigma} - \sigma^0)\|_2 = \mathrm{O}(h)$ follows from Theorem 1 and Lemma 2.

It remains to find a representation and estimates for $\Delta^2(\hat{\sigma} - \sigma^0)$:

$$\begin{aligned}(\Delta^2\hat{\sigma})_i &= h^{-1}B^T\Big((\Delta^1\hat{p})_i - (\Delta^1\hat{p})_{i-1}\Big)\\ &= -B^T\Big((\Delta^1[\nabla\hat{f}])_i^T\,\hat{p}_{i+1} + \nabla\hat{f}_i^T(\Delta^1\hat{p})_i\Big)\\ &= B^T\nabla\hat{f}_i^T\nabla\hat{f}_{i+1}^T\hat{p}_{i+1} - B^T h^{-1}\int_{\omega_i}\nabla_{xx}^2(\hat{p}_{i+1}^T\hat{f}[t])\,dt\cdot(\Delta^1\hat{x})_i\\ &=: \hat{P}_i + \hat{R}_i\hat{u}_i.\end{aligned}$$

where $\hat{f}[t] = f(\hat{x}(t))$, and $\hat{x}(t)$ stands for the linear interpolation to $\hat{x}_i, \hat{x}_{i+1}$ on ω_i. Therefore, the formula

$$\hat{R}_i - R_i = -B^T h^{-1}\int_{\omega_i}\Big(\nabla_{xx}^2(\hat{p}_{i+1}^T\hat{f}[t]) - \nabla_{xx}^2((p^0)^T f^0)_i\Big)\,dt\cdot B$$

leads to the desired estimates for the (discrete) L_2 and L_∞ norms. Similarly, $\|\hat{P} - P\|_r$ can be estimated for $r \in \{2, \infty\}$. □

References

1. Alt, W., Seydenschwanz, M.: Regularization and discretization of linear-quadratic control problems. Control Cybern. **40**(4), 903–920 (2011)
2. Alt, W., Seydenschwanz, M.: An implicit discretization scheme for linear-quadratic control problems with bang-bang solutions. Optim. Methods Softw. **29**(3), 535–560 (2014)
3. Alt, W., Baier, R., Gerdts, M., Lempio, F.: Error bounds for Euler approximation of linear-quadratic control problems with bang-bang solutions. Numer. Algebra Control Optim. **2**(3), 547–570 (2012)
4. Alt, W., Baier, R., Lempio, F., Gerdts, M.: Approximations of linear control problems with bang-bang solutions. Optimization **62**(1), 9–32 (2013)
5. Aronna, S.M., Bonnans, J.F., Dmitruk, A.V., Lotito, P.A.: Quadratic order conditions for bang-singular extremals. Numer. Algebra Control Optim. **2**(3), 511–546 (2012)
6. Aronna, M.S., Bonnans, J.F., Martinon, P.: A shooting algorithm for optimal control problems with singular arcs. J. Optim. Theory Appl. **158**(2), 419–459 (2013)
7. Dmitruk, A.V.: Quadratic conditions for a weak minimium for singular regimes in optimal control problems. Sov. Math. Dokl. **18**(2), 1 (1977)
8. Dmitruk, A.V.: Conditions of Jacobi type for Bolza's problem with inequalities. Math. Notes **35**, 427–435 (1984)
9. Dmitruk, A.V.: Jacobi type conditions for singular extremals. Control Cybern. **37**(2), 285–306 (2008)
10. Dmitruk, A.V., Milyutin, A.A.: A condition of Legendre type for optimal control problems, linear in control. In: Loffe, A., Reich, S., Shafrir, I. (eds.) Calculus of Variations and Optimal Control (Haifa, 1998), pp. 49–61. Chapman & Hall/CRC, Boca Raton (2000)
11. Dobrowolski, M.: Angewandte Funktionalanalysis. Springer, Berlin (2006)
12. Dontchev, A.L., Hager, W.W.: The Euler approximation in state constrained optimal control. Math. Comput. **70**(233), 173–203 (2001)
13. Dontchev, A.L., Hager, W.W., Malanowski, K.: Error bounds for Euler approximation of a state and control constrained optimal control. Numer. Funct. Anal. Optim. **21**, 653–682 (2000)
14. Felgenhauer, U.: Structural stability investigation of bang-singular-bang optimal controls. J. Optim. Theory Appl. **152**(3), 605–631 (2012)
15. Felgenhauer, U.: Stability analysis of variational inequalities for bang-singular-bang controls. Control Cybern. **42**(3), 557–592 (2013)

