On the shooting algorithm for optimal control problems with state constraints

Audrey Hermant

Thèse pour l'obtention du titre de

DOCTEUR DE L'ÉCOLE POLYTECHNIQUE

Spécialité : Mathématiques appliquées

présentée par

Audrey Ledoux ép. HERMANT

Sur l'algorithme de tir pour les problèmes de commande optimale avec contraintes sur l'état

Soutenue le 5 septembre 2008.

Composition du jury :

Grégoire Allaire	CMAP	*Président du jury*
J. Frédéric Bonnans	CMAP et INRIA	*Directeur de thèse*
Nicolas Petit	ENSMP	*Examinateur*
Marc Quincampoix	U. Brest	*Rapporteur*
Emmanuel Trélat	U. Orléans	*Examinateur*
David Vissière	DGA	*Examinateur*
Vera Zeidan	Michigan State U.	*Rapporteur*

Résumé

Cette thèse s'intéresse au problème de commande optimale (déterministe) d'une équation différentielle ordinaire soumise à une ou plusieurs contraintes sur l'état, d'ordres quelconques, dans le cas où la condition forte de Legendre-Clebsch est satisfaite. Le principe du minimum de Pontryaguine fournit une condition d'optimalité nécessaire bien connue. Dans cette thèse, on obtient premièrement une condition d'optimalité suffisante du second ordre la plus faible possible, c'est-à-dire qu'elle est aussi proche que possible de la condition nécessaire du second ordre et caractérise la croissance quadratique. Cette condition nous permet d'obtenir une caractérisation du caractère bien posé de l'algorithme de tir en présence de contraintes sur l'état. Ensuite on effectue une analyse de stabilité et de sensibilité des solutions lorsque l'on perturbe les données du problème. Pour des contraintes d'ordre supérieur ou égal à deux, on obtient pour la première fois un résultat de stabilité des solutions ne faisant aucune hypothèse sur la structure de la trajectoire. Par ailleurs, des résultats sur la stabilité structurelle des extrémales de Pontryaguine sont donnés. Enfin, ces résultats d'une part sur l'algorithme de tir et d'autre part sur l'analyse de stabilité nous permettent de proposer, pour des contraintes sur l'état d'ordre un et deux, un algorithme d'homotopie dont la nouveauté est de déterminer automatiquement la structure de la trajectoire et d'initialiser les paramètres de tir associés.

Mots clés Commande optimale, contrainte sur l'état, condition d'optimalité du second ordre nécessaire ou suffisante, algorithme de tir, analyse de stabilité et sensibilité, méthode d'homotopie.

Abstract

This thesis deals with (deterministic) optimal control problems of an ordinary differential equation subject to one or several state constraints, of arbitrary orders, in the case when the strengthened Legendre-Clebsch condition is satisfied. Pontryagin's minimum principle provides us with a well-known first-order optimality condition. In this thesis we first obtain a second-order sufficient optimality condition which is the weakest possible, i.e. which is as close as possible to the second-order necessary condition and characterizes quadratic growth. This condition allows us to obtain a characterization of the well-posedness of the shooting algorithm in presence of state constraints. Then stability and sensitivity analysis of solutions under perturbation of the data is investigated. We obtain for the first time stability results for state constraints of order greater than or equal to two that make no assumption on the structure of the trajectory. Moreover, results on structural stability of Pontryagin's extremals are given. Finally, the above results on the well-posedness of the shooting algorithm and on stability analysis allow us to design a new continuation method, for state constraints of first- and second-order, whose novelty is to automatically detect the structure of the trajectory and initialize the associated shooting parameters.

Keywords Optimal control, state constraint, necessary or sufficient second-order optimality condition, shooting algorithm, stability and sensitivity analysis, continuation method.

Remerciements

Je remercie très sincèrement mon directeur de thèse Frédéric Bonnans pour m'avoir accueillie dans son équipe et accompagnée pendant mon stage de master et ces trois années de thèse. Je le remercie vivement pour tout le temps qu'il m'a consacré et pour nos nombreuses discussions. J'ai pu bénéficier de sa grande culture et j'ai beaucoup appris à ses côtés, des mathématiques bien sûr mais aussi tout un métier. Cette thèse ne serait pas ce qu'elle est sans ses conseils et son soutien.

Je souhaite exprimer toute ma gratitude à Marc Quincampoix et Vera Zeidan pour avoir accepté la tâche de rapporteur. Je les remercie pour leur lecture attentive du manuscrit et leurs remarques intéressantes.

Je suis très reconnaissante à Grégoire Allaire, Nicolas Petit, Emmanuel Trélat et David Vissière d'avoir accepté de faire partie de mon jury de thèse. C'est un grand honneur pour moi de les avoir dans mon jury.

Je remercie également Laurent Bourgeois, Frédéric Jean, Francis Maisonneuve et tout particulièrement Hasnaa Zidani pour m'avoir permis de faire de l'enseignement pendant ma thèse. Cette expérience a été très enrichissante pour moi.

Mes remerciements vont également à tous les doctorants et post-doctorants de l'ex projet SYDOCO à l'INRIA et du CMAP à l'Ecole Polytechnique pour la bonne ambiance qui régnait parmi nous. Leur présence a été très importante pour moi tout au long de ces trois années de thèse. Plus particulièrement, merci à Stefania, Elisabeth, Mohamed et Pierre.

Par-dessus tout, je voudrais remercier mon mari Benjamin pour son amour et sa patience. Enfin, je souhaite remercier mon fils Ulysse qui nous a fait l'immense joie de venir nous rejoindre pendant ma dernière année de thèse.

Table des matières

Introduction

0.1 Présentation du sujet

0.1.1 Introduction générale à la commande optimale

Dans cette thèse, on s'intéresse au problème de commande optimale déterministe d'une équation différentielle ordinaire, pouvant s'écrire sous la forme suivante :

$$\min_{(u,y)} \quad \int_0^T \ell(u(t), y(t))\mathrm{d}t + \phi(y(T)) \tag{0.1}$$

$$\text{sous contrainte} \quad \dot{y}(t) = f(u(t), y(t)) \quad \text{p.p. sur } [0,T], \quad y(0) = y_0. \tag{0.2}$$

L'*état*, c'est-à-dire le système physique que l'on souhaite commander, est représenté par la variable y, appartenant à l'espace $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$. On peut agir sur cet état indirectement, via la *commande* (ou contrôle), représentée par la variable u, que l'on peut choisir dans un ensemble de commandes admissibles

$$u \in \mathcal{U}_{ad} := \{u \in L^\infty(0,T;\mathbb{R}^m) \ : \ u(t) \in U \text{ pour p.p. } t \in [0,T]\} \tag{0.3}$$

avec U un convexe fermé (éventuellement compact) de $\mathbb{R}^m$. L'action de la commande sur l'état est modélisée, ici, par une équation différentielle ordinaire (0.2). Parmi toutes les trajectoires (u,y) admissibles (c'est-à-dire qui satisfont l'équation d'état (0.2) avec $u \in \mathcal{U}_{ad}$), on en cherche une qui minimise une certaine fonction de coût (0.1).

Le problème peut de plus être soumis à un certain nombre de contraintes, par exemple : des contraintes sur l'état initial et/ou final, sous la forme $\Psi(y(0), y(T)) \leq 0$, des contraintes mixtes sur la commande et sur l'état, du type $c(u(t), y(t)) \leq 0$ p.p. sur $[0,T]$, ou encore des contraintes (dites pures) sur l'état $g(y(t)) \leq 0$ pour tout $t \in [0,T]$. Ces dernières font l'objet de cette thèse.

Les problèmes de commande optimale ont des applications dans de nombreux domaines, par exemple optimisation de trajectoire, robotique, chimie, biologie, économie... Pour résoudre ces problèmes, deux grandes théories ont émergées indépendamment depuis une cinquantaine d'années : le principe du minimum de Pontryaguine et le principe de la programmation dynamique de Bellman. Avant de présenter brièvement ces théories, introduisons le *Hamiltonien* $H : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$,

$$H(u,y,p) \ := \ \ell(u,y) + pf(u,y). \tag{0.4}$$

Principe du Minimum de Pontryaguine

La première théorie, basée sur le *Principe du minimum* de Pontryaguine (PMP) [116] à la fin des années 50, donne une condition *nécessaire* d'optimalité. Si $(u,y) \in \mathcal{U}_{ad} \times \mathcal{Y}$ est une

solution optimale du problème (0.1)-(0.2), alors il existe $p \in W^{1,\infty}(0,T;\mathbb{R}^{n*})$, appelé état adjoint, tel que, p.p. sur $[0,T]$,

$$\begin{aligned} \dot{y}(t) &= f(u(t),y(t)), \qquad y(0)=y_0, && (0.5)\\ -\dot{p}(t) &= H_y(u(t),y(t),p(t)), \qquad p(T)=\phi_y(y(T)) && (0.6)\\ u(t) &\in \operatorname*{argmin}_{w\in U} H(w,y(t),p(t)). && (0.7) \end{aligned}$$

Des versions du principe du minimum existent bien sûr en présence des diverses contraintes mentionnées plus haut, et aussi lorsque les données ne sont pas différentiables, voir à ce sujet Clarke [40], Vinter [126].

Principe de programmation dynamique de Bellman

La deuxième théorie est basée sur le principe de la programmation dynamique de Bellman [8] dans les années 60. La fonction valeur ϑ du problème, définie par

$$\begin{aligned} \vartheta(x,t) := \inf_{(u,y)} \{ \int_t^T \ell(u(s),y(s))\mathrm{d}s + \phi(y(T)) : \\ \dot{y}(s) = f(u(s),y(s)) \text{ p.p. } s\in[t,T],\ y(t)=x,\ u(s)\in U\} \end{aligned} \tag{0.8}$$

est solution d'une équation aux dérivées partielles non linéaire, dite équation de Hamilton-Jacobi-Bellman (HJB)

$$\begin{cases} \dfrac{\partial\vartheta}{\partial t}(x,t) + \inf\limits_{w\in U} H(w,x,\dfrac{\partial\vartheta}{\partial x}(x,t)) = 0 \qquad (x,t)\in\mathbb{R}^n\times(0,T), \\ \vartheta(x,T)=\phi(x). \end{cases} \tag{0.9}$$

Cette condition d'optimalité est *nécessaire et suffisante*. L'équation HJB est bien posée au sens de viscosité (Crandall-Lions [43]). Cette thèse n'aborde pas du tout cette approche, mais des références classiques sur le sujet sont Barles [7] et Bardi et Capuzzo-Dolcetta [6].

0.1.2 Méthodes numériques de résolution

Il existe différentes méthodes pour résoudre les problèmes de commande optimale, chacune avec ses avantages et inconvénients. Le choix de la méthode dépend du problème considéré.

Méthodes directes

La méthode la plus couramment employée consiste à discrétiser les équations du problème, et ainsi on se ramène à un problème de programmation non linéaire (NLP), c'est-à-dire un problème d'optimisation non linéaire en dimension finie. Le problème discrétisé peut ensuite être résolu par n'importe quel algorithme d'optimisation en dimension finie, par exemple par programmation quadratique séquentielle (SQP), voir par exemple Betts [12], Bonnans et Launay [22], ou par une méthode de points intérieurs, voir Laurent-Varin et al. [84].

L'avantage des méthodes directes est qu'elles sont très faciles à appliquer, et relativement robustes à l'initialisation. On peut traiter un système avec un grand nombre de variables d'état. Leur précision est limitée par la précision de la discrétisation, donc le nombre de variables utilisées, et peut s'avérer insuffisante pour certains problèmes, par exemple pour calculer des trajectoires aérospatiales, fortement instables et requérant une grande précision.

Les méthodes directes fournissent une trajectoire et une commande en boucle ouverte (u en fonction du temps). Des algorithmes récents basés sur les méthodes directes permettent un calcul temps réel de type feedback et robuste aux perturbations et sont efficaces en pratique, voir par exemple [46, 47] (Nonlinear Model Predictive Control), [33, 32] (algorithme basé sur le calcul 'offline' des dérivées directionnelles des solutions par rapport au paramètre de perturbation).

Méthodes de tir

Les méthodes de tir exploitent la forme particulière des conditions d'optimalité données par le PMP. Sous certaines hypothèses (Hamiltonien fortement convexe par rapport à la commande), le principe du minimum (0.7) permet d'exprimer la commande comme une fonction de l'état et de l'état adjoint

$$u(t) = \Upsilon(y(t), p(t)) \qquad t \in [0, T]. \tag{0.10}$$

La condition nécessaire d'optimalité se réduit alors aux équations d'état et d'état adjoint (0.5)-(0.6), desquelles u est éliminé par (0.10). On obtient alors un système au deux bouts, puisqu'on a une condition initiale en y et une condition finale en p. L'idée de l'algorithme de tir (voir par exemple Stoer et Bulirsch [125]) est d'introduire une inconnue, la valeur initiale de l'adjoint p_0, et de considérer la fonction de tir qui à p_0 associe la condition finale $p(T) - \phi_y(y(T))$, où (y, p) est solution du problème de Cauchy sur $[0, T]$:

$$\left\{ \begin{array}{rcll} \dot{y} & = & f(\Upsilon(y,p), y), & y(0) = y_0 \\ -\dot{p} & = & H_y(\Upsilon(y,p), y, p), & p(0) = p_0. \end{array} \right.$$

On se ramène donc par cette méthode à chercher un zéro d'une fonction de $\mathbb{R}^n$ dans $\mathbb{R}^n$, en utilisant par exemple un algorithme de Newton.

La méthode de tir a l'avantage d'être très précise, et son coût numérique est faible. Cependant la convergence nécessite un bon point initial p_0, qui est parfois difficile à obtenir dans la pratique. De plus, pour un problème avec contraintes, une connaissance a priori de la structure de la trajectoire optimale est requise, comme on le verra dans la suite. La méthode de tir fournit une trajectoire boucle ouverte et donc non robuste aux perturbations ; dans la pratique cette trajectoire peut être suivie en utilisant les techniques de suivi de trajectoire de l'automatique.

Dans cette thèse on s'intéresse tout particulièrement aux méthodes de tir avec contraintes.

Résolution de l'équation HJB

Cette méthode consiste à résoudre numériquement l'équation HJB (0.9) (voir par exemple [44, 124] et aussi [34]). Une fois la fonction valeur ϑ calculée, il faut ensuite reconstruire les trajectoires optimales.

Alors que les deux méthodes précédentes (méthodes directes et méthodes de tir) sont des méthodes locales, c'est-à-dire qu'elles peuvent converger vers un minimum local, la résolution de l'équation HJB possède l'avantage de fournir un minimum global. De plus, cette dernière méthode permet de calculer la trajectoire optimale en boucle fermée, c'est-à-dire d'obtenir la commande u en fonction de l'état y, et elle est donc robuste. Cependant son coût numérique très élevé la rend difficile à appliquer lorsque la dimension de l'espace d'état est élevée (typiquement supérieure à six). De plus, comme pour les méthodes directes, la précision obtenue est limitée.

0.1.3 Contraintes sur l'état

Dans cette thèse, on considère le problème de commande optimale (0.1)-(0.2) soumis à une contrainte distribuée sur l'état du type

$$g(y(t)) \leq 0 \qquad \text{pour tout } t \in [0,T], \tag{0.11}$$

où $g : \mathbb{R}^n \to \mathbb{R}$ est une fonction régulière. On note $(\mathcal{P})$ ce problème

$$(\mathcal{P}) \qquad \text{Minimiser (0.1) sous contraintes (0.2) et (0.11).} \tag{0.12}$$

Sur le plan numérique, la prise en compte de contraintes sur l'état n'introduit pas de difficulté supplémentaire lorsque l'on utilise une méthode directe. Ces contraintes peuvent également être prises en compte lors de la résolution de l'équation HJB. La fonction valeur vaut alors $+\infty$ à l'extérieur du domaine admissible.

Pour appliquer des méthodes de tir, les contraintes sur l'état posent des difficultés théoriques. L'algorithme de tir est basé sur le PMP dont l'énoncé en présence de contrainte sur l'état du type (0.11) est le suivant. Le multiplicateur associé à la contrainte sur l'état (vue comme une contrainte dans l'espace des fonctions continues) est une mesure de Radon, et l'état adjoint une fonction à variation bornée.

Théorème 0.1 (Principe du minimum avec contraintes sur l'état). *Soit (u,y) une solution optimale de $(\mathcal{P})$, qui satisfait la condition de qualification* [1]*. Alors il existe $(p, \mathrm{d}\eta) \in BV(0,T;\mathbb{R}^{n*}) \times \mathcal{M}[0,T]$ tels que, p.p. sur $[0,T]$,*

$$\begin{aligned}
\dot{y}(t) &= f(u(t),y(t)), \quad y(0) = y_0, && (0.13)\\
-\mathrm{d}p(t) &= H_y(u(t),y(t),p(t))\mathrm{d}t + \mathrm{d}\eta(t) g_y(y(t)), \quad p(T^+) = \phi_y(y(T)) && (0.14)\\
u(t) &\in \operatorname*{argmin}_{w\in U} H_u(w,y(t),p(t)) && (0.15)\\
g(y(t)) &\leq 0, \quad \mathrm{d}\eta \geq 0, \quad \int_{[0,T]} g(y(t))\mathrm{d}\eta(t) = 0. && (0.16)
\end{aligned}$$

Si (u,y) vérifie le PMP (0.13)-(0.16), on dit que c'est une *extrémale de Pontryaguine*. On parle de *point stationnaire* lorsque (u,y) vérifie (0.13), (0.14), (0.16) et la condition ci-dessous, plus faible que (0.15) (lorsque $U = \mathbb{R}^m$)

$$0 = H_u(u(t),y(t),p(t)) \qquad \text{p.p. } t \in [0,T]. \tag{0.17}$$

Ce principe du minimum ne permet pas d'appliquer directement un algorithme de tir (voir ci-après la section 0.2.3). Une reformulation du principe du miminum est pour cela nécessaire, voir [28, 68, 98]. Deux notions sont utilisées dans cette reformulation : la *structure* de la trajectoire, supposée connue *a priori*, et l'*ordre* de la contrainte sur l'état.

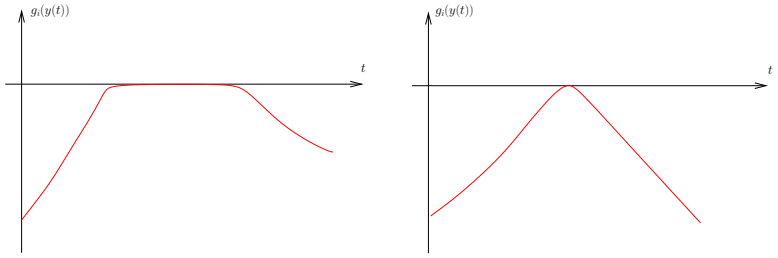
Structure de la trajectoire

Par structure de la trajectoire, on entend la structure de l'*ensemble de contact* de la contrainte

$$I(g(y)) := \{t \in [0,T] : g(y(t)) = 0\}, \tag{0.18}$$

[1] Il existe $v \in \mathcal{U}_{ad}$ tel que $g_y(y(t))z_{v-u}(t) < 0$ pour tout t tel que $g(y(t)) = 0$, où z_{v-u} est solution de l'équation d'état linéarisée $\dot{z}_{v-u} = f_u(u,y)(v-u) + f_y(u,y)z_{v-u}$ sur $[0,T]$, $z_{v-u}(0) = 0$.

i.e. l'ensemble des temps pour lesquels la contrainte est saturée. On appelle *arc frontière* (resp. *arc intérieur*) un intervalle de temps maximal (de mesure non nulle) sur lequel $g(y(t)) = 0$ (resp. $g(y(t)) < 0$). Les extrémités d'un arc frontière $[\tau_{en}, \tau_{ex}]$ sont appelées *point d'entrée* et *point de sortie*. Si la contrainte est active en un point localement isolé, on parle de *point de contact isolé*. Ceci est représenté sur la figure 0.1. On appelle les points d'entrée, de sortie et de contact isolé des *instants de jonction* entre arcs.



(a) Arc frontière (b) Point de contact isolé

FIG. 0.1 – Structure d'une trajectoire

Ordre de la contrainte sur l'état

L'*ordre* de la contrainte sur l'état (voir par exemple Bryson et al. [29]) est le plus petit nombre de dérivations de $t \mapsto g(y(t))$, lorsque y satisfait la dynamique (0.2), permettant de faire apparaître une dépendance explicite en la variable de commande u. Dans toute la thèse, on notera $q \in \mathbb{N}^*$ l'ordre de la contrainte sur l'état.

Par exemple, si la dynamique et la contrainte se mettent sur la forme canonique suivante :

$$\left\{\begin{array}{rcl} \dot{y}_1(t) &=& y_2(t) \\ &\vdots& \\ \dot{y}_{q-1}(t) &=& y_q(t) \\ \dot{y}_q(t) &=& u_1(t) \\ \dot{y}_j(t) &=& f_j(u(t), y(t)),\ j = q+1, \ldots, n \end{array}\right. , \qquad g(y(t)) = y_1(t) \leq 0 \tag{0.19}$$

alors la contrainte est d'ordre q. On voit que $\frac{\mathrm{d}^j}{\mathrm{d}t^j} g(y(t)) = y_{1+j}(t)$ ne dépend pas de u pour tout $j < q$ et $\frac{\mathrm{d}^q}{\mathrm{d}t^q} g(y(t)) = y_1^{(q)} = u_1(t)$. Plus généralement, pour une contrainte d'ordre q on peut écrire que

$$\frac{\mathrm{d}^j}{\mathrm{d}t^j} g(y(t)) =: g^{(j)}(y(t)), \quad 1 \leq j < q, \qquad \frac{\mathrm{d}^q}{\mathrm{d}t^q} g(y(t)) =: g^{(q)}(u(t), y(t)) \tag{0.20}$$

pour des certaines fonctions $g^{(j)} : \mathbb{R}^n \to \mathbb{R}$ et $g^{(q)} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$.

Relation entre structure de la trajectoire et ordre de la contrainte

Il est connu que la structure d'une trajectoire dépend fortement de l'ordre de la contrainte. Considérons par exemple les problèmes

$$\min \int_0^1 \left(\frac{u(t)^2}{2} - y(t) \right) \mathrm{d}t, \qquad y^{(q)}(t) = u(t), \qquad y(t) \leq h \tag{0.21}$$

pour des conditions initiales et finales données, et $h > 0$ un paramètre. Des exemples voisins ont été résolus (analytiquement) dans Bryson et al. [29] et Jacobson et al. [75]. Pour $q = 1, 2, 3$, nous avons tracé sur la figure 0.2 la famille de solutions lorsque h diminue (la trajectoire est de plus en plus contrainte). Les solutions ont été obtenues par une méthode de tir. On constate que l'évolution de la structure des trajectoires diffère selon l'ordre de la contrainte.

Pour $q = 1$ (figure 0.2(a)), la contrainte devient active pour $h = 1/8$ en un point de contact isolé en $t = 1/2$ qui se transforme immédiatement en un arc frontière pour $h < 1/8$.

Pour $q = 2$ (figure 0.2(b)), la contrainte est active d'abord en un point de contact isolé en $t = 1/2$ pour $0.167535 \leq h \leq 0.252604$, puis en un arc frontière pour $h < 0.167535$.

Pour $q = 3$ (figure 0.2(c)), la contrainte est active d'abord en un point de contact isolé en $t = 1/2$ pour $0.1750022 \leq h \leq 0.2500217$, puis en deux points de contact isolés pour $h < 0.1750022$.

0.2 Résumés des résultats de la thèse

0.2.1 Cadre de travail et hypothèses

Dans toute cette thèse, on étudie le problème de commande optimale avec contrainte sur l'état $(\mathcal{P})$ défini en (0.12). On s'intéresse à l'approche basée sur le principe du minimum de Pontryaguine et à la résolution par des méthodes de tir. On suppose de plus dans un premier temps que $\mathcal{U}_{ad} = L^\infty(0,T) =: \mathcal{U}$, i.e. la commande est non contrainte et à valeur scalaire. Dans le chapitre 4 on considérera le cas d'une commande et d'une contrainte g à valeurs vectorielles, et de contraintes mixtes sur la commande et sur l'état. Noter qu'en considérant le temps t comme variable d'état (toujours possible en introduisant une nouvelle variable d'état y_{n+1} vérifiant $\dot{y}_{n+1} = 1$, $y_{n+1}(0) = 0$ et donc $y_{n+1}(t) = t$) les problèmes non autonomes (avec données dépendant du temps) sont pris en compte.

Comme le principe de Pontryaguine ne fournit qu'une condition *nécessaire* d'optimalité (du premier ordre), on est amené naturellement à travailler sur des conditions *suffisantes* d'optimalité, et en particulier des conditions du second ordre. Ces dernières, comme nous le verrons dans cette thèse, sont au coeur de nombreux autres résultats, comme l'analyse de stabilité et sensibilité des solutions (i.e. comment se comportent les solutions si l'on perturbe les données du problème), à laquelle une grande partie de la thèse est consacrée, ainsi que l'analyse de convergence des algorithmes (par exemple, l'algorithme de tir, mais également la convergence des schémas de discrétisation, voir par exemple [54]).

On fera les hypothèses suivantes.

(A0) Les données $\ell : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, $\phi : \mathbb{R}^n \to \mathbb{R}$, $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}$ sont différentiables autant de fois que nécessaire (typiquement, de classe C^{2q} où q est l'ordre de la contrainte sur l'état, à dérivées secondes localement lipchitziennes si $q = 2$) et la dynamique f est lipschitzienne.

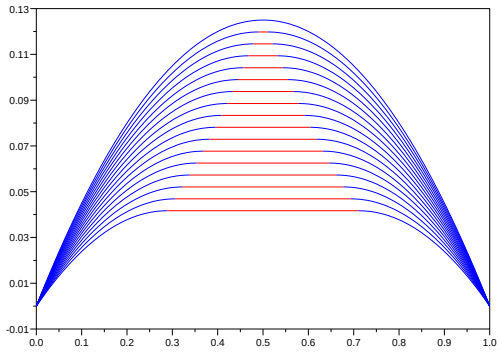
(A1) La condition initiale (fixée) $y_0 \in \mathbb{R}^n$ satisfait $g(y_0) < 0$.

(A2) Le Hamiltonien est uniformément fortement convexe par rapport à la commande le long de la trajectoire, i.e.

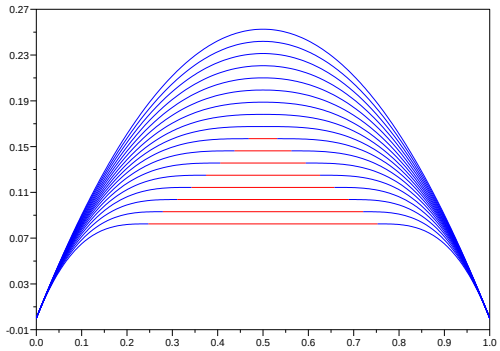
$$\exists\, \alpha > 0, \quad H_{uu}(\hat{u}, y(t), p(t)) \geq \alpha \quad \text{pour tout } \hat{u} \in \mathbb{R}^m \text{ et tout } t \in [0,T].$$

(A3) La contrainte sur l'état est d'ordre fini $q \in \mathbb{N}^*$ et régulière sur un voisinage de l'ensemble de contact $I(g(y))$, i.e.

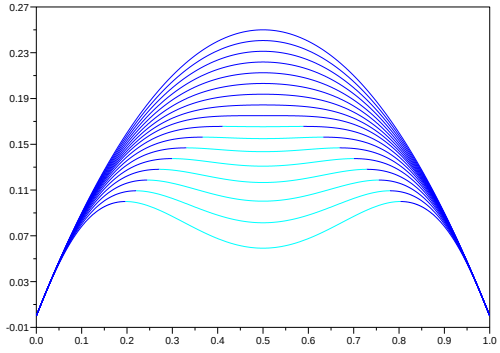
$$\exists\, \gamma, \varepsilon > 0, \qquad |g_u^{(q)}(u(t), y(t))| \geq \gamma, \quad \text{pour tout } t : \operatorname{dist}\{t, I(g(y))\} < \varepsilon.$$



(a) $q = 1$



(b) $q = 2$



(c) $q = 3$

FIG. 0.2 – Évolution de la structure de la trajectoire optimale du problème (0.21) en fonction de l'ordre q de la contrainte.

(A4) La trajectoire est composée d'un nombre fini d'arcs frontière et de points de contact isolés, et la contrainte est inactive au temps final, i.e. $g(y(T)) < 0$.

Sous les hypothèses ci-dessus, on a le résultat de régularité suivant, dû à Jacobson et al. [75] et Maurer [98].

Proposition 0.2. *Soit (u, y) une extrémale de Pontryaguine satisfaisant (A0)-(A4). Alors :*

(i) *La commande u est continue sur $[0, T]$, et sur l'intérieur des arcs de la trajectoire (frontière ou intérieur), u est C^q et η est continûement différentiable.*

(ii) *A un point de jonction τ :*

(a) *Si τ est un point d'entrée-sortie, alors u et ses dérivées jusqu'à l'ordre $q-2$ sont continues en τ, et si q est impair, alors la dérivée d'ordre $q-1$ de u est aussi continue et $[\eta(\tau)] = 0$.*

(b) *Si τ est un point de contact isolé, u et ses dérivées jusqu'à l'ordre $q-2$ sont continues en τ. De plus, si $q = 1$, alors $[\eta(\tau)] = 0$ et u et $\dot{u}$ sont continus en τ.*

Par cette proposition, à l'entrée et à la sortie d'un arc frontière, la fonction (du temps) $g(y(t))$ et ses dérivées jusqu'à un certain ordre $\hat{q}$ sont continues, où $\hat{q} := 2q - 2$ si q est pair et $\hat{q} := 2q - 1$ si q est impair. On dit qu'un point de contact isolé τ_{to} est *essentiel*, si

$$[\eta(\tau_{to})] > 0. \tag{0.22}$$

On fera également les hypothèses suivantes :

(A5) (i) (Conditions de tangentialité) En tout point d'entrée τ_{en} ou de sortie τ_{ex},

$$\frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}} g(y(t))|_{t=\tau_{en}^-, \tau_{en}^+} \neq 0. \tag{0.23}$$

(ii) Les points de contact isolés *essentiels* τ_{to} sont *réductibles*, i.e.

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(y(t))|_{t=\tau_{to}} < 0. \tag{0.24}$$

(A6) Complémentarité stricte sur les arcs frontière :

$$\operatorname{int} I(g(y)) \subset \operatorname{supp}(\mathrm{d}\eta).$$

Discussion des hypothèses Terminons cette section par quelques commentaires sur les hypothèses. Les problèmes avec données non régulières sont exclus de l'analyse par l'hypothèse (A0). Les hypothèses (A1) et (A3) de régularité de la contrainte et son analogue (hypothèse d'indépendance linéaire (4.30)) dans le cas de plusieurs contraintes sont classiques (voir dans la conclusion la section 7.1.4 pour un affaiblissement de cette dernière hypothèse).

L'hypothèse (A2), sans doute la plus restrictive, exclut un certain nombre de problèmes, fréquemment rencontrés dans les applications, pour lesquels la commande entre linéairement dans le coût et dans la dynamique. Cette classe de problèmes inclut les cas du contrôle bang-bang et des arcs singuliers. On peut affaiblir (A2) en supposant la commande u continue sur [0,T] et la condition forte de Legendre-Clebsch satisfaite

$$\exists\, \alpha > 0, \quad H_{uu}(u(t), y(t), p(t)) \geq \alpha \quad \text{pour tout } t \in [0, T]. \tag{0.25}$$

En revanche, cette dernière hypothèse est absolument essentielle pour les résultats de cette thèse. Si elle n'est plus satisfaite, les méthodes utilisées ne s'appliquent plus. Pour les conditions

du second ordre dans le cas de contrôle bang-bang et/où arcs singuliers, voir par exemple Dmitruk [49, 51], Osmolovskii et al. [109, 106, 110], Maurer et Osmolovskii [100].

L'hypothèse (A4) stipulant que la structure de la trajectoire est composée d'un nombre fini d'arcs frontières et points de contact isolés peut paraître, à juste titre, restrictive. Pour $q = 1, 2$, cela semble être une hypothèse raisonnable, dans le sens où elle est vérifiée en général sur les applications. De plus, dans [56] pour une contrainte d'ordre un et dans [48, chap. 2] pour une contrainte d'ordre deux, il a été montré que pour un problème de commande optimale linéaire quadratique autonome (i.e. le coût est quadratique, l'équation d'état est linéaire et la contrainte sur l'état linéaire, les données ne dépendant pas du temps), la solution est analytique par morceaux (et donc a effectivement un nombre fini d'arcs frontière et de points de contact isolés). Par contre, pour une contrainte d'ordre $q \geq 3$, on peut avoir une infinité de points de contact isolés, même pour un problème linéaire quadratique autonome, voir [118]. L'hypothèse (A4) paraît donc restrictive surtout pour des contraintes sur l'état d'ordre élevé (voir la conclusion section 7.1.3 où le cas d'un nombre infini de points de contact isolés est discuté).

Les hypothèses (A5) et (A6) sont 'génériquement' satisfaites, où 'génériquement' s'entend ici dans le sens où l'on peut toujours perturber légèrement les données pour que ces hypothèses soient satisfaites. Comme ces hypothèses (A5)-(A6) sont liées à la stabilité de la structure des solutions, elles deviennent restrictives dès lors que l'on s'intéresse justement aux cas où la structure des trajectoires n'est pas stable, cas qui se rencontrent inévitablement au cours des méthodes d'homotopie (voir la section 0.2.5 ci-après).

Enfin, l'extension des résultats dans le cas de contraintes sur l'état initial et/ou final est discutée dans la section 4.8.

0.2.2 Conditions d'optimalité du second ordre

Introduction

Comme cela a déjà été dit, le principe du minimum de Pontryaguine ne fournit qu'une condition nécessaire d'optimalité. Il est donc important de savoir si une extrémale de Pontryaguine est un optimum local ou non. C'est l'objet des *conditions suffisantes d'optimalité du second ordre*. Une théorie générale sur les conditions du second ordre est présentée dans Bonnans et Shapiro [24]. Par rapport à la théorie classique, les problèmes de commande optimale non linéaire font apparaître une difficulté bien connue, la divergence des deux-normes (*two-norms discrepancy*), voir [99]. Ceci est illustré sur l'exemple suivant [88, p.126].

Exemple 0.3. Soit

$$J(u) := \int_0^T (u^2(t) - 1)^2 \mathrm{d}t.$$

Alors J est deux fois différentiable sur $L^\infty(0, T)$ (mais pas sur $L^2(0, T)$), et sa dérivée seconde

$$D^2_{uu} J(u)(v, v) = 4 \int_0^T (3u^2(t) - 1)v^2(t) \mathrm{d}t$$

est une forme quadratique qui s'étend continûment à $v \in L^2(0, T)$. Tout minimiseur $\bar{u}$ de J sur $L^\infty(0, T)$ est tel que $|\bar{u}(t)| = 1$ pour p.p. $t \in [0, 1]$. En un quelconque de ces minimiseurs, on a

$$D^2_{uu} J(\bar{u})(v, v) = 8 \int_0^T v^2(t) \mathrm{d}t = 8\|v\|_2^2. \tag{0.26}$$

On constate que

$$\text{il n'existe pas } \alpha > 0 \text{ tel que } D^2_{uu}J(\bar{u})(v,v) \geq \alpha\|v\|^2_\infty \text{ pour tout } v \in L^\infty(0,T).$$

En revanche, par (0.26), la condition ci-dessus est naturellement satisfaite si on remplace la norme de L^∞ par celle de L^2, i.e.

$$\text{il existe } \alpha > 0 \text{ tel que } D^2_{uu}J(\bar{u})(v,v) \geq \alpha\|v\|^2_2 \text{ pour tout } v \in L^2(0,T). \tag{0.27}$$

On peut alors montrer que la condition suffisante du second ordre (0.27) implique que

$$\exists\, c, \rho > 0, \quad J(u) \geq J(\bar{u}) + c\|u-\bar{u}\|^2_2, \quad \forall u,\ \|u-\bar{u}\|_\infty < \rho,$$

i.e. $\bar{u}$ est un *minimiseur strict* de J sur un petit voisinage de L^∞ et satisfait la *condition de croissance quadratique* pour la norme L^2 (mais $\bar{u}$ n'est pas un minimum strict dans un voisinage L^2).

On voit sur cet exemple que l'on est amené à utiliser les deux espaces et les deux normes pour formuler des conditions du second ordre : L^∞, espace dans lequel les applications sont différentiables, et L^2, norme intervenant naturellement dans les conditions du second ordre et dans la croissance quadratique.

Les *conditions suffisantes du second ordre* servent à vérifier l'optimalité locale d'une trajectoire extrémale de Pontryaguine ou d'un point stationnaire, et elles jouent aussi un rôle important dans l'analyse des algorithmes (preuve de convergence et estimations d'erreurs) et dans l'analyse de stabilité et sensibilité des solutions. Il est donc intéressant d'avoir une condition suffisante la plus faible possible, et cela peut être réalisé en se rapprochant autant que possible de la *condition nécessaire du second ordre*. On parle en particulier de conditions "no-gap", lorsque les conditions nécessaire et suffisante du second ordre sont le plus proche possible, c'est-à-dire qu'elles ne diffèrent qu'entre une inégalité large et une inégalité stricte.

Pour les problèmes de commande optimale, des conditions du second ordre "no-gap" étaient connues pour les contraintes mixtes sur la commande et sur l'état [108, 105], mais pas pour les contraintes pures sur l'état. Pour ces dernières, des conditions suffisantes ont été obtenues dans [99, 89, 94, 95], et, indépendamment, des conditions nécessaires dans [80, 112, 113].

La difficulté propre aux contraintes sur l'état est la présence d'un terme supplémentaire apparaissant dans la condition nécessaire, appelé *terme de courbure*. Ce terme, découvert par Kawasaki [77] (voir aussi [114]), est dû à la présence d'un nombre infini de contraintes d'inégalités. Pour les contraintes sur la commande ou les contraintes mixtes, on peut montrer que ce terme de courbure est nul (c'est la théorie de la polyédricité, voir [24, section 3.2.3] et dans cette thèse la section 4.6). Pour les contraintes sur l'état, ce terme est a priori non nul. Par contre, seules des conditions suffisantes sans ce terme supplémentaire étaient connues. Dans cette thèse nous obtenons, pour la première fois, des conditions "no-gap" pour les contraintes sur l'état.

Résultat

Il est utile de réécrire le problème de commande optimale $(\mathcal{P})$ en fonction de la variable de commande uniquement, c'est-à-dire que l'état est vu comme une fonction de la commande, plus précisément $y = y_u$ où y_u désigne la solution (unique) de l'équation d'état (0.2). On obtient alors la forme abstraite suivante :

$$\min_{u\in\mathcal{U}} J(u), \qquad G(u) \in K \tag{0.28}$$

où le coût $J(u)$ est donné par (0.1) avec $y = y_u$, et la contrainte sur l'état (0.11) se réécrit $G(u) \in K$ avec $G(u) := g(y_u)$ et K le cône convexe fermé des fonctions continues à valeurs négatives sur $[0,T]$. On dit que u est une solution locale de $(\mathcal{P})$ satisfaisant la condition de *croissance quadratique*, si

$$\exists\, \alpha, \rho > 0, \quad J(\tilde{u}) \geq J(u) + \alpha \|\tilde{u} - u\|_2^2, \quad \forall\, \tilde{u} \in \mathcal{U}\ ;\ \|\tilde{u} - u\|_\infty \leq \rho, \ \ G(\tilde{u}) \in K. \tag{0.29}$$

La dérivée (Fréchet) au point u de $\mathcal{U} \to \mathcal{Y}$, $u \mapsto y_u$ est l'application $\mathcal{U} \to \mathcal{Y}$, $v \mapsto z_v$ où z_v est solution de l'équation d'état linéarisée

$$\dot{z}_v = f_u(u, y_u)v + f_y(u, y_u)z_v \quad \text{p.p. sur } [0,T], \quad z_v(0) = 0. \tag{0.30}$$

La forme quadratique, définie sur $L^2(0,T)$, impliquée dans les conditions du second ordre est la suivante. Ici p et η sont les multiplicateurs du Principe du minimum (0.13)-(0.16), et on note $\mathcal{T}_{to}^{ess}$ l'ensemble (supposé fini) des points de contact isolés essentiels (i.e. satisfaisant (0.22)) supposés réductibles (i.e. satisfaisant (0.24)) de la trajectoire (u, y) :

$$\begin{aligned} Q(v) \ := \ & \int_0^T [H_{uu}(u, y_u, p)(v, v) + 2H_{uy}(u, y_u, p)(v, z_v) + H_{yy}(u, y_u, p)(z_v, z_v)]\mathrm{d}t \\ & + \phi_{yy}(y_u(T))(z_v(T), z_v(T)) \ + \ \int_0^T g_{yy}(y_u)(z_v, z_v)\mathrm{d}\eta \\ & - \sum_{\tau \in \mathcal{T}_{to}^{ess}} [\eta(\tau)] \frac{(g_y^{(1)}(y_u(\tau))z_v(\tau))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(y_u(t))|_{t=\tau}}. \end{aligned} \tag{0.31}$$

Le terme apparaissant sur la dernière ligne de l'équation ci-dessus est le terme de courbure. On voit que ce terme ne fait intervenir que les points de contact isolés essentiels. Les arcs frontière n'ont pas de contribution. Par la proposition 0.2(ii)(b), ce terme de courbure est toujours nul pour les contraintes du premier ordre (n'ayant pas de points de contact isolés essentiels).

Enfin, le cône critique $C_{L^2}(u)$ dans L^2 utilisé dans les conditions du second ordre est défini comme l'ensemble des $v \in L^2(0,T)$ vérifiant les deux conditions ci-dessous :

$$g_y(y_u(t))z_v(t) \ = \ 0, \quad t \in \mathrm{supp}(\mathrm{d}\eta) \tag{0.32}$$

$$g_y(y_u(t))z_v(t) \ \leq \ 0, \quad t \in I(g(y)) \setminus \mathrm{supp}(\mathrm{d}\eta). \tag{0.33}$$

Le résultat principal est le suivant (voir les théorème 1.12, corollaire 1.15, théorèmes 1.18 et 1.27 ainsi que [18, Th. 2.2]).

Théorème 0.4. (i) *Soit (u, y) une solution locale de $(\mathcal{P})$, satisfaisant* (A0)-(A6)*. Alors :*

$$Q(v) \geq 0 \qquad \forall\, v \in C_{L^2}(u). \tag{0.34}$$

(ii) *Soit (u, y) une extremale de Pontryaguine satisfaisant* (A0)-(A6)*. Alors (u, y) est une solution locale de $(\mathcal{P})$ satisfaisant la condition de croissance quadratique* (0.29) *si et seulement si*

$$Q(v) > 0 \qquad \forall\, v \in C_{L^2}(u) \setminus \{0\}. \tag{0.35}$$

Pour la preuve de la condition nécessaire, on utilise la condition nécessaire du second ordre obtenue par Kawasaki [77]. On explicite le terme de courbure, dont une première expression avait été obtenue par Kawasaki [79] pour les contraintes de positivité dans l'espace des fonctions continues. Ce calcul est basé sur des développements de Taylor des fonctions (du temps)

$g(y_u(t))$ et $g_y(y_u(t))z_v(t)$. Les hypothèses (A4)-(A6) et les résultats cruciaux de la proposition 0.2(ii) permettent alors de calculer explicitement ce terme de courbure. Ainsi on obtient (0.34).

Pour obtenir la condition suffisante la plus proche possible de la condition nécessaire, on doit prendre en compte le terme de courbure, qui peut être non nul lorsque la contrainte est d'ordre $q \geq 2$ et qu'il y a des points de contact isolés essentiels. Pour cela, on utilise une méthode de *réduction*, connue en programmation semi-infinie [72] dans un cadre C^2. Cette approche est étendue à l'espace $W^{2,\infty}(0,T)$. Ceci consiste à dire que pour $\varepsilon, \delta > 0$ suffisamment petits, l'application

$$B_\infty(u,\delta) \to \mathbb{R}, \qquad \tilde{u} \mapsto g(y_{\tilde{u}}(\tau_{\tilde{u}}))$$

où $B_\infty(u,\delta)$ désigne la boule ouverte de centre u et rayon δ dans L^∞ et où $\tau_{\tilde{u}}$ est l'*unique point de maximum* de $g(y_{\tilde{u}})$ sur $(\tau-\varepsilon, \tau+\varepsilon)$ (τ étant un point de contact isolé de (u, y_u) satisfaisant (0.24)), est bien définie, de classe C^1, deux fois Fréchet différentiable en u. Reformulant ainsi la contrainte sur l'état au voisinage des points de contact isolés essentiels, on obtient dans la condition suffisante un terme supplémentaire, correspondant exactement au terme de courbure. Ceci nous permet d'obtenir la condition suffisante (0.35).

0.2.3 Étude de l'algorithme de tir

Introduction

L'algorithme de tir a été appliqué avec succès dans la littérature aux problèmes avec contraintes sur l'état, voir par exemple [115, 11, 27], mais des difficultés d'ordre théorique subsistent néanmoins. Pour appliquer l'algorithme de tir, une reformulation du PMP avec contraintes sur l'état est nécessaire. En effet, le principe de l'algorithme de tir est d'exprimer les variables algébriques (en l'occurence, u et $\dot{\eta}$ — noter que η est bien differentiable sur l'intérieur de chaque arc en vertu de la proposition 0.2(i)) en fonction des variables différentielles y et p. Sur un arc intérieur, $\dot{\eta} = 0$ et u s'obtient comme fonction de (y,p) par application du théorème des fonctions implicites à la relation (0.15) sous l'hypothèse (A2). Sur un arc frontière $[\tau_{en}, \tau_{ex}]$, on peut exprimer u comme fonction de y en appliquant sous l'hypothèse (A3) le théorème des fonctions implicites à la relation

$$g^{(q)}(u(t), y(t)) = 0, \qquad t \in [\tau_{en}, \tau_{ex}], \tag{0.36}$$

mais l'équation algébrique restante (0.15) ne permet pas d'exprimer $\dot{\eta}$ en fonction de (y,p), puisque $\dot{\eta}$ n'apparaît pas dans cette équation. C'est pourquoi on est amené à reformuler la condition d'optimalité, en considérant la relation (0.36) comme une contrainte mixte sur les arcs frontières $[\tau_{en}, \tau_{ex}]$, avec les q contraintes aux points d'entrée

$$g^{(j)}(y(\tau_{en})) = 0, \qquad j = 0, \ldots, q-1 \tag{0.37}$$

pour avoir l'équivalence avec la condition $g(y(t)) = 0$ sur $[\tau_{en}, \tau_{ex}]$. De même, à un point de contact isolé τ_{to}, on écrit que

$$g(y(\tau_{to})) = 0. \tag{0.38}$$

Ainsi, si l'on connaît *a priori* la structure de la trajectoire (nombre et ordre des arcs frontières et des points de contact isolés), on se ramène à un problème avec contraintes d'égalité, les inégalités $g(y(t)) < 0$ sur les arcs intérieurs, signe du multiplicateur) devant être vérifiées *a posteriori*. Le système d'optimalité de ce problème avec contraintes d'égalités (0.36)-(0.38)

fournit alors une *formulation alternative*, qui elle peut être résolue à l'aide d'un algorithme de tir. L'idée originale de cette reformulation est dûe à Bryson et al. [29, 28].

Le problème de cette reformulation est qu'elle ne prend en compte qu'une partie des conditions nécessaires d'optimalité. Ceci provient du fait que dans la formulation alternative, les instants de jonction sont considérés comme étant fixés, alors qu'ils sont en fait inconnus et doivent donc satisfaire certaines conditions d'optimalité. De plus, des conditions comme la continuité des dérivées de u aux points d'entrée/sortie, qui sont des conditions nécessaires d'optimalité par la proposition 0.2, ne sont pas prises en compte dans la formulation alternative. Ainsi, une partie des conditions d'optimalité est perdue, comme cela été montré par Jacobson et al. [75]. Des *conditions supplémentaires*, conditions nécessaires d'optimalité, non prises en compte dans la formulation alternative, et donc pas non plus dans l'algorithme de tir, doivent être vérifiées *a posteriori*. Si ce n'est pas le cas, on peut d'ores et déjà éliminer la solution trouvée qui n'est pas solution de la condition nécessaire d'optimalité, et donc a fortiori n'est pas une solution locale du problème.

La description précise de l'algorithme de tir est donnée dans la section 2.2. Disons seulement que pour chaque arc frontière et chaque point de contact isolé, des paramètres supplémentaires (dont les instants de jonction —inconnus— et des "paramètres de saut" de l'adjoint aux points d'entrée et de contact isolés, ces derniers pouvant être vu comme des multiplicateurs associés aux contraintes ponctuelles (0.37)-(0.38)) sont ajoutés comme inconnus de la fonction de tir, en plus de la valeur initiale de l'adjoint p_0.

Résultat

Les questions que nous nous sommes posées sont alors les suivantes : sous quelles conditions supplémentaires a-t-on précisément l'équivalence entre le PMP avec contraintes sur l'état et la formulation alternative (il en existe de nombreuses versions différentes dans la littérature, voir le tour d'horizon [68]) ? Certaines de ces conditions supplémentaires sont-elles automatiquement satisfaites ? Et enfin, cet algorithme de tir ne prenant en compte qu'une partie des conditions d'optimalité est-il bien posé (au sens où le jacobien de la fonction de tir est inversible) ?

On établit dans la proposition 2.10 l'équivalence entre la formulation alternative et les conditions supplémentaires d'une part, et le principe du minimum d'autre part. On montre de plus dans la proposition 2.15 que certaines des conditions supplémentaires sont automatiquement vérifiées. Enfin, nous obtenons le résultat principal suivant.

Théorème 0.5 (Th. 2.23). *Soit (u, y) une solution locale de $(\mathcal{P})$ satisfaisant (A0)-(A5) et la condition de complémentarité stricte $I(g(y)) = \mathrm{supp}(\mathrm{d}\eta)$. Alors l'algorithme de tir est bien posé (Jacobien de la fonction de tir inversible) au voisinage de (u, y) si et seulement si*

(i) *Si $q \geq 3$, il n'y a pas d'arcs frontière ;*

(ii) *La condition suffisante du second ordre* (0.35) *est satisfaite.*

Pour la preuve, on exprime la forme quadratique (0.31) en fonction des multiplicateurs utilisés dans la formulation alternative, et non plus des multiplicateurs du PMP p et η. On obtient une expression équivalente de (0.31) notée $Q^q(v)$. On calcule ensuite le Jacobien de la fonction de tir, et on montre qu'un élément dans le noyau est associé à une solution du système d'optimalité du problème linéaire-quadratique

$$\min_{v \in L^2(0,T)} Q^q(v), \quad \text{s.t. } v \in C_{L^2}(u) \tag{0.39}$$

(noter que l'on a supposé la complémentarité stricte, et donc le cône critique $C_{L^2}(u)$ se réduit aux $v \in L^2$ qui satisfont (0.32)). Ensuite on utilise les conditions de jonction de la proposition 0.2 pour montrer que (i) est nécessaire, puis les conditions du second ordre nécessaire et suffisante du théorème 0.4 pour conclure.

Le fait que l'algorithme de tir soit mal posé s'il y a des arcs frontière pour des contraintes d'ordre ≥ 3 semble venir du fait que pour les contraintes d'ordre élevé, les arcs frontières seraient en général précédés et suivis d'une infinité de points de contact isolés. Ceci est conjecturé d'après un exemple de Robbins [118] (cet exemple est rappelé dans l'annexe de la conclusion, section 7.2.2). Ainsi, les arcs frontière avec points d'entrée/sortie réguliers comme considérés dans cette thèse semblent être pour une contrainte d'ordre $q \geq 3$ un cas "pathologique", pour lequel l'algorithme de tir est mal posé.

0.2.4 Analyse de stabilité et de sensibilité

Introduction

Pour un problème d'optimisation avec contraintes d'égalités et données régulières (de classe C^2), lorsque les dérivées des contraintes sont "surjectives", un outil fondamental pour l'analyse de stabilité et sensibilité est le théorème des fonctions implicites, appliqué à la condition d'optimalité du premier ordre, sous une hypothèse de condition suffisante du second ordre [61, 60]. Ainsi, on peut montrer que les solutions sont C^1 par rapport au paramètre. Pour un problème avec contraintes d'inégalités, lorsqu'une hypothèse de "complémentarité stricte" est satisfaite, on peut parfois se ramener à un problème avec contraintes d'égalité et donc au cas précédent.

Pour des problèmes plus généraux dans des espaces de Banach du type (0.28), avec contrainte dans un cône convexe fermé K, un outil pour l'analyse de stabilité des systèmes d'optimalité est la théorie de la régularité forte de Robinson [121]. Cette théorie permet en particulier de s'affranchir de l'hypothèse de complémentarité stricte pour les problèmes avec contraintes d'inégalités. Lorsque l'hypothèse de complémentarité stricte n'est pas vérifiée, on sait en général que les solutions sont au mieux directionnellement différentiables. Le principe de la régularité forte de Robinson est le suivant. Si l'on peut montrer qu'un problème linéaire-quadratique, obtenu en linéarisant le problème non linéaire de départ, et perturbé d'une certaine manière, admet une unique solution qui est localement lipschitzienne par rapport au paramètre, alors on peut en déduire que localement, le problème non linéaire admet lui aussi une solution, localement unique, lipschitzienne par rapport au paramètre. Ainsi dans l'analyse on se ramène à étudier la stabilité des problèmes linéaire-quadratique (i.e. le coût est quadratique et la contrainte linéaire).

Pour les problèmes de commande optimale, le phénomène de "divergence des deux normes" (*two-norm discrepancy*, voir l'exemple 0.3) ne permet pas de pouvoir appliquer directement le résultat de régularité forte de Robinson [121]. On doit donc en utiliser des variantes. Une adaptation de ce résultat prenant en compte le problème des deux normes a été proposée par Malanowski [87]. Il faut encore travailler pour pouvoir prendre en compte les contraintes sur l'état, en raison de la faible régularité des multiplicateurs (à variation bornée). Ceci a été fait dans Malanowski [88] et dans Dontchev et Hager [53], *pour les contraintes du premier ordre uniquement* (pour lesquelles les multiplicateurs sont lipschitziens). Bien que les cadres théoriques utilisés dans ces deux articles diffèrent, les arguments qui permettent d'obtenir le résultat sont les mêmes. Les idées principales sont d'exploiter la régularité supplémentaire des solutions et des multiplicateurs du problème de départ (lipschitziens), et de considérer des perturbations du problème quadratique qui sont elles aussi plus régulières.

Pour l'analyse de stabilité et sensibilité des problèmes de commande optimale, une autre approche, plus simple au premier abord, peut être utilisée. Il s'agit de paramétriser le problème par un nombre fini de paramètres de tir et d'appliquer le théorème des fonctions implicites classique à la fonction de tir. Cette méthode a été appliquée par Malanowski et Maurer aux problèmes avec contraintes sur l'état du premier ordre dans [93] et d'ordre supérieur dans [94]. L'inconvénient de cette méthode, au contraire de l'approche 'régularité forte' présentée dans le paragraphe précédent, est qu'elle nécessite des hypothèses sur la structure de la trajectoire (A4) ainsi que des hypothèses de *complémentarité stricte uniforme* pour assurer la stabilité de la structure des solutions du problème perturbé.

Résultats

Nous avons considéré une classe de perturbations $(\mathcal{P}^\mu)$ de $(\mathcal{P})$ régulières (ici, μ désigne le paramètre de perturbation, et on suppose que $(\mathcal{P}) \equiv (\mathcal{P}^{\bar\mu})$ pour une certaine valeur $\bar\mu$ du paramètre). Plus précisément, les données dépendent de façon régulière (C^{2q}) du paramètre, et sont telles que l'ordre de la contrainte du problème perturbé reste le même que celui du problème de départ. Pour une analyse lorsque l'ordre de la contrainte varie, voir [81, 82].

Nous avons tout d'abord utilisé l'approche par le tir dans l'analyse de stabilité et sensibilité, en affaiblissant l'hypothèse de complémentarité stricte aux *points de contact isolés*. Dans ce cas *la structure des solutions n'est pas stable*, à la différence de [93, 94]. Cependant, nous avons montré que sous les hypothèses précédentes, si l'hypothèse de complémentarité stricte uniforme sur les arcs frontières est satisfaite, i.e.

$$\exists\, \beta > 0, \qquad \frac{\mathrm{d}\eta}{\mathrm{d}t} \geq \beta \quad \text{sur l'intérieur des arcs frontières} \tag{0.40}$$

alors les arcs frontières sont stables pour des contraintes d'ordre un et deux. Sous l'hypothèse supplémentaire suivante

$$\text{Tous les points de contact isolés } \tau_{to} \text{ sont } \textit{réductibles}\text{, i.e. satisfont (0.24)}, \tag{0.41}$$

nous obtenons le résultat suivant, le premier de ce type sur la stabilité structurelle des points stationnaires. Ce résultat est basé sur les théorèmes 3.4 et 6.8 pour la stabilité des arcs frontière pour les contraintes respectivement du premier ordre et du second ordre.

Théorème 0.6. *Soit $(\bar u, \bar y)$ un point stationnaire de $(\mathcal{P})$ satisfaisant (A0)-(A4),* (0.40)-(0.41) *et si la contrainte sur l'état est d'ordre $q \geq 3$ sans arc frontière. Alors il existe des voisinages $\mathcal{V}_\infty$ de $\bar u$ dans L^∞ et $\mathcal{W}$ de $\bar\mu$ tels que* tout point stationnaire (u, y) *de $(\mathcal{P}^\mu)$ avec $u \in \mathcal{V}_\infty$ et $\mu \in \mathcal{W}$ vérifie les propriétés suivantes :*

(i) *La contrainte sur l'état n'est pas active en dehors d'un voisinage de l'ensemble de contact $I(g(\bar y))$.*

(ii) *Au voisinage d'un arc frontière de $(\bar u, \bar y)$ $(q = 1, 2)$, (u, y) a un unique arc frontière.*

(iii) *Au voisinage d'un point de contact isolé essentiel de $(\bar u, \bar y)$ $(q \geq 2)$, (u, y) a un unique point de contact isolé (essentiel).*

(iv) *Au voisinage d'un point de contact isolé non essentiel de $(\bar u, \bar y)$,*

  (a) *si $q = 1$, ou bien la contrainte sur l'état n'est pas active, ou bien (u, y) a un unique point de contact isolé (non essentiel)* ou *un unique arc frontière,*

  (b) *si $q \geq 2$, ou bien la contrainte sur l'état n'est pas active, ou bien (u, y) a un unique point de contact isolé (essentiel ou non).*

Sous les hypothèses du théorème ci-dessus, il existe donc un nombre fini de structures possibles pour les solutions du problème perturbé. Des différences de structure peuvent se produire seulement là où la complémentarité stricte n'est pas satisfaite, i.e. aux points de contact isolés non essentiels. Avec le théorème 0.6, on peut alors conduire l'analyse de stabilité et de sensibilité par une méthode de tir, car il est possible, dans ce cas, de construire une même fonction de tir englobant les différentes structures possibles. Le cas le plus simple est celui des contraintes d'ordre ≥ 2 pour lesquelles un point de contact isolé peut seulement disparaître. Pour les contraintes d'ordre 1, la situation est plus compliquée car un point de contact isolé peut aussi se transformer en arc frontière. L'idée est alors de considérer dans la fonction de tir un point de contact isolé comme un arc frontière de longueur nulle. Nous appliquons ensuite la théorie de la régularité forte de Robinson [121] à la formulation de tir en résultant (sous forme d'une équation généralisée en dimension finie avec contraintes de complémentarité). Nous obtenons ainsi le résultat suivant, basé sur le théorème 0.6 et les théorèmes 3.11 et 2.34 pour les contraintes respectivement du premier ordre et d'ordre supérieur ou égal à deux.

Théorème 0.7. *Soit $(\bar{u}, \bar{y})$ un point stationnaire de $(\mathcal{P})$ satisfaisant (A0)-(A4), (0.40)-(0.41) et si la contrainte est d'ordre $q \geq 3$ sans arc frontière. Alors les propositions suivantes sont équivalentes.*

(i) *Pour toute perturbation suffisamment régulière $(\mathcal{P}^\mu)$ de $(\mathcal{P})$, il existe $\alpha, \rho > 0$ et un voisinage $\mathcal{W}$ de $\bar{\mu}$ tels que pour tout $\mu \in \mathcal{W}$, il existe un unique point stationnaire (u^μ, y^μ) de $(\mathcal{P}^\mu)$ avec $\|u^\mu - \bar{u}\|_\infty < \rho$ (et d'unique multiplicateurs associés (p^μ, η^μ)), et ce point stationnaire vérifie la condition de croissance quadratique uniforme*

$$J^\mu(u) \geq J^\mu(u^\mu) + \alpha \|u - u^\mu\|_2^2, \quad \forall\, u \in \mathcal{U}\ ;\ \|\tilde{u} - u\|_\infty \leq \rho, \quad G^\mu(u) \in K. \tag{0.42}$$

(ii) *La condition suffisante du second ordre forte ci-dessous est vérifiée :*

$$Q(v) > 0, \quad \forall\, v \in L^2(0,T) \setminus \{0\} \text{ satisfaisant } (0.32). \tag{0.43}$$

De plus, si (i) ou (ii) est satisfait, alors l'application $\mathcal{W} \to \mathcal{U} \times \mathcal{Y}$, $\mu \to (u^\mu, y^\mu)$ est lipschitzienne, et directionnellement différentiable dans l'espace $L^r(0,T) \times W^{1,r}(0,T;\mathbb{R}^n)$ pour tout $r \in [1, +\infty[$.

Les dérivées directionnelles des solutions sont obtenues comme solution d'un problème linéaire quadratique avec contraintes d'égalité et d'inégalité. L'équivalence du Th. 0.7 montre que la condition du second ordre (0.43) est la plus faible possible pour avoir la stabilité des solutions. La stabilité des multiplicateurs est un peu plus délicate à énoncer, car à l'exception des contraintes du premier ordre, les multiplicateurs p et η ne sont pas stables pour la norme L^∞ (en raison de la présence de sauts dont l'instant varie).

Pour les contraintes d'ordre deux (ou d'ordre supérieur à deux), les résultats précédents ne s'appliquent plus si un point de contact isolé non réductible τ_{to} apparaît, c'est-à-dire que $g^{(2)}(u(\tau_{to}), y(\tau_{to})) = 0$. Or ce cas peut se produire au cours des méthodes d'homotopie (voir section suivante). Dans ce cas, un analogue du Th. 0.6 est obtenu (théorème 6.13) pour les contraintes du second ordre, qui explicite les différents changements de structure possibles lorsqu'il y a un point de contact isolé non réductible. Parmi ces différentes possibilités, un arc frontière ou un second point de contact isolé peuvent apparaître. Ces différents changements de structure ne permettent plus d'utiliser une approche tir pour l'analyse de stabilité.

Pour cette raison, dans le chapitre 5 de la thèse, nous étendons la théorie de la régularité forte de Robinson et les résultats de stabilité obtenus pour les contraintes du premier ordre [88, 53] à des contraintes du second ordre. Seules les hypothèses (A0)-(A3) et une condition

du second ordre sont utilisées dans l'analyse. Ce résultat ne fait donc *aucune hypothèse sur la structure de la trajectoire*. Sous une condition du second ordre du type (0.43) (à la différence que la forme quadratique Q utilisée (0.31) ne fait plus intervenir le terme de courbure), on montre que localement, le problème perturbé a une solution locale, localement unique, vérifiant (0.42), qui est lipschitzienne par rapport au paramètre pour la norme L^2 et höldérienne pour la norme L^∞ (théorème 5.12).

On obtient donc un résultat plus faible que le Th. 0.7 (dans la mesure où on ne montre pas la stabilité (lipschitz) L^∞, ni que les solutions sont directionnellement différentiables, et on perd l'implication (i) $\Rightarrow$ (ii) du Th. 0.7), mais c'est le premier résultat de stabilité (L^2) obtenu pour les contraintes d'ordre 2 sans hypothèse sur la structure de la trajectoire. La preuve est basée sur la définition de multiplicateurs alternatifs, obtenus par intégration du multiplicateur η et donc plus réguliers, et l'application du théorème des fonctions implicites généralisé dans les espaces métriques de Dontchev et Hager [53] à la condition d'optimalité en résultant dans un cadre fonctionnel convenable. Ce résultat se généralise aisément à une contrainte d'ordre $q \geq 3$.

0.2.5 Méthodes d'homotopie

Introduction

Une difficulté pour appliquer l'algorithme de tir en présence de contraintes est la nécessité de connaître *a priori* la structure de la trajectoire optimale, qui en général n'est pas connue. De plus, l'algorithme de tir ayant un domaine de convergence restreint, même lorsque la structure est connue, initialiser tous les paramètres de tir (instants de jonction et sauts de l'adjoint) de façon à se trouver dans la zone de convergence de l'algorithme est souvent difficile. Une possibilité pour pallier cette difficulté est d'utiliser une méthode d'homotopie (ou continuation), voir [1] et [45, Chap. 5]. La méthode d'homotopie consiste a résoudre une suite de problèmes dépendant continûement d'un paramètre, telle que le premier problème est "facile" à résoudre, et le dernier problème est notre problème d'origine. Ainsi, partant par exemple du problème sans contrainte sur l'état, il devient possible de déterminer de proche en proche la structure de la trajectoire. C'est ce qui a été fait sur la figure 0.2 pour arriver à la solution du problème le plus contraint.

Cette méthode, bien connue, a été appliquée avec succès sur un problème non trivial issu de l'aéronautique avec contrainte sur l'état d'ordre 3 dans [11]. Dans cet article, les changements de structure étaient gérés "à la main". Récemment, des méthodes d'homotopie qui réalisent automatiquement le suivi du chemin ont été appliquées à des problèmes avec contraintes sur la commande [63, 97]. Plus précisément, pour trouver la structure d'une commande discontinue (bang-bang ou avec arc singulier), un terme de perturbation quadratique $(1-\mu)|u(t)|^2$ est ajouté au coût distribué. La commande solution du problème pour $\mu < 1$ est alors continue et converge (faiblement-* dans L^∞) vers la solution discontinue du problème de départ pour $\mu = 1$. Ce type d'homotopie est différent de celui que nous décrivons dans la suite car le problème de changement de la structure au cours de l'homotopie ne se pose pas.

Résultats

Grâce aux résultats de stabilité présentés dans la section précédente, nous pouvons proposer une méthode d'homotopie qui, sous certaines hypothèses (en particulier la complémentarité stricte uniforme sur les arcs frontière (0.40)), détermine automatiquement la structure de la trajectoire pour une contrainte d'ordre 1 ou 2. Partant du problème sans la contrainte sur l'état,

on introduit progressivement celle-ci. La structure des trajectoires va donc changer au cours des itérations, de même que la dimension des paramètres de tir. La méthode d'homotopie détecte automatiquement l'apparition ou la disparition d'un arc frontière ou d'un point de contact isolé pour une contrainte du second ordre, et, en cas d'apparition d'un arc frontière ou d'un point de contact isolé, initialise les paramètres de tir associés (saut de l'adjoint et instants d'entrée et de sortie ou de contact isolé). Ainsi, l'utilisateur n'a plus besoin de connaître à l'avance la structure de la trajectoire et doit seulement initialiser l'adjoint initial p_0 pour le problème sans contrainte sur l'état. Une méthode de type prédicteur-correcteur est utilisée le long du chemin d'homotopie lorsque la structure est constante, ce qui améliore la convergence de l'algorithme. C'est la première fois qu'une méthode d'homotopie prend en compte automatiquement des changements de structure avec changement de la dimension des paramètres de tir.

La convergence théorique de l'algorithme d'homotopie a été montrée, sous certaines hypothèses, dans la proposition 3.44 pour les contraintes du premier ordre et dans la proposition 6.28 pour les contraintes du second ordre. L'algorithme a été appliqué numériquement à un exemple académique avec contrainte du premier ordre dans la section 3.9.4. Pour les contraintes du second ordre, l'analyse est plus complexe en raison du nombre plus élevé de changements de structure pouvant se produire (points de contact isolés essentiels et arcs frontière sont tous deux possibles, alors que seuls des arcs frontière peuvent se produire pour les contraintes du premier ordre). Une difficulté théorique (non résolue) pour les contraintes du second ordre est liée à la possible transformation d'un point de contact isolé non réductible en deux points de contact isolés (car la fonction de tir devient singulière), alors que l'apparition d'un arc frontière au voisinage d'un point de contact isolé non réductible se traite comme pour une contrainte du premier ordre.

Les deux points clés dans l'analyse de l'algorithme d'homotopie sont les théorèmes 0.7 et 5.12, qui donnent l'existence et l'unicité locale d'une solution locale au problème perturbé, ce qui permet d'assurer localement l'existence du chemin d'homotopie, et les théorèmes 0.6 et 6.13, qui donnent l'évolution qualitative de la structure des solutions du problème perturbé (nombre fini de possibilités), et permettent ainsi à la méthode d'homotopie de déterminer automatiquement les changements de structure de la trajectoire.

0.2.6 Cas de plusieurs contraintes sur l'état et de contraintes mixtes

Introduction

Les résultats énoncés précédemment s'appliquent au cas d'une commande scalaire et d'une contrainte sur l'état scalaire. Qu'en est-il lorsqu'on a une commande et une contrainte g à valeurs vectorielles ? Alors que de nombreux articles se sont intéressés au cas de plusieurs contraintes du premier ordre et de contraintes mixtes, par exemple [65, 88, 53, 54] (en particulier des résultats de régularité sont connus), la seule référence que nous connaissons traitant le cas de plusieurs contraintes d'ordre supérieur est un article non publié de Maurer [98]. Dans [75, 68, 94], seule une contrainte scalaire d'ordre élevée est considérée.

L'extension des résultats précédents dans le cas de plusieurs contraintes sur l'état d'ordre arbitraire n'est pas triviale. Le cas d'une contrainte et d'une commande scalaires est particulier, car la commande s'obtient comme fonction implicite de l'état sur un arc frontière par (0.36), et on peut ensuite en déduire la régularité du multiplicateur sur les arcs frontières en différentiant autant de fois que nécessaire la relation (0.15). Cet argument ne s'étend pas au cas où la dimension de la commande est différente du nombre de contraintes actives. Par ailleurs, la proposition 0.2 sur les conditions de jonctions, qui joue un rôle important dans la preuve des

théorèmes 0.4 et 0.5, ne s'étend pas non plus trivialement au cas vectoriel. Ainsi la première question qui se pose est celle de la régularité des solutions et multiplicateurs.

Résultats

Ce sont ces questions qui sont traitées dans le chapitre 4 où l'on s'intéresse au cas de plusieurs contraintes sur l'état, d'ordres arbitraires, et d'une commande à valeurs dans $\mathbb{R}^m$, $m > 1$. Ce chapitre inclut aussi des contraintes mixtes sur la commande et sur l'état, qui dans l'analyse peuvent être vues comme des contraintes sur l'état d'ordre zéro.

Le premier résultat que nous obtenons est un résultat de régularité de la commande et des multiplicateurs (section 4.3) analogue à celui connu dans le cas scalaire. Dans la proposition 4.8 nous donnons une condition suffisante assurant la continuité de la commande, puis dans la proposition 4.13 nous montrons que la commande et les multiplicateurs sont réguliers sur l'intérieur d'un arc ayant un ensemble de contraintes actives constant. Ensuite nous étendons la proposition 0.2 au cas vectoriel dans la proposition 4.22. La preuve utilise la mise du système sous forme normale (section 4.4), c'est-à-dire que la dynamique de chaque composante de la contrainte peut, après un changement de variables, être mise localement sous la forme canonique (0.19) (lemme 4.19).

Une fois ces premiers résultats de régularité obtenus, nous sommes en mesure d'étendre au cas de plusieurs contraintes sur l'état et de contraintes mixtes sur la commande et sur l'état les conditions du second ordre no-gap du théorème 0.4 dans les théorème 4.24 et corollaire 4.25 et l'analyse de l'algorithme de tir ainsi que le théorème 0.5 dans la section 4.7 et le théorème 4.33.

0.3 Plan de la thèse

Le chapitre 1 correspond à l'article [21]

> J.F. Bonnans et A. Hermant. No-gap second-order optimality conditions for optimal control problems with a single state constraint and control. *Mathematical Programming*, Ser. B., 117 :21–50, 2009.

Les résultats sur les conditions d'optimalité du second ordre y sont présentés, dans le cas d'une commande et d'une contrainte sur l'état scalaires.

Le chapitre 2 correspond à l'article [19]

> J.F. Bonnans et A. Hermant. Well-Posedness of the shooting algorithm for state constrained optimal control problems with a single constraint and control. *SIAM Journal on Control and Optimization*, 46(4) :1398–1430, 2007.

Les résultats sur l'algorithme de tir y sont présentés, toujours pour une commande et une contrainte sur l'état scalaires, ainsi que l'analyse de stabilité et de sensibilité par l'approche tir en présence de points de contact isolés non essentiels pour une contrainte d'ordre supérieur ou égal à deux.

Le chapitre 3 correspond à l'article [20]

> J.F. Bonnans et A. Hermant. Stability and sensitivity analysis for optimal control problems with a first-order state constraint and application to continuation methods. *ESAIM Control, Optimization and Calculus of Variations*, 14(4) :825–863, 2008.

L'analyse de stabilité et de sensibilité en présence de points de contact isolés non essentiels pour une contrainte du premier ordre y est présentée, ainsi que la méthode d'homotopie.

Le chapitre 4 correspond à l'article [17]

> J.F. Bonnans et A. Hermant. Second-order analysis for optimal control problems with pure state constraints and mixed control-state constraints. *Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire*. À paraître.

Les résultats sur les conditions de jonction (proposition 0.2), sur les conditions du second ordre et sur l'algorithme de tir des chapitres 1 et 2 y sont étendus pour une commande à valeurs vectorielles, plusieurs contraintes sur l'état et des contraintes mixtes sur la commande et sur l'état.

Le chapitre 5 correspond à l'article [71]

> A. Hermant. Stability analysis of optimal control problems with a second-order state constraint. *SIAM Journal on Optimization*. A paraître.

On y présente les résultats de stabilité pour les contraintes d'ordre deux utilisant une variante de la théorie de la régularité forte sans hypothèse sur la structure de la trajectoire.

Le chapitre 6 correspond à l'article [69]

> A. Hermant. Homotopy algorithm for optimal control problems with a second-order state constraint. Rapport de recherche INRIA RR-6626 (2008). Soumis.

Ce chapitre est également consacré aux contraintes sur l'état d'ordre deux. On y présente des résultats étendant partiellement ceux du chapitre 3 aux contraintes d'ordre deux, résultats portant sur la stabilité structurelle des points stationnaires (stabilité des arcs frontières ; le cas des points de contact isolés non réductibles est également traité) et sur la méthode d'homotopie.

Enfin, dans le chapitre 7 (conclusion) quelques problèmes ouverts dans la continuité des travaux de cette thèse sont présentés (vérification de la condition suffisante du second ordre, extension des conditions du second ordre aux équations aux dérivées partielles, cas d'un nombre infini de points de contact isolés et cas de contraintes linéairement dépendantes).

Les six premiers chapitres, rédigés sous forme d'article, et présentés dans l'ordre chronologique, peuvent être lu indépendamment les uns des autres. Les notations, hypothèses, définitions et résultats utilisés y sont rappelés à chaque fois. Le chapitre 1 contient les conditions du second ordre, clé de voûte des autres résultats de la thèse. Le chapitre 2 utilise les résultats du chapitre 1. Les chapitres 3 et 4 utilisent indépendamment les résultats des chapitres 1 et 2. Le chapitre 5 utilise quelques résultats du chapitre 4. Le chapitre 6 utilise les chapitres 2, 3 et 5. Le chapitre 7 utilise les chapitres 1 et 4.

Chapitre 1

Conditions d'optimalité du second ordre*

Abstract The paper deals with optimal control problems with only one control variable and one state constraint, of arbitrary order. We consider the case of finitely many boundary arcs and touch times. We obtain a no-gap theory of second-order conditions, allowing to characterize second-order quadratic growth.

Résumé Dans cet article, nous étudions un problème de commande optimale avec une commande scalaire et une contrainte sur l'état scalaire d'ordre quelconque. Les instants de jonction sont supposés en nombre fini. Nous obtenons des conditions d'optimalité du second ordre nécessaires ou suffisantes, qui permettent de caractériser la croissance quadratique.

1.1 Introduction

Considerable efforts have been done in the past for reducing the gap between second-order necessary and sufficient optimality conditions for optimization problems in Banach spaces, with so-called cone constraint (i.e. the constraint mapping must be in a convex cone, or more generally in a convex set). This framework includes many optimal control problems. The theory of second-order necessary optimality conditions involves a term taking into account the curvature of the convex set, see Kawasaki [77], Cominetti [41]. By contrast, second-order sufficient optimality conditions typically involve no such term; see e.g. Maurer and Zowe [102]. We say that a no-gap condition holds, when the only change between necessary or sufficient second-order optimality conditions is between a strict and non strict inequality. In that case it is usually possible to obtain a characterization of the second-order growth condition. There are essentially two cases when no-gap conditions were obtained: (i) the polyhedric framework, in the case when the Hessian of Lagrangian is a Legendre form, originating in the work by Haraux [67] and Mignot [103], applied to optimal control problems in e.g. Sokolowski [123] and Bonnans [14], and the extended polyhedricity framework in [24, Section 3.2.3]; this framework essentially covers the case of control constraints (and finitely many final state constraints); and (ii) the second-order regularity framework, introduced in [16] and [15], with applications to semi definite optimization. We refer to [24] for an overview of these theories.

*Joint work with J.F. Bonnans. Published in Mathematical Programming Ser. B, 117 :21–50 (2009), under the title *No-gap second-order optimality conditions for optimal control problems with a single state constraint and control.*

Our paper deals with state-constrained optimal control problems. This occurs in many applications, see e.g. [11, 12, 5, 27, 9]. In optimal control theory, no-gap second-order optimality conditions were known for *mixed* control-state constraints, see e.g. Milutyin-Osmolovskii [105, Part. 2], Osmolovskii [108, 109], and Zeidan [127], whose results use conjugate point theory and Riccati equations.

Generally speaking, problems with non positivity constraints in spaces of continuous functions do not fit into these frameworks, where no-gap second-order conditions were obtained. The expression of the curvature term in this case was obtained by Kawasaki [79, 78] in the one dimensional case, and generalized in Cominetti and Penot [42]. Necessary conditions for variational problems with state constraints taking into account the curvature term can be found in Kawasaki and Zeidan [80]. However, only sufficient conditions without curvature terms were known. Two exceptions are a quite specific situation studied in [16] (with applications to some eigenvalue problems), and the case of finitely many contact points, when the problem can be reduced locally to finitely many inequality constraints in semi-infinite programming, see e.g. Hettich and Jongen [72].

Our main result is the following. By a localization argument, we split the curvature term into a finite number of contributions of boundary arcs and touch points. Using the theory of junction conditions in Jacobson et al. [75] and Maurer [98], we are able to prove that, under quite weak assumptions, the contribution of boundary arcs to the curvature term is zero. For touch points, we use a reduction argument for those that are essential (i.e. that belong to the support of the multiplier) and we make no hypotheses for the non essential ones. The only delicate point is to compute the expansion of the minimum value of a function in $W^{2,\infty}$. Since it is not difficult to state sufficient conditions taking into account essential reducible touch points, we obtain in this way no-gap conditions, that in addition characterize quadratic growth in a convenient two-norms setting.

The paper is organized as follows. In section 1.2, we recall the material needed, in both points of view of abstract optimization and junction conditions analysis. The main contributions of the paper are in sections 1.3-1.5 where the no-gap second-order condition is established. Section 1.3 states the second-order necessary condition (computation of the curvature term). Section 1.4 handles the second-order sufficient condition. In section 1.5, a reduction approach is presented in order to deal with the non-zero part of the curvature term.

1.2 Framework

We consider the following optimal control problem with a scalar state constraint and a scalar control:

$$(\mathcal{P}) \qquad \min_{u,y} \int_0^T \ell(u(t), y(t))\mathrm{d}t + \phi(y(T)) \tag{1.1}$$

$$\text{s.t.} \qquad \dot{y}(t) = f(u(t), y(t)) \quad \text{a.e. } t \in [0,T] \quad ; \quad y(0) = y_0 \tag{1.2}$$

$$g(y(t)) \leq 0 \qquad \forall t \in [0,T]. \tag{1.3}$$

The data of the problem are the distributed cost $\ell : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, the final cost $\phi : \mathbb{R}^n \to \mathbb{R}$, the dynamics $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, the state constraint $g : \mathbb{R}^n \to \mathbb{R}$, the final time $T > 0$, and the initial condition $y_0 \in \mathbb{R}^n$. We make the following assumptions on the data:

(A0) The mappings ℓ, ϕ, f and g are k-times continuously differentiable (C^k) with $k \geq 2$ and have locally Lipschitz continuous second-order derivatives, and the dynamics f is Lipschitz continuous.

(A1) The initial condition satisfies $g(y_0) < 0$.

Throughout the paper, it is assumed that assumption (A0) holds.

1.2.1 Abstract Optimization

For $1 \leq p \leq \infty$, $L^p(0,T)$ denotes the Banach space of measurable functions such that

$$\|u\|_p := \left(\int_0^T |u(t)|^p \mathrm{d}t \right)^{1/p} < \infty \ \text{ for } p < \infty; \quad \|u\|_\infty := \operatorname{supess} |u(t)| < \infty,$$

and $W^{1,p}(0,T)$ denotes the Sobolev space of functions having a weak derivative in L^p. The space of continuous functions over $[0,T]$ is denoted by $C[0,T]$, with the norm $\|x\|_\infty = \sup |x(t)|$.

Denote by $\mathcal{U} := L^\infty(0,T;\mathbb{R})$ (resp. $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$) the control (resp. state) space. A *trajectory* is an element $(u,y) \in \mathcal{U} \times \mathcal{Y}$ satisfying the state equation (1.2). Given $u \in \mathcal{U}$, denote by $y_u \in \mathcal{Y}$ the (unique) solution of (1.2). Under assumption (A0), by the Cauchy-Lipschitz Theorem, this mapping is well-defined and of class C^k. We may write problem $(\mathcal{P})$ as:

$$\min_{u \in \mathcal{U}} J(u) \quad ; \quad G(u) \in K \tag{1.4}$$

where $J : \mathcal{U} \to \mathbb{R}$ and $G : \mathcal{U} \to C[0,T]$ are defined, respectively, by $J(u) = \int_0^T \ell(u(t), y_u(t))\mathrm{d}t + \phi(y_u(T))$ and $G(u) = g(y_u)$. These mappings are C^k. Here $K = C_-[0,T]$ is the set of continuous functions over $[0,T]$, with values in $\mathbb{R}_-$.

We say that $u \in \mathcal{U}$ is a (weak) local solution of (1.4) that satisfies the *quadratic growth condition*, if there exist $\alpha > 0$ and $\rho > 0$ such that:

$$J(\tilde{u}) \geq J(u) + \alpha \|\tilde{u} - u\|_2^2 \qquad \text{for all } \ \tilde{u} \in B_\infty(u,\rho), \ G(\tilde{u}) \in K \tag{1.5}$$

where $B_\infty(u,\rho)$ denotes the open ball in $L^\infty(0,T)$ with center u and radius ρ. This condition involves two norms, $L^\infty(0,T)$ for the neighborhood, and $L^2(0,T)$ for the growth condition.

The space of row vectors is denoted by $\mathbb{R}^{n*}$. The space of Radon measures, the dual space to $C[0,T]$, is denoted by $\mathcal{M}[0,T]$ and identified with functions of bounded variation vanishing at zero. The cone of nonnegative measures is denoted by $\mathcal{M}_+[0,T]$ and is equal to K^-, the polar cone of K. The duality product over $\mathcal{M}[0,T] \times C[0,T]$ is denoted by $\langle \eta, x \rangle = \int_0^T x(t)\mathrm{d}\eta(t)$. Adjoint operators (and transpose in $\mathbb{R}^n$) are denoted by a star *. Fréchet derivatives of f, etc. w.r.t. arguments $u \in \mathbb{R}$, $y \in \mathbb{R}^n$, are denoted by a subscript, for instance $f_u(u,y) = D_u f(u,y)$, $f_{uu}(u,y) = D^2_{uu} f(u,y)$, etc.

Define the classical *Hamiltonian* and *Lagrangian* functions of problem $(\mathcal{P})$, respectively $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$ and $L : \mathcal{U} \times \mathcal{M}[0,T] \to \mathbb{R}$ by:

$$H(u,y,p) := \ell(u,y) + pf(u,y) \quad ; \quad L(u,\eta) := J(u) + \langle \eta, G(u) \rangle \,. \tag{1.6}$$

Denote by $BV[0,T]$ the space of functions of bounded variation. Given $u \in \mathcal{U}$ and $\eta \in \mathcal{M}_+[0,T]$, let the costate $p_{u,\eta}$ be the unique solution in $BV([0,T];\mathbb{R}^{n*})$ of:

$$-\mathrm{d}p_{u,\eta} = (\ell_y(u,y_u) + p_{u,\eta} f_y(u,y_u))\mathrm{d}t + g_y(y_u)\mathrm{d}\eta \ ; \quad p_{u,\eta}(T) = \phi_y(y_u(T)). \tag{1.7}$$

Given $v \in \mathcal{U}$, let the linearized state $z_{u,v} \in \mathcal{Y}$ be solution of:

$$\dot{z}_{u,v} = f_y(u, y_u) z_{u,v} + f_u(u, y_u) v \qquad ; \qquad z_{u,v}(0) = 0. \tag{1.8}$$

The mapping $\mathcal{U} \to \mathcal{Y}$, $v \mapsto z_{u,v}$ is the Fréchet derivative of the mapping $u \mapsto y_u$ at point u.

The next lemma gives the expressions of derivatives of Lagrangian, with respect to the control. For simplicity of notation, we write in the sequel $D^2 H_{(u,y)^2}(u, y, p)(v, z)^2$ instead of $D^2_{(u,y),(u,y)} H(u, y, p)((v, z), (v, z))$.

Lemma 1.1. *Let $\eta \in \mathcal{M}_+[0, T]$. Then $u \mapsto L(u, \eta)$ is of class C^2 over $\mathcal{U}$, with first and second derivatives given by, for all $v \in \mathcal{U}$ (omitting time argument):*

$$D_u L(u, \eta) v = \int_0^T H_u(u, y_u, p_{u,\eta}) v \mathrm{d}t, \tag{1.9}$$

$$\begin{aligned} D^2_{uu} L(u, \eta)(v, v) \;&=\; \int_0^T D^2 H_{(u,y)^2}(u, y_u, p_{u,\eta})(v, z_{u,v})^2 \mathrm{d}t \\ &\quad + z_{u,v}(T)^* \phi_{yy}(y_u(T)) z_{u,v}(T) \;+\; \int_0^T z^*_{u,v} g_{yy}(y_u) z_{u,v} \mathrm{d}\eta, \end{aligned} \tag{1.10}$$

where H is given by (1.6), $z_{u,v}$ and $p_{u,\eta}$ are the solutions, respectively, to (1.8) and (1.7).

Proof. Since $u \mapsto y_u$ is C^2, the Cauchy-Lipschitz Theorem ensures the existence of the second-order expansion of the state

$$y_{u+v} = y_u + z_{u,v} + \frac{1}{2} z_{u,vv} + o\left(\|v\|^2_\infty\right). \tag{1.11}$$

It is easily seen, substituting (1.11) into the state equation and keeping the terms of second-order, that $z_{u,vv}$ is solution of:

$$\dot{z}_{u,vv} = f_y(u, y_u) z_{u,vv} + D^2 f_{(u,y)^2}(u, y_u)(v, z_{u,v})^2 \quad ; \quad z_{u,vv}(0) = 0. \tag{1.12}$$

Using costate equation (1.7) and linearized state equations (1.8) and (1.12), we get easily (omitting arguments):

$$\begin{aligned} D_u L(u, \eta) v \;&=\; -\int_0^T (\mathrm{d}p_{u,\eta} z_{u,v} + p_{u,\eta} \dot{z}_{u,v} \mathrm{d}t) + \phi_y(y_u(T)) z_{u,v}(T) \\ &\qquad + \int_0^T H_u v \mathrm{d}t; \\ D^2_{uu} L(u, \eta)(v, v) \;&=\; \int_0^T D^2 H_{(u,y)^2}(v, z_{u,v})^2 \mathrm{d}t + z_{u,v}(T)^* \phi_{yy}(y_u(T)) z_{u,v}(T) \\ &\qquad + \int_0^T z^*_{u,v} g_{yy}(y_u) z_{u,v} \mathrm{d}\eta \\ &\quad - \int_0^T (\mathrm{d}p_{u,\eta} z_{u,vv} + p_{u,\eta} \dot{z}_{u,vv} \mathrm{d}t) + \phi_y(y_u(T)) z_{u,vv}(T). \end{aligned}$$

To obtain (1.9) and (1.10) it suffices, in view of Lemma 1.33 in the Appendix, to integrate by parts in the above expressions $p_{u,\eta}$ with $z_{u,v}$ and with $z_{u,vv}$, respectively. □

First Order Necessary Condition. For $x \in K = C_-(0,T)$, define the *first order contact set* $I(x) := \{t \in [0,T] \ ; \ x(t) = 0\}$. The expression of the *tangent* and *normal* cones (in the sense of convex analysis) to K at point x, respectively $T_K(x)$ and $N_K(x)$, are well-known (see e.g. [24]) and given, for $x \in K$ (these sets being empty if $x \notin K$), by:

$$\begin{aligned} T_K(x) &= \{h \in C[0,T] \ ; \ h(t) \leq 0 \text{ on } I(x)\}, \\ N_K(x) &= \{\eta \in \mathcal{M}_+[0,T] \ ; \ \mathrm{supp}(\mathrm{d}\eta) \subset I(x)\}. \end{aligned}$$

Here by $\mathrm{supp}(\mathrm{d}\eta)$ we denote the *support* of the measure $\eta \in \mathcal{M}[0,T]$, i.e. the complement in $[0,T]$ of the largest open set $W \subset [0,T]$ that satisfies: $\int_0^T x(t)\mathrm{d}\eta(t) = 0$, for all functions $x \in C[0,T]$ vanishing on $[0,T] \setminus W$.

Let $u \in \mathcal{U}$. We say that $\eta \in \mathcal{M}_+[0,T]$ is a *Lagrange multiplier* associated with u if the following first order necessary optimality condition holds:

$$D_u L(u,\eta) = DJ(u) + DG(u)^*\eta = 0 \quad ; \quad \eta \in N_K(G(u)). \tag{1.13}$$

The set of Lagrange multipliers associated with u is denoted by $\Lambda(u)$.

Robinson's constraint qualification (see [119, 120]) for problem (1.4) is as follows:

$$\exists\, \varepsilon > 0, \qquad \varepsilon B_C \subset G(u) + DG(u)\mathcal{U} - K. \tag{1.14}$$

Here B_C denotes the unit (open) ball of $C[0,T]$.

The next theorem is well-known (see e.g. [24], Lemma 2.99 and Theorem 3.9). Note that for $v \in \mathcal{U}$, we have $DG(u)v = g_y(y_u)z_{u,v}$, i.e., $(DG(u)v)(t) = g_y(y_u(t))z_{u,v}(t)$, for all $t \in [0,T]$.