16. Fischer, A.: An NCP-function and its use for the solution of complementarity problems. In: Du, D.Z., Qi, L., Womersley, R.S. (eds.) Recent Advances in Nonsmooth Optimization, pp. 88–105. World Scientific Publishers, Minneapolis (1995)
17. Fischer, A.: A Newton-type method for positive-semidefinite linear complementarity problems. J. Optim. Theory Appl. **46**(3), 585–608 (1995)
18. Fischer, A., Jiang, H.: Merit functions for complementarity and related problems: a survey. Comput. Optim. Appl. **17**, 159–182 (2000)
19. Gabasov, R., Kirillova, F.M.: High order necessary conditions for optimality. SIAM J. Control **10**(1), 127–168 (1972)
20. Goh, B.S.: The second variation for the singular Bolza problem. SIAM J. Control **4**(2), 309–325 (1966)
21. Goh, B.S.: Necessary conditions for singular extremals involving multiple control variables. SIAM J. Control **4**(4), 716–731 (1966)
22. Grigorieff, R.D.: Diskrete Approximation von Eigenwertproblemen, I. Qualitative Konvergenz. Numer. Math. **24**(4), 355–374 (1975)
23. Gronwall, T.H.: Note on the derivatives with respect to a parameter of the solutions of a system of differential equations. Ann. Math. (2) **20**(4), 292–296 (1919)
24. Haunschmied, J.L., Pietrus, A., Veliov, V.M.: The Euler method for linear control systems revisited. In: Lirkov, I., Margenov, S., Waśniewski, J. (eds.) Large-Scale Scientific Computing. Lecture Notes in Computer Science, pp. 90–97. Springer, Heidelberg (2014)
25. Jiang, H.: Smoothed Fischer-Burmeister equation methods for the complementarity problem, Technical Report, Department of Mathematics, The University of Melbourne, Melbourne (1997)
26. Kelley, H.J.: A second variation test for singular extremals. AIAA J. **2**(8), 1380–1382 (1964)
27. Kinderlehrer, D., Stampacchia, G.: An Introduction to Variational Inequalities and their Applications. Academic Press, New York (1980)
28. Korytowski, A., Szymkat, M., Maurer, H., Vossen, G.: Optimal control of a fedbatch fermentation process: numerical methods, sufficient conditions, and sensitivity analysis, In: 47th IEEE Conference on Decision and Control, Cancun (Mexico) 2008, IEEE, Control Systems Society, 1551–1556 (2008). doi:10.1109/CDC.2008.4739078
29. Krener, A.J.: The high order maximum principle and its application to singular extremals. SIAM J. Control Optim. **15**(2), 256–292 (1977)
30. Ledzewicz, U., Schättler, H.: Antiangiogenic therapy in cancer treatment as an optimal control problem. SIAM J. Control Optim. **46**(3), 1052–1079 (2007)
31. Ledzewicz, U., Maurer, H., Schättler, H.: Bang-bang and singular controls in a mathematical model for combined anti-angiogenic and chemotherapy treatments, In: 48th IEEE Conference on Decision and Control, Shanghai, China, IEEE Control Systems Society, 2280–2285 (2009). doi:10.1109/CDC.2009.5400735
32. Malanowski, K., Büskens, C., Maurer, H.: Convergence of approximations to nonlinear control problems. In: Fiacco, A.V. (ed.) Mathematical Programming with Data Perturbation. Lecture Notes in Pure and Applied Mathematics. Marcel Dekker, Inc., New York (1997)
33. Maurer, H.: Numerical solution of singular control problems using multiple shooting techniques. J. Optim. Theory Appl. **18**(2), 235–257 (1976)
34. Maurer, H., Tarnopolskaya, T., Fulton, N.L.: Singular controls in optimal collision avoidance for participants with unequal linear speeds. ANZIAM J. Electron. Suppl. **53**(C), C1–C18 (2011)
35. Milyutin, A.A., Osmolovskii, N.P.: Calculus of Variations and Optimal Control. American Mathematical Society, Providence (1998)
36. Moyer, H.G.: Sufficient conditions for a strong minimum in singular control problems. SIAM J. Control **11**(4), 620–636 (1973)
37. Poggiolini, L., Stefani, G.: Minimum time optimality for a bang-singular arc: second-order sufficient conditions, In: 44th IEEE Conference on Decision and Control and European Control Conference, CDC-ECC 2005, Sevilla, 1433–1438 (2005). doi:10.1109/CDC.2005.1582360
38. Poggiolini, L., Stefani, G.: Sufficient optimality conditions for a bang-singular extremal in the minimum time problem. Control Cybern. **37**(2), 469–490 (2008)
39. Poggiolini, L., Stefani, G.: Bang-singular-bang extremals: sufficient optimality conditions. J. Dyn. Control Syst. **17**(4), 469–514 (2011)
40. Poggiolini, L., Stefani, G.: Structural stability for bang-singular-bang extremals in the minimum time problem. SIAM J. Control Optim. **51**(5), 3511–3531 (2013)

41. Robinson, S.M.: Generalized equations and their solutions, I. Basic theory. In: Robinson, S.M. (ed.) Point-to-Set Maps and Mathematical Programming, Mathematical Programming Study, pp. 128–141. Dekker, New York (1979)
42. Robinson, S.M.: Strongly regular generalized equations. Math. Oper. Res. **5**, 43–62 (1980)
43. Robinson, S.M.: Generalized equations and their solutions, II. Applications to nonlinear programming. In: Robinson, S.M. (ed.) Optimality and Stability in Mathematical Programming. Mathematical Programming Studies, pp. 200–221. Dekker, New York (1982)
44. Stefani, G.: Strong optimality of singular trajectories. In: Ancona, F., et al. (eds.) Geometric Control and Nonsmooth Analysis. Applied Mathematical Sciences Series, pp. 300–327. World Science Publisher, Hackensack (2008)
45. Veliov, V.: Error analysis of discrete approximations to bang-bang optimal control problems: the linear case. Control Cybern. **34**(3), 967–982 (2005)
46. Veliov, V.: Approximations with error estimates for optimal control problems for linear system. In: Lirkov, I., Margenov, S., Waśniewski, J. (eds.) Large-Scale Scientific Computing. Lecture Notes in Computer Science, pp. 263–270. Springer, Berlin (2006)
47. Vossen, G.: Numerische Lösungsmethoden, hinreichende Optimalitätsbedingungen und Sensitivitätsanalyse für optimale bang-bang und singuläre Steuerungen, Ph.D. Thesis, Faculty of Mathematics and Science, University Münster (2005)
48. Vossen, G.: Switching time optimization for bang-bang and singular controls. J. Optim. Theory Appl. **144**, 409–429 (2010)
49. Zelikin, M.I., Borisov, V.F.: Theory of Chattering Control. Birkhäuser, Boston (1994)