Theorem 1.2. (i) *A characterization of (1.14) is:*

$$\textit{There exists } v \in \mathcal{U}; \quad g_y(y_u(t))z_{u,v}(t) < 0, \quad \textit{for all } t \in I(g(y_u)). \tag{1.15}$$

(ii) *Let u be a local solution of (1.4), satisfying (1.15). Then with u is associated a non empty and bounded set of Lagrange multipliers.*

Second Order Analysis. Let the *critical cone* be defined by:

$$C(u) = \{v \in \mathcal{U} \ ; \ DG(u)v \in T_K(G(u)) \ ; \ DJ(u)v \leq 0\}. \tag{1.16}$$

For $h \in T_K(x)$, the *second-order contact set* is defined by:

$$I^2(x,h) = \{t \in I(x) \ ; \ h(t) = 0\}. \tag{1.17}$$

If (1.13) holds, then $DJ(u)v \geq 0$ for all v such that $DG(u)v \in T_K(G(u))$ and $DJ(u)v = 0$ iff $\eta \perp DG(u)v$. Since η is a nonnegative measure with support in $I(G(u))$, and $DG(u)v \leq 0$ on $I(G(u))$, we obtain the following (classical) statement:

Lemma 1.3. *Let (u,η) satisfy the first order necessary condition (1.13). Then:*

$$C(u) = \{v \in \mathcal{U};\ DG(u)v \in T_K(G(u));\ \mathrm{supp}(\mathrm{d}\eta) \subset I^2(G(u), DG(u)v)\}. \tag{1.18}$$

The *inner* and *outer second-order tangent sets*, respectively $T_K^{2,i}(x,h)$ and $T_K^2(x,h)$, are defined by:

$$\begin{aligned} T_K^{2,i}(x,h) &:= \{w \in C[0,T];\ \mathrm{dist}(x + \varepsilon h + \tfrac{1}{2}\varepsilon^2 w, K) = o(\varepsilon^2),\ \varepsilon \geq 0\}, \\ T_K^2(x,h) &:= \{w \in C[0,T];\ \exists \varepsilon_n \downarrow 0, \mathrm{dist}(x + \varepsilon_n h + \tfrac{1}{2}\varepsilon_n^2 w, K) = o(\varepsilon_n^2)\}. \end{aligned}$$

We recall the characterization of the inner second-order tangent set $T_K^{2,i}(x,h)$ due to Kawasaki [79, 78] (see also Cominetti [42]): if $x \in K$ and $h \in T_K(x)$, then

$$T_K^{2,i}(x,h) = \{w \in C[0,T] \ ; \ w(t) \leq \varsigma_{x,h}(t) \text{ on } [0,T]\}, \tag{1.19}$$

where $\varsigma_{x,h} : [0,T] \to \overline{\mathbb{R}}$ is given by:

$$\varsigma_{x,h}(t) = \begin{cases} 0 & \text{if } t \in (\operatorname{int} I(x)) \cap I^2(x,h) \\ \liminf\limits_{t' \to t \, ; \, x(t') < 0} \dfrac{(h(t')_+)^2}{2x(t')} & \text{if } t \in \partial I(x) \cap I^2(x,h) \\ +\infty & \text{otherwise.} \end{cases} \tag{1.20}$$

Here $h(t)_+ := \max\{h(t), 0\}$, and $\operatorname{int} S$ and ∂S denote respectively the interior and boundary of set S. Set $\mathcal{T}(x,h) := \partial I(x) \cap I^2(x,h)$. We have $\varsigma_{x,h}(\tau) \leq 0$ for $\tau \in \mathcal{T}(x,h)$ and it is not difficult to check that $t \mapsto \varsigma_{x,h}(t)$ is lower semi-continuous. Consequently, $T_K^{2,i}(x,h) \neq \emptyset$ iff $\varsigma_{x,h}(t) > -\infty$ for all t. In that case, $\varsigma_{x,h}$ is the upper limit of a increasing sequence of continuous functions (ς_n). Given $\eta \in \mathcal{M}_+[0,T]$, we may define (see e.g. [79]):

$$\int_0^T \varsigma_{x,h}(t) \mathrm{d}\eta(t) := \sup \left\{ \int_0^T \varsigma(t) \mathrm{d}\eta(t); \ \varsigma \leq \varsigma_{x,h} \right\} \ \in \ \mathbb{R} \cup \{+\infty\}.$$

Then:

$$\sigma(\eta, T_K^{2,i}(x,h)) = \int_0^T \varsigma_{x,h}(t) \mathrm{d}\eta(t), \tag{1.21}$$

where $\sigma(\eta, S) = \sup_{w \in S} \langle \eta, w \rangle$ denotes the support function of the set S. If the support of η satisfies $\operatorname{supp}(\mathrm{d}\eta) \subset I^2(x,h)$, then

$$\sigma(\eta, T_K^{2,i}(x,h)) \leq 0. \tag{1.22}$$

A second-order necessary condition due to Kawasaki [77] is:

Theorem 1.4. *Let u be a local solution of (1.4) satisfying (1.14). Then, for all $v \in C(u)$, the following holds:*

$$\sup_{\eta \in \Lambda(u)} \left\{ D^2_{uu} L(u,\eta)(v,v) - \sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) \right\} \ \geq \ 0. \tag{1.23}$$

Remark 1.5. The above second-order necessary condition was improved by Cominetti in [41], by stating that for all convex set $\mathcal{S}_{u,v} \subset T_K^2(G(u), DG(u)v)$,

$$\sup_{\eta \in \Lambda(u)} \left\{ D^2_{uu} L(u,\eta)(v,v) - \sigma(\eta, \mathcal{S}_{u,v}) \right\} \ \geq \ 0. \tag{1.24}$$

Th. 1.4 is obtained for the particular choice of $\mathcal{S}_{u,v} = T_K^{2,i}(G(u), DG(u)v)$. For the problem considered in the present paper, we gain sufficient information from (1.23) (see Proposition 1.14).

1.2.2 Junction Condition Analysis

We first recall some classical definitions. A *boundary* (resp. *interior*) *arc* is a maximal interval of positive measure $\mathcal{I} \subset [0,T]$ such that $g(y(t)) = 0$ (resp. $g(y(t)) < 0$) for all $t \in \mathcal{I}$. If $[\tau_{en}, \tau_{ex}]$ is a boundary arc, τ_{en} and τ_{ex} are called *entry* and *exit* point, respectively. Entry and exit points are said to be *regular* if they are endpoints of an interior arc. A *touch* point τ in $(0,T)$ is an isolated contact point (endpoint of two interior arcs). Entry, exit and touch points are called *junction points* (or *times*). We say that the junctions are regular, when the entry and exit points are regular. In this paper, only the case of finitely many regular junctions is dealt with.

The first-order time derivative of the state constraint when y satisfies the state equation (1.2), i.e., $g^{(1)}(u,y) = \frac{\mathrm{d}}{\mathrm{d}t} g(y(t)) = g_y(y) f(u,y)$, is denoted by $g^{(1)}(y)$ if the function $\mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$; $(u,y) \mapsto g_y(y) f(u,y)$ does not depend on u (that is, the function $(u,y) \mapsto g_u^{(1)}(u,y)$ is identically zero). We may define similarly $g^{(2)}, \dots, g^{(q)}$ if g, f are C^q and if $g_u^{(j)} \equiv 0$, for all $j = 1, \dots, q-1$, and we have $g^{(j)}(u,y) = g_y^{(j-1)}(y) f(u,y)$, for $j = 1, \dots, q$.

Let $q \geq 1$ be the smallest number of time derivations of the state constraint, so that a dependence w.r.t. u appears, i.e. $g_u^{(q)} \not\equiv 0$. If q is finite, we say that q is the *order* of the state constraint (see e.g. Bryson et al. [29]).

Let $u \in \mathcal{U}$ be a solution of the first order necessary condition (1.13), with Lagrange multiplier η and costate $p_{u,\eta}$ solution of (1.7). Since η and $p_{u,\eta}$ are of bounded variation, they have at most countably many discontinuity times, and are everywhere on $[0,T]$ left and right continuous. We denote by $[\eta(\tau)] = \eta(\tau^+) - \eta(\tau^-)$ where $\eta(\tau^\pm) = \lim_{t \to \tau^\pm} \eta(t)$ the jump of η at time $\tau \in [0,T]$. We make the following assumptions:

(A2) The Hamiltonian is strongly convex w.r.t. the control variable, uniformly w.r.t. $t \in [0,T]$:

$$\exists\, \gamma > 0, \qquad H_{uu}(\hat{u}, y_u(t), p_{u,\eta}(t^\pm)) \geq \gamma \qquad \forall \hat{u} \in \mathbb{R},\ \forall t \in [0,T]. \tag{1.25}$$

(A3) (Constraint regularity) The data of the problem are C^{2q}, i.e. $k \geq 2q$ in (A0), the state constraint is of order q and the condition below holds:

$$\exists\, \beta > 0, \qquad |g_u^{(q)}(u(t), y_u(t))| \geq \beta, \qquad \forall t \in [0,T]. \tag{1.26}$$

(A4) The trajectory (u, y_u) has a *finite set of junction times*, that will be denoted by $\mathcal{T} =: \mathcal{T}_{en} \cup \mathcal{T}_{ex} \cup \mathcal{T}_{to}$, with $\mathcal{T}_{en}$, $\mathcal{T}_{ex}$ and $\mathcal{T}_{to}$ the *disjoint* (and possibly empty) subsets of respectively regular entry, exit and touch points, and we assume that $g(y_u(T)) < 0$.

The above hypotheses imply the continuity of the control variable and of some of its derivatives at junction points (see Proposition 1.7 below).

Remark 1.6. 1) An assumption weaker than (A2), that is enough for the sufficient conditions in section 1.4 and 1.5, is

(A2') (Strengthened Legendre-Clebsch condition)

$$\exists\, \gamma > 0, \qquad H_{uu}(u(t), y_u(t), p_{u,\eta}(t)) \geq \gamma \qquad \text{a.e. } t \in [0,T]. \tag{1.27}$$

Condition (1.27) does not imply the continuity of the control.
2) In assumption (A3), it is in fact sufficient to assume that (1.26) holds for t in the neighborhood of the contact set $I(g(y_u))$. In the definition of the order of the constraint q, it is

sufficient as well to restrict the variable y to a neighborhood in $\mathbb{R}^n$ of $\{y_u(t)\ ;\ t \in I(g(y_u))\}$.
3) The various results of this paper (Theorems 1.12, 1.18, 1.27 and Corollaries 1.13 and 1.15) as well as Prop. 1.7 below, are still true, replacing the assumption (A2) by the weaker assumptions that the control is continuous on $[0,T]$ and (1.27) holds.

A touch point $\tau \in \mathcal{T}_{to}$ is said to be *essential*, if the Lagrange multiplier η satisfies $[\eta(\tau)] > 0$. The set of essential touch points of the trajectory (u, y_u) will be denoted by $\mathcal{T}_{to}^{ess}$.

The next proposition is due to Jacobson et al. [75]. Its proof was later clarified in Maurer [98], see also the survey by Hartl et al. [68].

Proposition 1.7. *Let $u \in \mathcal{U}$ satisfying (1.13) with Lagrange multiplier η and assume that (A2)-(A4) hold. Then:*

(i) *The control u is continuous over $[0,T]$ (in particular at junction points $\tau \in \mathcal{T}$) and C^q on $[0,T] \setminus \mathcal{T}$. The multiplier η is continuously differentiable on $[0,T] \setminus \mathcal{T}$.*

(ii) *If $\tau \in \mathcal{T}_{en} \cup \mathcal{T}_{ex}$ is a regular entry or exit point, then:* (a) *if q is* odd, *η and the $q-1$ first time derivatives of u are continuous at τ;* (b) *if q is* even, *the $q-2$ first time derivatives of u are continuous at τ.*

(iii) *If $\tau \in \mathcal{T}_{to}$ is a touch point, then:* (a) *the $q-2$ first derivatives of u are continuous at τ;* (b) *if $q = 1$, then η and $\dot{u}$ are also continuous at τ (that is, if $q = 1$, then (u, y_u) does not have essential touch point).*

Remark 1.8. Under the assumptions of Prop. 1.7, we have the following decomposition: $\mathrm{d}\eta(t) = \eta_0(t)\mathrm{d}t + \sum_{\tau\in\mathcal{T}} \nu_\tau \delta_\tau(t)$ where δ_τ denotes the Dirac measure at time τ, the density $\eta_0 \in L^1(0,T)$ is equal to $\frac{\mathrm{d}\eta}{\mathrm{d}t}$ on $[0,T] \setminus \mathcal{T}$ and $\nu_\tau := [\eta(\tau)] \geq 0$. We have $\nu_\tau = 0$ if q is odd and τ is a regular entry/exit point, and if $q = 1$ and τ is a touch point.

We end this section by a result on constraint qualification and uniqueness of the multiplier. For this we need the expression of the time derivatives of $DG(u)v$.

Lemma 1.9. *Assume that f, g are C^q and that $g_u^{(j)} \equiv 0$, for $j = 1, \ldots, q-1$. Then: (i) For all $v \in \mathcal{U}$, the following relations hold:*

$$\frac{\mathrm{d}^j}{\mathrm{d}t^j} g_y(y_u) z_{u,v} = g_y^{(j)}(y_u) z_{u,v}, \qquad j = 1, \ldots, q-1, \tag{1.28}$$

$$\frac{\mathrm{d}^q}{\mathrm{d}t^q} g_y(y_u) z_{u,v} = g_y^{(q)}(u, y_u) z_{u,v} + g_u^{(q)}(u, y_u) v. \tag{1.29}$$

(ii) If in addition, (1.26) *is satisfied, then $DG(u)$ is an isomorphism between $L^\infty(0,T)$ and the space $\mathcal{W}$ defined by:*

$$\mathcal{W} := \{\varphi \in W^{q,\infty}(0,T)\ ;\ \varphi^{(j)}(0) = 0\ ;\ j = 0, \ldots, q-1\}. \tag{1.30}$$

Proof. (i) By (1.8), we have:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} g_y(y_u) z_{u,v} &= g_{yy}(y_u) f(u, y_u) z_{u,v} + g_y(y_u) f_y(u, y_u) z_{u,v} + g_y(y_u) f_u(u, y_u) v \\ &= g_y^{(1)}(u, y_u) z_{u,v} + g_u^{(1)}(u, y_u) v. \end{aligned}$$

Since $g_u^{(j)} \equiv 0$ for $j = 1$ to $q-1$, we obtain by induction that $\frac{\mathrm{d}^j}{\mathrm{d}t^j} g_y(y_u) z_{u,v} = g_y^{(j)}(y_u) z_{u,v}$ is independent of v, and that the derivative of order q has the expression in (1.29).

(ii) If in addition (1.26) is satisfied, it is easily seen by (1.29) that for all $\varphi \in \mathcal{W}$, there exists a unique $v \in \mathcal{U}$ such that $g_y(y_u)z_{u,v} = \varphi$. The conclusion follows from the open mapping theorem.

□

Proposition 1.10. *Assume that (A1) holds, and let $u \in \mathcal{U}$ satisfy (A3). Then: (i) Robinson's constraint qualification* (1.14) *holds; (ii) if $\Lambda(u) \neq \emptyset$, the Lagrange multiplier η associated with u is unique.*

Proof. It is obvious by Lemma 1.9(ii) and Th. 1.2(i) that (1.14) holds iff (A1) does. This proves (i). Assume that $\eta_1, \eta_2 \in \Lambda(u)$ and set $\mu := \eta_2 - \eta_1 \in \mathcal{M}[0,T]$. Since $DG(u)^*\mu = 0$, it follows that $\int_0^T \varphi(t)\mathrm{d}\mu(t) = 0$, for all $\varphi \in \mathcal{W}$, with $\mathcal{W}$ defined by (1.30). Since $g(y_0) < 0$, we have $\mathrm{supp}(\mathrm{d}\mu) \subset [2\varepsilon, T]$ for some $\varepsilon > 0$. Taking the restriction to $[\varepsilon, T]$ of functions in $DG(u)\mathcal{U}$, we obtain the whole space $W^{q,\infty}(\varepsilon, T)$. By density of the latter in $C[\varepsilon, T]$ we deduce that for all $\varphi \in C[0,T]$, $\int_0^T \varphi(t)\mathrm{d}\mu(t) = \int_\varepsilon^T \varphi(t)\mathrm{d}\mu(t) = 0$. Hence $\mathrm{d}\mu \equiv 0$, which achieves the proof of (ii). □

1.3 Second-order Necessary Conditions

1.3.1 Basic Second-order Necessary Conditions

Let $u \in \mathcal{U}$ satisfy assumptions (A2)-(A4) and $\eta \in \Lambda(u)$. We make the following assumptions. Let $\hat{q} := 2q - 1$ if q is *odd* and $\hat{q} := 2q - 2$ if q is *even*.

(A5) (Non Tangentiality Condition)

(i) For all entry times $\tau_{en} \in \mathcal{T}_{en}$ and all exit times $\tau_{ex} \in \mathcal{T}_{ex}$:

$$(-1)^{\hat{q}+1}\frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}}g(y_u(t))|_{t=\tau_{en}^-} < 0 \; ; \quad \frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}}g(y_u(t))|_{t=\tau_{ex}^+} < 0. \tag{1.31}$$

(ii) For all *essential* touch points $\tau_{to} \in \mathcal{T}_{to}^{ess}$:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y_u(t))|_{t=\tau_{to}} < 0. \tag{1.32}$$

(A6) (Strict Complementarity *on boundary arcs*): $\mathrm{int}\, I(G(u)) \subset \mathrm{supp}(\mathrm{d}\eta)$.

Remark 1.11. 1) By Proposition 1.7, the expressions appearing in assumption (A5)(i)-(ii) are well-defined, and $\hat{q} + 1$ is the smallest possible order for which the corresponding time derivative of $g(y_u)$ may be discontinuous at an entry or exit point. Therefore assumption (A5) does not contradict the junction conditions in Prop. 1.7. Note that $\hat{q} = q$ for $q = 1, 2$.
2) Only the assumption (A6') below, weaker than (A6), is used in necessary condition of Theorem 1.12, in order to ensure that the second-order tangent set $T_K^{2,i}(G(u), DG(u)v)$ is not empty, for all $v \in C(u)$:

(A6') (Strict Complementarity *near entry/exit* of boundary arcs): For all entry points $\tau_{en} \in \mathcal{T}_{en}$ and exit points $\tau_{ex} \in \mathcal{T}_{ex}$, there exists $\varepsilon > 0$ such that:

$$(\tau_{en}, \tau_{en} + \varepsilon) \; \subset \; \mathrm{supp}(\mathrm{d}\eta) \quad ; \quad (\tau_{ex} - \varepsilon, \tau_{ex}) \; \subset \; \mathrm{supp}(\mathrm{d}\eta). \tag{1.33}$$

Actually assumption (A6') is needed only when q is *even*, since it follows from (A2)-(A4) and (A5)(i) whenever q is odd, see e.g. [19, Lemma A.2][1].

Note that we do not assume strict complementarity at touch points.

Theorem 1.12. *Assume that (A1) holds. Let $u \in \mathcal{U}$ be a local solution of (1.4), with its Lagrange multiplier η, satisfying (A2)-(A5) and (A6'). Let $\mathcal{T}_{to}^{ess}$ denote the (finite) set of essential touch points of the trajectory (u, y_u) and $\nu_\tau = [\eta(\tau)] > 0$, for $\tau \in \mathcal{T}_{to}^{ess}$. Then, for all $v \in C(u)$:*

$$D^2_{uu}L(u,\eta)(v,v) - \sum_{\tau\in\mathcal{T}_{to}^{ess}} \nu_\tau \frac{(g_y^{(1)}(y_u(\tau))z_{u,v}(\tau))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y_u(t))|_{t=\tau}} \geq 0. \tag{1.34}$$

Corollary 1.13. *Under the assumptions of Theorem 1.12, if the trajectory (u, y_u) has no essential touch point (in particular, if the state constraint is of* first order *$q = 1$), then $D^2_{uu}L(u,\eta)(v,v) \geq 0$, for all $v \in C(u)$.*

In the sequel, we denote $I^2(G(u), DG(u)v)$ by $I^2_{u,v}$. For all $v \in C(u)$, by (1.18), we have $\mathcal{T}_{to}^{ess} \subset (\mathcal{T}_{to} \cap I^2_{u,v})$. Let us denote the subset of critical directions that "avoid" non essential touch point (i.e., such that $g(y_u(\tau))z_{u,v}(\tau) < 0$, for all $\tau \in \mathcal{T}_{to} \setminus \mathcal{T}_{to}^{ess}$) by:

$$C_0(u) := \{v \in C(u) \ ; \ \mathcal{T}_{to} \cap I^2_{u,v} = \mathcal{T}_{to}^{ess}\}.$$

The first step of the proof of Theorem 1.12 consists in computing the sigma-term for the critical directions in $C_0(u)$.

Proposition 1.14. *Let $v \in C_0(u)$. Under the assumptions of Theorem 1.12, we have that*

$$\sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) = \sum_{\tau\in\mathcal{T}_{to}^{ess}} \nu_\tau \frac{(g_y^{(1)}(y_u(\tau))z_{u,v}(\tau))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y_u(t))|_{t=\tau}}. \tag{1.35}$$

Proof. The proof is divided into 3 steps. We first analyse the contribution of entry/exit points, then the one of touch points, and finally conclude.

Remind that by (1.20), only the points in $\partial I(G(u)) \cap I^2_{u,v}$ have a contribution to the sigma term. Note that $\partial I(G(u)) = \mathcal{T}$. Set $\varsigma_{u,v} := \varsigma_{g(y_u),g_y(y_u)z_{u,v}} = \varsigma_{G(u),DG(u)v}$ and let $\tau \in \mathcal{T} \cap I^2_{u,v}$. By (1.20), we have:

$$\varsigma_{u,v}(\tau) \ = \ \liminf_{t\to\tau;\ g(y_u(t))<0} \frac{(\{g_y(y_u(t))z_{u,v}(t)\}_+)^2}{2g(y_u(t))}. \tag{1.36}$$

1) (Entry/exit point). Assume that $\tau \in \mathcal{T}_{en} \cup \mathcal{T}_{ex}$. According to Prop. 1.7(ii), time derivatives of the control at regular entry/exit points are continuous until order $q-2$ if q is even, and $q-1$ if q is odd. Consequently, by definition of the order of the state constraint, the time derivatives of $g(y_u)$ are continuous at τ until order $2q-2$ if q is even, and $2q-1$ if q is odd. Hence they all vanish at entry/exit time τ of a boundary arc. It follows that for t in a neighborhood of τ on the interior arc side, a Taylor expansion gives, by definition of $\hat{q}$:

$$g(y_u(t)) = \frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}}g(y_u)|_{t=\tau^\pm}\frac{(t-\tau)^{\hat{q}+1}}{(\hat{q}+1)!} + o((t-\tau)^{\hat{q}+1}), \tag{1.37}$$

where, for the sake of simplicity, we denote by $\tau^\pm$ either τ^- if $\tau \in \mathcal{T}_{en}$ or τ^+ if $\tau \in \mathcal{T}_{ex}$.

[1] Lemma 2.44 of this thesis.

Combining Lemma 1.3 and (A6'), we see that for all $v \in C(u)$, the function (of time) $g_y(y_u)z_{u,v}$ vanishes just after entering or before leaving a boundary arc on a small interval $[\tau, \tau \pm \varepsilon]$, and so do its first $q-1$ time derivatives since the latter are continuous by Lemma 1.9(i). The derivative of order q of $g_y(y_u)z_{u,v}$ being a bounded function by (1.29), we have, on the interior arc side:

$$|g_y(y_u(t))z_{u,v}(t)| \leq C|t-\tau|^q. \tag{1.38}$$

If q is *odd*, combining (1.37) with $\hat{q} = 2q-1$ and (1.38) and by tangentiality assumption (A5)(i), we deduce from (1.36) that:

$$\varsigma_{u,v}(\tau) \;\geq\; \lim_{t\to\tau^\pm} \frac{C^2(t-\tau)^{2q}}{\frac{\mathrm{d}^{2q}}{\mathrm{d}t^{2q}}g(y_u)|_{t=\tau^\pm}\frac{(t-\tau)^{2q}}{(2q)!} + o((t-\tau)^{2q})} \;>\; -\infty.$$

If q is *even*, (1.37) with $\hat{q} = 2q-2$, (1.38) and (A5)(i) in (1.36) give:

$$\varsigma_{u,v}(\tau) \;\geq\; \lim_{t\to\tau^\pm} \frac{C^2(t-\tau)^{2q}}{\frac{\mathrm{d}^{2q-1}}{\mathrm{d}t^{2q-1}}g(y_u)|_{t=\tau^\pm}\frac{(t-\tau)^{2q-1}}{(2q-1)!} + o((t-\tau)^{2q-1})} \;=\; 0.$$

Since $\varsigma_{u,v}(\tau) \leq 0$ by (1.20) at an entry or exit point, it follows that (when q is even) $\varsigma_{u,v}(\tau) = 0$.

2) (Touch point). Assume now that $\tau \in \mathcal{T}_{to} \cap I^2_{u,v}$. If that case happens, since $v \in C_0(u)$, our hypotheses imply that τ is an essential touch point satisfying (1.32), and hence, that $q \geq 2$. Since $g(y_u)$ has an isolated local maximum at τ, $g(y_u)$ and $g^{(1)}(y_u)$ vanish at τ while $\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y_u) = g^{(2)}(u, y_u)$ is nonpositive and continuous at τ since u is continuous by Prop. 1.7(i). We thus have:

$$g(y_u(t)) = \frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y_u)|_{t=\tau}\frac{(t-\tau)^2}{2} + o((t-\tau)^2). \tag{1.39}$$

Since $\tau \in I^2_{u,v}$, we also have $g_y(y_u(\tau))z_{u,v}(\tau) = 0$. The function $g_y(y_u)z_{u,v}$ being C^1 (since $q \geq 2$) with almost everywhere a bounded second derivative, we get by (1.28), taking the nonnegative part:

$$(g_y(y_u(t))z_{u,v}(t))_+ = (g^{(1)}_y(y_u(\tau))z_{u,v}(\tau)(t-\tau))_+ + o(t-\tau). \tag{1.40}$$

From (1.39), (1.40) and (A5)(ii), $(g_y(y_u)z_{u,v})^2_+/g(y_u)$ is left-and right continuous when $t \to \tau$. Therefore, taking the lim inf when $t \to \tau$ comes to take the min of both limits when $t \to \tau^+$ and $t \to \tau^-$, thus we obtain:

$$\varsigma_{u,v}(\tau) = \min\left\{\frac{(g^{(1)}_y(y_u(\tau))z_{u,v}(\tau))^2}{g^{(2)}(u(\tau), y_u(\tau))}\;;\; 0\right\} = \frac{(g^{(1)}_y(y_u(\tau))z_{u,v}(\tau))^2}{g^{(2)}(u(\tau), y_u(\tau))} > -\infty. \tag{1.41}$$

3) (Conclusion). For all $\tau \in \mathcal{T} \cap I^2_{u,v}$, we showed that $\varsigma_{u,v}(\tau) > -\infty$. Therefore we may apply (1.21). Set $I_0 := \operatorname{int} I(G(u))$. By (1.18), we have $\operatorname{supp}(\mathrm{d}\eta) \subset I^2_{u,v}$ and in view of remark 1.8 we may write that:

$$\sigma(\eta, T^{2,i}_K(G(u), DG(u)v)) = \int_{I_0} \varsigma_{u,v}(t)\eta_0(t)\mathrm{d}t + \sum_{\tau\in\mathcal{T}\cap I^2_{u,v}} \nu_\tau \varsigma_{u,v}(\tau) \tag{1.42}$$

where $\eta_0 \in L^1(I_0)$ and $\nu_\tau = [\eta(\tau)]$. By (1.20), $\varsigma_{u,v}$ vanishes on $I_0 \cap I^2_{u,v}$ and thus on $I_0 \cap \operatorname{supp}(\eta_0)$. Hence, $\int_{I_0} \varsigma_{u,v}(t)\eta_0(t)\mathrm{d}t = 0$. If $\tau \in \mathcal{T}_{en} \cup \mathcal{T}_{ex}$, we have, if q is *odd*, $\nu_\tau = 0$ by Prop. 1.7(ii)(a)

and we showed that $\varsigma_{u,v}(\tau) > -\infty$. If q is *even*, we showed in point 1) that $\varsigma_{u,v}(\tau) = 0$ (and we have $\nu_\tau < +\infty$). In both cases, we deduce that $\nu_\tau \varsigma_{u,v}(\tau) = 0$.

It remains only in (1.42), when $q \geq 2$, the contribution of finitely many touch points τ in $\mathcal{T}_{to} \cap I^2_{u,v} = \mathcal{T}^{ess}_{to}$ with $\varsigma_{u,v}(\tau)$ given by (1.41). Hence (1.35) follows. □

Proof of Theorem 1.12. Combining Theorem 1.4 and Propositions 1.10 and 1.14, we obtain that (1.34) holds, for all $v \in C_0(u)$. Since the left-hand-side of (1.34) is a continuous quadratic form, it remains nonnegative on the closure of $C_0(u)$. We end the proof by checking that the latter is equal to $C(u)$, the cone of critical directions.

Since $C(u)$ is closed and contains $C_0(u)$, we have of course $\overline{C_0(u)} \subset C(u)$. We prove the converse relation. Let $v_0 \in C(u)$. We remind that $v \in C(u)$ iff $g_y(y_u)z_{u,v} \leq 0$ on $I(g(y_u))$ and $g_y(y_u)z_{u,v} = 0$ on the support of the Lagrange multiplier η. Let $\rho : \mathbb{R} \to \mathbb{R}$ be a function of class C^∞ having support on $[-1, 1]$ which is positive on $(-1, 1)$. For $\varepsilon > 0$, set $\rho_\varepsilon(t) := \varepsilon^{q+1}\rho(t/\varepsilon)$, thus we have $\rho_\varepsilon \to 0$ in $W^{q,\infty}$. By Lemma 1.9(ii), for $\varepsilon > 0$ small enough, there exists a unique $v_\varepsilon \in L^\infty(0, T)$ such that $g(y_u)z_{u,v_\varepsilon} = g(y_u)z_{u,v_0} - \sum_{t\in\mathcal{T}_{to}\setminus\mathcal{T}^{ess}_{to}} \rho_\varepsilon(t - \tau) \in W^{q,\infty}(0, T)$. Then we have $g_y(y_u)z_{u,v_\varepsilon} = g_y(y_u)z_{u,v_0}$ outside $(\tau - \varepsilon, \tau + \varepsilon)$, for all non essential touch point τ, $g_y(y_u(\tau))z_{u,v_\varepsilon}(\tau) < 0$ for such τ, and hence, the touch points being isolated, for $\varepsilon > 0$ small enough, $v_\varepsilon \in C_0(u)$. Since $DG(u)v_\varepsilon \to DG(u)v_0$ in $\mathcal{W}$, where $\mathcal{W}$ was defined in (1.30), and $DG(u)$ has a bounded inverse by Lemma 1.9(ii), we have $v_\varepsilon \to v_0$ in $L^\infty(0, T)$ when $\varepsilon \downarrow 0$. The conclusion follows. □

1.3.2 Extended Second-order Necessary Conditions

The solution $z_{u,v}$ of the linearized state equation (1.8) when $v \in L^2(0, T)$, is well-defined and belongs to $H^1(0, T) \subset C[0, T]$. Thus we may extend continuously $DJ(u)$ and $DG(u)$ over $L^2(0, T)$ (we keep the same notations for the extensions). Since $DG(u) : L^2(0, T) \to C[0, T]$, it makes sense to extend the critical cone $C(u)$ defined in (1.16) to critical directions in L^2, as follows:

$$C_{L^2}(u) = \{v \in L^2(0, T) \ ; \ DG(u)v \in T_K(G(u)) \ ; \ DJ(u)v \leq 0\}. \tag{1.43}$$

Note that when (u, η) satisfies (1.13), relation (1.18) remains true with $C_{L^2}(u)$ and $L^2(0, T)$ instead of respectively $C(u)$ and $\mathcal{U}$.

The necessary and sufficient second-order conditions involve respectively $C(u)$ and $C_{L^2}(u)$ (see sections 1.4 and 1.5). Therefore, to obtain the no-gap second-order conditions, we need the following variant of Theorem 1.12.

Corollary 1.15. *The statements of Theorem 1.12 and Corollary 1.13 still hold replacing assumption (A6') and $C(u)$ respectively by (A6) and $C_{L^2}(u)$.*

Corollary 1.15 is obtained as a consequence of Th. 1.12, the continuity of the left-hand side of (1.34) w.r.t. $v \in L^2$, and the density of $C(u)$ in $C_{L^2}(u)$ (Lemma 1.17). To prove the latter, we first need a general result.

Lemma 1.16. *Let $q \geq 1$ and $a < b \in \mathbb{R}$. Then for all $\hat{x} \in H^q(a, b) = W^{q,2}(a, b)$, there exists a sequence (x_n) of $W^{q,\infty}(a, b)$ such that $x_n^{(j)}(a) = \hat{x}^{(j)}(a)$, $x_n^{(j)}(b) = \hat{x}^{(j)}(b)$ for all $j = 0, \ldots, q-1$, $n \in \mathbb{N}$ and $\|x_n - \hat{x}\|_{q,2} \to 0$.*

Proof. Set $\hat{x}_a := (\hat{x}(a), \ldots, \hat{x}^{(q-1)}(a))^*$, $\hat{x}_b := (\hat{x}(b), \ldots, \hat{x}^{(q-1)}(b))^* \in \mathbb{R}^q$ and $\hat{u} := \hat{x}^{(q)} \in L^2(a, b)$. For $u \in L^2(a, b)$, let $x_u \in H^q(a, b)$ be the solution of:

$$x_u^{(q)}(t) = u(t) \quad \text{a.e. on } [a, b] \quad ; \quad (x_u(a), \ldots, x_u^{(q-1)}(a)) = \hat{x}_a^*. \tag{1.44}$$

For $n \in \mathbb{N}$, consider the following problem:

$$(\mathcal{P}_n) \qquad \min \tfrac{1}{2}\|u-\hat{u}\|_2^2 \quad ; \quad \mathcal{A}u = \hat{x}_b \quad ; \quad u \in \mathcal{U}_n, \tag{1.45}$$

where $\mathcal{U}_n := \{u \in L^2(0,T) \ ; \ |u(t)| \leq n \text{ a.e.}\}$ and $\mathcal{A} : L^2 \to \mathbb{R}^q \ ; \ u \mapsto (x_u(b), \ldots, x_u^{(q-1)}(b))^*$. By construction, $\mathcal{A}\hat{u} = \hat{x}_b$. It is readily seen that the mapping $L^2(a,b) \to H^q(a,b); \ u \mapsto x_u$ solution of (1.44) is continuous. Since $H^q(a,b)$ has a continuous inclusion into $C^{q-1}[a,b]$, it follows that the linear mapping $\mathcal{A}$ is also continuous.

Let us first show that for n large enough, the problems $(\mathcal{P}_n)$ are feasible and uniformly qualified, that is there exist $n_0 \in \mathbb{N}$ and $\delta_0 > 0$ such that

$$\hat{x}_b + \delta_0 B_{\mathbb{R}^q} \subset \mathcal{A}\mathcal{U}_{n_0} \subset \mathcal{A}\mathcal{U}_n \qquad \forall n \geq n_0, \tag{1.46}$$

with $B_{\mathbb{R}^q}$ the unit ball in $\mathbb{R}^q$. Indeed, consider e.g. for $\delta \in \mathbb{R}^q$ the (unique) polynomial function x_δ of degree $2q-1$ that takes with its $q-1$ first derivatives the values $\hat{x}_a$ and $\hat{x}_b + \delta$ at a and b. It is easily seen that its coefficients are solution of a full-rank linear system with $\hat{x}_b - \hat{x}_a + \delta$ as right-hand side, hence, taking the sup over $(t,\delta) \in [a,b] \times B_{\mathbb{R}^q}(0,\delta_0)$ of the functions $u_\delta(t) = x_\delta^{(q)}(t)$ that are C^∞ w.r.t. t and δ provides an uniform bound n_0 such that (1.46) holds.

Since Robinson's constraint qualification holds for n large enough, there exists a (unique) optimal solution u_n of $(\mathcal{P}_n)$ and a normal Lagrange multiplier $\lambda_n \in \mathbb{R}^{q*}$, such that (throughout the proof, $\langle \cdot, \cdot \rangle$ denotes the scalar product over L^2):

$$0 \leq \langle u_n - \hat{u} + \mathcal{A}^*\lambda_n, v - u_n \rangle \qquad \forall v \in \mathcal{U}_n. \tag{1.47}$$

Since the feasible set of problem $(\mathcal{P}_n)$ is increasing for inclusion when $n \to +\infty$, the cost function is decreasing, thus $\|u_n - \hat{u}\|_2$ is bounded. Hence the sequence (u_n) converges weakly to some $\bar{u} \in L^2$. We may rewrite (1.47) as:

$$\|u_n - \hat{u}\|_2^2 + \lambda_n(\hat{x}_b - \mathcal{A}v) \leq \langle u_n - \hat{u}, v - \hat{u} \rangle \qquad \forall v \in \mathcal{U}_n. \tag{1.48}$$

Qualification property (1.46) implies that $\delta_0|\lambda_n| \leq \sup_{v \in \mathcal{U}_{n_0}} \lambda_n(\hat{x}_b - \mathcal{A}v)$, hence, taking the sup for $v \in \mathcal{U}_{n_0}$ successively in the right and left hand side of (1.48), we deduce that for some constant $K(n_0) > 0$ that depends on n_0, we have $\delta_0|\lambda_n| \leq K(n_0)$, for all $n \geq n_0$. Therefore the sequence (λ_n) is uniformly bounded. Define now $v_n \in \mathcal{U}_n$ as $v_n(t) = \max\{-n; \min\{n, \hat{u}(t)\}\}$ a.e. By the Lebesgue dominated convergence Theorem, $v_n \to \hat{u}$ in L^2 and by (1.48):

$$\|u_n - \hat{u}\|_2^2 \ \leq \ \langle u_n - \hat{u}, v_n - \hat{u} \rangle \ + \ \lambda_n(\mathcal{A}v_n - \hat{x}_b) \ \longrightarrow \ 0,$$

since $u_n - \hat{u} \rightharpoonup \bar{u} - \hat{u}$ weakly in L^2, $v_n - \hat{u} \to 0$ strongly in L^2, λ_n is bounded and $\mathcal{A}v_n \to \mathcal{A}\hat{u} = \hat{x}_b$ by continuity of $\mathcal{A}$. It follows that $\|u_n - \hat{u}\|_2 \to 0$ and the sequence $x_n := x_{u_n}$ satisfies all the required properties, so the proof is completed. □

Lemma 1.17. *Let $u \in \mathcal{U}$ and $\eta \in \Lambda(u)$ such that (A3), (A4) and (A6) are satisfied. Then $C(u)$ is a dense subset of $C_{L^2}(u)$.*

Proof. Since (A4) holds, denote by $0 < \tau_1 < \ldots < \tau_N < T$ the junction times of the trajectory (u, y_u), and set $\tau_0 := 0$, $\tau_{N+1} := T$. Let $v \in C_{L^2}(u)$ and set $x := DG(u)v$. By Lemma 1.16 applied on intervals $[\tau_k, \tau_{k+1}]$ that are not boundary arcs, there exists a sequence $x_n \in W^{q,\infty}(0,T)$ such that $x_n = 0 = x$ by (A6) on boundary arcs, $x_n^{(j)}(\tau_k) = x^{(j)}(\tau_k)$ for all $j = 0, \ldots, q-1$ and $k = 0, \ldots, N+1$, and $x_n \to x$ in H^q. By (A3) and Lemma 1.9(ii), we may define $v_n \in L^\infty(0,T)$ such that $DG(u)v_n = x_n$ for all n. It is readily seen that $v_n \in C(u)$ for all n and $v_n \to v$ in L^2, which achieves the proof. □

1.4 Second-order Sufficient Conditions

The second-order sufficient conditions theory classically involves two norms, namely L^2 and L^∞, see Ioffe [73, Part III] and Maurer [99].

Assume that X, Z are Banach spaces endowed with the norms $\|\cdot\|_X$ and $\|\cdot\|_Z$, respectively, such that $Z \subset X$ with continuous embedding. Let $k \in \mathbb{N}$. We say that $r(x) = \mathcal{O}_Z(\|x\|_X^k)$ if $|r(x)| \leq C\|x\|_X^k$ for some $C > 0$ when $\|x\|_Z$ is small enough. We say that $r(x) = o_Z(\|x\|_X^k)$ if $|r(v)|/\|x\|_X^k$ goes to zero when $\|x\|_Z$ goes to zero. In the sequel, $\|\cdot\|_p$ (resp. $\|\cdot\|_{r,p}$) denotes the norm of the space $L^p(0,T)$ (resp. the Sobolev space $W^{r,p}(0,T)$), for $1 \leq p \leq \infty$ and $r = 1, \ldots < +\infty$. We write $\mathcal{O}_p$ and $\mathcal{O}_{r,p}$ for respectively $\mathcal{O}_{\|\cdot\|_{L^p}}$ and $\mathcal{O}_{\|\cdot\|_{W^{r,p}}}$, and we use the same convention for o_p and $o_{r,p}$. Similarly, B_p and $B_{r,p}$ denote open balls in L^p and $W^{r,p}$, respectively.

We remind that a quadratic form $Q(v)$ on a Hilbert space is a *Legendre form* (Ioffe and Tihomirov [74]), if it is weakly lower semi-continuous (w.l.s.c.) and if $v_n \rightharpoonup v$ weakly and $Q(v_n) \to Q(v)$ imply that $v_n \to v$ strongly.

The next theorem gives the second-order sufficient condition in its well-known form (i.e. without the curvature term).

Theorem 1.18. *Let $u \in \mathcal{U}$ satisfy (1.13) with Lagrange multiplier η and assume that (A2') holds. If the following second-order sufficient condition is satisfied:*

$$D^2_{uu}L(u,\eta)(v,v) > 0 \qquad \forall\, v \in C_{L^2}(u) \setminus \{0\} \tag{1.49}$$

then u is a local solution of (1.4) satisfying the quadratic growth condition (1.5).

Conversely, if (A1)-(A6) hold and if (u, y_u) has no essential touch point (in particular, if the state constraint is of first order $q = 1$*), then the second-order sufficient condition (1.49) is satisfied* iff *the quadratic growth condition (1.5) is satisfied.*

The proof of Theorem 1.18 will be given after a sequence of short lemmas.

Lemma 1.19. *Let $(u,\eta) \in \mathcal{U} \times \mathcal{M}_+[0,T]$ and $v \in \mathcal{U}$. The following holds, for all $\sigma \in [0,1]$:*

$$\|y_{u+\sigma v} - y_u\|_\infty = \mathcal{O}_\infty(\|v\|_1) \tag{1.50}$$
$$\|p_{u+\sigma v,\eta} - p_{u,\eta}\|_\infty = \mathcal{O}_\infty(\|v\|_1) \tag{1.51}$$
$$\|z_{u+\sigma v,v}\|_\infty = \mathcal{O}_\infty(\|v\|_1) \tag{1.52}$$
$$\|z_{u+\sigma v,v} - z_{u,v}\|_\infty = \mathcal{O}_\infty(\|v\|_2^2). \tag{1.53}$$

Proof. Set $u_\sigma := u + \sigma v$, and let C denote a positive constant. Since f is Lipschitz continuous by (A0), (1.50) is an easy consequence of Lemma 1.32. Thus, u and v being essentially bounded, u_σ and y_{u_σ} take values a.e. in a compact set of type

$$V_\delta = \{(\hat{u},\hat{y}) \in \mathbb{R} \times \mathbb{R}^n \ ;\ |\hat{u}| + |\hat{y}| \leq \delta\}, \tag{1.54}$$

for some $\delta > 0$. The mappings f, ℓ and g as well as their first order derivatives are C^1, and hence Lipschitz continuous over the compact set V_δ. Lemma 1.32, applied to the costate equation (1.7), ensures that $p_{u_\sigma,\eta}$ also remains uniformly bounded. The derivation of (1.51) and (1.52) being similar to the one of (1.53), we detail only the latter. We have (omitting time argument):

$$\begin{aligned} |\dot{z}_{u_\sigma,v}(t) - \dot{z}_{u,v}(t)| \ \leq\ & \|f_y\|_\infty |z_{u_\sigma,v} - z_{u,v}| \\ & + \left(|Df(u_\sigma, y_{u_\sigma}) - Df(u, y_u)|\right)\left(|z_{u,v}| + |v(t)|\right). \end{aligned}$$

Since Df is Lipschitz on V_δ, we have by (1.50) $|Df(u_\sigma, y_{u_\sigma}) - Df(u, y_u)| \leq C(\|v\|_1 + |v|)$. Combining with (1.52) and the inequality $ab \leq \frac{1}{2}(a^2 + b^2)$, we deduce from the above display that

$$|\dot{z}_{u_\sigma,v}(t) - \dot{z}_{u,v}(t)| \quad \leq \quad \|f_y\|_\infty |z_{u_\sigma,v} - z_{u,v}| + C\left(\|v\|_1^2 + |v(t)|^2\right).$$

We conclude with Lemma 1.32 and the inequality $\|v\|_1 \leq \sqrt{T}\|v\|_2$. □

Lemma 1.20. *Let $(u,\eta) \in \mathcal{U} \times \mathcal{M}_+[0,T]$ and $v \in \mathcal{U}$. Then:*

$$L(u+v,\eta) = L(u,\eta) + D_u L(u,\eta)v + \frac{1}{2} D^2_{uu} L(u,\eta)(v,v) + r(v) \tag{1.55}$$

with $r(v) = \mathcal{O}_\infty(\|v\|_3^3)$. In particular, $r(v) = o_\infty(\|v\|_2^2)$.

Proof. For $\sigma \in [0,1]$, set again $u_\sigma := u + \sigma v$ and $p_{u_\sigma} := p_{u_\sigma,\eta}$. By Lemma 1.1:

$$\begin{aligned} r(v) &= \left[\int_0^1 (1-\sigma)\left(D^2_{uu}L(u+\sigma v,\eta) - D^2_{uu}L(u,\eta)\right)\mathrm{d}\sigma\right](v,v) \\ &= \int_0^1\int_0^T \Delta_1(t)\mathrm{d}t\mathrm{d}\sigma + \int_0^1\int_0^T \Delta_2(t)\mathrm{d}\eta(t)\mathrm{d}\sigma + \int_0^1 \Delta_3 \mathrm{d}\sigma, \end{aligned} \tag{1.56}$$

with (omitting time argument)

$$\begin{aligned} \Delta_1(t) &= D^2 H_{(u,y)^2}(u_\sigma, y_{u_\sigma}, p_{u_\sigma})(v, z_{u_\sigma,v})^2 - D^2 H_{(u,y)^2}(u, y_u, p_u)(v, z_{u,v})^2 \\ \Delta_2(t) &= z^*_{u_\sigma,v} g_{yy}(y_{u_\sigma}) z_{u_\sigma,v} - z^*_{u,v} g_{yy}(y_u) z_{u,v} \\ \Delta_3 &= z_{u_\sigma,v}(T)^* \phi_{yy}(y_{u_\sigma}(T)) z_{u_\sigma,v}(T) - z_{u,v}(T)^* \phi_{yy}(y_u(T)) z_{u,v}(T). \end{aligned}$$

Under assumption (A0), second-order derivatives g_{yy}, etc. are Lipschitz continuous over a compact set V_δ defined in (1.54) for some $\delta > 0$. By Lemma 1.19 we get, for some constant $C > 0$:

$$\begin{aligned} \Delta_2(t) &\leq C\left(|y_{u_\sigma} - y_u||z_{u_\sigma,v}|^2 + (|z_{u_\sigma,v}| + |z_{u,v}|)|z_{u_\sigma,v} - z_{u,v}|\right) \\ &\leq \mathcal{O}_\infty(\|v\|_1^3 + \|v\|_1 \|v\|_2^2) \leq \mathcal{O}_\infty(\|v\|_3^3), \end{aligned}$$

since by the Cauchy-Schwarz and Hölder inequalities, that give respectively $\|\cdot\|_2^2 \leq \|\cdot\|_3^{3/2}\|\cdot\|_1^{1/2}$ and $\|\cdot\|_1 \leq T^{2/3}\|\cdot\|_3$, we have $\|\cdot\|_2^2\|\cdot\|_1 \leq T\|\cdot\|_3^3$. Since the measure $\mathrm{d}\eta$ is bounded and the $\mathcal{O}_\infty$ are uniform w.r.t. time, we obtain $\int_0^T \Delta_2(t)\mathrm{d}\eta(t) = \mathcal{O}_\infty(\|v\|_3^3)$. The same upper bound holds for $\Delta_3(T)$. As for $\Delta_1(t)$, we have in the same way, by Lemma 1.19:

$$\begin{aligned} \Delta_1(t) &\leq C(|y_{u_\sigma} - y_u| + |p_{u_\sigma} - p_u| + \sigma|v|)(|z_{u_\sigma,v}|^2 + |v|^2) \\ &\quad + C(|z_{u_\sigma,v}| + |z_{u,v}| + |v|)|z_{u_\sigma,v} - z_{u,v}| \\ &\leq C(\|v\|_1^3 + \|v\|_1^2|v(t)| + \|v\|_1|v(t)|^2 + |v(t)|^3 + \|v\|_1\|v\|_2^2 + \|v\|_2^2|v(t)|). \end{aligned}$$

Hence, $\int_0^T \Delta_1(t)\mathrm{d}t = \mathcal{O}_\infty(\|v\|_3^3)$. Finally, since the $\mathcal{O}_\infty$ do not depend on $\sigma \in [0,1]$, we obtain after integration over $[0,1]$ that $r(v) = \mathcal{O}_\infty(\|v\|_3^3)$. Since $\|\cdot\|_3^3 \leq \|\cdot\|_2^2\|\cdot\|_\infty$, it follows that $r(v) = o_\infty(\|v\|_2^2)$. □

Lemma 1.21. *Let $(u,\eta) \in \mathcal{U} \times \mathcal{M}_+[0,T]$ satisfy (A2'). Then the quadratic form $\mathcal{U} \to \mathbb{R}$, $v \mapsto D^2_{uu}L(u,\eta)(v,v)$ has a unique extension to a continuous quadratic form over $L^2(0,T)$, and the latter is a Legendre form.*

Proof. Since L^∞ is a dense subset of L^2 and $v \mapsto D^2_{uu}L(u,\eta)(v,v)$ is continuous for the norm of L^2, it has a unique continuous extension Q over L^2. Set $p := p_{u,\eta}$. By (1.10), we can write $Q(v) = Q_0(v) + Q_1(v) + Q_2(v)$ with:

$$\begin{array}{rcl} Q_2(v) & = & \int_0^T H_{yy}(u,y_u,p)(z_{u,v},z_{u,v})\mathrm{d}t \\ & & + z_{u,v}(T)^*\phi_{yy}(y_u(T))z_{u,v}(T) + \int_0^T z^*_{u,v}g_{yy}(y_u)z_{u,v}\mathrm{d}\eta \\ Q_1(v) & = & 2\int_0^T H_{yu}(u,y_u,p)(z_{u,v},v)\mathrm{d}t \\ Q_0(v) & = & \int_0^T H_{uu}(u,y_u,p)(v,v)\mathrm{d}t. \end{array}$$

Let $v_n \rightharpoonup \bar v \in L^2(0,T)$. The mapping $L^2(0,T) \to H^1(0,T)$; $v \mapsto z_{u,v}$ being linear continuous, $z_n := z_{u,v_n}$ converges weakly to $\bar z := z_{u,\bar v}$. Since (z_n) is bounded in $H^1(0,T)$ and the inclusion of the latter in $C[0,T]$ is compact, (z_n) is strongly convergent to $\bar z$, and thus $Q_2(v_n)$ converges strongly to $Q_2(\bar v)$. The term $Q_1(v_n)$, bilinear in (z_n, v_n), also converges strongly to $Q_1(\bar v)$ when z_n converges strongly and v_n weakly. Therefore, Q is a Legendre form iff Q_0 is one.

Since $H_{uu}(u(t), y_u(t), p(t))$ is essentially bounded and, by (1.27), is uniformly invertible for almost all $t \in [0,T]$, $v \mapsto \sqrt{Q_0(v)}$ is a norm equivalent to the one of $L^2(0,T)$. Hence by [24, Prop. 3.76(i)], Q_0 is a Legendre form, and therefore so is Q. □

Proof of Theorem 1.18. Assume that (1.49) holds but that the quadratic growth condition (1.5) is not satisfied. Then there exist a sequence $u_n \to u$ in L^∞, $u_n \neq u$, such that $G(u_n) \in K$ for all n and

$$J(u_n) \leq J(u) + o(\|u_n - u\|_2^2). \tag{1.57}$$

Since $G(u_n) \in K$ and $\eta \in N_K(G(u))$, we have:

$$J(u_n) - J(u) = L(u_n,\eta) - L(u,\eta) - \langle \eta, G(u_n) - G(u)\rangle \geq L(u_n,\eta) - L(u,\eta).$$

Since $u_n - u \to 0$ in L^∞, Lemma 1.20 yields $r(u_n - u) = o(\|u_n - u\|_2^2)$. As $D_uL(u,\eta) = 0$, we have:

$$o(\|u_n - u\|_2^2) \geq J(u_n) - J(u) \geq \frac{1}{2}D^2_{uu}L(u,\eta)(u_n - u, u_n - u) + o(\|u_n - u\|_2^2).$$

Let (v_n, ϵ_n) be such that $u_n - u = \epsilon_n v_n$ with $\|v_n\|_2 = 1$ and $\epsilon_n = \|u_n - u\|_2 \to 0$. Dividing by $\epsilon_n^2 > 0$ the above inequality, we get:

$$D^2_{uu}L(u,\eta)(v_n,v_n) + o(1) \leq o(1). \tag{1.58}$$

The sequence (v_n) being bounded in $L^2(0,T)$, taking if necessary a subsequence, we may assume that (v_n) converges weakly to some $\bar v \in L^2(0,T)$. Since $D^2_{uu}L(u,\eta)$ is weakly l.s.c., we get passing to the limit:

$$D^2_{uu}L(u,\eta)(\bar v,\bar v) \leq 0. \tag{1.59}$$

From (1.57), we derive that $J(u+\epsilon_n v_n) - J(u) = \epsilon_n DJ(u)v_n + r_n \leq o(\epsilon_n^2)$, where $r_n = \mathcal{O}(\epsilon_n^2)$ (by the same arguments as in the proof of Lemma 1.20). Thus $DJ(u)v_n + \mathcal{O}(\epsilon_n) \leq o(\epsilon_n)$, and passing to the limit, since the mapping $v \mapsto DJ(u)v = \int_0^T (\ell_y(u,y_u)z_{u,v} + \ell_u(u,y_u)v)\mathrm{d}t + \phi_y(y_u(T))z_{u,v}(T)$ is weakly continuous, we obtain:

$$DJ(u)\bar v \leq 0. \tag{1.60}$$

Since $K \ni G(u_n) = G(u) + \epsilon_n DG(u)v_n + \epsilon_n r_n$, where r_n is a continuous function satisfying $\|r_n\|_\infty = \mathcal{O}(\epsilon_n)$, we deduce that

$$DG(u)v_n + r_n \in T_K(G(u)). \tag{1.61}$$

Since the mapping $DG(u) : L^2 \to C[0,T]$ is linear and continuous for the strong topologies, it is also continuous for the weak topologies, which implies that $DG(u)v_n \rightharpoonup DG(u)\bar{v}$. The set K being closed and convex, so is $T_K(G(u))$, and hence the latter is weakly closed. Therefore, passing to the weak limit in (1.61), and using (1.60), we obtain that $\bar{v} \in C_{L^2}(u)$. Thus (1.49) and (1.59) imply that $\bar{v} = 0$. On the other hand, (1.58) gives (with $Q := D^2_{uu}L(u,\eta)$):

$$0 = Q(\bar{v}) \leq \liminf Q(v_n) \leq \limsup Q(v_n) \leq 0$$

therefore $Q(v_n) \to Q(v)$. But Q is a Legendre form by Lemma 1.21 and $v_n \rightharpoonup \bar{v}$, which implies that $v_n \to \bar{v}$ in $L^2(0,T)$, hence $\|v_n\|_2 \to \|\bar{v}\|_2$. The expected contradiction arises since $\|v_n\|_2 = 1$ for all n whereas $\|\bar{v}\|_2 = 0$.

The converse, that holds under stronger assumptions, is a consequence of Corollaries 1.13 and 1.15. For convenience, we prove it later with Theorem 1.27. □

1.5 Reduction Approach

There is still a gap between statements of Corollary 1.15 of Theorem 1.12 and Theorem 1.18, whenever essential touch points occur. We show in this section how to deal with this case, using a reduction approach in order to reformulate the constraint.

The idea of reduction methods (see e.g. [72] and [24, section 3.4.4]) is, when the constraint has finitely many contact points, to replace it by finitely many inequality constraints. The Hessian of Lagrangian of the corresponding reduced problem has an additional term that matches the curvature term. We obtain thus a no-gap second-order condition.

1.5.1 General results on reduction

It is known that the Sobolev spaces $W^{1,\infty}(0,T)$ and $W^{2,\infty}(0,T)$, endowed with the norms $\|x\|_{1,\infty} = \|x\|_\infty + \|\dot{x}\|_\infty$ and $\|x\|_{2,\infty} = \|x\|_{1,\infty} + \|\ddot{x}\|_\infty$, coincide with the spaces of Lipschitz continuous functions and the one of functions having a Lipschitz continuous derivative, respectively. For all $t, t_0 \in [0,T]$, $h \in W^{1,\infty}(0,T)$ and $x \in W^{2,\infty}(0,T)$, we have:

$$|h(t) - h(t_0)| \leq |t - t_0| \|\dot{h}\|_\infty, \tag{1.62}$$

$$|x(t) - x(t_0) - \dot{x}(t_0)(t - t_0)| \leq \tfrac{1}{2}|t - t_0|^2 \|\ddot{x}\|_\infty. \tag{1.63}$$

We now give some general results about zeros of functions of $W^{1,\infty}(0,T)$, and local minima/maxima of functions of $W^{2,\infty}(0,T)$.

Lemma 1.22. *Let $h_0 \in W^{1,\infty}(0,T)$ and $\tau_0 \in (0,T)$ satisfy the three following conditions: $h_0(\tau_0) = 0$; $\dot{h}_0$ is continuous at τ_0 ; $\dot{h}_0(\tau_0) \neq 0$. Then for some $\delta,\ \varepsilon > 0$, the mapping:*

$$\Xi \ : \ B_{1,\infty}(h_0, \delta) \to (\tau_0 - \varepsilon, \tau_0 + \varepsilon) \quad ; \quad h \mapsto \tau_h \ \text{ such that } \ h(\tau_h) = 0, \tag{1.64}$$

is well-defined and Lipschitz continuous on $B_{1,\infty}(h_0,\delta)$, and Fréchet differentiable at h_0, with derivative given by:

$$D\Xi(h_0)d = -d(\tau_0)/\dot{h}_0(\tau_0), \qquad \text{for all } d \in W^{1,\infty}. \tag{1.65}$$

More precisely, we have for all $h, h_i \in B_{1,\infty}(h_0,\delta)$, $i = 1,2$ and $\tau_i = \tau_{h_i}$:

$$\tau_2 - \tau_1 = \mathcal{O}_{1,\infty}(\|h_2 - h_1\|_\infty), \tag{1.66}$$

$$\dot{h}_0(\tau_0)(\tau_h - \tau_0) + h(\tau_0) = o_{1,\infty}\left(\|h - h_0\|_\infty\right). \tag{1.67}$$

Proof. Assume w.l.o.g that $\beta := \dot{h}_0(\tau_0) > 0$, and denote by $c(\cdot)$ the modulus of continuity of $\dot{h}_0$ at τ_0. Fix $\varepsilon > 0$ such that $c(\varepsilon) < \frac{1}{4}\beta$. Thus, $\dot{h}_0 \geq \frac{3}{4}\beta$ on $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$ and it follows that $h_0(\tau_0 - \varepsilon) < -\frac{3}{4}\beta\varepsilon$ and $h_0(\tau_0 + \varepsilon) > \frac{3}{4}\beta\varepsilon$. Set $\delta := \min\{\frac{1}{4}\beta\varepsilon; \frac{1}{4}\beta\}$ and let $h \in B_{1,\infty}(h_0, \delta)$. Thus, $h(\tau_0 - \varepsilon) < 0 < h(\tau_0 + \varepsilon)$ and h is continuous, so h has at least one zero τ_h in $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$. Let $(h_1, h_2) \in B_{1,\infty}(h_0, \delta)$ and τ_i such that $h_i(\tau_i) = 0$, $i = 1, 2$. By the definition of δ, we have $\dot{h}_1 \geq \frac{1}{2}\beta$ a.e. on $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$, and, in consequence,

$$\frac{\beta}{2}|\tau_2 - \tau_1| \leq |h_1(\tau_2)| = |h_1(\tau_2) - h_2(\tau_2)| \leq \|h_2 - h_1\|_\infty. \tag{1.68}$$

Hence $|\tau_2 - \tau_1| \leq \frac{2}{\beta}\|h_2 - h_1\|_\infty$, which shows the uniqueness of the zero (take $h_1 = h_2$), Lipschitz continuity and (1.66).

By continuity of Ξ and h_0, and (1.62) applied to $h - h_0$, we have:

$$\begin{aligned} h_0(\tau_h) - \dot{h}_0(\tau_0)(\tau_h - \tau_0) &= o(|\tau_h - \tau_0|) \\ (h - h_0)(\tau_h) - (h - h_0)(\tau_0) = -h_0(\tau_h) - h(\tau_0) &= \mathcal{O}(\|\dot{h} - \dot{h}_0\|_\infty |\tau_h - \tau_0|). \end{aligned}$$

Since $\tau_h - \tau_0 = \mathcal{O}_{1,\infty}(\|h - h_0\|_\infty)$ by (1.68), summing the above expansions yields (1.67), from which (1.65) follows. □

Lemma 1.23. *Let $x_0 \in W^{2,\infty}(0,T)$ and $\tau_0 \in (0,T)$ be such that $\dot{x}_0(\tau_0) = 0$, $\ddot{x}_0$ is continuous at τ_0 and $\ddot{x}_0(\tau_0) < 0$. Thus x_0 has a local maximum at τ_0, and for $\varepsilon > 0$ and $\delta > 0$ small enough, $x \in B_{2,\infty}(x_0, \delta)$ attains its maximum over $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$ at a unique point τ_x. The mapping $\Theta : B_{2,\infty}(x_0, \delta) \to (\tau_0 - \varepsilon, \tau_0 + \varepsilon)$; $x \mapsto \tau_x$ is Lipschitz continuous over $B_{2,\infty}(x_0, \delta)$, Fréchet differentiable at x_0, with derivative given by:*

$$D\Theta(x_0)w = -\dot{w}(\tau_0)/\ddot{x}_0(\tau_0) \qquad \forall\, w \in W^{2,\infty}. \tag{1.69}$$

Furthermore, the mapping

$$\Phi : B_{2,\infty}(x_0, \delta) \to \mathbb{R} \ ; \ x \mapsto x(\tau_x), \tag{1.70}$$

that associates with x the value of its maximum on $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$, is C^1 over $B_{2,\infty}(x_0, \delta)$ and twice Fréchet differentiable at x_0 with first and second derivatives given by, for all $x \in B_{2,\infty}(x_0, \delta)$ and $d \in W^{2,\infty}$:

$$D\Phi(x)d = d(\tau_x) \qquad ; \qquad D^2\Phi(x_0)(d,d) = -\frac{\dot{d}(\tau_0)^2}{\ddot{x}_0(\tau_0)}. \tag{1.71}$$

More precisely, for all $x, x_i \in B_{2,\infty}(x_0, \delta)$, $i = 1, 2$ and $\tau_i = \tau_{x_i}$, we have:

$$x_2(\tau_2) = x_2(\tau_1) + \mathcal{O}_{2,\infty}(\|x_2 - x_1\|_{1,\infty}^2), \tag{1.72}$$

$$x(\tau_x) = x(\tau_0) - \frac{\dot{x}(\tau_0)^2}{2\ddot{x}_0(\tau_0)} + o_{2,\infty}(\|x - x_0\|_{1,\infty}^2). \tag{1.73}$$

Proof. Define δ as in the proof of Lemma 1.22, with h_0 replaced by $-\dot{x}_0$. It follows that for all $x \in B_{2,\infty}(x_0, \delta)$, there exists a unique τ_x satisfying $\dot{x}(\tau_x) = 0$, and we have $\ddot{x}(t) \leq \ddot{x}_0(\tau_0)/2 < 0$ a.e. on $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$. Hence $\dot{x}$ is decreasing on $(\tau_0 - \varepsilon, \tau_0 + \varepsilon)$, and x has unique maximum over $[\tau_0 - \varepsilon, \tau_0 + \varepsilon]$ attained at time τ_x. By composition of the mapping Ξ of Lemma 1.22 by the mapping $x \mapsto h = \dot{x} \in W^{1,\infty}$, Θ is well-defined, continuous over $B_{2,\infty}(x_0, \delta)$ and Fréchet differentiable at x_0, and (1.69) follows from (1.65).

By (1.63) applied to x_2, introducing the term $\dot{x}_1(\tau_1)$ equal to zero and since $\tau_2 - \tau_1 = \mathcal{O}_{2,\infty}(\|x_2 - x_1\|_{1,\infty})$ by (1.66), we get:

$$\begin{aligned} x_2(\tau_2) &= x_2(\tau_1) + (\dot{x}_2(\tau_1) - \dot{x}_1(\tau_1))(\tau_2 - \tau_1) + \mathcal{O}(|\tau_2 - \tau_1|^2) \\ &= x_2(\tau_1) + \mathcal{O}_{2,\infty}(\|x_2 - x_1\|^2_{1,\infty}) \end{aligned}$$

which shows (1.72) and proves that Φ is C^1 with first order derivative given by (1.71). By continuity of $\ddot{x}_0$ and (1.63) applied to $x - x_0$, we have, as $\dot{x}_0(\tau_0) = 0$:

$$\begin{aligned} x_0(\tau_x) &= x_0(\tau_0) + \ddot{x}_0(\tau_0)\tfrac{(\tau_x - \tau_0)^2}{2} + o(|\tau_x - \tau_0|^2), \\ (x - x_0)(\tau_x) &= (x - x_0)(\tau_0) + \dot{x}(\tau_0)(\tau_x - \tau_0) + \mathcal{O}(\|\ddot{x} - \ddot{x}_0\|_\infty |\tau_x - \tau_0|^2). \end{aligned}$$

Summing the above expansions, and since by (1.67),

$$\tau_x - \tau_0 = -\frac{\dot{x}(\tau_0)}{\ddot{x}_0(\tau_0)} + o_{2,\infty}\left(\|x - x_0\|_{1,\infty}\right),$$

we obtain (1.73). Hence Φ is twice Fréchet differentiable at x_0 with second-order derivative given by (1.71). □

1.5.2 Application to optimal control problems.

If the state constraint is of first order $q = 1$, then Theorem 1.18 gives a no-gap second-order condition, that characterizes the quadratic growth. We show in this section how to extend this no-gap condition to the case when the trajectory has essential touch points (see Theorem 1.27).

Therefore, we assume in this section that the state constraint *is not of first order*, that is, the function $g^{(1)}(u, y) = g_y(y)f(u, y)$ does not depend on u (which means $g^{(1)}_u(u, y) \equiv 0$). Note that this implies that $G(u) = g(y_u) \in W^{2,\infty}$, for all $u \in \mathcal{U}$.

Definition 1.24. Assume that $g^{(1)}_u \equiv 0$ (the state constraint is not of order one). Let $u \in G^{-1}(K)$. We say that a touch point τ of the trajectory (u, y_u) is reducible, if the following conditions are satisfied: (i) the function $t \mapsto g^{(2)}(u(t), y_u(t))$ is continuous at τ; (ii) non-tangentiality condition (1.32) is satisfied at τ.

Remark 1.25. 1) Point (i) in the above definition is always satisfied if the state constraint is of order $q > 2$, since in that case $g^{(2)}(u, y_u) = g^{(2)}(y_u)$.
2) If $q = 2$ and $\eta \in \Lambda(u) \neq \emptyset$, sufficient conditions for point (i) are assumptions (A2)-(A4), since by Prop. 1.7(i) they imply the continuity of u.

Let $u \in G^{-1}(K)$, and let $\mathcal{T}_{red}$ be a *finite* subset of reducible touch points of the trajectory (u, y_u). By definition of touch points, there exists $\varepsilon > 0$ such that $(\tau - 2\varepsilon, \tau + 2\varepsilon) \subset (0, T)$ and $(\tau - 2\varepsilon, \tau + 2\varepsilon) \cap I(g(y_u)) = \{\tau\}$, for all $\tau \in \mathcal{T}_{red}$. Set $I_a = \cup_{\tau \in \mathcal{T}_{red}}(\tau - \varepsilon, \tau + \varepsilon)$ and $I_b = [0, T] \setminus I_a$. Note that I_b is closed. Let N be the cardinal of $\mathcal{T}_{red}$ and denote by $\tau^1_u, \ldots, \tau^N_u$ the elements of $\mathcal{T}_{red}$. By definition of reducible touch points and continuity of the mapping $\mathcal{U} \to W^{2,\infty}$, $u \mapsto g(y_u)$, we may apply Lemma 1.23. Reducing ε if necessary, there exists $\delta > 0$, such that for all $i = 1, \ldots, N$, the mappings

$$\mathcal{R}^i : B_\infty(u, \delta) \to \mathbb{R} \quad ; \quad \tilde{u} \mapsto g(y_{\tilde{u}}(\tau^i_{\tilde{u}})),$$

such that $g(y_{\tilde{u}})$ attains its (unique) maximum over $[\tau^i_u - \varepsilon, \tau^i_u + \varepsilon]$ at time $\tau^i_{\tilde{u}}$, are well-defined. It follows that for all $\tilde{u} \in B_\infty(u, \delta)$,

$$G(\tilde{u}) \in K \quad \text{iff} \quad g(y_{\tilde{u}}(t)) \leq 0 \ \forall t \in I_b \ \text{ and } \ \mathcal{R}^i(\tilde{u}) \leq 0 \ \forall i = 1, \ldots, N. \tag{1.74}$$

Denote by $g(y_{\tilde u})|_b$ the restriction of $g(y_{\tilde u})$ to I_b and $\mathcal{R} : \tilde u \mapsto (\mathcal{R}^i(\tilde u))_{1\le i\le N}$. The *reduced problem* is defined as follows:

$$\min_{\tilde u\in\mathcal{B}_\infty(u,\delta)} J(\tilde u) \quad ; \quad \mathcal{G}(\tilde u) = \begin{pmatrix} g(y_{\tilde u})|_b \\ \mathcal{R}(\tilde u)\end{pmatrix} \in \mathcal{K} := C_-[I_b]\times\mathbb{R}^N_-. \tag{1.75}$$

From (1.74), it follows that (1.75) is locally equivalent to problem (1.4) in a L^∞ neighborhood of u. The Lagrangian $\mathcal{L}$ of the reduced problem (1.75) is given, for $\tilde u\in B_\infty(u,\delta)$ and $\lambda = (\eta_b,\nu)\in\mathcal{M}_+[I_b]\times\mathbb{R}^N_+$, by:

$$\mathcal{L}(\tilde u,\lambda) = J(\tilde u) + \int_{I_b} g(y_{\tilde u}(t))\mathrm{d}\eta_b(t) + \sum_{i=1}^N \nu_i\mathcal{R}^i(\tilde u). \tag{1.76}$$

The next lemma shows how the Lagrangian, multipliers and critical cone of the reduced problem (1.75) are related to the ones of problem (1.4).

Lemma 1.26. *Assume that $g_u^{(1)}\equiv 0$, and let $u\in G^{-1}(K)$ and $\mathcal{T}_{red}$, I_a, I_b, $\mathcal{R}$, $\mathcal{G}$ and $\mathcal{L}$ be defined as above. Let $\lambda=(\eta_b,\nu)\in\mathcal{M}_+[I_b]\times\mathbb{R}^N_+$. For $\delta>0$ small enough, the function $\tilde u\mapsto\mathcal{L}(\tilde u,\lambda)$ is C^1 on $B_\infty(u,\delta)$ and twice Fréchet differentiable at u. Define $\eta\in\mathcal{M}_+[0,T]$ by:*

$$\mathrm{d}\eta(t) = \mathrm{d}\eta_b(t) \text{ on } I_b \qquad ; \qquad \mathrm{d}\eta(t) = \sum_{i=1}^N \nu_i\delta_{\tau_u^i}(t) \text{ on } I_a. \tag{1.77}$$

Then we have: $\mathcal{L}(u,\lambda) = L(u,\eta)$, $D_u\mathcal{L}(u,\lambda) = D_uL(u,\eta)$,

$$\begin{array}{rcl} D\mathcal{G}(u)^{-1}T_{\mathcal{K}}(\mathcal{G}(u)) & = & DG(u)^{-1}T_K(G(u)), \\ \lambda\in N_{\mathcal{K}}(\mathcal{G}(u)) & \text{iff} & \eta\in N_K(G(u)),\end{array} \tag{1.78}$$

$$D^2_{uu}\mathcal{L}(u,\lambda)(v,v) \;=\; D^2_{uu}L(u,\eta)(v,v) - \sum_{i=1}^N \nu_i \frac{(g_y^{(1)}(y_u(\tau_u^i))z_{u,v}(\tau_u^i))^2}{g^{(2)}(u(\tau_u^i),y_u(\tau_u^i))}. \tag{1.79}$$

Proof. Note that $\mathcal{R}^i = \Phi^i\circ G$, $i=1,\dots,N$, where the mappings Φ^i are defined by (1.70) in Lemma 1.23 applied to $(x_0,\tau_0) = (g(y_u),\tau_u^i)$. It follows from Lemma 1.23 that $\mathcal{R}$ is C^1 over a small ball $B_\infty(u,\delta)$. By (1.71), the second-order expansion of the state (1.11) and (1.28) (since $g_u^{(1)}\equiv 0$), that gives $\frac{\mathrm{d}}{\mathrm{d}t}DG(u)v = g_y^{(1)}(y_u)z_{u,v}$, we see that, for all $v\in\mathcal{U}$:

$$D\mathcal{R}^i(u)v = D\Phi^i(G(u))DG(u)v = g_y(y_u(\tau_u^i))z_{u,v}(\tau_u^i), \tag{1.80}$$

$$\begin{array}{rcl} D^2\mathcal{R}^i(u)(v,v) & = & D\Phi^i(G(u))D^2G(u)(v,v) + D^2\Phi^i(G(u))(DG(u)v,DG(u)v) \\ & = & z_{u,v}(\tau_u^i)^* g_{yy}(y_u(\tau_u^i))z_{u,v}(\tau_u^i) + g_y(y_u(\tau_u^i))z_{u,vv}(\tau_u^i) \\ & & - \dfrac{(g_y^{(1)}(y_u(\tau_u^i))z_{u,v}(\tau_u^i))^2}{g^{(2)}(u(\tau_u^i),y_u(\tau_u^i))}.\end{array}$$

The conclusion follows easily from the above expressions (see the proof of Lemma 1.1), (1.78) is obtained as a consequence of (1.80). □

It follows that if $u\in\mathcal{U}$ and $\Lambda(u)\neq\emptyset$, the Lagrange multipliers λ and η associated with u in problems (1.75) and (1.4) respectively, are related by (1.77). By (1.78), it follows also that the critical cone $\mathcal{C}(u)$ for problem (1.75) is equal to $C(u)$. We shall show that the statement of Th. 1.18 remains true by replacing $L(u,\eta)$ by $\mathcal{L}(u,\lambda)$. That is, the main result of this paper, with Th. 1.12 (and Th. 1.18 for first-order state constraint), is the next theorem.

Theorem 1.27. *Assume that $g_u^{(1)} \equiv 0$ (the state constraint is not of first order). Let $u \in \mathcal{U}$ satisfy (1.13) with Lagrange multiplier η, and assume that (A2') holds. Let $\mathcal{T}_{red}$ be a finite set of reducible touch points of u, and $\nu_\tau := [\eta(\tau)]$. If the following second-order sufficient condition is satisfied:*

$$D^2_{uu}L(u,\eta)(v,v) - \sum_{\tau\in\mathcal{T}_{red}} \nu_\tau \frac{(g_y^{(1)}(y_u(\tau))z_{u,v}(\tau))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y_u(t))|_{t=\tau}} > 0 \quad \forall v \in C_{L^2}(u)\setminus\{0\} \tag{1.81}$$

then u is a local solution of (1.4) satisfying the quadratic growth condition (1.5).

Conversely, if (A1)-(A6) hold, then the finitely many essential touch points of the trajectory (u, y_u) are all reducible, and the second-order sufficient condition (1.81) *is satisfied with $\mathcal{T}_{red} = \mathcal{T}_{to}^{ess}$* iff *the quadratic growth condition (1.5) is satisfied.*

Remark 1.28. Note that if $\mathcal{T}_{red} = \emptyset$, (1.81) coincides with (1.49). If $\mathcal{T}_{red}$ contains essential touch points, then by (1.32) the contribution in (1.81) of points in $\mathcal{T}_{red}$ is such that the sum is nonpositive, and therefore the sufficient condition (1.81) is in general weaker than (1.49).

We first need to extend Lemma 1.20 to the Lagrangian $\mathcal{L}$. Note that $\mathcal{L}$ is not C^2 in a L^∞ neighborhood of u, thus (1.56) does not hold with $\mathcal{L}$.

Lemma 1.29. *Assume that $g_u^{(1)} \equiv 0$. For $\delta > 0$ small enough and all $v \in B_\infty(0,\delta)$,*

$$\mathcal{L}(u+v,\lambda) = \mathcal{L}(u,\lambda) + D_u\mathcal{L}(u,\lambda)v + \frac{1}{2}D^2_{uu}\mathcal{L}(u,\lambda)(v,v) + \tilde{r}(v), \tag{1.82}$$

with $\tilde{r}(v) = o_\infty(\|v\|_2^2)$.

Proof. It is easily seen from (1.76) and (1.77) that

$$\mathcal{L}(u+v,\lambda) = L(u+v,\eta) + \sum_{i=1}^N \nu_i(g(y_{u+v}(\tau^i_{u+v})) - g(y_{u+v}(\tau^i_u))).$$

We may write $\tilde{r}(v) = r(v) + \hat{r}(v)$, where $r(v)$ is given by (1.55) and satisfies $r(v) = \mathcal{O}(\|v\|_3^3)$ by Lemma 1.20, and by (1.79) we have $\hat{r}(v) = \sum_{i=1}^N \nu_i \hat{r}_i(v)$ with, for $i = 1,\dots,N$:

$$\hat{r}_i(v) := g(y_{u+v}(\tau^i_{u+v})) - g(y_{u+v}(\tau^i_u)) + \frac{(g_y^{(1)}(y_u(\tau^i_u))z_{u,v}(\tau^i_u))^2}{2g^{(2)}(u(\tau^i_u), y_u(\tau^i_u))}. \tag{1.83}$$

Fix $i = 1,\dots,N$, and set $x_0 := g(y_u)$ and $\tau_0 := \tau^i_u$. By definition of reducible touch points, (x_0,τ_0) satisfies the assumptions of Lemma 1.23. Set $x := g(y_{u+v}) \in W^{2,\infty}$, then $\tau_x = \tau^i_{u+v}$, and since the state constraint is not of first order, we have $\dot{x} = g^{(1)}(y_{u+v})$, $\ddot{x} = g^{(2)}(u+v, y_{u+v})$ and hence, by (1.50):

$$\|x - x_0\|_{1,\infty} = \mathcal{O}_\infty(\|v\|_1) \quad ; \qquad \|\ddot{x} - \ddot{x}_0\|_\infty = \mathcal{O}_\infty(\|v\|_\infty). \tag{1.84}$$

Since

$$g^{(1)}(y_{u+v}) - g^{(1)}(y_u) - g_y^{(1)}(y_u)z_{u,v} = \int_0^1 (g_y^{(1)}(y_{u+\sigma v})z_{u+\sigma v,v} - g_y^{(1)}(y_u)z_{u,v})\mathrm{d}\sigma,$$

we also have by (1.50) and (1.52)-(1.53), setting $h := g_y^{(1)}(y_u)z_{u,v}$, that

$$\|\dot{x} - \dot{x}_0 - h\|_\infty = \mathcal{O}_\infty(\|v\|_2^2). \tag{1.85}$$

We may now write $\hat{r}_i(v) = \hat{r}_{i,1}(v) + \hat{r}_{i,2}(v)$ with:

$$\hat{r}_{i,1}(v) = x(\tau_x) - x(\tau_0) + \frac{\dot{x}(\tau_0)^2}{2\ddot{x}_0(\tau_0)} \quad ; \quad \hat{r}_{i,2}(v) = \frac{h(\tau_0)^2 - \dot{x}(\tau_0)^2}{2\ddot{x}_0(\tau_0)}.$$

By (1.73) and (1.84), we have $\hat{r}_{i,1}(v) = o_\infty(\|v\|_1^2)$. From $|a^2 - b^2| \leq (2|a| + |a-b|)|a-b|$, $\|h\|_\infty = \mathcal{O}_\infty(\|v\|_1)$ by (1.52), (1.85) with $\dot{x}_0(\tau_0) = 0$, and $\|\cdot\|_2^2 \leq \|\cdot\|_1 \|\cdot\|_\infty$, we see that $\hat{r}_{i,2}(v) = \mathcal{O}_\infty(\|v\|_1\|v\|_2^2) \leq \mathcal{O}_\infty(\|v\|_1^2\|v\|_\infty)$. It follows that $\hat{r}_i(v) = o_\infty(\|v\|_1^2)$ for all i and finally that $\tilde{r}(v) = o_\infty(\|v\|_2^2)$, which achieves the proof. □

Proof of Theorem 1.27. Since the sum of a Legendre form and of a weakly continuous quadratic form remains a Legendre form, we deduce easily from (1.79) and Lemma 1.21, since the additional terms

$$v \;\mapsto\; z_{u,v}(\tau_u^i)^* \frac{g_y^{(1)}(y_u(\tau_u^i))^* g_y^{(1)}(y_u(\tau_u^i))}{g^{(2)}(u(\tau_u^i), y_u(\tau_u^i))} z_{u,v}(\tau_u^i)$$

are weakly continuous quadratic forms, that the unique continuous extension of $D_{uu}\mathcal{L}(u,\lambda)$ over L^2 is a Legendre form. In addition, since $\tilde{r}(v) = o_\infty(\|v\|_2^2)$ by Lemma 1.29, the proof of Theorem 1.18 still applies, replacing $L(u,\eta)$ by $\mathcal{L}(u,\lambda)$. It follows that (1.81) implies the quadratic growth condition (1.5).

Conversely, if (A1)-(A6) hold, there are finitely many essential touch points of (u, y_u), all being reducible. Assume that (1.5) holds. Then for sufficiently small $\varepsilon > 0$, u is solution of the following problem:

$$(\mathcal{P}_\varepsilon) \qquad \min_{\tilde{u}\in\mathcal{U}} \{\, J^\varepsilon(\tilde{u}) := J(\tilde{u}) - \tfrac{1}{2}\varepsilon\|\tilde{u} - u\|_2^2 \,\} \quad ; \quad G(\tilde{u}) \in K, \tag{1.86}$$

with the same (unique) Lagrange multiplier η, since $D_u J^\varepsilon(u) = D_u J(u)$. Since in addition $(\mathcal{P}_\varepsilon)$ and (1.4) have the same constraints, they have the same critical cone. Denote the Lagrangian of $(\mathcal{P}_\varepsilon)$ by $L^\varepsilon(u,\eta)$. Note that since only the cost function has been perturbed, Theorem 1.12 and Corollary 1.15 have an immediate extension to the non-autonomous problem $(\mathcal{P}_\varepsilon)$. Therefore, noticing that $D_{uu}^2 L^\varepsilon(u,\eta)(v,v) = D_{uu}^2 L(u,\eta)(v,v) - \varepsilon\|v\|_2^2$, we obtain:

$$D_{uu}^2 L(u,\eta)(v,v) - \sum_{\tau\in\mathcal{T}_{to}^{ess}} \nu_\tau \frac{(g_y^{(1)}(y_u(\tau))z_{u,v}(\tau))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(y_u(t))|_{t=\tau}} \geq \varepsilon\|v\|_2^2, \quad \forall v \in C_{L^2}(u). \tag{1.87}$$

Hence (1.81) is satisfied with $\mathcal{T}_{red} = \mathcal{T}_{to}^{ess}$.

Note that taking $\mathcal{T}_{red} = \emptyset = \mathcal{T}_{to}^{ess}$ proves the converse in Th. 1.18, when (u, y_u) has no essential touch point (including the case $q = 1$). □

Remark 1.30. The second-order sufficient condition in (1.81) remains in quite an abstract form, of little help to check the optimality of a trajectory in application to real life problems. Some *verifiable* second-order sufficient conditions exist in the literature that are based on Riccati equations, see e.g. Maurer [99]. They may be too strong, however, since they ensure in general the coercivity of the Hessian of the Lagrangian over a space that is larger than the critical cone $C_{L^2}(u)$. See also Malanowski et al. [89, 95] for first order state constraints.

Remark 1.31. Handling an infinite number of junction points remains an open problem. It was shown indeed by Robbins in [118], on an example involving a *third order* state constraint, and though satisfying all regularity assumptions (A0)-(A3), that the optimal trajectory has a boundary arc, but except for a nowhere dense subset of initial conditions y_0, the entry point

of the boundary arc is not regular, being the cluster point of an infinite sequence of touch points.

It happens that boundary arcs with regular entry and exit points may occur for any order of the state constraint q, see for instance the example given in [19, Rem. 4.11][2]. However, when q is greater than or equal to three, it seems that boundary arcs with regular entry and exit points occur only in degenerate (i.e., non generic) situations, and that generically, as Robbins' example suggests, the junctions at boundary arcs are irregular with an infinite sequence of touch points.

1.6 Conclusion

Our main result is a no-gap condition for an optimal control problem with a single state constraint of any order and only one control. The main hypotheses are that there are finitely many junction points, the essential touch points being reducible, the entry/exit points being regular, and strict complementarity on boundary arcs. The extension of the result to the case when $g(y_u(T)) = 0$ should present no difficulty.

In our recent work [19], we relate these second-order conditions to the study of the well-posedness of the shooting algorithm, and to the characterization of strong regularity in the sense of Robinson [121] (see also related results [24, Section 5.1] and Malanowski [86]).

We hope in the future to extend some of the results of these papers to the case of several state constraints and control variables.

Acknowlegments The authors thank two anonymous referees for their useful suggestions.

1.7 Appendix

Lemma 1.32 (Extension of Gronwall Lemma). *Let $p \in BV([0,T];\mathbb{R}^n)$ be such that:*

$$|\mathrm{d}p(t)| \leq \kappa |p(t)| \mathrm{d}t + \mathrm{d}\mu(t), \quad \forall t \in [0,T], \tag{1.88}$$

for some positive constant κ, and a nonnegative bounded measure μ. Then:

$$\|p\|_\infty \leq e^{\kappa T}|p(0)| + \int_0^T e^{\kappa(T-t)}\mathrm{d}\mu(t).$$

Proof. Set $\rho(t) = |p(t)|$. Then ρ is a nonnegative bounded measure, and for all $t \in [0,T)$ and $s \to 0^+$, we have:

$$\begin{aligned}\int_t^{t+s} \mathrm{d}\rho(\sigma) &= \rho(t+s) - \rho(t) = |p(t+s)| - |p(t)| \\ &\leq |p(t+s) - p(t)| = |\int_t^{t+s} \mathrm{d}p(\sigma)| \leq \int_t^{t+s} |\mathrm{d}p(\sigma)|.\end{aligned}$$

From (1.88) it follows that $\rho(t) \leq \varphi(t)$ for all $t \in [0,T]$, where φ is solution of

$$\varphi(t) = |p(0)| + \kappa \int_0^t \varphi(s)\mathrm{d}s + \int_0^t \mathrm{d}\mu(s), \quad \text{for all } \ t \in [0,T].$$

[2] Remark 2.42 of this thesis

Then

$$\mathrm{d}(e^{-\kappa t}\varphi(t)) = e^{-\kappa t}\mathrm{d}\varphi(t) - \kappa e^{-\kappa t}\varphi(t)\mathrm{d}t = e^{-\kappa t}\mathrm{d}\mu(t).$$

Therefore, $e^{-\kappa t}\rho(t) \leq |p(0)| + \int_0^t e^{-\kappa s}\mathrm{d}\mu(s)$. The result follows. □

Lemma 1.33 (Integration by parts). *The following relation holds, for any $p \in BV([0,T],\mathbb{R}^{n*})$ and $z \in BV(0,T;\mathbb{R}^n) \cap C([0,T];\mathbb{R}^n)$:*

$$\int_0^T \mathrm{d}p(t)z(t) = -\int_0^T p(t)\mathrm{d}z(t) + p(T)z(T) - p(0)z(0). \tag{1.89}$$

Proof. See e.g. [58, p.154]. □

Chapitre 2

Application à l'étude de l'algorithme de tir*

Abstract This paper deals with the shooting algorithm for optimal control problems with a scalar control and a regular scalar state constraint. Additional conditions are displayed, under which the so-called alternative formulation is equivalent to Pontryagin's minimum principle. The shooting algorithm appears to be well-posed (invertible Jacobian), iff (i) the no-gap second order sufficient optimality condition holds, and (ii) when the constraint is of order $q \geq 3$, there is no boundary arc. Stability and sensitivity results without strict complementarity at touch points are derived using Robinson's strong regularity theory, under a minimal second-order sufficient condition. The directional derivatives of the control and state are obtained as solutions of a linear quadratic problem.

Résumé Dans cet article, on étudie l'algorithme de tir pour les problèmes de commande optimale avec contraintes sur l'état. On donne les conditions supplémentaires nécessaires, sous lesquelles la formulation alternative est équivalente au Principe de Pontryaguine. On montre que l'algorithme de tir est bien posé, ssi (i) une condition suffisante minimale du second ordre est satisfaite, et (ii) lorsque la contrainte est d'ordre $q \geq 3$, il n'y a pas d'arc frontière. Enfin, une analyse de stabilité et de sensibilité est effectuée, sans hypothèse de complémentarité stricte aux points de contacts isolés. On utilise pour ceci la théorie de la forte régularité de Robinson, dont on donne une caractérisation par une condition suffisante du second ordre. Les dérivées directionnelles sont obtenues comme solution d'un problème linéaire quadratique.

2.1 Introduction

For optimal control problems satisfying the strengthened Legendre-Clebsch condition, Pontryagin's principle allows us to express the control as a function of the state and the costate. For unconstrained problems, the resulting two-points boundary value problem reduces to a finite-dimensional "shooting" equation whose unknown is the initial costate (see e.g. [125]). The extension to control constrained problems is relatively easy, assuming nontangentiality conditions when a constraint becomes active or inactive. This approach allows us to compute accurate solutions at low cost, once the *structure* of active constraints is known, and reasonable

*Joint work with J.F. Bonnans. Published in SIAM Journal on Control and Optimization, 46(4) :1398–1430 (2007), under the title *Well-posedness of the shooting algorithm for state constrained optimal control problems with a single constraint and control*.

initial values of unknowns can be guessed. For state constrained optimal control problems, a reformulation of the optimality conditions is needed, and the shooting equations take into account only some of the optimality conditions. Therefore, checking that the shooting equations are well-posed under minimal hypotheses becomes challenging.

An alternative formulation, suitable for the shooting algorithm in the presence of state constraints, was first introduced by Bryson, Denham and Dreyfus [29], (see also [28]), in an heuristic manner. Some additional conditions (necessary for optimality) were missing, as shown in Jacobson, Lele and Speyer [75], where the first results on the regularity of the multiplier and on junction conditions are stated. A significant clarification of their work can be found in the unpublished paper by Maurer [98], where the link between the results of [75] and the alternative formulation of [29, 28] is established. Numerous different versions of Pontryagin's principle with state constraints were given in the literature; see the survey by Hartl, Sethi and Vickson [68].

Stability results for *first*-order state constraints and directional differentiability of solutions in L^2 were first obtained by Malanowski [88] using an infinite-dimensional implicit function theorem and differentiation of the projection on a convex set [67]. The (strong) second-order sufficient condition used in the analysis was later weakened by Malanowski [89], taking into account the strictly active constraints. These results require *no assumptions on the structure* of the trajectory. However, no extensions of this method for higher-order state constraints are known. Dontchev and Hager [53] derived, still for first-order constraints, L^∞ stability results under an additional assumption on the structure of the contact set. Malanowski and Maurer obtain sensitivity results in [93] (first-order) and [94] (higher order), when there are finitely many nontangential junction points and strict complementarity holds, by application of the implicit function theorem to the shooting mapping. They obtain derivatives as the solution of an equality constrained linear quadratic problem, but when the order of the constraint is $q \geq 2$, the data of the latter depend on the (precomputed) variation of entry times. Numerical applications of the shooting algorithm to state constrained problems in the aerospace field are presented e.g. in [30, 11] and in [115], where the role of additional conditions appears crucial to eliminate nonoptimal solutions; numerical examples of sensitivity analysis are given in [4]. Discretization errors are studied in e.g. [54].

This paper handles the case of a scalar control and a regular scalar state constraint, for which regularity and junction conditions results are known. We assume that the Hamiltonian is uniformly strongly convex w.r.t. the control variable, that there are finitely many nontangential junction times, and that strict complementarity on boundary arcs holds.

We express the additional conditions under which the alternative formulation is equivalent to Pontryagin's principle. When strict complementarity holds at touch points as well, we prove that the shooting algorithm is well-posed (invertible Jacobian) iff (i) the no-gap second-order sufficient condition in [21] holds, and (ii) when the constraint is of order $q \geq 3$, there is no boundary arc. Then stability and sensitivity results, removing the strict complementarity hypothesis at touch points, are derived, applying Robinson's strong regularity theory [121] to the shooting mapping. We give a necessary and sufficient second-order condition characterizing the strong regularity property. The directional derivatives of the control and state are obtained as solutions of an inequality constrained linear quadratic problem, independent of the variations of junction times.

The paper is organized as follow. In section 2.2, we give the characterization of Pontryagin extremals as solutions of the shooting equations under some minimal additional conditions. Then, in section 2.3, we give the characterization of the well-posedness of the shooting algorithm and its relation to the no-gap second-order optimality conditions obtained in [21, 18].

Finally, in section 2.4, we give stability and sensitivity analysis results.

The results of sections 2.2 and 2.3 of this paper are extended to the case of vector-valued state constraints and control in the report [17]. The main difficulty is the extension of the junction conditions result of Jacobson, Lele and Speyer [75] (Prop. 2.5 below). The latter plays a crucial role in the proof of the *necessity* of the condition claimed in this paper as necessary and sufficient for the well-posedness of the shooting algorithm (see Th. 2.23).

2.2 Junction Conditions

The section is organized as follows. After introducing notation, definitions, assumptions, and basic results needed in the paper, we recall in subsection 2.2.1 an alternative formulation for optimality conditions (Def. 2.7), which is useful for the shooting algorithm. This is one of the various formulations existing in the literature (see e.g. the survey [68]). Therefore, one of the main concerns of this paper is to investigate, in subsection 2.2.2, the equivalence with Pontryagin's minimum principle (Prop. 2.10). Finally, in subsection 2.2.3 we formulate the shooting algorithm and show that some of the additional conditions are automatically satisfied by a solution of the shooting equations (Prop. 2.15).

Denote by $L^\infty(0,T)$ the Banach space of measurable and essentially bounded functions and by $W^{1,\infty}(0,T)$ the Sobolev space of functions having a weak derivative in $L^\infty(0,T)$. Let the control and state spaces be respectively $\mathcal{U} := L^\infty(0,T)$ and $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$. We consider the following optimal control problem with a scalar state constraint and a scalar control:

$$(\mathcal{P}) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell(u(t),y(t))\mathrm{d}t + \phi(y(T)) \tag{2.1}$$

$$\text{subject to} \qquad \dot{y}(t) \;=\; f(u(t),y(t)) \quad \text{a.e. } t\in[0,T] \quad ; \quad y(0) = y_0 \tag{2.2}$$

$$g(y(t)) \;\le\; 0 \qquad \forall\, t\in[0,T]. \tag{2.3}$$

The data of the problem are the distributed cost $\ell : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$, final cost $\phi : \mathbb{R}^n \to \mathbb{R}$, dynamics $f : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^n$, state constraint $g : \mathbb{R}^n \to \mathbb{R}$, final time $T > 0$, and initial condition $y_0 \in \mathbb{R}^n$.

We assume throughout the paper that the following hold:

(A0) The mappings ℓ, ϕ, f and g are k-times continuously differentiable (C^k) with $k \ge 2$, and have locally Lipschitz continuous second-order derivatives when $k = 2$. The dynamics f is Lipschitz continuous.

(A1) The initial condition satisfies $g(y_0) < 0$.

The space of row vectors is denoted by $\mathbb{R}^{n*}$. The space of continuous functions over $[0,T]$ is denoted by $C[0,T]$. The dual space of Radon measures, denoted by $\mathcal{M}[0,T]$, is identified with the space of functions of bounded variation $BV(0,T)$ vanishing at zero. The transposition operator in $\mathbb{R}^n$ is denoted by a star *. Fréchet derivatives of f, ℓ, etc., w.r.t. arguments $u \in \mathbb{R}$, $y \in \mathbb{R}^n$, are denoted by a subscript, for instance $f_u(u,y) = D_u f(u,y)$, $f_{uu}(u,y) = D^2_{uu} f(u,y)$. One exception to this rule, which should not be a source of confusion, is that we denote by y_u the (unique) solution in $\mathcal{W}$ of the state equation (2.2) associated with the control $u \in \mathcal{U}$. Total derivation w.r.t. time is denoted by a dot, i.e. $\dot{y}(t) = \frac{\mathrm{d}y(t)}{\mathrm{d}t}$.

A *trajectory* is an element (u,y) of $\mathcal{U}\times\mathcal{Y}$ satisfying the state equation (2.2). A trajectory (u,y) is said to be *feasible* if it satisfies the state constraint (2.3). Define the classical (resp.

generalized) *Hamiltonian* functions of $(\mathcal{P})$, $H : \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^{n*} \to \mathbb{R}$ (resp. $\mathcal{H} : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^{n*} \to \mathbb{R}$) by

$$H(u,y,p) := \ell(u,y) + pf(u,y)\ ; \quad \mathcal{H}(p_0,u,y,p) := p_0\ell(u,y) + pf(u,y). \tag{2.4}$$

First-order necessary optimality conditions for $(\mathcal{P})$ are given by *Pontryagin's minimum principle*.

Definition 2.1. A trajectory (u,y) is a Pontryagin extremal if there exists $p_0 \in \mathbb{R}^+$, $p \in BV([0,T];\mathbb{R}^{n*})$, and $\eta \in \mathcal{M}[0,T]$, with $(p_0,\mathrm{d}\eta) \neq 0$, such that

$$\begin{aligned}
\dot{y}(t) &= \mathcal{H}_p(p_0,u(t),y(t),p(t)) \quad \text{a.e. } t\in[0,T] \quad ; \quad y(0)=y_0 && (2.5)\\
-\mathrm{d}p(t) &= \mathcal{H}_y(p_0,u(t),y(t),p(t))\mathrm{d}t + g_y(y(t))\mathrm{d}\eta(t) \quad \text{in } \mathcal{M}([0,T];\mathbb{R}^{n*}) && (2.6)\\
p(T) &= p_0\phi_y(y(T)) && (2.7)\\
u(t) &\in \operatorname{argmin}_{w\in\mathbb{R}}\mathcal{H}(p_0,w,y(t),p(t)) \quad \text{a.e. } t\in[0,T] && (2.8)\\
g(y(t)) &\leq 0,\ \forall t\in[0,T]\ ; \quad \mathrm{d}\eta \geq 0\ ; \quad \int_0^T g(y(t))\mathrm{d}\eta(t) = 0. && (2.9)
\end{aligned}$$

By $\mathrm{d}\eta \geq 0$, we mean that $\int_0^T \varphi(t)\mathrm{d}\eta(t) \geq 0$ for all nonnegative continuous functions $\varphi \in C[0,T]$, or equivalently, that η is nondecreasing. The costate equation (2.6) with final condition (2.7) are equivalent to

$$p(t) = \int_t^T \mathcal{H}_y(p_0,u(s),y(s),p(s))\mathrm{d}s + \int_t^T g_y(y(s))\mathrm{d}\eta(s) + p_0\phi_y(y(T)).$$

The next theorem is well known (see [39, 62] for nondifferentiable versions).

Theorem 2.2. *A trajectory (u,y) solution of $(\mathcal{P})$ is a Pontryagin extremal.*

A trajectory $(\bar{u},\bar{y})$ is a *local solution* of $(\mathcal{P})$ if it minimizes (2.1) subject to (2.2)-(2.3) and $\|u-\bar{u}\|_\infty \leq \rho$ for some $\rho > 0$. We say that $(u,y) \in \mathcal{U}\times\mathcal{Y}$ is a *stationary point* of $(\mathcal{P})$ if there exists a nonzero $(p_0,p,\eta) \in \mathbb{R}^+ \times BV(0,T;\mathbb{R}^{n*}) \times \mathcal{M}(0,T)$ such that (2.5)-(2.7), (2.9) are satisfied and

$$\mathcal{H}_u(p_0,u(t),y(t),p(t)) = 0 \qquad \text{for a.a. } t\in[0,T].$$

It is well known that a local solution of $(\mathcal{P})$ is a stationary point. Obviously a Pontryagin extremal is a stationary point, but the converse is in general false. An exception is when the (generalized) Hamiltonian is convex with respect to the control variable along the trajectory (see also our assumption (A2) below). Whenever this holds, definitions of both Pontryagin extremals and stationary points are equivalent.

Definitions A *boundary* (resp. *interior*) *arc* is a maximal interval of positive measure $\mathcal{I} \subset [0,T]$ such that $g(y(t)) = 0$ (resp. $g(y(t)) < 0$) for all $t \in \mathcal{I}$. If $[\tau_{en},\tau_{ex}]$ is a boundary arc, τ_{en} and τ_{ex} are called an *entry* and an *exit* point, respectively. Entry and exit points are said to be *regular* if they are endpoint of an interior arc. A *touch* point τ in $(0,T)$ is an isolated contact point (endpoint of two interior arcs). Entry, exit and touch points are called *junction points* (or *times*). We say that the junctions are regular when the entry/exit points are regular.

The first-order time derivative of the state constraint along a trajectory (u,y), defined by $g^{(1)}(u,y) = \frac{\mathrm{d}}{\mathrm{d}t}g(y(t)) = g_y(y)f(u,y)$, is denoted by $g^{(1)}(y)$ if the function $\mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$, $(u,y) \mapsto$

$g_y(y)f(u,y)$ does not depend on u (that is, the function $(u,y)\mapsto g_u^{(1)}(u,y)$ is identically zero). If f and g are C^q, we may define similarly $g^{(2)},\dots,g^{(q)}$ if $g_u^{(j)}\equiv 0$, for all $j=1,\dots,q-1$, and we have $g^{(j)}(u,y)=g_y^{(j-1)}(y)f(u,y)$, for $j=1,\dots,q$.

Let $q\geq 1$ be the smallest number of time derivations of the state constraint, so that a dependence w.r.t. u appears, i.e. $g_u^{(q)}\not\equiv 0$. If q is finite, we say that q is the *order* of the state constraint (see e.g. [29]). A state constraint of order q is said to be *regular* along the trajectory (u,y) if the condition below holds:

$$\exists\,\gamma>0,\quad |g_u^{(q)}(\hat u,y(t))|\geq\gamma\quad\text{ for all } t\in[0,T] \text{ and all } \hat u\in\mathbb{R}. \tag{2.10}$$

Note that the set of generalized multipliers (p_0,p,η) is a cone. When $p_0=0$, we say that the multiplier is singular; otherwise it is regular. Dividing then (p,η) by p_0, we obtain the qualified version of Pontryagin's principle, substituting the generalized Hamiltonian with the classical Hamiltonian. It is easily seen that a Pontryagin extremal satisfying (2.10) (and (A1)) has no singular multiplier, and that the multiplier (p,η) in the qualified version of Pontryagin's principle ($p_0=1$) is unique. The same is true for a stationary solution.

Being of bounded variation, p has at most countably many discontinuity times and has everywhere on $[0,T]$ left and right limits, denoted by $p(t^\pm)=\lim_{t'\to t^\pm}p(t')$. The jump at $\tau\in(0,T)$ is denoted by $[p(\tau)]=p(\tau^+)-p(\tau^-)$. Similar observations hold for η.

Assumptions We say that (u,y) is a *regular Pontryagin extremal* if it satisfies Def. 2.1 with $p_0=1$, with costate p and multiplier η, and if assumptions (A2)-(A4) below are satisfied.

(A2) The Hamiltonian is strongly convex w.r.t. the control variable, uniformly w.r.t. $t\in[0,T]$:

$$\exists\,\alpha>0,\quad H_{uu}(\hat u,y(t),p(t^\pm))\geq\alpha\quad\text{for all } t\in[0,T] \text{ and all } \hat u\in\mathbb{R}. \tag{2.11}$$

(A3) The data of the problem are C^{2q}, i.e. $k\geq 2q$ in (A0), and the state constraint is of order q and regular, i.e. (2.10) holds.

(A4) The trajectory (u,y) has a *finite set of junction times*, that will be denoted by $\mathcal{T}=:\mathcal{T}_{en}\cup\mathcal{T}_{ex}\cup\mathcal{T}_{to}$, with $\mathcal{T}_{en}$, $\mathcal{T}_{ex}$, and $\mathcal{T}_{to}$ the *disjoint* (and possibly empty) subsets of respectively entry, exit and touch points, and we assume that $g(y(T))<0$.

Hypothesis (A4) implies that all entry and exit points are regular. In what follows, we denote by $\mathcal{I}_b$ the union of boundary arcs, i.e. $\mathcal{I}_b:=\cup_{i=1}^{N_b}[\tau_{en}^i,\tau_{ex}^i]$ for $\mathcal{T}_{en}:=\{\tau_{en}^1<\dots<\tau_{en}^{N_b}\}$ and $\mathcal{T}_{ex}:=\{\tau_{ex}^1<\dots<\tau_{ex}^{N_b}\}$.

Remark 2.3. Troughout the paper, (A3) can be weakened, replacing (2.10) by

$$\exists\,\gamma,\varepsilon>0,\quad |g_u^{(q)}(\hat u,y(t))|\geq\gamma\quad\text{for all } t,\ \operatorname{dist}(t,\mathcal{I}_b\cup\mathcal{T}_{to})<\varepsilon,\ \text{and all } \hat u\in\mathbb{R}. \tag{2.12}$$

Notation Given a finite subset $\mathcal{S}$ of $(0,T)$, we denote by $PC^k_{\mathcal{S}}[0,T]$ the set of functions over $[0,T]$ that are of class C^k outside $\mathcal{S}$ (PC stands for piecewise continuous), and have, as well as their first k derivatives, a left and right limit over $\mathcal{S}$ and a right (resp. left) limit at 0 (resp. T).

Let φ be a real-valued function over $[0,T]$. Assuming w.l.o.g. the elements of $\mathcal{S}$ in increasing order, we may define $\varphi(\mathcal{S}):=(\varphi(\tau))_{\tau\in\mathcal{S}}\in\mathbb{R}^{\operatorname{Card}\mathcal{S}}$. We adopt a similar convention for

vectors, $\nu_{\mathcal{S}} := (\nu_\tau)_{\tau\in\mathcal{S}} \in \mathbb{R}^{\mathrm{Card}\,\mathcal{S}}$, and will also use the following notation:

$$\nu_{\mathcal{S}}^{1:q} := \begin{pmatrix} \nu_{\mathcal{S}}^1 \\ \vdots \\ \nu_{\mathcal{S}}^q \end{pmatrix} \in \mathbb{R}^{q\,\mathrm{Card}\,\mathcal{S}} \;;\; g^{(0:q-1)}(y(\mathcal{S})) := \begin{pmatrix} g(y(\mathcal{S})) \\ \vdots \\ g^{(q-1)}(y(\mathcal{S})) \end{pmatrix} \in \mathbb{R}^{q\,\mathrm{Card}\,\mathcal{S}}.$$

2.2.1 Alternative Formulation of Optimality Conditions

Under assumption (A4) we have a finite number of arcs and we can show, with regularity assumptions (A2)-(A3), that the multiplier η is differentiable on the interior of each arc [75, 98]. An analysis of the optimality system on interiors of arcs shows then that a regular Pontryagin extremal satisfies the conditions stated in Prop. 2.4 below. An analysis at junction times leads afterwards to the junction conditions given in Prop. 2.5.

Proposition 2.4. *Let (u, y) be a regular Pontryagin extremal, satisfying (A2)-(A4). Then we have $u \in PC^q_{\mathcal{T}}[0,T]$, $y \in PC^{q+1}_{\mathcal{T}}([0,T];\mathbb{R}^n)$ and there exists $p \in PC^1_{\mathcal{T}}([0,T];\mathbb{R}^{n*})$, $\eta_0 \in PC^0_{\mathcal{T}}[0,T]$, and jump parameters $\nu_{\mathcal{T}}$, such that the following optimality system is satisfied:*

$$\begin{aligned}
\dot y(t) &= H_p(u(t),y(t),p(t)) = f(u(t),y(t)) \quad \text{on } [0,T]\;; \qquad y(0) = y_0 && (2.13)\\
-\dot p(t) &= H_y(u(t),y(t),p(t)) + g_y(y(t))\eta_0(t) \quad \text{on } [0,T]\setminus\mathcal{T} && (2.14)\\
p(T) &= \phi_y(y(T)) && (2.15)\\
0 &= H_u(u(t),y(t),p(t)) \quad \text{on } [0,T]\setminus\mathcal{T} && (2.16)\\
g(y(t)) &= 0 \quad \text{on } \mathcal{I}_b \qquad ; \quad \eta_0(t) = 0 \quad \text{on } [0,T]\setminus\mathcal{I}_b && (2.17)\\
g(y(t)) &< 0 \quad \text{on } [0,T]\setminus(\mathcal{I}_b\cup\mathcal{T}_{to}) \quad ; \quad \eta_0(t) \ge 0 \quad \text{on } \operatorname{int}\mathcal{I}_b && (2.18)\\
g(y(\tau)) &= 0 \qquad \forall\tau\in\mathcal{T}_{to} && (2.19)\\
[p(\tau)] &= -\nu_\tau g_y(y(\tau))\;;\; \nu_\tau \ge 0 \qquad \forall\tau\in\mathcal{T}. && (2.20)
\end{aligned}$$

We denote by $\operatorname{int}\mathcal{I}_b$ the interior of $\mathcal{I}_b$. A touch point $\tau \in \mathcal{T}_{to}$ is said to be *essential* if $\nu_\tau > 0$ in (2.20); otherwise it is nonessential. We denote by $\mathcal{T}_{to}^{ess}$ the set of essential touch points. Hypotheses (A2)-(A4) also imply the continuity of the control variable and of some of its time derivatives at junction points. The next proposition is due to Jacobson et al. [75].

Proposition 2.5. *Let (u, y) be a regular Pontryagin extremal, satisfying (A2)-(A4). Then:*
(i) *For all entry or exit point $\tau \in \mathcal{T}_{en}\cup\mathcal{T}_{ex}$:* (a) *if q is* odd*, u and its $q-1$ first derivatives are continuous at τ, $\nu_\tau = 0$ and p is continuous at τ;* (b) *if q is* even*, u and its $q-2$ first derivatives are continuous at τ.*
(ii) *For all touch points $\tau \in \mathcal{T}_{to}$:* (a) *$u$ and its $q-2$ first derivatives are continuous at τ;* (b) *if τ is nonessential (i.e. $\nu_\tau = 0$), u and its q first derivatives and p are continuous at τ;* (c) *if $q = 1$, then τ is a nonessential touch point.*

Remark 2.6. If (u, y) satisfies (A2)-(A4) and (2.13)-(2.20), the multiplier $\eta \in \mathcal{M}[0,T]$ such that (u, y) satisfies Definition 2.1 is given by:

$$\mathrm{d}\eta(t) = \sum_{\tau\in\mathcal{T}} \nu_\tau\delta_\tau(t) + \eta_0(t)\mathrm{d}t, \tag{2.21}$$

where δ_τ denotes the Dirac measure at time τ, $\nu_\tau = [\eta(\tau)]$ is the nonnegative jump at $\tau \in \mathcal{T}$, and the density $\eta_0 \in PC^0_{\mathcal{T}}[0,T]$ equals $\frac{\mathrm{d}\eta}{\mathrm{d}t}$ on $[0,T]\setminus\mathcal{T}$.

We now present the alternative formulation that will be used in the shooting algorithm. First introduced heuristically in [29], it is based on the use of the mixed explicit constraint $g^{(q)}(u(t), y(t)) = 0$ on boundary arcs. Let the *augmented Hamiltonian* $\tilde{H} : \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^{n*}\times\mathbb{R} \to \mathbb{R}$ be defined by

$$\tilde{H}(u, y, p_q, \eta_q) = H(u, y, p_q) + \eta_q g^{(q)}(u, y), \tag{2.22}$$

where q denotes the order of the state constraint and H is the classical Hamiltonian (2.4).

Definition 2.7. We say that a trajectory (u, y) in $PC^q_{\mathcal{T}}[0,T] \times PC^{q+1}_{\mathcal{T}}([0,T];\mathbb{R}^n)$ satisfying (A3)-(A4) is solution of the alternative formulation, if there exist $p_q \in PC^{q+1}_{\mathcal{T}}([0,T];\mathbb{R}^{n*})$, $\eta_q \in PC^q_{\mathcal{T}}[0,T]$, alternative jump parameters $\nu^j_{\mathcal{T}_{en}}$, $j = 1,\dots,q$, and $\nu_{\mathcal{T}_{to}}$ such that the following relations are satisfied (we omit dependence in time):

$$\begin{aligned}
\dot{y} &= \tilde{H}_p(u, y, p_q, \eta_q) = f(u, y) \quad \text{on } [0,T] \quad ; \quad y(0) = y_0 && (2.23)\\
-\dot{p}_q &= \tilde{H}_y(u, y, p_q, \eta_q) = H_y(u, y, p_q) + \eta_q g^{(q)}_y(u, y) \ \text{ on } [0,T]\setminus\mathcal{T} && (2.24)\\
p_q(T) &= \phi_y(y(T)) && (2.25)\\
0 &= \tilde{H}_u(u, y, p_q, \eta_q) = H_u(u, y, p_q) + \eta_q g^{(q)}_u(u, y) \ \text{ on } [0,T]\setminus\mathcal{T} && (2.26)\\
g^{(j)}(y(\tau)) &= 0 \quad \text{for } j = 0, 1, \dots, q-1 \ ; \quad \tau \in \mathcal{T}_{en} && (2.27)\\
g^{(q)}(u, y) &= 0 \quad \text{on } \mathcal{I}_b && (2.28)\\
g(y(\tau)) &= 0 \quad \text{for all } \tau \in \mathcal{T}_{to} && (2.29)\\
\eta_q(t) &= 0 \quad \text{on } [0,T]\setminus\mathcal{I}_b && (2.30)\\
[p_q(\tau)] &= -\sum_{j=1}^{q} \nu^j_\tau g^{(j-1)}_y(y(\tau)) \qquad \text{for all } \tau \in \mathcal{T}_{en} && (2.31)\\
[p_q(\tau)] &= 0 \quad \text{for all } \tau \in \mathcal{T}_{ex} && (2.32)\\
[p_q(\tau)] &= -\nu_\tau g_y(y(\tau)) \qquad \text{for all } \tau \in \mathcal{T}_{to}. && (2.33)
\end{aligned}$$

In the heuristic formulation of [29], (2.23)-(2.33) are interpreted as necessary optimality conditions for the problem of minimizing (2.1) subject to (2.2) and *equality constraints* (2.27)-(2.29) for a *fixed* set of junction times $\mathcal{T}$. Alternative jump parameters $\nu^{1:q}_{\tau_{en}}$ appearing in (2.31) are seen as multipliers associated with the q interior point constraints in (2.27) at a regular entry time τ_{en}.

The assumption equivalent to (A2) for the alternative formulation, is the following, see Remark 2.11(ii):

(A2$_q$) $\exists\, \alpha > 0,\ \tilde{H}_{uu}(\hat{u}, y(t), p_q(t^\pm), \eta_q(t^\pm)) \geq \alpha$ for all $t \in [0,T]$ and all $\hat{u} \in \mathbb{R}$.

We will write in what follows (A2)-(A4) (resp. (A2$_q$)-(A4)) to denote the assumptions (A2) (resp. (A2$_q$)), (A3) and (A4).

2.2.2 Additional Conditions

Relations (2.23)-(2.33) due to [29] are necessary, but not sufficient, conditions for regular Pontryagin extremals. This was underlined in [75], where some additional necessary conditions were provided, that allowed the authors to show that a trajectory (with a fourth-order state constraint) was not a Pontryagin extremal. We state in Prop. 2.10 the characterization of regular Pontryagin extremals based on the alternative formulation. We need some preliminary lemmas.

Lemma 2.8. *Let (u,y) be a trajectory, and let $(p_q,\eta_q)\in PC^1_{\mathcal{T}}([0,T];\mathbb{R}^{n*})\times PC^0_{\mathcal{T}}[0,T]$ satisfying (A2$_q$)-(A4) and (2.23)-(2.24), (2.26), (2.28). Then (u,y,p_q,η_q) belongs to the set $PC^q_{\mathcal{T}}[0,T]\times PC^{q+1}_{\mathcal{T}}([0,T];\mathbb{R}^n)\times PC^{q+1}_{\mathcal{T}}([0,T];\mathbb{R}^{n*})\times PC^q_{\mathcal{T}}[0,T]$.*

Proof. By the implicit function theorem, applied to (2.26) on interior arcs, and to (2.26) and (2.28) on boundary arcs, the algebraic variables (u,η_q) can be expressed, on the interior of each arc, as C^q functions of (y,p_q). The result follows. □

Lemma 2.9. *If constraint regularity (A3) holds along a trajectory (u,y), and if $u\in PC^q_{\mathcal{T}}[0,T]$, then, for all $t\in[0,T]$, vectors $(g_y(y(t)),\dots,g_y^{(q-1)}(y(t)))$ are linearly independent (and hence, $q\le n$).*

Proof. Since $u\in PC^q_{\mathcal{T}}[0,T]$, the mappings $(A_l)_{0\le l\le q}:[0,T]\setminus\mathcal{T}\to\mathbb{R}^n$ defined inductively by

$$\left\{\begin{array}{l} A_0(t):=f_u(u(t),y(t)),\\ A_l(t):=f_y(u(t),y(t))A_{l-1}(t)-\dot A_{l-1}(t),\qquad l=1,\dots,q,\end{array}\right. \tag{2.34}$$

are well-defined, and $A_l\in PC^{q-l}_{\mathcal{T}}([0,T];\mathbb{R}^n)$ for $l=0,\dots,q$. It has been shown in [98] that the following relations hold for all $t\in[0,T]$:

$$\left\{\begin{array}{l} g_y^{(j)}(y(t))A_l(t^\pm)=0\qquad\text{for } j=0,\dots,q-2,\quad l=0,\dots,q-2-j,\\ g_u^{(q)}(u(t^\pm),y(t))=g_y^{(q-l-1)}(y(t))A_l(t^\pm)\qquad\text{for } l=0,\dots,q-1,\end{array}\right. \tag{2.35}$$

where $t^\pm$ denotes, on both sides of the equality, either t^- or t^+. Denote by C the $n\times q$ matrix $(g_y(y(t))^*,\dots,g_y^{(q-1)}(y(t))^*)$. The above relations imply that the $q\times q$ matrix $D:=C^\top(A_{q-1}(t^\pm),\dots,A_0(t^\pm))$ is lower triangular with nonzero diagonal elements equal to $g_u^{(q)}(u(t^\pm),y(t))$, hence has rank q. Therefore C has rank at least q. The conclusion follows. □

Proposition 2.10. *Let (u,y) be a trajectory satisfying (A2$_q$)-(A4) and the alternative formulation (2.23)-(2.33). Define the functions η_j, $0\le j\le q-1$, the costate p and the jump parameters $\nu_{\mathcal{T}_{en}}$ and $\nu_{\mathcal{T}_{ex}}$ by*

$$\eta_j(t) = (-1)^{q-j}\frac{\mathrm{d}^{q-j}}{\mathrm{d}t^{q-j}}\eta_q(t)\qquad \textit{for } j=0,\dots,q-1,\quad t\in[0,T]\setminus\mathcal{T}, \tag{2.36}$$

$$p(t) = p_q(t)+\sum_{j=1}^{q}\eta_j(t)g_y^{(j-1)}(y(t))\qquad t\in[0,T]\setminus\mathcal{T}, \tag{2.37}$$

$$\nu_{\tau_{en}} = \nu^1_{\tau_{en}}-\eta_1(\tau_{en}^+),\quad\forall\,\tau_{en}\in\mathcal{T}_{en};\qquad \nu_{\tau_{ex}}=\eta_1(\tau_{ex}^-),\quad\forall\,\tau_{ex}\in\mathcal{T}_{ex}. \tag{2.38}$$

Then (u,y) is a regular Pontryagin extremal that satisfies (2.13)-(2.20) iff all the following additional conditions are satisfied:

$$g(y(t)) < 0\qquad \textit{on } [0,T]\setminus(\mathcal{I}_b\cup\mathcal{T}_{to}) \tag{2.39}$$

$$\eta_0(t)=(-1)^q\frac{\mathrm{d}^q}{\mathrm{d}t^q}\eta_q(t) \ge 0\qquad \textit{on } \operatorname{int}\mathcal{I}_b \tag{2.40}$$

At all entry time τ_{en}:

$$\left\{\begin{array}{ll}\nu^1_{\tau_{en}}=\eta_1(\tau_{en}^+) & \textit{if } q \textit{ is odd};\\ \nu^1_{\tau_{en}}\ge\eta_1(\tau_{en}^+) & \textit{if } q \textit{ is even};\end{array}\right.\qquad \nu^j_{\tau_{en}}=\eta_j(\tau_{en}^+);\quad j=2,\dots,q \tag{2.41}$$

At all exit time τ_{ex}:

$$\begin{cases} \eta_1(\tau_{ex}^-) = 0 & \text{if } q \text{ is odd;} \\ \eta_1(\tau_{ex}^-) \geq 0 & \text{if } q \text{ is even;} \end{cases} \qquad \eta_j(\tau_{ex}^-) = 0; \quad j = 2, \ldots, q. \tag{2.42}$$

At all touch times τ_{to}:

$$\nu_{\tau_{to}} \geq 0. \tag{2.43}$$

Remark 2.11. (i) If (u, y) is a regular Pontryagin extremal solution of (2.13)-(2.20), the functions η_j, $1 \leq j \leq q$, costate p_q and alternative jump parameters $\nu_{\mathcal{T}_{en}}^{1:q}$ such that (u, y) satisfies the alternative formulation (2.23)-(2.33) and additional conditions (2.39)-(2.43), can be recovered from p, η_0 and $\nu_{\mathcal{T}}$ as follows. The functions η_j are given by (2.36) by successive integrations of η_0 over boundary arcs, with integration constants determined by the exit time conditions (2.38) for $j = 1$ and (2.42) for $j = 2, \ldots, q$. Costate p_q follows then from (2.37), and jump parameters at entry times $\nu_{\tau_{en}}^j$ are given by (2.38) for $j = 1$ and (2.41) for $j = 2, \ldots, q$. Jump parameters $\nu_{\mathcal{T}_{to}}$ associated with touch points are the same in both formulations.
(ii) Assumptions (A2) and (A2$_q$) are equivalent, when (2.36)-(2.37) hold, since the constraint are of order q, and hence we have

$$\begin{aligned} \tilde{H}_{uu}(u, y, p_q, \eta_q) &= H_{uu}(u, y, p) - \textstyle\sum_{j=1}^{q} \eta_j(t) g_y^{(j-1)}(y) f_{uu}(u, y) + \eta_q g_{uu}^{(q)}(u, y) \\ &= H_{uu}(u, y, p) - \textstyle\sum_{j=1}^{q-1} \eta_j(t) g_{uu}^{(j)}(y)(u, y) \;=\; H_{uu}(u, y, p). \end{aligned}$$

Proof of Proposition 2.10. Since η_q is piecewise C^q by Lemma 2.8, the functions η_j, $0 \leq j \leq q-1$ are well-defined. We show the equivalence between (2.13)-(2.20) and (2.23)-(2.33) augmented with (2.39)-(2.43).

Equivalence between state equations (2.13) and (2.23); final costate conditions (2.15) and (2.25); state constraint equations (2.17) and (2.27), (2.28), (2.30) on boundary arcs, and (2.19) and (2.29) at touch points, is obvious. Equivalence between costate equations (2.14) and (2.24), and between control equations (2.16) and (2.26), follows from calculation, using the relations between the functions η_j, p, and p_q and the fact that the state constraint is of order q (see e.g. [98]).

Additional conditions are necessary to ensure equivalence between complementarity and junction conditions. Obviously, (2.39)-(2.40) are equivalent to (2.18); as well, (2.33) and (2.43) are equivalent to to (2.20) for touch points. It remains to check that (2.20) is also equivalent to (2.31)-(2.32) and (2.41)-(2.42) at entry/exit points. Let $\tau_{en} \in \mathcal{T}_{en}$. Expressing $[p_q(\tau_{en})]$, using on the one hand the relationship (2.37) between p and p_q, as well as (2.20), and using on the other hand jump condition (2.31), we obtain

$$[p_q(\tau_{en})] = -\nu_{\tau_{en}} g_y(y(\tau_{en})) - \sum_{j=1}^{q} \eta_j(\tau_{en}^+) g_y^{(j-1)}(y(\tau_{en})), \tag{2.44}$$

$$[p_q(\tau_{en})] = -\sum_{j=1}^{q} \nu_{\tau_{en}}^j g_y^{(j-1)}(y(\tau_{en})). \tag{2.45}$$

By Lemma 2.9 at $t = \tau_{en}$, the right-hand sides of (2.44) and (2.45) are equal iff the coefficients of $g_y^{(j-1)}(y(\tau_{en}))$ for $j = 1, \ldots, q$ are equal. Eliminating $\nu_{\tau_{en}}$, which must be nonnegative (and equals zero for odd-order state constraints by Prop. 2.5(i)), we deduce (2.41). Proceeding similarly at exit points, (2.42) follows. □

Remark 2.12. Proposition 2.10 slightly improves section 5 of [98], in the sense that we give the complete set of additional conditions for which equivalence between regular Pontryagin extremals and the alternative formulation holds.

Remark 2.13. The sign condition of $\eta_q^{(q)}$ on boundary arcs (2.40) and exit point conditions (2.42) implies that the necessary condition

$$(-1)^{q-j}\frac{\mathrm{d}^{q-j}}{\mathrm{d}t^{q-j}}\eta_q(t) = \eta_j(t) \geq 0 \ \text{ on } \mathcal{I}_b \qquad \text{for } j = 1,\dots,q \tag{2.46}$$

holds as a consequence of (2.40) and (2.42). It is easily seen by induction, since $\dot\eta_j = -\eta_{j-1} \leq 0$ on $\mathcal{I}_b$ and $\eta_j(\tau_{ex}^-) \geq 0$ for all $\tau_{ex} \in \mathcal{T}_{ex}$. By (2.41), we deduce also that $\nu^j_{\tau_{en}} \geq 0$ for all $\tau \in \mathcal{T}_{en}$ and $j = 1,\dots,q$.

2.2.3 The shooting algorithm

The shooting algorithm extracts from the necessary optimality conditions a finite-dimensional set of equations (the shooting equations). If its Jacobian is invertible, we obtain a locally convergent algorithm by solving the shooting equations using, say, Newton's method.

In the unconstrained case, the initial value of the costate p_0 is mapped into the final condition (2.25). To handle alternative formulation of Def. 2.7, jump parameters and junction times are introduced as *shooting parameters*. A given set of shooting parameters determines a unique trajectory and multipliers (u, y, p_q, η_q) solution of the coupled state-costate system (2.23)-(2.24) with initial condition $p_q(0) = p_0$; algebraic equations (2.26), (2.28) and (2.30) that give u and η_q as implicit functions of (y, p_q) by (A2)-(A3); and jump conditions (2.31)-(2.33).

We use the shooting formulation of Malanowski and Maurer [93, 94]. Jump parameters $\nu^{1:q}_{\tau_{en}}$ at an entry time τ_{en} are associated with the q interior points conditions (2.27). Necessary optimality conditions for entry and exit points τ_{en} and τ_{ex} and touch points τ_{to} (when $q \geq 2$) are as follows:

$$g^{(q)}(u(\tau_{en}^-), y(\tau_{en})) = 0; \qquad g^{(q)}(u(\tau_{ex}^+), y(\tau_{ex})) = 0, \tag{2.47}$$

$$g^{(1)}(y(\tau_{to})) = 0. \tag{2.48}$$

By Proposition 2.5, the control is continuous along a regular Pontryagin extremal, so that (2.47) is a necessary optimality condition for entry/exit times. For a first order state constraint, we assume in what follows that $\mathcal{T}_{to} = \emptyset$ (see remark 2.19 below). Since a touch point τ_{to} is a local maximum of $g(y)$, when $q \geq 2$ (2.48) is a necessary optimality condition. Therefore, (2.48) together with the interior point constraint (2.29) provide two conditions associated with τ_{to} and its jump parameter $\nu_{\tau_{to}}$, for each $\tau_{to} \in \mathcal{T}_{to}$.

Definition 2.14. A trajectory (u, y) is a shooting extremal if it satisfies both the alternative formulation (Def. 2.7) and conditions (2.47)-(2.48).

Let us show how (2.47) relates to the additional conditions of Prop. 2.10.

Proposition 2.15. *Let (u, y) be a trajectory solution of the alternative formulation (2.23)-(2.33) and satisfying (A2$_q$)-(A4). Then the two following conditions are equivalent:*
(i) *The control u is continuous at entry/exit times τ_{en}, τ_{ex} (i.e., (2.47) holds);*
(ii) *Those additional conditions in (2.41)-(2.42) involving η_q are satisfied, i.e.*

$$\eta_q(\tau_{en}^+) - \nu^q_{\tau_{en}} = 0; \qquad \eta_q(\tau_{ex}^-) = 0. \tag{2.49}$$

Proof. Let $\tau_{en} \in \mathcal{T}_{en}$. By assumption (A3), the function $\hat{u} \mapsto g^{(q)}(\hat{u}, y(\tau_{en}))$ is one-to-one. Since $g^{(q)}(u(\tau_{en}^+), y(\tau_{en})) = 0$, we have that $g^{(q)}(u(\tau_{en}^-), y(\tau_{en})) = 0$ iff the control is continuous at time τ_{en}; the same type of arguments holds for exit points. It follows that (2.47) is equivalent to the continuity of the control at entry/exit points.

By (2.26), we have

$$\tilde{H}_u(u(\tau_{en}^-), y(\tau_{en}), p_q(\tau_{en}^-), 0) \;=\; 0 \;=\; \tilde{H}_u(u(\tau_{en}^+), y(\tau_{en}), p_q(\tau_{en}^+), \eta_q(\tau_{en}^+)).$$

We abbreviate $u(\tau_{en}^-)$ to u^- and so on. Using the jump condition of the costate (2.31), it follows that

$$\tilde{H}_u(u^+, y, p_q^+, \eta_q^+) = H_u(u^+, y, p_q^-) - \sum_{j=1}^{q} \nu_{\tau_{en}}^j g_y^{(j-1)}(y) f_u(u^+, y) + \eta_q^+ g_u^{(q)}(u^+, y).$$

The state constraint being of order q, we have $g_y^{(j-1)}(y) f_u(u, y) = g_u^{(j)}(y) = 0$ for $j = 1, \ldots, q-1$, and hence, we obtain

$$0 \;=\; H_u(u^+, y, p_q^-) + (\eta_q^+ - \nu_{\tau_{en}}^q) g_u^{(q)}(u^+, y).$$

Since $g_u^{(q)}(u^+, y) \neq 0$ by (A3), it follows that $H_u(u^+, y, p_q^-) = 0$ iff $\eta_q^+ = \nu_{\tau_{en}}^q$. Since by ($\mathrm{A2}_q$), $H_u(u^+, y, p_q^-) = 0$ iff $u^+ = u^-$, we deduce that u is continuous at time τ_{en} iff $\eta_q^+ = \nu_{\tau_{en}}^q$. Similar arguments hold for exit points. The conclusion follows. □

Remark 2.16. We can also check that if (u, y) is a shooting extremal satisfying ($\mathrm{A2}_q$)-(A4), then u is continuous at touch points $\tau \in \mathcal{T}_{to}$, if $q \geq 2$. Indeed, (2.26), (2.30), and (2.33) lead to

$$H_u(u^-, y, p_q^-) = 0 = H_u(u^+, y, p_q^+) = H_u(u^+, y, p_q^-) - \nu_\tau g_y(y) f_u(y, u^+).$$

Since $g_y f_u = g_u^{(1)} \equiv 0$ and $H_u(\cdot, y, p_q^-)$ is one-to-one by ($\mathrm{A2}_q$), we obtain $u^+ = u^-$.

It follows that if (u, y) is a shooting extremal satisfying ($\mathrm{A2}_q$)-(A4), then u is continuous on $[0, T]$, provided that we still assume that $\mathcal{T}_{to} = \emptyset$ if $q = 1$ (see Remark 2.19).

The *structure* of a feasible trajectory is defined as the (finite) *number* of boundary arcs and touch points of the trajectory, and the *order* in which they occur w.r.t. time. Assuming the structure of the optimal trajectory is known, we define the shooting mapping as follows. Denote by N_b and N_{to} the number of boundary arcs and touch points of the trajectory, respectively. The space of shooting parameters is

$$\Theta := \mathbb{R}^n \times \mathbb{R}^{qN_b} \times \mathbb{R}^{N_{to}} \times \mathbb{R}^{N_b} \times \mathbb{R}^{N_b} \times \mathbb{R}^{N_{to}}.$$

With the above notations, and for a given order of boundary arcs and touch points, the shooting mapping $\mathcal{F}$ is defined over a neighborhood in Θ of shooting parameters associated with a regular Pontryagin extremal, into Θ, by

$$\theta = \begin{pmatrix} p_0^* \\ \nu_{\mathcal{T}_{en}}^{1:q} \\ \nu_{\mathcal{T}_{to}} \\ \mathcal{T}_{en} \\ \mathcal{T}_{ex} \\ \mathcal{T}_{to} \end{pmatrix} \;\mapsto\; \begin{pmatrix} p_q(T)^* - \phi_y(y(T))^* \\ g^{(0:q-1)}(y(\mathcal{T}_{en})) \\ g(y(\mathcal{T}_{to})) \\ g^{(q)}(u(\mathcal{T}_{en}^-), y(\mathcal{T}_{en})) \\ g^{(q)}(u(\mathcal{T}_{ex}^+), y(\mathcal{T}_{ex})) \\ g^{(1)}(y(\mathcal{T}_{to})) \end{pmatrix}. \tag{2.50}$$

By construction, a zero of the shooting mapping $\mathcal{F}$ provides a trajectory (u, y) that is a shooting extremal. In view of Propositions 2.10 and 2.15, the following holds.

Corollary 2.17. *A shooting extremal satisfying (A2$_q$)-(A4) is a regular Pontryagin extremal iff it satisfies the following* minimal additional conditions*: (2.39) on interior arcs, (2.40) on boundary arcs, (2.43) at touch points, and for all entry points $\tau_{en} \in \mathcal{T}_{en}$ and exit points $\tau_{ex} \in \mathcal{T}_{ex}$:*

$$\text{if } q \geq 2 \text{ is even: } \nu^1_{\tau_{en}} - (-1)^{q-1}\eta_q^{(q-1)}(\tau_{en}^+) \geq 0 \ ; \ (-1)^{q-1}\eta_q^{(q-1)}(\tau_{ex}^-) \geq 0; \tag{2.51}$$

$$\left\{ \begin{array}{c} \text{if } q \geq 3 \text{ is odd, } j = 1, \dots, q-1, \text{ and if } q \geq 4 \text{ is even, } j = 2, \dots, q-1\text{:} \\ \nu^j_{\tau_{en}} - (-1)^{q-j}\eta_q^{(q-j)}(\tau_{en}^+) = 0 \quad ; \quad (-1)^{q-j}\eta_q^{(q-j)}(\tau_{ex}^-) = 0. \end{array} \right. \tag{2.52}$$

Note that (2.51)-(2.52) is only a reformulation of (2.41)-(2.42), from which we removed the condition corresponding to $j = q$, namely (2.49), since the latter is automatically satisfied by Prop. 2.15. Consequently, when $q = 1$, there remain no additional conditions at entry/exit points for shooting extremals.

Remark 2.18. It follows that for first- and second-order state constraints, and for constraints of order $q > 2$ having no boundary arcs (see Remark 2.42 concerning existence of boundary arcs for state constraints of order $q \geq 3$), the additional conditions reduce to the *inequalities* (2.39), (2.40), (2.43), and also (2.51) when $q = 2$ at entry/exit points.

Remark 2.19. For a first-order state constraint, jump parameters $\nu_{\mathcal{T}_{to}}$ associated with touch points are equal to zero along a regular Pontryagin extremal by Prop. 2.5. For this reason, we assume in this paper that $\mathcal{T}_{to} = \emptyset$ if $q = 1$.

Remark 2.20. The nonlocal hypotheses (A2) (or (A2$_q$)) as well as (2.10) (or (2.12)) are essential in order to prove that the control is continuous. Some of our results remain valid, substituting everywhere *stationary point* for *(regular) Pontryagin extremal*, when the assumptions (A2) and (2.10) in (A3) are replaced by the weaker assumptions that u is continuous over $[0, T]$ and that there exists $\alpha, \gamma > 0$ such that

$$H_{uu}(u(t), y(t), p(t)) \geq \alpha \quad \text{and} \quad |g_u^{(q)}(u(t), y(t))| \geq \gamma \quad \text{for all } t \in [0, T]. \tag{2.53}$$

This holds in particular for Propositions 2.4, 2.5, 2.10, 2.15, Remark 2.16, and Corollary 2.17. The same remark applies for the other results of this paper, i.e. Theorems 2.22, 2.23, 2.34; Corollary 2.41, and Lemmas 2.43 and 2.44 in the appendix.

2.3 Well-Posedness of the Shooting Algorithm

We say that the shooting algorithm is locally *well-posed* if the Jacobian of the shooting mapping (2.50) is invertible at some local solution of $(\mathcal{P})$. This allows us to apply locally a Newton method in order to find a shooting extremal; the additional conditions for a Pontryagin extremal have to be checked afterwards.

Let us first give some definitions. Given $u \in \mathcal{U}$, recall that we denote by y_u the (unique) solution in $\mathcal{Y}$ of the state equation (2.2). This well-defined mapping is of class C^k under assumption (A0). Let the cost function be

$$J(u) = \int_0^T \ell(u(t), y_u(t))\mathrm{d}t + \phi(y_u(T)). \tag{2.54}$$

We say that a feasible trajectory $(u, y = y_u)$ is a local solution of $(\mathcal{P})$ satisfying the *quadratic growth condition* if there exists $c, r > 0$ such that

$$J(\tilde{u}) \geq J(u) + c\,\|\tilde{u} - u\|_2^2 \qquad \forall\, \tilde{u} \in B_\infty(u, r); \quad g(y_{\tilde{u}}(t)) \leq 0 \text{ on } [0, T], \tag{2.55}$$

where B_∞ denotes the open ball in $L^\infty(0,T)$ with center u and radius r. This condition involves two norms, $L^\infty(0,T)$ for the neighborhood and $L^2(0,T)$ for the growth condition.

Let (u,y) be a regular Pontryagin extremal. We make the following strict complementarity assumption (compare to (2.40), (2.51), and (2.43), where large inequalities are replaced by strict inequalities):

(A5) **(i)** For all boundary arcs $[\tau_{en},\tau_{ex}]$:

$$(-1)^q \frac{\mathrm{d}^q}{\mathrm{d}t^q}\eta_q(t) > 0 \qquad \text{a.e. on } (\tau_{en},\tau_{ex}), \tag{2.56}$$

$$\text{If } q \text{ is } \textit{odd}: \qquad \frac{\mathrm{d}^q}{\mathrm{d}t^q}\eta_q(\tau_{en}^+) < 0; \qquad \frac{\mathrm{d}^q}{\mathrm{d}t^q}\eta_q(\tau_{ex}^-) < 0, \tag{2.57}$$

$$\text{If } q \text{ is } \textit{even}: \qquad \nu^1_{\tau_{en}} + \frac{\mathrm{d}^{q-1}}{\mathrm{d}t^{q-1}}\eta_q(\tau_{en}^+) > 0; \qquad \frac{\mathrm{d}^{q-1}}{\mathrm{d}t^{q-1}}\eta_q(\tau_{ex}^-) < 0. \tag{2.58}$$

(ii) For all touch points $\tau_{to} \in \mathcal{T}_{to}$:

$$\nu_{\tau_{to}} > 0. \tag{2.59}$$

Recall that $(-1)^q \frac{\mathrm{d}^q}{\mathrm{d}t^q}\eta_q(t)$ equals η_0, the density of η (see Prop. 2.10). Let $\hat{q} := 2q-1$ if q is *odd* and $\hat{q} := 2q-2$ if q is *even*. By Prop. 2.5, $\hat{q}+1$ is the smallest possible order for which the corresponding time derivative of $g(y(t))$ may be discontinuous at an entry/exit point. Note that $\hat{q} = q$ for $q = 1, 2$.

Lemma 2.21. *Let (u,y) be a regular Pontryagin extremal satisfying (A2)-(A4). For odd (resp. even) q, assumption* (2.57) *(resp.* (2.58)*) holds iff the following non-tangentiality condition at order $\hat{q}+1$ holds: for all entry times $\tau_{en} \in \mathcal{T}_{en}$ and all exit times $\tau_{ex} \in \mathcal{T}_{ex}$,*

$$(-1)^{\hat{q}+1}\frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}}g(y(t))|_{t=\tau_{en}^-} < 0; \qquad \frac{\mathrm{d}^{\hat{q}+1}}{\mathrm{d}t^{\hat{q}+1}}g(y(t))|_{t=\tau_{ex}^+} < 0. \tag{2.60}$$

Proof. By Prop. 2.10 (see (2.38)), (2.58) is equivalent, when q is even, to the strict positivity of ν_τ at entry/exit points $\tau \in \mathcal{T}_{en} \cup \mathcal{T}_{ex}$. The conclusion is then a consequence of Prop. 2.10 and of Lemma 2.44 whose (technical) proof is given in the appendix. □

Assumption (A5)(ii) implies that if $q = 1$, then $\mathcal{T}_{to} = \emptyset$ by Prop. 2.5(ii). When $q \geq 2$, we assume that all touch points of (u,y) are *reducible*, in the following sense:

(A6) For all touch points $\tau_{to} \in \mathcal{T}_{to}$:

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y(t))|_{t=\tau_{to}} < 0. \tag{2.61}$$

This makes sense, since when $q \geq 2$, we have $\frac{\mathrm{d}^2}{\mathrm{d}t^2}g(y(t)) = g^{(2)}(u,y)$ and u is continuous by Prop. 2.5.

2.3.1 Statement of main results

Define the quadratic cost function:

$$\begin{aligned}\mathcal{J}_q(v,z) \;:=\; & \int_0^T \tilde{H}_{(u,y),(u,y)}(u,y,p_q,\eta_q)((v,z),(v,z))\mathrm{d}t \\ & + z(T)^* \phi_{yy}(y(T))z(T) \;+\; \sum_{\tau\in\mathcal{T}_{en}} \sum_{j=1}^{q} \nu_\tau^j z(\tau)^* g_{yy}^{(j-1)}(y(\tau))z(\tau) \\ & + \sum_{\tau\in\mathcal{T}_{to}} \nu_\tau \left(z(\tau)^* g_{yy}(y(\tau))z(\tau) \;-\; \frac{(g_y^{(1)}(y(\tau))z(\tau))^2}{\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(y(t))|_{t=\tau}} \right)\end{aligned} \tag{2.62}$$

where $\tilde{H}$ is the augmented Hamiltonian (2.22), and the set of constraints:

$$\dot{z} = f_y(u,y)z + f_u(u,y)v \qquad \text{on } [0,T]; \qquad z(0)=0 \tag{2.63}$$

$$g_y^{(j)}(y(\tau))z(\tau) = 0 \qquad \text{for } j=0,\dots,q-1; \qquad \tau\in\mathcal{T}_{en} \tag{2.64}$$

$$g_{(u,y)}^{(q)}(u(t),y(t))(v(t),z(t)) = 0 \qquad t\in\mathcal{I}_b \tag{2.65}$$

$$g_y(y(\tau))z(\tau) = 0 \qquad \tau\in\mathcal{T}_{to}. \tag{2.66}$$

Since the state equation and constraints are linear, the cost function is quadratic, and all have bounded coefficients, we may take as linearized control and state spaces $\mathcal{V} := L^2(0,T)$ and $\mathcal{Z} := H^1(0,T;\mathbb{R}^n)$, where $H^1(0,T)$ is the Sobolev space of functions in $L^2(0,T)$ with a weak derivative in $L^2(0,T)$. Let the linear quadratic problem (PQ_q) be defined by

$$(\mathbf{PQ}_q) \qquad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \frac{1}{2}\mathcal{J}_q(v,z) \qquad \text{subject to (2.63)-(2.66).} \tag{2.67}$$

Consider the following second-order conditions:

$$(v,z)=0 \text{ is a solution of } (PQ_q). \tag{2.68}$$

$$(v,z)=0 \text{ is the } \textit{unique} \text{ solution of } (PQ_q). \tag{2.69}$$

Theorem 2.22 (No-gap second-order optimality conditions). (i) *Let (u,y) be a local solution of $(\mathcal{P})$ satisfying (A2)-(A6). Then its associated multipliers in the alternative formulation are such that the second-order necessary condition* (2.68) *holds.*
(ii) *Let (u,y) be a Pontryagin extremal satisfying (A2)-(A6). Then the second-order sufficient condition* (2.69) *holds iff (u,y) is a local solution of $(\mathcal{P})$ satisfying the quadratic growth condition* (2.55).

Theorem 2.23 (Well-posedness of the shooting algorithm). *Let (u,y) be a local solution of $(\mathcal{P})$ satisfying (A2)-(A6). Then the shooting algorithm is locally well-posed (invertible Jacobian), iff the following two conditions hold:* (i) *If $q\geq 3$, the trajectory (u,y) does not have boundary arcs;* (ii) *The second-order sufficient condition* (2.69) *holds.*

In general, even for unconstrained problems, the invertibility of the Jacobian of the shooting mapping at a Pontryagin extremal does not imply that the second-order sufficient condition (2.69) holds. We comment on the ill-posedness of the shooting algorithm along boundary arcs of order $q\geq 3$ in Remark 2.42.

Combining Theorems 2.22(ii) and 2.23, we obtain that if (u,y) is a local solution of $(\mathcal{P})$ satisfying (A2)-(A6) and condition (i) of Th. 2.23, then the shooting algorithm is well-posed iff (u,y) satisfies the quadratic growth condition.

2.3.2 Proof of the no-gap Second-order Optimality Conditions (Theorem 2.22)

We use the no-gap second-order optimality conditions established in [18, 21]. Let (u, y) be a regular Pontryagin extremal, with the multiplier $\eta \in \mathcal{M}[0, T]$ given by (2.21). Consider the quadratic cost function:

$$\begin{aligned}\mathcal{J}(v,z) \;:=\; & \int_0^T H_{(u,y),(u,y)}(u,y,p)((v,z),(v,z))\mathrm{d}t + z(T)^*\phi_{yy}(y(T))z(T) \\ & + \int_0^T (z^* g_{yy}(y)z)\,\mathrm{d}\eta \;-\; \sum_{\tau\in\mathcal{T}_{to}} \nu_\tau \frac{(g_y^{(1)}(y(\tau))z(\tau))^2}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y(t))|_{t=\tau}},\end{aligned} \tag{2.70}$$

where H is the classical Hamiltonian (2.4), and consider the constraint

$$g_y(y(t))z(t) = 0 \;\text{ on } \mathcal{I}_b \cup \mathcal{T}_{to}. \tag{2.71}$$

The quadratic problem used in the formulation of the second-order optimality conditions in [21] is the following:

$$(\mathbf{PQ}) \qquad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \; \frac{1}{2}\mathcal{J}(v,z) \qquad \text{subject to (2.63) and (2.71).} \tag{2.72}$$

Theorem 2.24. (i) *If (u, y) is a local solution of $(\mathcal{P})$ such that (A2)-(A6) hold, then $(v, z) = 0$ is a solution of problem* (2.72).
(ii) *If (u, y) is a Pontryagin extremal such that (A2)-(A6) hold, it is a local solution of $(\mathcal{P})$ satisfying the quadratic growth condition* (2.55) *iff problem* (2.72) *has zero for unique solution.*

Proof. See Corollary 15 and Theorems 18 and 27 in [21][1], or Theorem 0.1 in [18]. For the sake of completeness, let us recall the main ideas. The proof of the second-order necessary condition is based on the computation of the curvature term obtained by Kawasaki [77, 79] in abstract optimization framework. With the junction conditions results of Prop. 2.5 and (A5)(i), we can show that boundary arcs have a zero contribution to the curvature term. For the second-order sufficient condition, a reduction method is used around the finitely many reducible touch points. In fact, the proof of the sufficient condition is very similar to the proof of Lemma 2.40 in the stability analysis below. □

We establish the link between Th. 2.24 and the second-order conditions (2.68)-(2.69) derived from the alternative formulation. In the end of this section we often omit the time argument when there is no ambiguity. The proof of the next lemma is easy and therefore omitted.

Lemma 2.25. *Assume that the state constraint is of order q. Then for every trajectory (u, y) and every linearized trajectory $(v, z) \in \mathcal{V} \times \mathcal{Z}$ satisfying* (2.63)*, the following holds:*

$$\frac{\mathrm{d}^j}{\mathrm{d}t^j} g_y(y(t))z(t) \;=\; g_y^{(j)}(y)z, \qquad j = 1, \ldots, q-1, \tag{2.73}$$

$$\frac{\mathrm{d}^q}{\mathrm{d}t^q} g_y(y(t))z(t) \;=\; g_y^{(q)}(u,y)z + g_u^{(q)}(u,y)v. \tag{2.74}$$

[1] Corollary 1.15 and Theorems 1.18 and 1.27 of this thesis.

Lemma 2.26. *Let (u,y) be a regular Pontryagin extremal satisfying (A2)-(A4), with classical and alternative multipliers (p,η) and $(p_q,\eta_q,\nu^{1:q}_{\mathcal{T}_{en}},\nu_{\mathcal{T}_{to}})$, respectively, related to each other by* (2.36)-(2.38)*,* (2.41)*, and* (2.21)*. Then the quadratic cost functions $\mathcal{J}$ and $\mathcal{J}_q$, defined respectively in* (2.70) *and* (2.62)*, are equal to each other over the space of linearized trajectories $(v,z)\in\mathcal{V}\times\mathcal{Z}$ satisfying* (2.63)*.*

Proof. Let $(v,z)\in\mathcal{V}\times\mathcal{Z}$ satisfy (2.63) and set $\Delta_{PQ}:=\mathcal{J}(v,z)-\mathcal{J}_q(v,z)$. Using (2.21), it is easily seen that the terms corresponding to the touch points and to the final time vanish, and hence we get

$$\begin{aligned}\Delta_{PQ} &= \int_0^T (p-p_q)D^2f(u,y)((v,z),(v,z))\mathrm{d}t+\int_0^T g_{yy}(y)(z,z)\eta_0(t)\mathrm{d}t\\ &\quad-\int_0^T D^2g^{(q)}(u,y)((v,z),(v,z))\eta_q(t)\mathrm{d}t+\sum_{\tau\in\mathcal{T}_{ex}}\nu_\tau g_{yy}(y)(z,z)(\tau)\\ &\quad+\sum_{\tau\in\mathcal{T}_{en}}\left(\nu_\tau g_{yy}(y)(z,z)(\tau)\ -\sum_{j=1}^{q}\nu_\tau^j g_{yy}^{(j-1)}(y)(z,z)(\tau)\right).\end{aligned}$$

In what follows we abbreviate the notation $((v,z),(v,z))$ by $((v,z))^2$. Relations (2.36)-(2.37) between p and p_q lead to

$$\begin{aligned}\Delta_{PQ} &= \sum_{j=1}^{q}\int_0^T g_y^{(j-1)}(y)D^2f(u,y)((v,z))^2\eta_j(t)\mathrm{d}t+\int_0^T g_{yy}(y)(z,z)\eta_0(t)\mathrm{d}t\\ &\quad-\int_0^T D^2g^{(q)}(u,y)((v,z))^2\eta_q(t)\mathrm{d}t+\sum_{\tau\in\mathcal{T}_{ex}}\nu_\tau g_{yy}(y)(z,z)(\tau)\\ &\quad+\sum_{\tau\in\mathcal{T}_{en}}\left(\nu_\tau g_{yy}(y)(z,z)(\tau)\ -\sum_{j=1}^{q}\nu_\tau^j g_{yy}^{(j-1)}(y)(z,z)(\tau)\right).\end{aligned}\tag{2.75}$$

The constraint being of order q, we have $g^{(j)}(u,y)=g_y^{(j-1)}(y)f(u,y)$ for $j=0$ to $q-1$. It follows that

$$\begin{aligned}D^2g^{(j)}(u,y)((v,z))^2 &= g_{yyy}^{(j-1)}(y)(f(u,y),z,z)+2g_{yy}^{(j-1)}(y)(z,Df(u,y)(v,z))\\ &\quad+g_y^{(j-1)}(y)D^2f(u,y)((v,z))^2.\end{aligned}\tag{2.76}$$

In addition, by the linearized state equation (2.63), we have, for all $j=1,\dots,q$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[g_{yy}^{(j-1)}(y(t))(z(t),z(t))\right]=g_{yyy}^{(j-1)}(y)(f(u,y),z,z)+2g_{yy}^{(j-1)}(y)(z,Df(u,y)(v,z)),$$

which gives by (2.76), for $j=1,\dots,q$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[g_{yy}^{(j-1)}(y(t))(z(t),z(t))\right]=D^2g^{(j)}(u,y)((v,z))^2-g_y^{(j-1)}(y)D^2f(u,y)((v,z))^2.\tag{2.77}$$

Since $g_u^{(j-1)}(u,y)\equiv 0$ for $j=1,\dots,q$, we have $g_{yy}^{(j-1)}(y)(z,z)=D^2g^{(j-1)}(u,y)((v,z))^2$ for $j=1,\dots,q$. Multiplying (2.77) by η_j, integrating over $[0,T]$, and integrating by parts the

left-hand side (recall that $\dot{\eta}_j = -\eta_{j-1}$), we obtain, for $j = 1, \dots, q$

$$\begin{aligned}\int_0^T D^2 g^{(j-1)}(u,y)((v,z))^2 \eta_{j-1}(t)\mathrm{d}t + \sum_{\tau\in\tau_{ex}} g_{yy}^{(j-1)}(y)(z,z)\eta_j(\tau^-) \\ - \sum_{\tau\in\tau_{en}} g_{yy}^{(j-1)}(y)(z,z)\eta_j(\tau^+) \\ = \int_0^T D^2 g^{(j)}(u,y)((v,z))^2 \eta_j(t)\mathrm{d}t - \int_0^T g_y^{(j-1)}(y) D^2 f(u,y)((v,z))^2 \eta_j(t)\mathrm{d}t.\end{aligned}$$

Adding the above equalities for $j = 1, \dots, q$, we get after simplification by $\int_0^T D^2 g^{(j)}(u,y)((v,z))^2\eta_j$ for $j = 1, \dots, q-1$ that

$$\begin{aligned}\int_0^T g_{yy}(y)(z,z)\eta_0(t)\mathrm{d}t + \sum_{j=1}^{q}\sum_{\tau\in\tau_{ex}} g_{yy}^{(j-1)}(y)(z,z)\eta_j(\tau^-) \\ - \sum_{j=1}^{q}\sum_{\tau\in\tau_{en}} g_{yy}^{(j-1)}(y)(z,z)\eta_j(\tau^+) \\ = \int_0^T D^2 g^{(q)}(u,y)((v,z))^2 \eta_q(t)\mathrm{d}t - \sum_{j=1}^{q}\int_0^T g_y^{(j-1)}(y) D^2 f(u,y)((v,z))^2 \eta_j(t)\mathrm{d}t.\end{aligned}$$

Substituting into (2.75) gives

$$\begin{aligned}\Delta_{PQ} &= \sum_{\tau\in\tau_{ex}} \left(\nu_\tau g_{yy}(y)(z,z)(\tau) - \sum_{j=1}^{q} g_{yy}^{(j-1)}(y)(z,z)\eta_j(\tau^-) \right) \\ &+ \sum_{\tau\in\tau_{en}} \left(\nu_\tau g_{yy}(y)(z,z)(\tau) + \sum_{j=1}^{q} \left(\eta_j(\tau^+) - \nu_\tau^j\right) g_{yy}^{(j-1)}(y)(z,z)(\tau) \right).\end{aligned}$$

Using (2.38) and additional conditions at entry and exit points (2.41)-(2.42), we obtain that $\Delta_{PQ} = 0$. Thus, the cost functions of the two quadratic problems coincide on the feasible set. □

Proof of Theorem 2.22. The state constraint being of order q, it follows from (2.73)-(2.74) that (2.64)-(2.66) and (2.71) are equivalent. By Lemma 2.26, problems (PQ_q) and (2.72) have the same feasible set and the same cost function on that feasible set, and hence they also have the same value and the same set of optimal solutions. The conclusion follows then from Theorem 2.24. □

2.3.3 Proof of the Well-posedness (Theorem 2.23)

We give a sequence of lemmas; some of them will also be used in section 2.4.

We denote e.g. by $g_y^{(j)}(y(\mathcal{T}_{en}))z(\mathcal{T}_{en})$, $g_{(u,y)}^{(q)}(u(\mathcal{T}_{en}), y(\mathcal{T}_{en}))(v(\mathcal{T}_{en}^-), z(\mathcal{T}_{en}))$, the vectors in $\mathbb{R}^{N_b}$ of components $g_y^{(j)}(y(\tau))z(\tau)$, $g_{(u,y)}^{(q)}(u(\tau), y(\tau))(v(\tau^-), z(\tau))$, respectively, for $\tau \in \mathcal{T}_{en}$. By $g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z(\mathcal{T}_{en})$ we denote the vector in $\mathbb{R}^{qN_b}$ of component $g_y^{(j)}(y(\tau))z(\tau)$, $0 \le j \le q-1$, $\tau \in \mathcal{T}_{en}$.

Lemma 2.27. *Let (u, y) be a shooting extremal satisfying (A2$_q$)-(A4), with the set of shooting parameters $\theta_0 = (p_0^*, \nu_{\mathcal{T}_{en}}^{1:q}, \nu_{\mathcal{T}_{to}}, \mathcal{T}_{en}, \mathcal{T}_{ex}, \mathcal{T}_{to}) \in \Theta$, such that $\mathcal{F}(\theta_0) = 0$ with the shooting mapping $\mathcal{F}$ defined in (2.50). Then $\mathcal{F}$ is of class C^1 on a neighborhood Θ_0 of θ_0, and at the direction*

$$\omega := (\pi_0^*, \gamma_{\mathcal{T}_{en}}^{1:q}, \gamma_{\mathcal{T}_{to}}, \sigma_{\mathcal{T}_{en}}, \sigma_{\mathcal{T}_{ex}}, \sigma_{\mathcal{T}_{to}}) \in \Theta, \tag{2.78}$$

the vector $\mathcal{M} := D\mathcal{F}(\theta_0)\omega$ can be split into $\mathcal{M} = (\mathcal{M}_{\mathcal{Q}}^*, \mathcal{M}_{\mathcal{T}}^*)^*$ given by

$$\mathcal{M}_{\mathcal{Q}} := \begin{pmatrix} \pi(T)^* - \phi_{yy}(y(T))z(T) \\ g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z(\mathcal{T}_{en}) \\ g_y(y(\mathcal{T}_{to}))z(\mathcal{T}_{to}) \end{pmatrix}, \tag{2.79}$$

$$\mathcal{M}_{\mathcal{T}} := \begin{pmatrix} g_{(u,y)}^{(q)}(u(\mathcal{T}_{en}), y(\mathcal{T}_{en}))(v(\mathcal{T}_{en}^-), z(\mathcal{T}_{en})) + \sigma_{\mathcal{T}_{en}} \dfrac{\mathrm{d}}{\mathrm{d}t} g^{(q)}(u, y)|_{t=\mathcal{T}_{en}^-} \\ g_{(u,y)}^{(q)}(u(\mathcal{T}_{ex}), y(\mathcal{T}_{ex}))(v(\mathcal{T}_{ex}^+), z(\mathcal{T}_{ex})) + \sigma_{\mathcal{T}_{ex}} \dfrac{\mathrm{d}}{\mathrm{d}t} g^{(q)}(u, y)|_{t=\mathcal{T}_{ex}^+} \\ g_y^{(1)}(y(\mathcal{T}_{to}))z(\mathcal{T}_{to}) + \sigma_{\mathcal{T}_{to}} \dfrac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(y)|_{t=\mathcal{T}_{to}} \end{pmatrix}, \tag{2.80}$$

where (v, z, π, ζ), the linearized control, state, costate and state constraint multiplier, are the solutions of (omitting arguments (u, y, p_q, η_q) and t)

$$\begin{aligned}
\dot{z} &= f_y z + f_u v && \text{on } [0,T]\ ; \qquad z(0) = 0 && (2.81)\\
-\dot{\pi} &= \tilde{H}_{yy} z + \tilde{H}_{yu} v + \pi f_y + \zeta g_y^{(q)} && \text{on } [0,T] \setminus \mathcal{T} && (2.82)\\
0 &= \tilde{H}_{uy} z + \tilde{H}_{uu} v + \pi f_u + \zeta g_u^{(q)} && \text{a.e. on } [0,T] && (2.83)\\
0 &= g_y^{(q)} z + g_u^{(q)} v && \text{a.e. on } \mathcal{I}_b && (2.84)\\
0 &= \zeta && \text{on } [0,T] \setminus \mathcal{I}_b && (2.85)
\end{aligned}$$

with initial condition of π given by $\pi(0) = \pi_0$ and and jump conditions of π given by

$$\begin{aligned}
[\pi(\tau)] &= -\sum_{j=1}^{q} \nu_\tau^j z(\tau)^* g_{yy}^{(j-1)}(y(\tau)) - \sum_{j=1}^{q} \gamma_\tau^j g_y^{(j-1)}(y(\tau)) \\
&\quad - \sigma_\tau \sum_{j=1}^{q-1} \nu_\tau^j g_y^{(j)}(y(\tau)); \qquad \tau \in \mathcal{T}_{en} && (2.86)\\
[\pi(\tau)] &= 0; \qquad \tau \in \mathcal{T}_{ex} && (2.87)\\
[\pi(\tau)] &= -\nu_\tau z(\tau)^* g_{yy}(y(\tau)) - \gamma_\tau g_y(y(\tau)) - \sigma_\tau \nu_\tau g_y^{(1)}(y(\tau)); \qquad \tau \in \mathcal{T}_{to}. && (2.88)
\end{aligned}$$

Proof. We detail only how we obtain the jump conditions of the linearized costate π at entry times; the other equations are obvious. In view of (2.31), it is easy to check that the jump of π at $\tau \in \mathcal{T}_{en}$ is given by

$$[\pi(\tau)] = -\sum_{j=1}^{q} \nu_\tau^j z(\tau)^* g_{yy}^{(j-1)}(y(\tau)) - \sum_{j=1}^{q} \gamma_\tau^j g_y^{(j-1)}(y(\tau)) + \sigma_\tau \Delta_\tau,$$

where the vector of sensitivity coefficients Δ_τ on junction time is given by

$$\Delta_\tau = -\sum_{j=1}^{q} \nu_\tau^j g_{yy}^{(j-1)}(y(\tau)) f(u(\tau^-), y(\tau)) + [\tilde{H}_y(u(\tau), y(\tau), p_q(\tau), \eta_q(\tau))].$$

By continuity of u at junction times (Prop. 2.15) and by (2.31), we have (omitting argument τ and setting $\eta_q^+ = \eta_q(\tau^+)$)

$$\Delta_\tau = -\sum_{j=1}^{q} \nu_\tau^j g_{yy}^{(j-1)}(y) f(u,y) - \sum_{j=1}^{q} \nu_\tau^j g_y^{(j-1)}(y) f_y(u,y) + \eta_q^+ g_y^{(q)}(u,y).$$

Since $g_y^{(j)}(u,y) = g_{yy}^{(j-1)}(y) f(u,y) + g_y^{(j-1)}(y) f_y(u,y)$ for $j = 1,\dots,q$, and since by Prop. 2.15, we have $\eta_q(\tau^+) = \nu_\tau^q$, we obtain (2.86). □

We recall that a continuous quadratic form defined over a Hilbert space is a *Legendre form* (see e.g. [74, 24]) if it is weakly lower semicontinuous and satisfies the following property: For all weakly convergent sequence $(v_n) \subset L^2(0,T)$, $v_n \rightharpoonup v$, we have that $v_n \to v$ strongly if $Q(v_n) \to Q(v)$.

Lemma 2.28. *Let (u,y) be a shooting extremal satisfying (A2$_q$)-(A4). For all $v \in \mathcal{V}$, define z_v as the (unique) solution in $\mathcal{Z}$ of the linearized state equation* (2.63)*, and define the operator $\mathcal{A} : \mathcal{V} \to W := L^2(\mathcal{I}_b) \times \mathbb{R}^{qN_b} \times \mathbb{R}^{N_{to}}$ by*

$$\mathcal{A}v = \begin{pmatrix} (g_y^{(q)}(u(\cdot),y(\cdot))z_v(\cdot) + g_u^{(q)}(u(\cdot),y(\cdot))v(\cdot))|_{\mathcal{I}_b} \\ g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z_v(\mathcal{T}_{en}) \\ g_y(y(\mathcal{T}_{to}))z_v(\mathcal{T}_{to}) \end{pmatrix}. \tag{2.89}$$

Then (i) *the continuous linear operator $\mathcal{A}$ is onto, and* (ii) *if in addition the second-order sufficient condition* (2.69) *holds, then there exists $\alpha > 0$, such that*

$$Q(v) := \mathcal{J}_q(v, z_v) \geq \alpha \|v\|_2^2, \quad \forall v \in \operatorname{Ker} \mathcal{A}. \tag{2.90}$$

By $\varphi|_{\mathcal{I}_b}$, we denote the restriction to $\mathcal{I}_b$ of function φ defined over $[0,T]$.

Proof. The continuity of $\mathcal{A}$ follows from that of $\mathcal{V} \to \mathcal{Z}$, $v \mapsto z_v$. By (2.10) and Lemma 2.25, the range of the mapping $\mathcal{V} \to \mathcal{Z}$, $v \mapsto g_y(y(\cdot))z_v(\cdot)$ is the subspace denoted by H_0^q of functions $\varphi \in H^q(0,T) = W^{q,2}(0,T)$ satisfying $\varphi^{(j)}(0) = 0$ for all $j = 0,\dots,q-1$. Points (i) follows, since by (A4), for all $(\psi(\cdot), b_{\mathcal{T}_{en}}^{1:q}, b_{\mathcal{T}_{to}}) \in W$, there exists $\varphi \in H_0^q$ such that $\varphi^{(q)}(t) = \psi(t)$ a.e. on $\mathcal{I}_b$, $\varphi^{(j-1)}(\mathcal{T}_{en}) = b_{\mathcal{T}_{en}}^j$, $j = 1,\dots,q$, and $\varphi(\mathcal{T}_{to}) = b_{\mathcal{T}_{to}}$.

By (A2$_q$), we can show that $Q(v)$ is a Legendre form over $L^2(0,T)$ (the proof is similar to that of Lemma 21 in [21][2]). By (2.69), we have $Q(v) > 0$ for all $v \in \operatorname{Ker} \mathcal{A} \setminus \{0\}$, which implies (2.90) by Lemma 2.45. □

Proposition 2.29. *Let (u,y) be a shooting extremal satisfying (A2$_q$)-(A4) and denote by $\theta_0 \in \Theta$ its set of shooting parameters. Assume that* (i) *the second-order sufficient condition* (2.69) *is satisfied; and* (ii) *the following holds at junction times:*

$$\frac{\mathrm{d}}{\mathrm{d}t} g^{(q)}(u,y)|_{t=\tau^-} \neq 0 \quad \forall \tau \in \mathcal{T}_{en}\ ; \qquad \frac{\mathrm{d}}{\mathrm{d}t} g^{(q)}(u,y)|_{t=\tau^+} \neq 0 \quad \forall \tau \in \mathcal{T}_{ex} \tag{2.91}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(y)|_{t=\tau} \neq 0 \qquad \forall \tau \in \mathcal{T}_{to}. \tag{2.92}$$

Then the Jacobian $D\mathcal{F}(\theta_0)$ of the shooting mapping is invertible, and for all

$$\delta = (a_T, b_{\mathcal{T}_{en}}^{1:q}, b_{\mathcal{T}_{to}}, c_{\mathcal{T}_{en}}, c_{\mathcal{T}_{ex}}, c_{\mathcal{T}_{to}}) \in \Theta,$$

[2] Lemma 1.21 of this thesis.

the (unique) solution $\omega \in \Theta$ of $D\mathcal{F}(\theta_0)\omega = \delta$, with ω given by (2.78), is as follows. With the notation of Lemma 2.28, denote by (v_δ, w_δ) with $w_\delta = (\zeta_\delta, \lambda^{1:q}_{\delta,\mathcal{T}_{en}}, \lambda_{\delta,\mathcal{T}_{to}})$ the unique solution in $L^2(0,T) \times W$ of the first-order optimality system of the problem

$$(\mathcal{P}^\delta) \quad \min_{v\in\mathcal{V}} \quad \frac{1}{2}\mathcal{J}_q(v, z_v) \; + \; a_T^* z_v(T) \; + \; \sum_{\tau\in\mathcal{T}_{to}} c_\tau \nu_\tau \frac{g_y^{(1)}(y(\tau))z_v(\tau)}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y)|_{t=\tau}}, \qquad (2.93)$$
$$\text{subject to} \qquad \mathcal{A}v = (0_{L_2(\mathcal{I}_b)}, b^{1:q}_{\mathcal{T}_{en}}, b_{\mathcal{T}_{to}})^*.$$

Then $\pi_0 = \pi_\delta(0)$, where π_δ is the solution on $[0,T] \setminus \mathcal{T}$ of (2.82) with $(v_\delta, \zeta_\delta, z_\delta := z_{v_\delta})$, with final and jump conditions of π_δ being given by

$$\pi_\delta(T) = z_\delta(T)^* \phi_{yy}(y(T)) + a_T^*, \qquad (2.94)$$
$$-[\pi_\delta(\tau)] = \sum_{j=1}^{q} \nu_\tau^j z_\delta(\tau)^* g_{yy}^{(j-1)}(y(\tau)) + \sum_{j=1}^{q} \lambda^j_{\delta,\tau} g_y^{(j-1)}(y(\tau)), \quad \tau \in \mathcal{T}_{en}, \qquad (2.95)$$
$$-[\pi_\delta(\tau)] = 0, \quad \tau \in \mathcal{T}_{ex}, \qquad (2.96)$$
$$-[\pi_\delta(\tau)] = \nu_\tau z_\delta(\tau)^* g_{yy}(y(\tau)) \; + \; \lambda_{\delta,\tau} g_y(y(\tau)) - \nu_\tau z_\delta(\tau)^* \frac{g_y^{(1)}(y(\tau))^* g_y^{(1)}(y(\tau))}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y)|_{t=\tau}} + c_\tau \nu_\tau \frac{g_y^{(1)}(y(\tau))}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y)|_{t=\tau}}, \quad \tau \in \mathcal{T}_{to}; \qquad (2.97)$$

and we have $\gamma_{\mathcal{T}_{to}} = \lambda_{\delta,\mathcal{T}_{to}}$,

$$\sigma_\tau = \frac{c_\tau - g_y^{(1)}(y(\tau))z_\delta(\tau)}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(y)|_{t=\tau}}, \quad \tau \in \mathcal{T}_{to}, \qquad (2.98)$$
$$\sigma_\tau = \frac{c_\tau - g^{(q)}_{(u,y)}(u(\tau), y(\tau))(v_\delta(\tau^+), z_\delta(\tau))}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(q)}(u,y)|_{t=\tau^+}}, \quad \tau \in \mathcal{T}_{ex}, \qquad (2.99)$$
$$\sigma_\tau = \frac{c_\tau - g^{(q)}_{(u,y)}(u(\tau), y(\tau))(v_\delta(\tau^-), z_\delta(\tau))}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(q)}(u,y)|_{t=\tau^-}}, \quad \tau \in \mathcal{T}_{en}, \qquad (2.100)$$
$$\gamma_\tau^1 = \lambda^1_{\delta,\tau}, \quad \gamma_\tau^j = \lambda^j_{\delta,\tau} - \nu_\tau^{j-1}\sigma_\tau, \quad j = 2, \dots, q, \quad \tau \in \mathcal{T}_{en}. \qquad (2.101)$$

Note that $(v_\delta, \zeta_\delta, z_\delta, \pi_\delta)$ satisfies (2.81)-(2.85). It follows by $(\mathrm{A2}_q)$ and (2.10) that $v_\delta, \zeta_\delta \in PC^q_{\mathcal{T}}[0,T]$, and hence v_δ has limits when $t \to \tau^-$ and $t \to \tau^+$ for τ in respectively $\mathcal{T}_{en}$ and $\mathcal{T}_{ex}$, so (2.99)-(2.100) make sense.

Remark 2.30. Note that (2.91) is equivalent to the discontinuity of $\dot{u}$ at entry/exit points and that, when $q = 1, 2$, (2.60) implies (2.91), since $\hat{q} = q$.

Remark 2.31. The above proposition is an explicit elimination property, valid for any order $q \geq 1$, that enables us to express the solution ω of $D\mathcal{F}(\theta_0)\omega = \delta$ as a function of the optimal solution and multipliers of the quadratic problem $(\mathcal{P}^\delta)$, *independent of the variations of junction times.* In the case $q = 1$, the term in the factor of the variation of entry time σ_τ in (2.86) is zero so that Lemma 2.29 is nothing but the block decoupling property of the Jacobian already established in [93]. In the case $q \geq 2$, our result differs from the one in [94], since its authors use a quadratic problem depending on the variation of the entry point, leading to an additional assumption, (A.11).

Proof. Let $\delta \in \Theta$. By (i) and Lemma 2.28, Lemma 2.46 (with $r = 0$) implies that the first-order optimality system of $(\mathcal{P}^\delta)$ has a unique solution and multipliers. One can easily check that (2.81)-(2.85) and (2.95)-(2.97), together with (2.94) and

$$g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z_\delta(\mathcal{T}_{en}) = b^{1:q}_{\mathcal{T}_{en}}, \qquad g_y(y(\mathcal{T}_{to}))z_\delta(\mathcal{T}_{to}) = b_{\mathcal{T}_{to}}, \tag{2.102}$$

constitute the first-order optimality system of $(\mathcal{P}_\delta)$, with $\lambda^{1:q}_{\delta,\mathcal{T}_{en}}$ and $\lambda_{\delta,\mathcal{T}_{to}}$ the multipliers associated with (2.102), and thus have a unique solution $(v_\delta, z_\delta, \pi_\delta, \zeta_\delta, \lambda^{1:q}_{\delta,\mathcal{T}_{en}}, \lambda_{\delta,\mathcal{T}_{to}})$.

By (ii), define now $\sigma_\mathcal{T}$ by (2.98)-(2.100), and let $\gamma^{1:q}_{\mathcal{T}_{en}}$ and $\gamma_{\mathcal{T}_{to}}$ be related to $\lambda^{1:q}_{\delta,\mathcal{T}_{en}}$ and $\lambda_{\delta,\mathcal{T}_{to}}$ by the invertible relations (2.101) and $\gamma_{\mathcal{T}_{to}} = \lambda_{\delta,\mathcal{T}_{to}}$. Using (2.98) and (2.101) in respectively (2.97) and (2.95), it follows that the system of equations (2.81)-(2.85), (2.86)-(2.88), (2.94), (2.102), and (2.98)-(2.100) has a unique solution $(v_\delta, z_\delta, \pi_\delta, \zeta_\delta, \gamma^{1:q}_{\mathcal{T}_{en}}, \gamma_{\mathcal{T}_{to}}, \sigma_\mathcal{T})$. With Lemma 2.27, this implies that $D\mathcal{F}(\theta_0)\omega = \delta$ iff $\pi_0 = \pi_\delta(0)$, and the remaining variables of ω are determined by (2.98)-(2.101). Lipschitz continuity of ω w.r.t. δ is obtained as an easy consequence of Lemma 2.46 and the above relations. □

Proof of Theorem 2.23. The proof is organized as follows. We first show the sufficiency of the conditions (i) and (ii) for the well-posedness of the shooting algorithm, which is an easy consequence of the above lemmas. After that we show that (i), and then (ii), are also necessary.

Since (A5)(i) implies, by Lemma 2.21, that (2.60) holds, (2.91) is satisfied when $q = 1, 2$ (see Rem. 2.30) or trivially when the trajectory (u, y) has no boundary arc, i.e. $\mathcal{T}_{en} = \mathcal{T}_{ex} = \emptyset$. With (A6) and the second-order sufficient condition (2.69), the invertibility of the Jacobian of the shooting mapping follows from Prop. 2.29.

Let us now show the converse. Assume first that (i) does not hold, i.e. $q \geq 3$ and (u, y) has a boundary arc. By Prop. 2.5(i), $\dot{u}$ is continuous at junction times τ_{en} and τ_{ex}. Therefore, the function $\frac{\mathrm{d}}{\mathrm{d}t} g^{(q)}(u(t), y(t))$ depending on $(y, u, \dot{u})$ is also continuous at entry and exit times and vanishes on the boundary arc, so that (2.91) *does not hold*, at any of the regular entry/exit times. Then it is easily seen by Lemma 2.27 that we can find some nonzero $\tilde{\omega} \in \Theta$ such that $D\mathcal{F}(\theta_0)\tilde{\omega} = 0$. Indeed, take e.g. $\tilde{\sigma}_\tau \neq 0$ for $\tau \in \mathcal{T}_{ex}$, and all other components of $\tilde{\omega}$ equal to zero. It follows that the Jacobian of the shooting mapping is singular.

Assume now that (i) is satisfied but (ii) is not. Since (u, y) is a local solution of $(\mathcal{P})$, by Th. 2.22 the *second-order necessary condition* (2.68) is satisfied. This says that $(v, z) = 0$ is a solution of problem (PQ_q), therefore the value of (PQ_q) is zero, the infimum is attained, and solutions of this problem do exist. If $(v, z) = 0$ is not the unique solution, that is, if the second-order sufficient condition (2.69) does not hold, this means that there exists another optimal solution $(\tilde{v}_0, \tilde{z}_0) \neq 0$ of (PQ_q), and hence a nonzero solution of its first-order optimality conditions (2.63)-(2.66), (2.81)-(2.85), with final and jump conditions of the associated costate $\tilde{\pi}_0$ given by (2.94)-(2.97) with $a_T = 0$ and $c_{\mathcal{T}_{to}} = 0$, and multipliers $(\tilde{\lambda}^{1:q}_{\mathcal{T}_{en}}, \tilde{\lambda}_{\mathcal{T}_{to}})$ associated respectively with (2.64) and (2.66).

Setting $\tilde{\pi}_0 := \tilde{\pi}_0(0)$, we claim that $(\tilde{\pi}_0, \tilde{\lambda}^{1:q}_{\mathcal{T}_{en}}, \tilde{\lambda}_{\mathcal{T}_{to}}) \neq 0$. Indeed, suppose that all of them were zero. Eliminating v by (2.83) as a linear function of (z, π), and integrating from $(z(0), \pi(0)) = 0$ over the first arc the linear differential equations (2.81)-(2.82), we would have $(z, \pi, v, \zeta) = 0$, until the first junction time. If all the jump parameters $\tilde{\lambda}^j_{\mathcal{T}_{en}}$ and $\tilde{\lambda}_{\mathcal{T}_{to}}$ are equal to zero, and (v, ζ) is given by (2.83)-(2.84) on boundary arcs, we obtain $(\tilde{z}_0, \tilde{\pi}_0, \tilde{v}_0, \tilde{\zeta}_0) = 0$ over $[0, T]$, which leads to a contradiction.

Now let $\tilde{\gamma}_{\mathcal{T}_{to}} = \tilde{\lambda}_{\mathcal{T}_{to}}$ and $(\tilde{\sigma}_\mathcal{T}, \tilde{\gamma}^{1:q}_{\mathcal{T}_{en}})$ be solution of (2.98)-(2.101) with $c_\mathcal{T} = 0$. We have $\tilde{\omega} := (\tilde{\pi}_0, \tilde{\gamma}^{1:q}_{\mathcal{T}_{en}}, \tilde{\gamma}_{\mathcal{T}_{to}}, \tilde{\sigma}_{\mathcal{T}_{en}}, \tilde{\sigma}_{\mathcal{T}_{ex}}, \tilde{\sigma}_{\mathcal{T}_{to}}) \neq 0$, and by Lemma 2.27, $D\mathcal{F}(\theta_0)\tilde{\omega} = 0$. Therefore, the Jacobian of the shooting mapping is singular, which achieves the proof. □

2.4 Sensitivity analysis without strict complementarity at touch points

In this section, we show how to conduct a sensitivity analysis, removing the strict complementarity hypothesis for touch points.

Let us first note that our framework allows us to deal with nonautonomous problems (i.e. when the data f, ℓ, g depend on t) as well, by introducing an additional state variable equal to the time, provided that the data are sufficiently smooth with respect to t. When the original problem (2.1)-(2.3) is autonomous, we still can add the time as a state variable. This transformation affects neither the assumptions nor the first- and second-order optimality conditions in sections 2.2 and 2.3 and the condition (ii) in Th. 2.34. Therefore, we will assume w.l.o.g. throughout this section that the problem $(\mathcal{P})$ is written such that the last component of the state variable y_n satisfies

$$\dot{y}_n(t) = 1 \qquad \text{for all } t \in [0,T]; \qquad y_n(0) = 0$$

(i.e. $y_n(t) = t$, for all t). The reason for doing so is to consider in our stability analysis a wide class of perturbations, including nonautonomous perturbations (and possibly a nonautonomous original problem). Allowing nonautonomous perturbations is indeed needed to obtain the equivalence in Th. 2.34, even when the original problem is autonomous. We shall not repeat in this section this assumption, which intervenes only in the proof of (i) $\Rightarrow$ (ii) in Th. 2.34.

Let M_0 be an open subset of a Banach space M (the perturbation space). Consider, for $\mu \in M_0$, the family of perturbed optimal control problems

$$(\mathcal{P}^\mu) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \tilde{\ell}(u(t),y(t),\mu)\mathrm{d}t + \tilde{\phi}(y(T),\mu) \quad \text{subject to}$$
$$\dot{y} = \tilde{f}(u(t),y(t),\mu), \quad \text{a.e. } t\in[0,T]; \qquad y(0) = \tilde{y}_0(\mu),$$
$$\tilde{g}(y(t),\mu) \le 0 \quad \text{for all } t \in [0,T],$$

where $\tilde{\ell} : \mathbb{R}\times\mathbb{R}^n\times M_0 \to \mathbb{R}$, $\tilde{\phi} : \mathbb{R}^n \times M_0 \to \mathbb{R}$, $\tilde{f} : \mathbb{R}\times\mathbb{R}^n\times M_0 \to \mathbb{R}^n$, $\tilde{g} : \mathbb{R}^n \times M_0 \to \mathbb{R}$, and $\tilde{y}_0 : M_0 \to \mathbb{R}^n$ are at least C^2 mappings. We denote $y_0^\mu := \tilde{y}_0(\mu)$, $\ell^\mu(u,y) := \tilde{\ell}(u,y,\mu)$, etc., and identify $(\ell^\mu, \phi^\mu, f^\mu, g^\mu, y_0^\mu)$ with problem $(\mathcal{P}^\mu)$.

We say that $(\mathcal{P}^\mu)$ is a *q-stable extension* of $(\mathcal{P})$ if (i) there exists $\mu_0 \in M_0$ such that $(\mathcal{P}^{\mu_0}) = (\mathcal{P})$ (i.e. $\ell^{\mu_0} \equiv \ell$, etc.); (ii) the mappings $\tilde{\ell}, \tilde{\phi}, \tilde{f}, \tilde{g}$ are C^{2q}, where q is the order of the state constraint of problem $(\mathcal{P})$; (iii) the state constraints are of order q for all $\mu \in M_0$; and (iv) the mappings f^μ are Lipschitz continuous over $\mathbb{R}\times\mathbb{R}^n$, uniformly over $\mu \in M_0$.

For each $\mu \in M_0$, problem $(\mathcal{P}^\mu)$ satisfies (A0); taking if necessary a smaller neighborhood of μ_0, we may assume that (A1) holds as well. Given $(\mu,u,v) \in M_0\times\mathcal{U}\times\mathcal{V}$, denote by $(y_u^\mu, z_{u,v}^\mu) \in \mathcal{Y}\times\mathcal{Z}$ the state and linearized state solution of

$$\dot{y}_u^\mu = f^\mu(u,y_u^\mu); \qquad y_u^\mu(0) = y_0^\mu, \tag{2.103}$$
$$\dot{z}_{u,v}^\mu = f_y^\mu(u,y_u^\mu)z_{u,v}^\mu + f_u^\mu(u,y_u^\mu)v; \qquad z_{u,v}^\mu(0) = 0, \tag{2.104}$$

and let $J^\mu(u) := \int_0^T \ell^\mu(u(t),y_u^\mu(t))\mathrm{d}t + \phi^\mu(y_u^\mu(T))$.

In what follows, $(\bar{u},\bar{y})$ denotes a Pontryagin extremal of $(\mathcal{P}) \equiv (\mathcal{P}^{\mu_0})$, with associated multipliers $(\bar{p},\bar{\eta})$. We denote by $\theta_0 \in \Theta$ the vector of shooting parameters associated with $(\bar{u},\bar{y})$.

We say that a feasible trajectory (u, y) for $(\mathcal{P}^\mu)$ has a *neighboring structure* to that of $(\bar{u}, \bar{y})$ if the structure of (u, y) (number and order of boundary arcs and touch points) differs from that of $(\bar{u}, \bar{y})$ only by possibly removing some nonessential touch points. With a trajectory (u, y) having a neighboring structure to that of $(\bar{u}, \bar{y})$ is naturally associated a set of shooting parameters $\hat{\theta}$, but the latter may have a lower dimension than θ_0 if (u, y) has (strictly) less touch points than $(\bar{u}, \bar{y})$. We can show (and this is precisely the idea of reduction methods, see further) that when $\|u - \bar{u}\|_\infty$ and $\|\mu - \mu_0\|$ are small enough and $q \geq 2$, for every touch point τ_{to} of $(\bar{u}, \bar{y})$ satisfying (2.61), the function $g^\mu(y(\cdot))$ reaches its maximum over a small neighborhood of τ_{to} at a unique time denoted τ'_{to}. Then adding to $\hat{\theta}$ this time τ'_{to} and a zero jump parameter, and doing so for each touch point of $(\bar{u}, \bar{y})$ that is inactive for (u, y), we obtain an *augmented* vector of shooting parameters θ having the same dimension as θ_0. Therefore the following definition makes sense.

Definition 2.32. We say that the uniform second-order quadratic growth condition holds if for every q-stable extension $(\mathcal{P}^\mu)$, there exists $c > 0$ and open neighborhoods $V_\mu \times V_u \times V_\theta$ of $(\mu_0, \bar{u}, \theta_0)$ in $M_0 \times \mathcal{U} \times \Theta$, such that for all $\mu \in V_\mu$, there exists a unique stationary point $(u^\mu, y^\mu := y^\mu_{u^\mu}) \in V_u \times \mathcal{Y}$ of $(\mathcal{P}^\mu)$ having a neighboring structure to that of $(\bar{u}, \bar{y})$ with its augmented shooting parameters in V_θ, and that point satisfies

$$J^\mu(u) \geq J^\mu(u^\mu) + c\|u - u^\mu\|_2^2, \qquad \forall\, u \in V_u, \quad g^\mu(y^\mu_u) \leq 0 \ \text{ on } [0, T]. \tag{2.105}$$

As a consequence of the definition of the uniform growth condition, we have $\bar{u} = u^{\mu_0}$ and $\bar{y} = y^{\mu_0}$.

Note that in the uniform growth condition (2.105), the neighborhood (in L^∞) on which u^μ satisfies the quadratic growth condition is independent on μ. Our definition of uniform quadratic growth is different from the one in [24, section 5.1], since the latter implies the local uniqueness of solutions of the first-order optimality system (stationary points). Here, since our stability analysis is based on the shooting formulation, we can argue only the uniqueness of the stationary point among the feasible trajectories that have their structure and shooting parameters "in the neighborhood" of those of $(\bar{u}, \bar{y})$. The uniqueness of the stationary point, in a certain sense, is needed to prove the implication (i) $\Rightarrow$ (ii) in Th. 2.34 below.

We will use the assumption below, which is a modification of (A5)

(A5') **(i)** If $q \leq 2$, the following strengthening of (2.56)-(2.57) holds:

$$\exists \beta > 0 \qquad (-1)^q \frac{\mathrm{d}^q}{\mathrm{d}t^q} \bar{\eta}_q(t) \geq \beta \qquad \text{for all } t \in \operatorname{int} \mathcal{I}_b; \tag{2.106}$$

if $q = 2$, (2.58) holds; if $q > 2$, the trajectory $(\bar{u}, \bar{y})$ has no boundary arc;
(ii) If $q = 1$, $(\bar{u}, \bar{y})$ has no (nonessential) touch points.

Assumption (A5')(i) is a strengthening of (A5)(i). It requires, in addition to (A5)(i), *uniform* strict complementarity on boundary arcs, which is stronger than (2.56) (and implies (2.57)), and that $(\bar{u}, \bar{y})$ have no boundary arc if $q \geq 3$. Assumption (A5')(ii) is weaker than (A5)(ii) since it allows nonessential touch points for constraints of order $q \geq 2$ only.

Define the set of increasing times in $(0, T)$ of cardinal N as

$$IT_N := \{\tau \in \mathbb{R}^N;\ 0 < \tau_1 < \cdots < \tau_N < T\}. \tag{2.107}$$

Set $\tau_0 := 0$ and $\tau_{N+1} := T$. Given $\mathcal{S} \subset IT_N$, we have a natural isomorphism between $PC^k_{\mathcal{S}}[0, T]$ and $C^k([0, 1]; \mathbb{R}^{N+1})$, defined by

$$\left\{ \begin{array}{ll} \hat{\varphi}_i(s) = & \varphi(\tau_i + (\tau_{i+1} - \tau_i)s) \\ \hat{\varphi}_i(0) = & \varphi(\tau_i^+),\ \hat{\varphi}_i(1) = \varphi(\tau_{i+1}^-) \end{array} \right. \text{for all } \ s \in (0, 1), \qquad i = 0, \ldots, N. \tag{2.108}$$

We may therefore identify the set $PC_N^k[0,T] := \cup\{PC_{\mathcal{S}}^k[0,T]; \mathcal{S} \in IT_N\}$ of all possible N-piecewise k times continuously differentiable functions, with $C^k([0,1];\mathbb{R}^{N+1}) \times IT_N$. The corresponding notion of convergence follows: A sequence $\varphi^n \in PC_{\mathcal{S}^n}^k[0,T]$ converges to $\varphi \in PC_{\mathcal{S}}^k[0,T]$ if $\mathcal{S}^n \to \mathcal{S}$ in $\mathbb{R}^N$ and $\hat{\varphi}^n \to \hat{\varphi}$ in $C^k([0,1];\mathbb{R}^{N+1})$. Similarly, a mapping defined over an open subset W of a Banach space, $W \to PC_N^k$, $w \mapsto \varphi^w \in PC_{\mathcal{S}^w}^k$ is of class C^k if the mapping $W \to C^k([0,1];\mathbb{R}^{N+1}) \times \mathbb{R}^N$, $w \mapsto (\hat{\varphi}^w, \mathcal{S}_w)$ is C^k. We denote by $PC_N^{k,r}[0,T] = PC_N^k[0,T] \cap C^r[0,T]$ the subset of $PC_N^k[0,T]$ of functions having continuous derivatives on $[0,T]$ until order $r \geq 0$. The next lemma is elementary and will be used at the end of this section.

Lemma 2.33. *Let W be an open subset of a Banach space, and $W \to PC_N^{1,0}$, $w \mapsto \varphi^w \in PC_{\mathcal{S}^w}^{1,0}$ a C^1 mapping. Then the mapping $w \mapsto \varphi^w$ is C^1 in $L^r(0,T)$ for all $1 \leq r < \infty$. More precisely, for $w \in W$, let $\mathcal{S}^w := \{\tau_1^w < \cdots < \tau_N^w\}$ and denote by $(\hat{\xi}^w, \sigma^w)$ the directional derivative in $C^1([0,1];\mathbb{R}^{N+1}) \times IT_N$ of the mapping $w \mapsto (\hat{\varphi}^w, \tau^w)$ at point w in direction $\delta w \in W$. Then the directional derivative $\tilde{\xi}^w$ in $L^r(0,T)$ is given by*

$$\tilde{\xi}^w(t) = \hat{\xi}_i^w \left(\frac{t - \tau_i^w}{\tau_{i+1}^w - \tau_i^w} \right) - \dot{\varphi}^w(t) \left(\sigma_i^w + \frac{t - \tau_i^w}{\tau_{i+1}^w - \tau_i^w} (\sigma_{i+1}^w - \sigma_i^w) \right) \text{ on } (\tau_i^w, \tau_{i+1}^w).$$

By Prop. 2.5, a regular Pontryagin extremal and its multipliers $(u^\mu, y^\mu, p^\mu, \eta^\mu)$ satisfying (A2)-(A4) belong to the product space

$$\mathcal{X}_{\mathcal{S}} := PC_{\mathcal{S}}^{q,0}[0,T] \times PC_{\mathcal{S}}^{q+1,1}([0,T];\mathbb{R}^n) \times PC_{\mathcal{S}}^1([0,T];\mathbb{R}^{n*}) \times PC_{\mathcal{S}}^1[0,T], \tag{2.109}$$

with here $\mathcal{S} = \mathcal{T}$, which is the finite set of its junction times assumed to be of cardinal N. So let us define the union $\mathcal{X}_N$ of all such spaces, and define as well some other sets needed later:

$$\begin{aligned}
\mathcal{X}_N &:= \cup\{\mathcal{X}_{\mathcal{S}};\ \mathcal{S} \in IT_N\},\\
\mathcal{X}_{\mathcal{S}}^q &:= PC_{\mathcal{S}}^q[0,T] \times PC_{\mathcal{S}}^{q+1,0}([0,T];\mathbb{R}^n) \times PC_{\mathcal{S}}^{q+1}([0,T];\mathbb{R}^{n*}) \times PC_{\mathcal{S}}^q[0,T],\\
\mathcal{X}_{\mathcal{S}}^1 &:= PC_{\mathcal{S}}^q[0,T] \times PC_{\mathcal{S}}^{q+1,0}([0,T];\mathbb{R}^n) \times PC_{\mathcal{S}}^1([0,T];\mathbb{R}^{n*}) \times PC_{\mathcal{S}}^1[0,T],\\
\mathcal{X}_N^q &:= \cup\{\mathcal{X}_{\mathcal{S}}^q;\ \mathcal{S} \in IT_N\}, \qquad \mathcal{X}_N^1 := \cup\{\mathcal{X}_{\mathcal{S}}^1;\ \mathcal{S} \in IT_N\}.
\end{aligned}$$

The main result of this section is the next theorem, which gives stability results for the optimal control problem $(\mathcal{P})$ without assuming strict complementarity at touch points. Therefore we cannot directly apply the implicit function theorem as it was done in [93, 94] and in our section 2.3.

Theorem 2.34. *Let $(\bar{u}, \bar{y})$ be a Pontryagin extremal of $(\mathcal{P})$ satisfying (A2)-(A4), (A5'), and (A6). Then the following statements are equivalent:*
(i) *The uniform second-order quadratic growth condition (Def. 2.32) holds. Denote by $u^\mu \in V_u$ the solution of (2.105) for $\mu \in V_\mu$, and set $y^\mu := y_{u^\mu}^\mu$. With (u^μ, y^μ) are associated a unique costate p^μ and state constraint multiplier η^μ, and the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu) \in \mathcal{X}_N$ is Lipschitz continuous over V_μ.*
(ii) *The following strong second-order sufficient condition holds:*

$$\begin{aligned}
\mathcal{J}(v,z) > 0, \quad &\text{for all } (v,z) \in \mathcal{V} \times \mathcal{Z} \setminus \{0\} \text{ satisfying (2.63) and}\\
&g_y(\bar{y}(t))z(t) = 0 \quad \text{for all } t \in \mathcal{I}_b \cup \mathcal{T}_{to}^{ess}.
\end{aligned} \tag{2.110}$$

Remark 2.35. Note that condition (ii) is stronger than the following second-order characterization of quadratic growth (2.55) (see [21]):

$$\begin{aligned}\mathcal{J}(v,z) > 0, \quad &\text{for all } (v,z) \in \mathcal{V} \times \mathcal{Z} \setminus \{0\} \text{ satisfying (2.63), (2.110) and} \\ &g_y(\bar{y}(\tau))z(\tau) \leq 0 \quad \text{for all } \tau \in \mathcal{T}_{to} \setminus \mathcal{T}_{to}^{ess}.\end{aligned}$$

We need the following notation. Denote by $\mathcal{T}_{to}^{nes} := \mathcal{T}_{to} \setminus \mathcal{T}_{to}^{ess}$ the subset of *nonessential touch points* of the trajectory $(\bar{u}, \bar{y})$. For μ close to μ_0, let $\mathcal{F}(\cdot, \mu)$ be the shooting mapping (2.50) for problem $(\mathcal{P}^\mu)$, with the *same structure* as the trajectory $(\bar{u}, \bar{y})$, i.e. the same number of boundary arcs and touch points and the same order of their occurrence w.r.t. time. Thus nonessential touch points are present in the shooting mapping and may be active or inactive for the perturbed problem. Let $\bar{N} := n + (q+2)N_b + 2N_{to}$ denote the dimension of the shooting mapping, with $N_b = \operatorname{Card} \mathcal{T}_{en} = \operatorname{Card} \mathcal{T}_{ex}$ and $N_{to} = \operatorname{Card} \mathcal{T}_{to}$, and denote by N_0 the cardinal of $\mathcal{T}_{to}^{nes}$, the set of nonessential touch points. Split $\mathcal{F}$ into two components such that $\mathcal{F}(\cdot, \mu) = (\Phi(\cdot, \mu)^*, \Psi(\cdot, \mu)^*)^*$ and Ψ corresponds to the component $g^\mu(y(\mathcal{T}_{to}^{nes})) \in \mathbb{R}^{N_0}$. We consider the following problem for μ close to μ_0: Find

$$\theta = (p_0^{\mu *}, \nu_{\mathcal{T}_{en}}^{\mu,1:q}, \nu_{\mathcal{T}_{to}^{ess}}^{\mu}, \nu_{\mathcal{T}_{to}^{nes}}^{\mu}, \mathcal{T}_{en}^{\mu}, \mathcal{T}_{ex}^{\mu}, \mathcal{T}_{to}^{\mu,ess}, \mathcal{T}_{to}^{\mu,nes}) \in \Theta \tag{2.111}$$

such that

$$\Phi(\theta, \mu) = 0; \qquad \Psi(\theta, \mu) \in \mathbb{R}_-^{N_0} \cap (\nu_{\mathcal{T}_{to}^{nes}}^{\mu})^\perp; \qquad \nu_{\mathcal{T}_{to}^{nes}}^{\mu} \in \mathbb{R}_+^{N_0}. \tag{2.112}$$

In (2.112), we express the complementarity condition for nonessential touch points only. The complementarity condition at essential touch points and boundary arcs, where strict complementarity is satisfied, will hold by continuity, since we perform a local analysis (see further Lemmas 2.37-2.38).

The point θ_0, solution of (2.112) for $\mu = \mu_0$, is said to be *strongly regular* (see Robinson [121]), if there exists a neighborhood $V_\theta' \times V_\delta$ in $\mathbb{R}^{\bar{N}} \times \mathbb{R}^{\bar{N}}$ of $(\theta_0, 0)$ such that for all $\delta \in V_\delta$, $\delta = (\delta_1, \delta_2) \in \mathbb{R}^{\bar{N}-N_0} \times \mathbb{R}^{N_0}$, there exists a unique solution θ in V_θ' of:

$$\begin{aligned}&D_\theta \Phi(\theta_0, \mu_0)(\theta - \theta_0) - \delta_1 = 0 \\ &D_\theta \Psi(\theta_0, \mu_0)(\theta - \theta_0) - \delta_2 \in \mathbb{R}_-^{N_0} \cap \nu_{\mathcal{T}_{to}^{nes}}^{\perp}; \qquad \nu_{\mathcal{T}_{to}^{nes}} \in \mathbb{R}_+^{N_0},\end{aligned} \tag{2.113}$$

and the mapping $\Xi : \delta \mapsto \theta(\delta)$ is Lipschitz continuous over V_δ. If θ_0 is strongly regular, then by [121], there exists a neighborhood $V_\theta \times V_\mu$ of (θ_0, μ_0), such that for each $\mu \in V_\mu$, (2.112) has in V_θ a unique solution θ^μ and there exists $\kappa > 0$ such that for all $\mu, \mu' \in V_\mu$,

$$|\theta^\mu - \theta^{\mu'}| \leq \kappa \|\mu - \mu'\|. \tag{2.114}$$

In addition, the following expansion of θ^μ holds (see [24, p.413] eq. (5.41)):

$$\theta^\mu = \Xi(-D_\mu \mathcal{F}(\theta_0, \mu_0)(\mu - \mu_0)) + o(\|\mu - \mu_0\|). \tag{2.115}$$

2.4.1 Stability Analysis (Proof of Th. 2.34)

The first step in the proof of (ii) $\Rightarrow$ (i) in Th. 2.34 is to show that (ii) implies the strong regularity property (Lemma 2.36). The existence of a (locally unique) shooting extremal (u^μ, y^μ) for problem $(\mathcal{P}^\mu)$ having its shooting parameters in the neighborhood of those of $(\bar{u}, \bar{y})$ follows (Lemma 2.37). The next step is to check the additional conditions of Cor. 2.17, implying that (u^μ, y^μ) is a stationary point (Lemma 2.38). We end the proof by checking that u^μ satisfies the uniform quadratic growth condition (2.105) (Lemmas 2.39-2.40).

Lemma 2.36. *Under the assumptions of Th. 2.34, condition (ii) of Th. 2.34 implies that θ_0 is a strongly regular solution of (2.112) for $\mu = \mu_0$.*

Proof. The proof is somewhat similar to that of Proposition 2.29. Let $\delta = (\delta_1, \delta_2) \in \mathbb{R}^{\bar{N}-N_0} \times \mathbb{R}^{N_0}$ with

$$\delta_1 = (a_T, b^{1:q}_{\mathcal{T}_{en}}, b_{\mathcal{T}_{to}^{ess}}, c_{\mathcal{T}_{en}}, c_{\mathcal{T}_{ex}}, c_{\mathcal{T}_{to}^{ess}}, c_{\mathcal{T}_{to}^{nes}}); \qquad \delta_2 = b_{\mathcal{T}_{to}^{nes}}.$$

Let us show that there exists a unique $\omega \in \Theta$,

$$\omega = (\pi_0^*, \gamma^{1:q}_{\mathcal{T}_{en}}, \gamma_{\mathcal{T}_{to}^{ess}}, \gamma_{\mathcal{T}_{to}^{nes}}, \sigma_{\mathcal{T}_{en}}, \sigma_{\mathcal{T}_{ex}}, \sigma_{\mathcal{T}_{to}^{ess}}, \sigma_{\mathcal{T}_{to}^{nes}}),$$

solution of the following relation, equivalent to (2.113) with $\omega = \theta - \theta_0$:

$$\begin{array}{l} D_\theta \Phi(\theta_0, \mu_0)\omega - \delta_1 = 0, \\ D_\theta \Psi(\theta_0, \mu_0)\omega - \delta_2 \in \mathbb{R}^{N_0}_- \cap \gamma^\perp_{\mathcal{T}_{to}^{nes}}; \qquad \gamma_{\mathcal{T}_{to}^{nes}} \in \mathbb{R}^{N_0}_+. \end{array} \tag{2.116}$$

Consider the following linear quadratic optimal control problem:

$$\begin{array}{lll} (\mathcal{P}^\delta) & \min\limits_{v \in \mathcal{V}} & \dfrac{1}{2}\mathcal{J}_q(v, z_v) \ + \ a_T^* z_v(T) \ + \ \displaystyle\sum_{\tau \in \mathcal{T}_{to}} c_\tau \nu_\tau \frac{g_y^{(1)}(y(\tau))z_v(\tau)}{\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(y)|_{t=\tau}} \\ & \text{subject to} & Av \ = \ (0_{L_2(\mathcal{I}_b)}, b^{1:q}_{\mathcal{T}_{en}}, b_{\mathcal{T}_{to}^{ess}})^*; \ \ Bv \ \le \ b_{\mathcal{T}_{to}^{nes}}, \end{array} \tag{2.117}$$

where $\mathcal{J}_q(v, z_v)$ is defined by (2.62) and the linear operators A, B are defined by

$$\begin{array}{rcl} Av & := & \begin{pmatrix} (g_y^{(q)}(u(\cdot), y(\cdot))z_v(\cdot) + g_u^{(q)}(u(\cdot), y(\cdot))v(\cdot))|_{\mathcal{I}_b} \\ g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z_v(\mathcal{T}_{en}) \\ g_y(y(\mathcal{T}_{to}^{ess}))z_v(\mathcal{T}_{to}^{ess}) \end{pmatrix} \\ Bv & := & g_y(y(\mathcal{T}_{to}^{nes}))z_v(\mathcal{T}_{to}^{nes}). \end{array} \tag{2.118}$$

Being equal to $\mathcal{A}$ defined in (2.89), the operator (A, B) is onto by Lemma 2.28. By Lemma 2.45, the Legendre form $\bar{Q}(v) := \mathcal{J}_q(v, z_v)$ is coercive over $\operatorname{Ker} A$. It follows from Lemma 2.46 that the first-order optimality system of problem $(\mathcal{P}^\delta)$ has a unique solution $v_\delta \in \mathcal{V}$, with a unique associated Lagrange multiplier $(\zeta_\delta, \lambda^{1:q}_{\delta,\mathcal{T}_{en}}, \lambda_{\delta,\mathcal{T}_{to}^{ess}}, \lambda_{\delta,\mathcal{T}_{to}^{nes}})$ in $L^2(\mathcal{I}_b) \times \mathbb{R}^{qN_b} \times \mathbb{R}^{N_{to}-N_0} \times \mathbb{R}^{N_0}$, and the mapping $\delta \mapsto (v_\delta, \zeta_\delta, \lambda^{1:q}_{\delta,\mathcal{T}_{en}}, \lambda_{\delta,\mathcal{T}_{to}^{ess}}, \lambda_{\delta,\mathcal{T}_{to}^{nes}})$ is Lipschitz continuous. Now, defining as in Prop. 2.29 $\sigma_\mathcal{T}$ by (2.98)-(2.100) and defining $\gamma^{1:q}_{\mathcal{T}_{en}}$, $\gamma_{\mathcal{T}_{to}^{ess}}$, $\gamma_{\mathcal{T}_{to}^{nes}}$ by the invertible relations (2.101), $\gamma_{\mathcal{T}_{to}^{ess}} = \lambda_{\delta,\mathcal{T}_{to}^{ess}}$ and $\gamma_{\mathcal{T}_{to}^{nes}} = \lambda_{\delta,\mathcal{T}_{to}^{nes}}$, this implies that the system of equations (2.81)-(2.85), (2.86)-(2.88), (2.94), (2.98)-(2.100), together with the constraints and complementarity conditions of $(\mathcal{P}^\delta)$

$$\begin{array}{c} g_y^{(0:q-1)}(y(\mathcal{T}_{en}))z_\delta(\mathcal{T}_{en}) = b^{1:q}_{\mathcal{T}_{en}}, \qquad g_y(y(\mathcal{T}_{to}^{ess}))z_\delta(\mathcal{T}_{to}^{ess}) = b_{\mathcal{T}_{to}^{ess}}, \\ g_y(y(\mathcal{T}_{to}^{nes}))z_\delta(\mathcal{T}_{to}^{nes}) \le b_{\mathcal{T}_{to}^{nes}}, \ \ \gamma_{\mathcal{T}_{to}^{nes}} \ge 0, \ \ (g_y(y(\mathcal{T}_{to}^{nes}))z_\delta(\mathcal{T}_{to}^{nes}) - b_{\mathcal{T}_{to}^{nes}}) \perp \gamma_{\mathcal{T}_{to}^{nes}}, \end{array}$$

has a unique solution $(v_\delta, z_\delta, \pi_\delta, \zeta_\delta, \gamma^{1:q}_{\mathcal{T}_{en}}, \gamma_{\mathcal{T}_{to}^{ess}}, \gamma_{\mathcal{T}_{to}^{nes}}, \sigma_\mathcal{T})$. Thus by Lemma 2.27, we obtain that ω is a solution of (2.116) iff $\pi_0 = \pi_\delta(0)$ and the other variables of ω are given as above. The existence and uniqueness of ω follows, and it is not difficult to check the Lipschitz continuity of ω w.r.t. δ. □

By strong regularity, there exist neighborhoods V_μ and V_θ of μ_0 and θ_0 such that, for all $\mu \in V_\mu$, there exists in V_θ a unique solution θ^μ of (2.112):

$$\theta^\mu = (p_0^{\mu *}, \nu^{\mu,1:q}_{\mathcal{T}_{en}}, \nu^{\mu}_{\mathcal{T}^{ess}_{to}}, \nu^{\mu}_{\mathcal{T}^{nes}_{to}}, \mathcal{T}^\mu_{en}, \mathcal{T}^\mu_{ex}, \mathcal{T}^{\mu,ess}_{to}, \mathcal{T}^{\mu,nes}_{to}) \in V_\theta \subset \mathbb{R}^{\bar N}.$$

Denote the associated trajectory and multipliers by $(u^\mu, y^\mu, p^\mu_q, \eta^\mu_q) \in \mathcal{X}^q_N$. Recall that $\Psi(\theta^\mu, \mu) = g^\mu(y^\mu(\mathcal{T}^{\mu,nes}_{to}))$ and set

$$\mathcal{T}^\mu_{to} := \mathcal{T}^{\mu,ess}_{to} \cup \{\tau \in \mathcal{T}^{\mu,nes}_{to} \ ; \ g^\mu(y^\mu(\tau)) = 0\}.$$

By the definition of (2.112), we have that $g^\mu(y^\mu(\tau)) < 0$ and $\nu^\mu_\tau = 0$ if $\tau \notin \mathcal{T}^\mu_{to}$. Hence $(u^\mu, y^\mu, p^\mu_q, \eta^\mu_q)$ is a shooting extremal for $(\mathcal{P}^\mu)$, with jump parameters $(\nu^{\mu,1:q}_{\mathcal{T}_{en}}, \nu^\mu_{\mathcal{T}^\mu_{to}})$ and junction times $(\mathcal{T}^\mu_{en}, \mathcal{T}^\mu_{ex}, \mathcal{T}^\mu_{to})$.

In order to show now that the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu)$ is Lipschitz continous, where (p^μ, η^μ) is given by (2.36)-(2.38) and (2.21), consider the mapping

$$V_\mu \times V_\theta \to \mathcal{X}^q_N, \qquad (\mu, \theta) \mapsto (u^{\mu,\theta}, y^{\mu,\theta}, p^{\mu,\theta}_q, \eta^{\mu,\theta}_q), \tag{2.119}$$

where $(u^{\mu,\theta}, y^{\mu,\theta}, p^{\mu,\theta}_q, \eta^{\mu,\theta}_q)$ is the solution of (2.23)-(2.24), (2.26), (2.28), (2.30), and (2.31)-(2.33) for $(\mathcal{P}^\mu)$, with initial value of the costate, jump parameters and junction times given by argument θ. By the Cauchy-Lipschitz theorem, this mapping is well-defined and of class C^q on neighborhoods $V_\mu \times V_\theta$ of (μ_0, θ_0). Therefore the mapping

$$V_\mu \times V_\theta \to \mathcal{X}^1_N, \qquad (\mu, \theta) \mapsto (u^{\mu,\theta}, y^{\mu,\theta}, p^{\mu,\theta}, \eta^{\mu,\theta}), \tag{2.120}$$

where $\eta^{\mu,\theta}_j$, $0 \le j \le q-1$, $p^{\mu,\theta}$, and $\eta^{\mu,\theta}$ are defined by (2.36)-(2.38) and (2.21), is of class C^1.

Lemma 2.37. *Under assumptions and condition (ii) of Th. 2.34, there exists a neighborhood V_μ of μ_0 such that the mapping $V_\mu \to \mathcal{X}_N$, $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu)$ is well-defined and Lipschitz continuous on V_μ.*

Proof. Since strong regularity holds by Lemma 2.36, the mapping $\mu \mapsto \theta^\mu$ solution of (2.112) is well-defined on a neighborhood of μ and Lipschitz continuous by (2.114). By continuity of the mappings (2.120) and $\mu \mapsto \theta^\mu$, the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu)$ is continuous $V_\mu \to \mathcal{X}^1_N$. Let us show now that u^μ is continuous. By (A2)-(A3), reducing V_μ if necessary, we have $H^\mu_{uu}(\hat u, y^\mu(t), p^\mu(t^\pm)) \ge \alpha/2$ and $|(g^\mu)^{(q)}_u(\hat u, y^\mu(t))| \ge \gamma/2$ for all t and all $\hat u$ in the segment $[u^\mu(t^-), u^\mu(t^+)] := \{\sigma u^\mu(t^+) + (1-\sigma)u(t^-), \ \sigma \in [0,1]\}$. By arguments similar to those used in the proof of Prop. 2.15(i) and in Rem. 2.16, this is enough to show that u^μ is continuous, and hence, $(u^\mu, y^\mu) \in PC^{q,0}_{\mathcal{T}^\mu}[0,T] \times PC^{q+1,1}_{\mathcal{T}^\mu}([0,T]; \mathbb{R}^n)$. Reducing V^μ if necessary, by composition of $\mu \mapsto \theta^\mu$ with the C^1-mapping (2.120), we deduce that the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu) \in \mathcal{X}_N$ is Lipschitz continuous on a neighborhood of μ. □

Lemma 2.38. *Under assumptions and condition (ii) of Th. 2.34, the shooting extremal (u^μ, y^μ) is a stationary point for problem $(\mathcal{P}^\mu)$.*

Proof. By Corollary 2.17 and Rem. 2.20, we need to check (2.39), (2.40), (2.43), and also, when $q = 2$, (2.51). By (A5') and Lemma 2.37, (2.40) follows from (2.106). If $q = 2$, (2.51) follows from (2.58). By continuity of jumps at essential touch points and the definition of (2.112), we obtain (2.43). It remains to prove (2.39). Near an entry/exit point τ^μ (when $q = 1$ or 2) this is a consequence of hypothesis (2.60) and continuity w.r.t. μ of $u(\tau^{\mu\pm})$. Similarly, near touch points, this follows from the reducibility hypothesis (2.61). Finally, outside a small neighborhood of contact points, we obtain that $g^\mu(y^\mu) < 0$ by a standard compactness argument. □

The next two lemmas extend those in [21, section 4][3] to the setting of perturbed optimal control problems. In what follows we denote by $\mathrm{supp}(\mathrm{d}\eta)$ the support of the measure η in $\mathcal{M}[0,T]$.

Lemma 2.39. *Assume that the assumptions and condition (ii) of Th. 2.34 hold. Let $(\mathcal{P}^\mu)$ be a q-stable extension, and $\mu_n \to \mu_0$ with its associated shooting extremal (u_n, y_n) and multipliers (p_n, η_n). For $v \in \mathcal{V}$, define $Q^n(v) := \mathcal{J}^{\mu_n}(v, z^{\mu_n}_{u_n,v})$, where $\mathcal{J}^{\mu_n}(\cdot,\cdot)$ is given by (2.70) for $(\mathcal{P}^{\mu_n})$ and $z^{\mu_n}_{u_n,v}$ is defined by (2.104). Define similarly $\bar{Q}(v) := \mathcal{J}^{\mu_0}(v, z^{\mu_0}_{\bar{u},v})$. Let $v_n \rightharpoonup \bar{v} \in L^2$. Then it holds that*

$$\bar{Q}(\bar{v}) \;\le\; \liminf Q^n(v_n) \quad \textit{and} \quad v_n \to \bar{v} \textit{ strongly if } Q^n(v_n) \to \bar{Q}(\bar{v}). \tag{2.121}$$

Set $z_n := z^{\mu_n}_{u_n,v_n}$, and assume in addition that $g^{\mu_n}_y(y_n(t))z_n(t) \le r_n$, where $\|r_n\|_\infty \to 0$, for all $t \in \mathrm{supp}(\mathrm{d}\eta_n)$ and all n. Let $\bar{z} := z^{\mu_0}_{\bar{u},\bar{v}}$. Then

$$g_y(\bar{y}(t))\bar{z}(t) \;\le\; 0 \quad \textit{on } \mathrm{supp}(\mathrm{d}\bar{\eta}). \tag{2.122}$$

Proof. Since by Lemma 2.37, (u_n, y_n) converges uniformly to $(\bar{u}, \bar{y})$, and $v_n \rightharpoonup v$, we have that (z_n) converges weakly in H^1 to $\bar{z}$, and hence uniformly. Relation (2.122) follows from the convergence of η_n in PC^1_N, strict complementarity (2.106), and uniform convergence of $g^{\mu_n}_y(y_n)z_n$. Let us now show (2.121).

Set $Q^0_n(v_n) := \int_0^T v_n^* H^{\mu_n}_{uu}(u_n, y_n, p_n)v_n \mathrm{d}t$. By Lemma 2.37, uniform convergence of z_n, and convergence in $\mathcal{X}_N$ of $H^{\mu_n}_{uy}(u_n, y_n, p_n)$ and $H^{\mu_n}_{yy}(u_n, y_n, p_n)$, it follows easily that $Q_n(v_n) - Q^0_n(v_n) \to \bar{Q}(\bar{v}) - \bar{Q}^0(\bar{v})$. Writing $Q^0_n(v_n) = \bar{Q}^0(v_n) + \epsilon_n$ with $\epsilon_n = \int_0^T v_n^*(H^{\mu_n}_{uu}(u_n, y_n, p_n) - H_{uu}(\bar{u}, \bar{y}, \bar{p}))v_n \mathrm{d}t$, by continuity of $H^{\mu_n}_{uu}$ at junction times (Lemma 2.43 and Rem. 2.20), Lemma 2.37 implies that $H^{\mu_n}_{uu}(u_n, y_n, p_n) \to H_{uu}(\bar{u}, \bar{y}, \bar{p})$ uniformly, and hence, $\epsilon_n \to 0$. Since by (A2), $\bar{Q}^0 : v \mapsto \int_0^T v^* H_{uu}(\bar{u}, \bar{y}, \bar{p})v$ is a Legendre form, (2.121) follows. □

We recall the reduction approach of [21, section 5.2][4]. When $q \ge 2$, with all touch points of the trajectory $(\bar{u}, \bar{y})$ being reducible by (A6), let $\varepsilon, \delta > 0$ and V_μ be small enough so that, for all $\|u - \bar{u}\|_\infty \le \delta$, all $\mu \in V_\mu$ and all $\tau_{to} \in \mathcal{T}_{to}$, the function $g^\mu(y^\mu_u)$ attains its maximum over $[\tau_{to} - \varepsilon, \tau_{to} + \varepsilon]$ at a unique point $\tau^\mu_u \in (\tau_{to} - \varepsilon, \tau_{to} + \varepsilon)$. Set $\bar{I}_{to} := \cup_{\tau_{to} \in \mathcal{T}_{to}}(\tau_{to} - \varepsilon, \tau_{to} + \varepsilon)$ and $\bar{I}_b := [0,T] \setminus \bar{I}_{to}$. When $q = 1$, set $\bar{I}_b := [0,T]$ and $\bar{I}_{to} := \emptyset$. Then the following *reduced problem* is well-defined and locally equivalent to $(\mathcal{P}^\mu)$:

$$\begin{aligned} (\mathcal{P}^\mu_{red}) \qquad & \min_{u \in B_\infty(\bar{u},\delta)} J^\mu(u) \qquad \text{subject to} \\ & \mathcal{G}^\mu(u) := \begin{pmatrix} g(y_u)|_{\bar{I}_b} \\ g^\mu(y^\mu_u(\tau^{\mu,1}_u)) \\ \vdots \\ g^\mu(y^\mu_u(\tau^{\mu,N_{to}}_u)) \end{pmatrix} \in \mathcal{K} := C_-[\bar{I}_b] \times \mathbb{R}^{N_{to}}_-. \end{aligned} \tag{2.123}$$

The Lagrangian $\mathcal{L}^\mu$ of the reduced problem (2.123) is given, for $u \in B_\infty(\bar{u}, \delta)$ and a multiplier $\lambda = (\eta_b, \nu) \in \mathcal{M}_+[\bar{I}_b] \times \mathbb{R}^{N_{to}}_+$, by

$$\mathcal{L}^\mu(u, \lambda) = J^\mu(u) + \int_{\bar{I}_b} g^\mu(y^\mu_u(t))\mathrm{d}\eta_b(t) + \sum_{i=1}^{N_{to}} \nu_i g^\mu(y^\mu_u(\tau^{\mu,i}_u)). \tag{2.124}$$

[3] Section 1.4 of this thesis.
[4] Section 1.5.2 of this thesis.

Multipliers η^μ and $\lambda^\mu = (\eta_b^\mu, \nu^\mu)$ associated with u^μ in respectively problem $(\mathcal{P}^\mu)$ and its reduced form $(\mathcal{P}^\mu_{red})$, are related by

$$\mathrm{d}\eta^\mu(t) = \mathrm{d}\eta_b^\mu(t) \text{ on } \bar{I}_b; \qquad \mathrm{d}\eta^\mu(t) = \sum_{i=1}^{N_{to}} \nu^\mu_{\tau_i} \delta_{\tau_u^{\mu,i}}(t) \text{ on } \bar{I}_{to}. \tag{2.125}$$

In addition, we can show that the reduced Lagrangian (2.124) is twice Fréchet differentiable at u^μ, and its second-order derivative satisfies, for $v \in \mathcal{V}$,

$$D^2_{uu}\mathcal{L}^\mu(u^\mu, \lambda^\mu)(v, v) = \mathcal{J}^\mu(v, z^\mu_{u,v}), \tag{2.126}$$

with $\mathcal{J}^\mu$ given by (2.70), and that the remainder $r(v)$ in the second-order expansion

$$\mathcal{L}^\mu(u^\mu + v, \lambda^\mu) = \mathcal{L}^\mu(u^\mu, \lambda^\mu) + D_u\mathcal{L}^\mu(u^\mu, \lambda^\mu)v + \tfrac{1}{2}D^2_{uu}\mathcal{L}^\mu(u^\mu, \lambda^\mu)(v, v) + r(v)$$

satisfies

$$r(v)/\|v\|_2^2 \;\to\; 0 \;\text{ when }\; \|v\|_\infty \;\to\; 0. \tag{2.127}$$

In what follows, $T_{\mathcal{K}}(x)$ and $N_{\mathcal{K}}(x)$ denote respectively the tangent and normal cones to $\mathcal{K}$ at point $x \in \mathcal{K}$ (in the sense of convex analysis).

Lemma 2.40. *Under assumptions and condition (ii) of Th. 2.34, there exists an open neighborhood V_μ of μ_0 such that the shooting extremal (u^μ, y^μ) associated with $(\mathcal{P}^\mu)$ for $\mu \in V_\mu$ satisfies the uniform quadratic growth condition, and hence, is a local solution of $(\mathcal{P}^\mu)$.*

Proof. If the conclusion does not hold, then there exists a q-stable extension $(\mathcal{P}^\mu)$, a sequence $\mu_n \to \mu_0$, with associated shooting extremal and multipliers (u_n, y_n, p_n, η_n) converging to $(\bar{u}, \bar{y}, \bar{p}, \bar{\eta})$ in $\mathcal{X}_N$ by Lemma 2.37 (which implies in particular $u_n \to \bar{u}$ in L^∞), and a point $\tilde{u}_n \in \mathcal{U}$ feasible for (P^{μ_n}), $\tilde{u}_n \neq u_n$, $\tilde{u}_n \to \bar{u}$ in L^∞, satisfying for all n,

$$J^{\mu_n}(\tilde{u}_n) \leq J^{\mu_n}(u_n) + o(\|\tilde{u}_n - u_n\|_2^2). \tag{2.128}$$

Since $\lambda_n \in N_{\mathcal{K}}(\mathcal{G}^{\mu_n}(u_n))$, we have (for the appropriate duality products)

$$\langle \lambda_n, \mathcal{G}^{\mu_n}(\tilde{u}_n) - \mathcal{G}^{\mu_n}(u_n) \rangle \;\leq\; 0,$$

and thus

$$\mathcal{L}^{\mu_n}(\tilde{u}_n, \lambda_n) - \mathcal{L}^{\mu_n}(u_n, \lambda_n) \;\leq\; o(\|\tilde{u}_n - u_n\|_2^2). \tag{2.129}$$

Let $0 < \varepsilon_n := \|\tilde{u}_n - u_n\|_2 \to 0$ and $v_n := \varepsilon_n^{-1}(\tilde{u}_n - u_n)$. Since $\|v_n\|_2 = 1$ for all n, taking a subsequence if necessary, we may assume that $v_n \rightharpoonup \bar{v} \in \mathcal{V}$. With the notation of Lemma 2.39, we deduce from this lemma that (2.121) holds. Combining $D_u\mathcal{L}^{\mu_n}(u_n, \lambda_n) = 0$ and (2.126) with (2.129) and (2.127), we get

$$Q^n(v_n) = D_{uu}\mathcal{L}^{\mu_n}(u_n, \lambda_n)(v_n, v_n) \;\leq\; o(1), \tag{2.130}$$

and thus $\bar{Q}(\bar{v}) \leq 0$ by (2.121). Now

$$\mathcal{K} \ni \mathcal{G}^{\mu_n}(\tilde{u}_n) = \mathcal{G}^{\mu_n}(u_n) + \varepsilon_n D\mathcal{G}^{\mu_n}(u_n)v_n + \varepsilon_n r_n,$$

where $\|r_n\|_\infty = o(1)$, and therefore $D\mathcal{G}^{\mu_n}(u_n)v_n + r_n \in T_{\mathcal{K}}(\mathcal{G}^{\mu_n}(u_n))$, implying $g_y^{\mu_n}(y_n)z_n + r_n \leq 0$ on $\mathrm{supp}(\mathrm{d}\eta_n)$. Thus (2.122) is satisfied by Lemma 2.39. Also, by (2.128), $DJ^{\mu_n}(u_n)v_n \leq o(1)$, and hence,

$$\langle \eta_n, g_y^{\mu_n}(y_n)z_n \rangle = \langle \lambda_n, D\mathcal{G}^{\mu_n}(u_n)v_n \rangle \geq o(1).$$

Passing to the limit, we obtain $\langle \bar{\eta}, g_y(\bar{y})\bar{z}\rangle \geq 0$. By (2.122) and $\mathrm{d}\bar{\eta} \geq 0$, we deduce that $g_y(\bar{y})\bar{z} \in \mathrm{supp}(\mathrm{d}\bar{\eta})^{\perp}$; thus $\bar{v}$ and its associated linearized state $\bar{z}$ satisfy (2.63) and (2.110). Therefore condition (ii) and $\bar{Q}(\bar{v}) \leq 0$ imply $\bar{v} = 0$. Since by (2.130), $\mathrm{limsup}\, Q^n(v_n) \leq 0$, it follows from (2.121) that $Q^n(v_n) \to 0 = \bar{Q}(\bar{v})$, and hence, $v_n \to \bar{v} = 0$, contradicting $\|v_n\|_2 = 1$ for all n. □

Proof of Theorem 2.34. (ii) ⇒ (i) is a consequence of Lemmas 2.36 to 2.40. Let us show (i) ⇒ (ii). Let ρ be a C^∞ function over $\mathbb{R}$ such that $\mathrm{supp}(\rho) \subset [-1,1]$ and ρ is positive over $(-1,1)$. The function ψ^μ defined by $\psi^\mu(s) := \sum_{\tau\in\mathcal{T}_{to}^{nes}} \mu^{4q+2}\rho\left(\frac{s-\tau}{\mu}\right)$ for $\mu \neq 0$ and $\psi^0(s) = 0$, for all $s \in [0,T]$, is of class C^{2q} with respect to its arguments s and μ and has support in $\cup_{\tau\in\mathcal{T}_{to}^{nes}}[\tau - |\mu|, \tau + |\mu|]$ for $\mu \neq 0$. Consider the perturbed constraint mapping $g^\mu(y) := g(y) - \psi^\mu(y_n)$ (recall that we assume that $(\mathcal{P})$ is written such that $y_n(t) = t$). Observe that $g^0 = g$ and g^μ is of order q for all μ; therefore $(\mathcal{P}^\mu) \equiv (\ell, \phi, f, g^\mu, y_0)$ is a q-stable extension of $(\mathcal{P}^0) = (\mathcal{P})$ with $\mu_0 = 0$. In addition, $g^\mu(y) = g(y)$ for all y such that $y_n \notin \cup_{\tau\in\mathcal{T}_{to}^{nes}}(\tau - |\mu|, \tau + |\mu|)$, and $g^\mu(\bar{y}(t)) < 0$ on $(\tau - |\mu|, \tau + |\mu|)$, for all $\tau \in \mathcal{T}_{to}^{nes}$. Since the touch points are isolated, we have for $|\mu| > 0$ small enough $g^\mu(\bar{y}) = g(\bar{y})$ on $\mathcal{I}_b \cup \mathcal{T}_{to}^{ess} = \mathrm{supp}(\mathrm{d}\bar{\eta})$, and it is easy to see that $(\bar{u}, \bar{y})$ is a stationary point for $(\mathcal{P}^\mu)$, with the same Lagrange multiplier $\bar{\eta}$ and the same costate $\bar{p}$. In addition, the stationary point $(\bar{u}, \bar{y})$ for $(\mathcal{P}^\mu)$ has a neighboring structure to that of $(\bar{u}, \bar{y})$ for $(\mathcal{P}^0)$ (all nonessential touch points are removed). Therefore, by (i) and Def. 2.32, for $|\mu|$ small enough, $(\bar{u}, \bar{y})$ satisfies the uniform quadratic growth condition (2.105) for $(\mathcal{P}^\mu)$. Since assumptions (A2)-(A6) are satisfied for $(\mathcal{P}^\mu)$, it follows from Th. 2.24(ii) that the sufficient condition (ii) holds, which achieves the proof. □

2.4.2 Sensitivity Analysis

If strong regularity holds, the mapping $\Xi : V_\delta \to V_\theta$, $\delta \mapsto \theta(\delta)$ is given by $\Xi(\delta) = \theta_0 + \omega(\delta)$, where $\omega(\delta)$ is the solution of (2.116). It follows then from (2.115) that

$$\theta^\mu = \theta_0 + \omega(-D_\mu\mathcal{F}(\theta_0, \mu_0)(\mu - \mu_0)) + o(\|\mu - \mu_0\|).$$

Since the mapping $\mathbb{R}^{\bar{N}} \to \Theta$, $\delta \mapsto \omega(\delta)$ is positively homogeneous of degree one, the mapping $\mu \mapsto \theta^\mu$ is Fréchet directionally differentiable. The directional derivatives in direction d are obtained by substituting $-D_\mu\mathcal{F}(\theta_0, \mu_0)d$ for δ in (2.116). Therefore,

$$\theta^{\mu_0+d} = \theta_0 + \omega_d + o(\|d\|), \tag{2.131}$$

where

$$\omega_d = (\pi^*_{d,0}, \gamma^{1:q}_{d,\mathcal{T}_{en}}, \gamma_{d,\mathcal{T}_{to}}, \sigma_{d,\mathcal{T}_{en}}, \sigma_{d,\mathcal{T}_{ex}}, \sigma_{d,\mathcal{T}_{to}})$$

is as follows. Denote by (v_d, z_d) and $(\zeta_d, \pi_d, \lambda^{1:q}_{d,\mathcal{T}_{en}}, \lambda_{d,\mathcal{T}_{to}})$ the (unique) optimal solution and multipliers of the quadratic problem below:

$$
\begin{aligned}
(\mathcal{P}_d) \quad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \; & \frac{1}{2}\int_0^T D^2_{(u,y,\mu),(u,y,\mu)}\tilde{H}(\bar{u},\bar{y},\bar{p}_q,\bar{\eta}_q,\mu_0)((v,z,d),(v,z,d))\mathrm{d}t \\
& + \frac{1}{2}D^2\tilde{\phi}(\bar{y}(T),\mu_0)((z(T),d),(z(T),d)) \\
& + \frac{1}{2}\sum_{\tau\in\mathcal{T}_{en}}\sum_{j=1}^{q}\nu_\tau^j D^2\tilde{g}^{(j-1)}(\bar{y}(\tau),\mu_0)((z(\tau),d),(z(\tau),d)) \\
+\frac{1}{2}\sum_{\tau\in\mathcal{T}_{to}}\nu_\tau & \left(D^2\tilde{g}(\bar{y}(\tau),\mu_0)((z(\tau),d),(z(\tau),d)) - \frac{(D\tilde{g}^{(1)}(\bar{y}(\tau),\mu_0)(z(\tau),d))^2}{\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}^{(1)}(\bar{y}(t),\mu_0)|_{t=\tau}} \right)
\end{aligned}
$$

$$
\text{subject to:} \quad \begin{cases}
\dot{z}(t) = D\tilde{f}(\bar{u},\bar{y},\mu_0)(v,z,d) & \text{on } [0,T], \quad z(0) = D\tilde{y}_0(\mu_0)d, \\
D\tilde{g}^{(0:q-1)}(\bar{y}(\tau),\mu_0)(z(\tau),d) = 0, & \tau\in\mathcal{T}_{en}, \\
D\tilde{g}(\bar{y}(\tau),\mu_0)(z(\tau),d) = 0, & \tau\in\mathcal{T}_{to}^{ess}, \\
D\tilde{g}(\bar{y}(\tau),\mu_0)(z(\tau),d) \le 0, & \tau\in\mathcal{T}_{to}^{nes}, \\
D\tilde{g}^{(q)}(\bar{u},\bar{y},\mu_0)(v,z,d) = 0 & \text{on } \mathcal{I}_b.
\end{cases}
$$

Then ω_d is given by $\pi_{d,0} = \pi_q(0)$, $\gamma_{d,\mathcal{T}_{to}} = \lambda_{d,\mathcal{T}_{to}}$,

$$
\sigma_{d,\tau} = -\frac{D\tilde{g}^{(1)}(\bar{y}(\tau),\mu_0)(z_d(\tau),d)}{\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}^{(1)}(\bar{y},\mu_0)|_{t=\tau}}, \; \tau\in\mathcal{T}_{to}, \tag{2.132}
$$

$$
\sigma_{d,\tau} = -\frac{D\tilde{g}^{(q)}(\bar{u}(\tau),\bar{y}(\tau),\mu_0)(v_d(\tau^+),z_d(\tau),d)}{\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}^{(q)}(\bar{u},\bar{y},\mu_0)|_{t=\tau^+}}, \; \tau\in\mathcal{T}_{ex}, \tag{2.133}
$$

$$
\sigma_{d,\tau} = -\frac{D\tilde{g}^{(q)}(\bar{u}(\tau),\bar{y}(\tau),\mu_0)(v_d(\tau^-),z_d(\tau),d)}{\frac{\mathrm{d}}{\mathrm{d}t}\tilde{g}^{(q)}(\bar{u},\bar{y},\mu_0)|_{t=\tau^-}}, \; \tau\in\mathcal{T}_{en}, \tag{2.134}
$$

$$
\gamma^1_{d,\tau} = \lambda^1_{d,\tau}, \quad \gamma^j_{d,\tau} = \lambda^j_{d,\tau} - \nu_\tau^{j-1}\sigma_{d,\tau}, \quad j = 2,\dots,q, \; \tau\in\mathcal{T}_{en}. \tag{2.135}
$$

Once we have the expressions for the directional derivatives of the shooting parameters, by composition with the Fréchet derivatives of the C^1 mapping (2.120) in direction (d,ω_d), we obtain the expressions of the directional derivatives, in $\mathcal{X}_N$, of the mapping $\mu\mapsto(u^\mu,y^\mu,p^\mu,\eta^\mu)$. By Lemma 2.33, we then easily obtain the expression of the directional derivatives of the control and state in $L^r(0,T)\times W^{1,r}(0,T;\mathbb{R}^n)$ for all $1\le r<\infty$.

Corollary 2.41. *If either point* (i) *or* (ii) *of Theorem 2.34 is satisfied, then there exists a neighborhood V_μ of μ such that the mapping $V_\mu\to\mathcal{X}_N$, $\mu\mapsto(u^\mu,y^\mu,p^\mu,\eta^\mu)$ is Fréchet-directionally differentiable on V_μ. In addition, the directional derivative in the space $L^r(0,T)\times W^{1,r}(0,T;\mathbb{R}^n)$, $1\le r<\infty$, of the mapping $\mu\mapsto(u^\mu,y^\mu)$ at point μ_0 in direction d, is the optimal solution (v_d,z_d) of problem $(\mathcal{P}_d)$.*

We end the paper with a remark related to the ill-posedness of the shooting algorithm for a state constraint of order $q\ge 3$ when boundary arcs are present (see Th. 2.23).

Remark 2.42. (Existence of regular boundary arcs for constraints of order $q\ge 3$.) Contrary to some conjectures in the literature, regular boundary arcs *can occur* for state constraints of *all* orders. Take, for example, the problem:

$$
\begin{aligned}
(P_q) \qquad & \min_{(u,y)\in L^\infty(0,T)\times W^{q,\infty}(0,T)} \int_0^T \left(y(t) + \frac{u^2(t)}{2} \right) \mathrm{d}t \\
\text{subject to} \quad & y^{(q)}(t) = u(t) \; ; \; y(0) = y^0_1; \quad \dot{y}(0) = y^0_2; \quad ...; \quad y^{(q-1)}(0) = y^0_q; \\
& y(t) \ge 0, \quad t\in[0,T].
\end{aligned}
$$

It is easy to check that, for $\tau \in (0,T)$, y defined by $y(t)=0$ on $[\tau,T]$ and

$$y(t)=\begin{cases} \dfrac{(t-\tau)^{2q}}{(2q)!} & \text{if } q \text{ is odd} \\ -\dfrac{(t-\tau)^{2q}}{(2q)!}-\nu\dfrac{(t-\tau)^{2q-1}}{(2q-1)!} & \text{if } q \text{ is even} \end{cases} \quad \text{on } [0,\tau],$$

is, for $\nu > \tau/2q$ if q is even and for *appropriate initial conditions* when $q \geq 3$, a solution that satisfies all necessary optimality conditions, and hence, by convexity of the problem, an optimal solution with a regular entry point τ. Moreover, strict complementarity holds since $\eta_0(t)=1$ on $(\tau,T]$.

Robbins in [118] studies this example when $q=3$ for generic initial conditions and shows that the optimal trajectory has a boundary arc, whose entry point is not regular, being the limit of an infinite number of touch points, with a geometric decreasing of the length of the interior arcs. Regular boundary arcs correspond to the case when the multiplier of the geometric sequence is equal to zero for a specific subset of initial conditions. Therefore, we see in that example, though satisfying all regularity assumptions (A0)-(A3), that the *structure of boundary arcs* is *not stable* under perturbations of the initial condition when $q \geq 3$, which illustrates why the shooting algorithm should be ill-posed in that case.

2.5 Appendix

The next two lemmas follow immediately from the junction conditions established in [75, 98].

Lemma 2.43. *Let (u,y) be a regular Pontryagin extremal satisfying (A2)-(A4). Then the function $t \mapsto H_{uu}(u(t),y(t),p(t))$ is continuous on $[0,T]$.*

Proof. Let $\tau \in \mathcal{T}$. Since u is continuous by Prop. 2.5, we have:

$$[H_{uu}(u(\tau),y(\tau),p(\tau))] = [p(\tau)]f_{uu}(u(\tau),y(\tau)) = -\nu_\tau g_{uu}^{(1)}(u(\tau),y(\tau)) = 0,$$

since either $\nu_\tau = 0$ when $q=1$ by Prop. 2.5, or $g_u^{(1)} \equiv 0$ when $q>1$. □

Lemma 2.44. *Let (u,y) be a regular Pontryagin extremal, satisfying (A2)-(A4), and let $\tau \in \mathcal{T}_{en} \cup \mathcal{T}_{ex}$ be an entry/exit time. The following conditions are equivalent:*
(i) (2.60) *holds at τ;* (ii) *if q is* odd, $\lim_{t\to\tau;\, t\in\mathcal{I}_b} \eta_0(t) > 0$*; if q is* even, $\nu_\tau > 0$.

Proof. Define the mappings $(A_l)_{0\leq l\leq q} : [0,T]\setminus\mathcal{T} \to \mathbb{R}^n$ by (2.34) and $(a_l)_{0\leq l\leq q} : [0,T]\setminus\mathcal{T} \to \mathbb{R}$ by

$$a_0(t) = \ell_u(u(t),y(t)); \qquad a_l(t) = \ell_y(u(t),y(t))A_{l-1}(t) - \dot{a}_{l-1}(t) \quad l=1,\dots,q.$$

Then it can be seen by (2.35) (see [98]) that for all $t \in [0,T]\setminus\mathcal{T}$, we have

$$0 = \frac{\mathrm{d}^j}{\mathrm{d}t^j}H_u(u(t),y(t),p(t)) = (-1)^j(a_j(t)+p(t)A_j(t));\ j=0,\dots,q-1, \tag{2.136}$$

$$0 = \frac{\mathrm{d}^q}{\mathrm{d}t^q}H_u(u(t),y(t),p(t)) = (-1)^q\left(a_q(t)+p(t)A_q(t)+\frac{\mathrm{d}\eta}{\mathrm{d}t}g_u^{(q)}(u(t),y(t))\right). \tag{2.137}$$

Since the derivatives of the control are continuous until order $q-2$, the functions a_j and A_j are continuous for $j=0,\dots,q-2$, and it is then easily seen, since u is continuous, that the jumps of A_{q-1} and a_{q-1} at $\tau \in \mathcal{T}$, when q is even, are given respectively by

$$[A_{q-1}(\tau)] = (-1)^{q-1}f_{uu}(u(\tau),y(\tau))[u^{(q-1)}(\tau)],$$
$$[a_{q-1}(\tau)] = (-1)^{q-1}\ell_{uu}(u(\tau),y(\tau))[u^{(q-1)}(\tau)].$$

Taking the jump in (2.136) at τ for $j = q-1$ then yields:

$$0 \;=\; (-1)^{q-1} H_{uu}(u(\tau), y(\tau), p(\tau^+))[u^{(q-1)}(\tau)] - \nu_\tau g_y(y(\tau)) A_{q-1}(\tau^-).$$

By (2.35), we have $g_y(y(\tau))A_{q-1}(\tau^\pm) = g_u^{(q)}(u(\tau), y(\tau))$, so we obtain, when q is even:

$$\nu_\tau = (-1)^{q-1} \frac{H_{uu}(u(\tau), y(\tau), p(\tau^+))[u^{(q-1)}(\tau)]}{g_u^{(q)}(u(\tau), y(\tau))}. \tag{2.138}$$

It follows that $\nu_\tau > 0$ iff $u^{(q-1)}$ is discontinuous at τ, which is equivalent to saying that (2.60) holds (when q is even). When q is odd, $u^{(q-1)}$, a_{q-1}, and A_{q-1} are continuous (and $\nu_\tau = 0$). Taking the jump in (2.137), we obtain

$$0 \;=\; (-1)^q H_{uu}(u(\tau), y(\tau), p(\tau))[u^{(q)}(\tau)] + [\eta_0(\tau)]\, g_u^{(q)}(u(\tau), y(\tau)).$$

Consequently, we have $\eta_0(\tau^\pm) > 0$ at an entry/exit point, where $\tau^\pm$ stands for τ^+ if $\tau \in \mathcal{T}_{en}$ and τ^- if $\tau \in \mathcal{T}_{ex}$ iff $u^{(q)}$ is discontinuous at τ, and hence iff (2.60) holds. □

The next two lemmas recall classical results. For the second one see related results by Aubin [3].

Lemma 2.45. *Let X be a Hilbert space and Q a Legendre form over X. Let A be a continuous linear operator over X. The following assertions are equivalent:*
(i) *$Q(v) > 0$ for all $v \in \operatorname{Ker} A \setminus \{0\}$;*
(ii) *There exists $\alpha > 0$ such that $Q(v) \geq \alpha \|v\|_2^2$, for all $v \in \operatorname{Ker} A$.*

Lemma 2.46. *Let X be a Hilbert space and Y a Banach space, $H : X \to X^* \equiv X$ a self-adjoint continuous linear operator, and $A : X \to Y$ and $B : X \to \mathbb{R}^r$, $r \in \mathbb{N}$, continuous linear operators. Assume that*

$$\begin{array}{ll} \text{(i)} & \exists \alpha > 0 \qquad \langle Hx, x\rangle \geq \alpha \|x\|^2, \quad \text{for all } x \in \operatorname{Ker} A, \\ \text{(ii)} & \text{The operator } (A, B) : X \to Y \times \mathbb{R}^r \text{ is onto.} \end{array}$$

Then, for all $(x^*, y, \delta) \in X^* \times Y \times \mathbb{R}^r$, there exists a unique $(x, y^*, \nu) \in X \times Y^* \times \mathbb{R}^{r*}$, solution of

$$\left\{ \begin{array}{l} Hx + A^*y^* + B^*\nu \;=\; x^* \\ Ax \;=\; y \\ Bx \;\leq\; \delta, \quad \nu \geq 0, \quad \nu(Bx - \delta) = 0, \end{array} \right. \tag{2.139}$$

and the mapping $(x^*, y, \delta) \mapsto (x, y^*, \nu)$, where (x, y^*, ν) is solution of (2.139), is Lipschitz continuous.

Acknowledgments The authors thank the anonymous referees for their useful remarks.

Chapitre 3

Stabilité et sensibilité pour des contraintes d'ordre 1 et méthodes d'homotopie*

Abstract The paper deals with an optimal control problem with a scalar first-order state constraint and a scalar control. In presence of (nonessential) touch points, the arc structure of the trajectory is not stable. Under some reasonable assumptions, we show that boundary arcs are structurally stable, and that touch point can either remain so, vanish or be transformed into a single boundary arc. Assuming a weak second-order optimality condition (equivalent to uniform quadratic growth), stability and sensitivity results are given. The main tools are the study of a quadratic tangent problem and the notion of strong regularity. Those results enables us to design a new continuation algorithm, presented at the end of the paper, that handles automatically changes in the structure of the trajectory.

Résumé Dans cet article, on s'intéresse aux problèmes de commande optimale avec une contrainte sur l'état du premier ordre. En présence de points de contact isolés (non essentiels), la structure en arcs de la solution n'est pas stable. Sous des hypothèses raisonnables, on montre que les arcs frontières sont stables et qu'un point de contact isolé devient inactif, reste point de contact, ou bien se transforme en arc frontière. Sous une condition du second ordre faible (équivalente à la croissance quadratique uniforme), une analyse de stabilité et sensibilité des solutions est présentée. Le résultat s'appuie sur l'étude du problème linéaire-quadratique tangent et sur la notion de régularité forte. Ces résultats nous permettent de concevoir un nouvel algorithme d'homotopie qui prend en compte automatiquement des changements de structure de la trajectoire.

3.1 Introduction

This paper deals with an optimal control problem (of an ordinary differential equation) with a scalar first-order state constraint and a scalar control, with a free final state and no control constraints. It is well-known that for *first-order* state constraints, when the strengthened Legendre-Clebsch condition holds and the state constraint is regular, touch points (locally

*Joint work with J.F. Bonnans. Published in ESAIM Control, Optimization and Calculus of Variations, 14(4) :825–863 (2008), under the title *Stability and sensitivity analysis for optimal control problems with a first-order state constraint and application to continuation methods.*

unique times where the constraint is active) are nonessential (the associated jump of the multiplier is null) (see e.g. [75, 68]). Situations where touch points are present may be encountered, for instance, when solving the optimal control problem by indirect approaches using an homotopy method in order to guess the arc structure of the trajectory, see e.g. the example in [11]. Therefore it is of interest to study sensitivity of solutions around touch points, when the constraint becomes active. Under a small perturbation, several events may occur. Among them, the constraint may locally become inactive, the touch point may remain a touch point, or it may give rise to a boundary arc. Our main result is that, under natural hypotheses, these are the only three possibilities, and that the boundary arcs have a length of the order of the perturbation, and satisfy a "strict complementarity" hypothesis. In addition, we show how to compute a first-order expansion of the solution. The analysis uses in a critical way a certain tangent quadratic problem, and at the same time is in the spirit of the shooting approach, in the sense that touch points are converted into boundary arcs of zero length, and we compute the first-order expansion of all entry and exit points. Fréchet directional derivatives are obtained as the solution of an inequality-constrained linear quadratic problem. The proof applies the notion of "strong regularity" in the sense of Robinson [121] to a system that happens to be equivalent to the optimality conditions of the tangent quadratic problem. Our formulation of the corresponding shooting formulation (of which all entry and exit times are variables, in addition to the initial costate and jumps of the alternative multiplier at entry times) allows exit times to be lower than entry times; however, we check that the solution of the shooting formulation is such that entry times are lower than or equal to corresponding exit times.

Optimal control problems with *first-order* state constraints were first studied in the book by Pontryagin et al. [116]. Numerous results have been obtained since for stability and sensitivity analysis of those problems. Two different approaches have been used. The first one is the use of implicit function theorems in infinite dimensional spaces (see [123, 87, 67, 103]), and the second one is to reduce the problem to a finite-dimensional one (a two- or multi points boundary value problem) using the so-called *shooting formulation* (see [125, 101]). With first-order state constraints, L^2-stability of solutions was first obtained by Malanowski [88], under strong second-order sufficient conditions, using an infinite-dimensional implicit function theorem based on two-norms approach, and later by Dontchev and Hager [53], using an implicit function theorem in metric spaces. In Malanowski [88], directional differentiability of solutions in L^2 was established, using the results on differentiability of projection onto a closed convex cone in Hilbert spaces [67]. The second-order sufficient condition used in the analysis was weakened by Malanowski [89]. All those results require no assumptions on the structure of the trajectory. In order to obtain L^∞-stability of solutions, Dontchev and Hager [53] needed an additional assumption on the structure of the contact set ("contact separation"). Using a finite dimensional approach, Malanowski and Maurer obtained in [93] differentiability of solutions in L^∞ by application of the implicit function theorem to the shooting mapping, under stronger assumptions (finitely many nontangential junction points, and strict complementarity) needed to ensure the stability of the structure of solutions.

The approach presented in this paper is different from the ones in [88, 89, 53] where the stability and sensitivity analysis was done in infinite dimensional spaces without any assumptions on the structure of the trajectory. On the contrary, our aim is to describe changes in the structure of the trajectory, both qualitatively and quantitatively. Thus the first step is to consider nonessential touch points. Indeed, as mentioned before, changes in the structure are likely to occur when performing continuation methods, therefore the more information we have on the continuity and/or differentiability of the homotopy path, the easier will be

the latter to follow. Our stability and sensitivity results generalize those of [93] to the case when (nonessential) touch points are present. However, in that case strict complementarity does not hold anymore, so we cannot apply the classical implicit function Theorem as done in [93]. This paper is related to our previous work: the study of no-gap second-order optimality conditions in [21], and the shooting formulation, allowing nonessential touch points for state constraints of order greater than one, and for which we also use the notion of strong regularity [19]. In both papers we assume also the state constraint and the control to be scalar-valued. Some of these results are extended to the case of vector-valued state constraints and control in [17]. We follow here the analysis in [19] where sensitivity results with nonessential touch points for state constraints of order *greater than one* were obtained. The contributions of this paper are the following:

- A stability result of the structure of *stationary points* (and not only the stability of the structure of locally optimal solutions) is proved. That is, if the nominal trajectory satisfies several assumptions, among which uniform strict complementarity on boundary arcs, then *any stationary point* in the neighborhood has a "neighboring structure", in a sense made precise in section 2.

- In the stability and sensitivity analysis we cover the case of the possible transformation of touch points into boundary arcs. This possibility was excluded from the analysis in [19] and in [93], and leads to technical complications. In particular we show that for first-order state constraints, the shooting algorithm remain well-posed when touch points are converted in boundary arcs, which is false for control constraints (see Remark 3.32).

- At the end of the paper, we present an application of those results to a preliminary homotopy algorithm whose novelty is to handle changes in the structure (appearance/disappearance of a boundary arc) automatically. Numerical application on a simple academic problem is presented.

The paper is organized at follows. The framework is presented in section 3.2. In section 3.3, the stability results of the structure of stationary points are given. In section 3.4, the main result is stated. In section 3.5, the problem is reduced to a generalized finite-dimensional equation, with a complementarity constraint. Robinson's strong regularity theory is applied to the latter in section 3.6, where the main result is proved. Section 3.7 deals with directional differentiability of solutions. In section 3.8, a basic illustrative example is presented. The homotopy method is described in section 3.9. Section 3.10 contains the proofs of the results of section 3.3.

3.2 Preliminaries

Let $\mathcal{U} := L^\infty(0,T)$ (resp. $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$) denote the control (resp. state) space. Let M be a Banach space (the space of perturbations parameter) and, for $\mu \in M$, the cost function $\ell^\mu : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, final cost function $\phi^\mu : \mathbb{R}^n \to \mathbb{R}$, dynamics $f^\mu : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, state constraint $g^\mu : \mathbb{R}^n \to \mathbb{R}$, initial condition $y_0^\mu \in \mathbb{R}^n$, and (fixed) final time $T > 0$. We consider

the following optimal control problem:

$$(\mathcal{P}^\mu) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell^\mu(u(t),y(t))\mathrm{d}t + \phi^\mu(y(T)) \tag{3.1}$$

$$\text{subject to} \qquad \dot{y}(t) = f^\mu(u(t),y(t)) \text{ for a.a. } t\in[0,T], \quad y(0)=y_0^\mu, \tag{3.2}$$

$$g^\mu(y(t)) \le 0, \text{ for all } t\in[0,T]. \tag{3.3}$$

This notation allow us to deal with *non autonomous* problems (i.e. when the data ℓ^μ, f^μ, g^μ depend on time t) as well, by assuming w.l.o.g. that the last component of the state variable y_n satisfies in (3.2)

$$\dot{y}_n(t) = 1 \quad \text{on } [0,T], \quad y_n(0)=0 \qquad\qquad (\text{i.e. } y_n(t)=t). \tag{3.4}$$

We shall assume in all the paper that $(\mathcal{P}^\mu)$ is written such that (3.4) holds. In this way our analysis will include non autonomous perturbations, even when the starting problem is autonomous. This assumption is only used in Th. 3.11 to obtain the implication (i) $\Rightarrow$ (ii).

We study perturbations of problem $(\mathcal{P}^\mu)$ around a given value of parameter $\mu_0\in M$, and we often omit the superscript μ when we refer to the problem and data associated with μ_0, i.e. $(\mathcal{P}) := (\mathcal{P}^{\mu_0})$ and $(\ell,\phi,f,g,y_0) := (\ell^{\mu_0},\phi^{\mu_0},f^{\mu_0},g^{\mu_0},y_0^{\mu_0})$.

We assume throughout the paper that the assumptions below hold:

(A0) The mappings ℓ, ϕ, f and g are of class C^2, with locally Lipschitz continuous second-order derivatives, and the dynamics f is Lipschitz continuous;

(A1) the initial condition satisfies $g(y_0)<0$.

These assumptions will not be repeated in the various results of the paper.

A parametrization $(\ell^\mu,\phi^\mu,f^\mu,g^\mu,y_0^\mu)$, identified with problem $(\mathcal{P}^\mu)$, is a *stable extension* of $(\mathcal{P})$, if there exists an open neighborhood M_0 of μ_0, such that (i) there exist C^2 mappings $\hat{\ell}:\mathbb{R}\times\mathbb{R}^n\times M_0\to\mathbb{R}$; $\hat{\phi}:\mathbb{R}^n\times M_0\to\mathbb{R}$; $\hat{f}:\mathbb{R}\times\mathbb{R}^n\times M_0\to\mathbb{R}^n$; $\hat{g}:\mathbb{R}^n\times M_0\to\mathbb{R}$ and $\hat{y}_0:M_0\to\mathbb{R}^n$, such that $\ell^\mu(u,y)=\hat{\ell}(u,y,\mu)$ for all $(u,y)\in\mathbb{R}\times\mathbb{R}^n$ and all $\mu\in M_0$ (and similarly for ϕ^μ, f^μ, g^μ, and y_0^μ); (ii) the mappings ℓ^μ, f^μ, ϕ^μ, g^μ have Lipschitz continuous second-order derivatives and f^μ is Lipschitz continuous, uniformly over $\mu\in M_0$.

In this paper, we always consider stable extensions $(\mathcal{P}^\mu)$, that satisfy (3.4) as said before.

Definitions and Notations

The space of row vectors is denoted by $\mathbb{R}^{n*}$, and the adjoint and transposition operator in $\mathbb{R}^n$ are denoted by a star *. Fréchet derivatives of f, ℓ, etc. w.r.t. arguments $u\in\mathbb{R}$, $y\in\mathbb{R}^n$, are denoted by a subscript, for instance $f_u(u,y)=D_uf(u,y)$. The space $L^r(0,T)$, $r\in[1,\infty]$, is the Lebesgue space of measurable functions such that $\|u\|_r := (\int_0^T|u(t)|^r)^{1/r} < \infty$ for $1\le r<\infty$ and $\|u\|_\infty := \mathrm{supess}_{t\in[0,T]}|u(t)| < \infty$, and $W^{1,r}(0,T)$ is the Sobolev space of functions in $L^r(0,T)$ with a weak derivative in $L^r(0,T)$. The space of continuous functions and its dual space, the space of bounded Borel measures, are denoted respectively by $C^0[0,T]$ and $\mathcal{M}[0,T]$. The cone of nonnegative measures is denoted by $\mathcal{M}_+[0,T]$, and $BV([0,T];\mathbb{R}^n)$ denotes the space of vector-valued functions of bounded variation over $[0,T]$. The elements of $\mathcal{M}[0,T]$ are identified with the derivative of functions of bounded variation vanishing at T. We denote by $\varphi(t^-)$ and $\varphi(t^+)$ the respectively left- and right limits of a function of bounded variation φ at a time $t\in[0,T]$. Jumps are denoted by $[\varphi(t)] := \varphi(t^+)-\varphi(t^-)$.

Given $\mu \in M_0$, a *trajectory* of $(\mathcal{P}^\mu)$ is an element $(u, y) \in \mathcal{U} \times \mathcal{Y}$ satisfying the state equation (3.2). A *feasible* trajectory is one satisfying the state constraint (3.3). The first-order time derivative of the state constraint is the function defined by $(g^\mu)^{(1)} : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R},\ (u, y) \mapsto g^\mu_y(y) f^\mu(u, y)$. In this paper, we consider state constraints of *first order*, that is, the function $(g^\mu)^{(1)}(u, y)$ depends explicitly on the control variable u in the neighborhood of the contact set of the constraint, see assumption (A3). It will be convenient to introduce the second-order time derivative of the state constraint by:

$$(g^\mu)^{(2)} \ : \ \mathbb{R} \times \mathbb{R} \times \mathbb{R}^n \ \to \ \mathbb{R}, \quad (\upsilon, u, y) \ \mapsto \ (g^\mu)^{(1)}_u(u, y)\upsilon + (g^\mu)^{(1)}_y(u, y) f^\mu(u, y). \tag{3.5}$$

Wherever u is differentiable, we have that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g^\mu(y(t)) = (g^\mu)^{(2)}(\dot{u}(t), u(t), y(t)). \tag{3.6}$$

The classical (resp. augmented) *Hamiltonian* functions $H^\mu : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$ (resp. $\tilde{H}^\mu : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n*} \times \mathbb{R} \to \mathbb{R}$) are defined by:

$$H^\mu(u, y, p) \ := \ \ell^\mu(u, y) + p f^\mu(u, y) \tag{3.7}$$

$$\tilde{H}^\mu(u, y, p^1, \eta^1) \ := \ H^\mu(u, y, p^1) + \eta^1 (g^\mu)^{(1)}(u, y). \tag{3.8}$$

For (u, y) a feasible trajectory of $(\mathcal{P}^\mu)$, define the *contact set* by:

$$I(g^\mu(y)) \ := \ \{t \in [0, T] \ ; \ g^\mu(y(t)) = 0\}. \tag{3.9}$$

We say that the constraint is *active* at time t, if $t \in I(g^\mu(y))$; otherwise it is said inactive at time t. A *boundary arc* (resp. *interior arc*) is a maximal interval of positive measure $\mathcal{I}$ such that $g^\mu(y(t)) = 0$ (resp. $g^\mu(y(t)) < 0$), for all $t \in \mathcal{I}$. Left and right endpoints of a boundary arc $[\tau_{en}, \tau_{ex}]$ are called *entry* and *exit* point, respectively. A *touch point* τ_{to} is an isolated contact point, satisfying $g^\mu(y(\tau_{to})) = 0$ and $g^\mu(y(t)) < 0$, for $t \neq \tau_{to}$ in the neighborhood of τ_{to}. The endpoints of interior arcs belonging to $(0, T)$ are called *junction points* (or *times*).

If the set of junction points of a trajectory is *finite*, then it is of the form

$$\mathcal{T} =: \mathcal{T}_{en} \cup \mathcal{T}_{ex} \cup \mathcal{T}_{to},$$

with $\mathcal{T}_{en}$, $\mathcal{T}_{ex}$ and $\mathcal{T}_{to}$ the *disjoint* (and possibly empty) subsets of respectively regular entry, exit and touch points. We denote by $\mathcal{I}_b$ the union of boundary arcs, i.e. $\mathcal{I}_b := \cup_{i=1}^{N_b} [\tau_i^{en}, \tau_i^{ex}]$ for $\mathcal{T}_{en} := \{\tau_1^{en} < \cdots < \tau_{N_b}^{en}\}$ and similar definition of $\mathcal{T}_{ex}$, and we have $I(g^\mu(y)) = \mathcal{T}_{to} \cup \mathcal{I}_b$. The *arc structure* (or simply *structure*) of a trajectory is the (finite) number of boundary arcs and touch points, and the order in which they occur.

Given a finite subset $\mathcal{S}$ of $(0, T)$, we denote by $PC^k_{\mathcal{S}}[0, T]$ the set of functions over $[0, T]$ that are of class C^k outside $\mathcal{S}$, and have, as well as their first k derivatives, a left and right limit over $\mathcal{S}$ and a left (resp. right) limit at T (resp. 0). The subset of functions in $PC^k_{\mathcal{S}}[0, T]$ having continuous derivatives on $[0, T]$ until order r, $0 \leq r \leq k$, is denoted by $PC^{k,r}_{\mathcal{S}}[0, T] := PC^k_{\mathcal{S}}[0, T] \cap C^r[0, T]$. We also use the notation $\nu_{\mathcal{S}} := (\nu_\tau)_{\tau \in \mathcal{S}} \ \in \mathbb{R}^{\mathrm{Card}\, \mathcal{S}}$.

Given $(\mu, u) \in M_0 \times \mathcal{U}$, we denote by y^μ_u the (unique) state solution in $\mathcal{Y}$ of:

$$\dot{y}^\mu_u(t) \ = \ f^\mu(u(t), y^\mu_u(t)) \qquad \text{a.e. on } [0, T], \qquad y^\mu_u(0) = y^\mu_0. \tag{3.10}$$

By definition of a stable extension, the mapping $\mathcal{U} \times M_0 \to \mathcal{Y}$, $(u, \mu) \mapsto y^\mu_u$ is C^2. A useful equivalent *abstract* formulation of $(\mathcal{P}^\mu)$ is

$$\min_{u \in \mathcal{U}} J^\mu(u), \qquad G^\mu(u) \in K, \tag{3.11}$$

with the cost function $J^\mu : \mathcal{U} \to \mathbb{R}$, $u \mapsto \int_0^T \ell^\mu(u(t), y_u^\mu(t))\mathrm{d}t + \phi^\mu(y_u^\mu(T))$, $K := C^0_-[0,T]$ the cone of continuous functions taking nonpositive values, and G^μ the mapping $\mathcal{U} \to C^0[0,T]$, $u \mapsto g^\mu(y_u^\mu)$. We write J and G for J^{μ_0} and G^{μ_0}, respectively.

Optimality Conditions

Let us first recall the definition of Pontryagin extremals.

Definition 3.1. A trajectory (u, y) is a *Pontryagin extremal* of $(\mathcal{P}^\mu)$, if there exist $\alpha \in \mathbb{R}_+$, $\mathrm{d}\eta \in \mathcal{M}[0,T]$ and $p \in BV([0,T];\mathbb{R}^{n*})$, $(\mathrm{d}\eta, p, \alpha) \neq 0$, such that:

$$\begin{aligned}
\dot{y}(t) &= f^\mu(u(t), y(t)) \quad \text{a.e. on } [0,T], \qquad y(0) = y_0^\mu && (3.12)\\
\mathrm{d}p(t) &= \{\alpha \ell_y^\mu(u(t), y(t)) + p(t) f_y^\mu(u(t), y(t))\}\mathrm{d}t + g_y^\mu(y(t))\mathrm{d}\eta(t) \quad \text{on } [0,T] && (3.13)\\
p(T^+) &= \alpha \phi_y^\mu(y(T)) && (3.14)\\
u(t) &\in \operatorname{argmin}_{\hat{u}\in\mathbb{R}}\{\alpha \ell^\mu(\hat{u}, y(t)) + p(t) f^\mu(\hat{u}, y(t))\} \quad \text{a.e. on } [0,T] && (3.15)\\
0 &\geq g^\mu(y(t)), \qquad \mathrm{d}\eta \geq 0, \qquad \int_0^T g^\mu(y(t))\mathrm{d}\eta(t) = 0. && (3.16)
\end{aligned}$$

When $\alpha > 0$, dividing p and η by α, we can take $\alpha = 1$ in the above equations, and in that case we say that (u, y) is a *regular* Pontryagin extremal.

It is well known that optimal solutions of $(\mathcal{P}^\mu)$ are Pontryagin extremals. A sufficient condition to ensure that $\alpha = 1$, i.e. that an optimal solution (u, y) of $(\mathcal{P}^\mu)$ is a regular Pontryagin extremal, is that Robinson's constraint qualification [119, 120] below is satisfied (recall (3.11)):

$$\exists \gamma > 0, \qquad \gamma B_{C^0[0,T]} \subset G^\mu(u) + DG^\mu(u)\mathcal{U} - K, \tag{3.17}$$

with $B_{C^0[0,T]}$ the unit (open) ball of the space of continuous functions.

A trajectory (u, y) is a *stationary point* of $(\mathcal{P}^\mu)$, if there exist $\mathrm{d}\eta \in \mathcal{M}[0,T]$ and $p \in BV([0,T];\mathbb{R}^{n*})$ such that (3.12)-(3.14) and (3.16) hold (with $\alpha = 1$), as well as

$$0 = \ell_u^\mu(u(t), y(t)) + p(t) f_u^\mu(u(t), y(t)) \qquad \text{for a.a. } t \in [0,T]. \tag{3.18}$$

The above condition is in general weaker than (3.15). However, when the Hamiltonian H^μ is convex w.r.t. the control variable along the trajectory (and in particular when assumption (3.22) below holds), then the definitions of regular Pontryagin extremals and stationary points are equivalent.

We say that (u, y) is a *local solution* (weak minimum) of $(\mathcal{P}^\mu)$, if it minimizes (3.1) over the set of feasible trajectories $(\tilde{u}, \tilde{y})$ satisfying $\|\tilde{u} - u\|_\infty \leq \delta$ for some $\delta > 0$. Local solutions of $(\mathcal{P}^\mu)$ satisfying (3.17) are stationary points.

Note that the complementarity conditions (3.16) can be equivalently rewritten as:

$$g^\mu(y) \in K, \qquad \mathrm{d}\eta \in \mathcal{M}_+[0,T], \qquad \mathrm{supp}(\mathrm{d}\eta) \subset I(g^\mu(y)), \tag{3.19}$$

where $\mathrm{supp}(\mathrm{d}\eta)$ denotes the support of the measure $\mathrm{d}\eta$. Another condition equivalent to (3.16) is $\mathrm{d}\eta \in N_K(G^\mu(u))$, where $N_K(G^\mu(u))$ denotes the normal cone (in the sense of convex analysis) to K at point $G^\mu(u)$.

Assumptions

We assume that problem $(\mathcal{P})$ has a local solution, denoted in the sequel by $(\bar{u}, \bar{y})$, and that the latter satisfies, with $\bar{p}$ and $\bar{\eta}$ its associated multipliers, the following assumptions:

(A2) The control $\bar{u}$ is continuous over $[0,T]$, and there exists $\alpha > 0$ such that

$$H_{uu}(\bar{u}(t), \bar{y}(t), \bar{p}(t^{\pm})) \geq \alpha, \quad \text{for all } t \in [0,T]. \tag{3.20}$$

(A3) Uniform regularity of the state constraint near the contact set, i.e., there exists $\beta, \varepsilon > 0$ such that

$$|g_u^{(1)}(\bar{u}(t), \bar{y}(t))| \geq \beta, \quad \text{for a.a. } t, \ \operatorname{dist}\{t; I(g(\bar{y}))\} \leq \varepsilon. \tag{3.21}$$

A condition stronger than (A2) which *implies* the continuity of the control is the uniform strong convexity of the Hamiltonian w.r.t. the control variable, i.e. there exists $\alpha > 0$, such that

$$H_{uu}(\hat{u}, \bar{y}(t), \bar{p}(t^{\pm})) \geq \alpha, \quad \text{for all } \hat{u} \in \mathbb{R} \text{ and all } t \in [0,T]. \tag{3.22}$$

It is well-know (see e.g. [65, 68]) that when (A2)-(A3) hold, then $\bar{u}$ and the multiplier $\bar{\eta}$ are Lipschitz continuous. In particular this implies that all touch points τ_{to} are *nonessential*, i.e. $[\bar{\eta}(\tau_{to})] = 0$. Furthermore, (A3) implies that (3.17) holds, and that the multipliers $(\bar{p}, \bar{\eta})$ associated with $(\bar{u}, \bar{y})$ are unique. This is a consequence of the lemma below. For $\delta > 0$, let $\Omega^{\delta} := \{t \in [0,T], \ \operatorname{dist}\{t; I(g(\bar{y}))\} < \delta\}$.

Lemma 3.2. *Assumption (A3) implies that for all $0 < \delta < \varepsilon$, with the ε of (3.21), assumed to be so small that $\Omega^{\varepsilon} \subset [a,T]$ for some $a > 0$, the linear mapping*

$$\mathcal{U} \to W^{1,\infty}(\Omega^{\delta}), \qquad v \mapsto (DG(\bar{u})v)|_{\Omega^{\delta}}, \tag{3.23}$$

where $|_{\Omega^{\delta}}$ denotes the restriction to the set Ω^{δ}, is onto.

Proof. Let us recall the proof of [21, Lemma 9][1]. For $v \in \mathcal{U}$, we have that $DG(\bar{u})v = g_y(\bar{y})z_v$, where z_v is the (unique) solution in $\mathcal{Y}$ of the linearized state equation:

$$\dot{z}_v = f_u(\bar{u}, \bar{y})v + f_y(\bar{u}, \bar{y})z_v, \qquad \text{a.e. on } [0,T], \qquad z_v(0) = 0. \tag{3.24}$$

It is easy to see that

$$\frac{\mathrm{d}}{\mathrm{d}t} g_y(\bar{y}(t))z_v(t) \ = \ g_u^{(1)}(\bar{u}, \bar{y})v + g_y^{(1)}(\bar{u}, \bar{y})z_v,$$

and since by (3.21) and (A1), $g_u^{(1)}(\bar{u}, \bar{y})$ is uniformly invertible on a neighborhood of Ω^{δ} for small $\delta > 0$, the result follows as a consequence of Gronwall's Lemma. □

We will also make in addition to (A2)-(A3) the following assumptions:

(A4) The trajectory $(\bar{u}, \bar{y})$ has a *finite set of junction times* $\bar{\mathcal{T}}$, and we assume that $g(\bar{y}(T)) < 0$.

(A5) *Uniform strict complementarity* on *boundary arcs*:

$$\exists \beta > 0 \qquad \frac{\mathrm{d}\bar{\eta}}{\mathrm{d}t}(t) \geq \beta \qquad \text{for all } t \text{ in the interior of boundary arcs;} \tag{3.25}$$

[1] Lemma 1.9 of this thesis.

(A6) *Non tangentiality at second-order* at (nonessential) *touch points*: for all touch point $\bar\tau_{to}$,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\bar y(t))|_{t=\bar\tau_{to}} < 0. \tag{3.26}$$

Note that (3.26) makes sense, since $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\bar y(t))$ is by (3.6) a continuous function of $(\bar y, \bar u, \dot{\bar u})$, and $\bar u$ and $\dot{\bar u}$ are continuous at a touch point $\bar\tau_{to}$ (indeed, $\bar\tau_{to}$ being a nonessential touch point, $(\bar\tau_{to}-\varepsilon, \bar\tau_{to}+\varepsilon) \cap \mathrm{supp}(\mathrm{d}\bar\eta) = \emptyset$ for some small $\varepsilon > 0$, so the continuity of $\dot{\bar u}$ follows from the implicit function theorem applied to the relation $H_u(\bar u, \bar y, \bar p) = 0$). This condition is similar to the reducibility hypothesis when the state constraint is of order $q \geq 2$ (see [19]). The lemma below will be proved later (see Lemma 3.22), the proof being based on the *alternative formulation* (Def. 3.14).

Lemma 3.3. *Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)-(A4). Then assumption (A5) implies that the following non-tangentiality condition at second-order holds at entry and exit points:*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\bar y(t))|_{t=\bar\tau_{en}^-} < 0, \quad \bar\tau_{en} \in \bar{\mathcal{T}}_{en}\ ; \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\bar y(t))|_{t=\bar\tau_{ex}^+} < 0, \quad \bar\tau_{ex} \in \bar{\mathcal{T}}_{ex}. \tag{3.27}$$

3.3 Structural stability of stationary points

Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)-(A6). Assume that $(\bar u, \bar y)$ has N_{ba} boundary arcs and N_{to} touch points, and let $N := N_{ba} + N_{to}$. Number the boundary arcs and touch points of $(\bar u, \bar y)$ by $i = 1, \dots, N$, and denote by I_{ba} and I_{to} the (disjoint) sets of index in $\{1, \dots, N\}$ corresponding respectively to boundary arcs and touch points. Denote the junction times of $(\bar u, \bar y)$ by $\bar{\mathcal{T}}_{en} = \{\bar\tau_{en}^i\}_{i\in I_{ba}}$, $\bar{\mathcal{T}}_{ex} = \{\bar\tau_{ex}^i\}_{i\in I_{ba}}$, and $\bar{\mathcal{T}}_{to} = \{\bar\tau_{to}^i\}_{i\in I_{to}}$. For $\delta > 0$, define

$$\Omega_i^\delta := (\bar\tau_{en}^i - \delta, \bar\tau_{ex}^i + \delta), \quad i \in I_{ba}, \qquad \Omega_i^\delta := (\bar\tau_{to}^i - \delta, \bar\tau_{to}^i + \delta), \quad i \in I_{to}. \tag{3.28}$$

In view of (A4), (A6) and (3.27), we may fix $\kappa, \bar\delta > 0$ satisfying the conditions below:

$$\bar\delta \leq \varepsilon \text{ with the } \varepsilon \text{ of (3.21)}, \tag{3.29}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(\bar y(t)) \leq -\kappa < 0 \ \text{ on } \Omega_i^{\bar\delta} \setminus [\bar\tau_{en}^i, \bar\tau_{ex}^i] \ \text{ for all } i \in I_{ba} \ \text{ and on } \Omega_i^{\bar\delta} \ \text{ for all } i \in I_{to}, \tag{3.30}$$

$$\text{the sets } (\Omega_i^{\bar\delta})_{1\leq i\leq N} \text{ are pairwise disjoint and contained in } [a, T] \text{ for some } a > 0. \tag{3.31}$$

The next theorem gives a direct result (i.e. without using a shooting formulation) of the stability of structure of stationary points, when assumptions (A2)-(A6) are satisfied.

Theorem 3.4. *Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P}^{\mu_0})$ satisfying (A2)-(A6), and let $\bar\delta$ satisfy* (3.29)-(3.31)*. Then for all $0 < \delta < \bar\delta$ and all stable extensions $(\mathcal{P}^\mu)$ of $(\mathcal{P}^{\mu_0})$, there exists a neighborhood $V_u \times V_\mu$ of $(\bar u, \mu_0)$ in $\mathcal{U} \times M$, such that all stationary points (u, y) of $(\mathcal{P}^\mu)$ with $(u, \mu) \in V_u \times V_\mu$ satisfy the following properties, with the contact set $I(g^\mu(y))$ defined by* (3.9)*:*

(S1) *$I(g^\mu(y)) \subset \cup_{i=1}^N \Omega_i^\delta$,*

(S2) *for all $i \in I_{ba}$, $I(g^\mu(y)) \cap \Omega_i^\delta$ is an interval of positive measure;*

(S3) *for all $i \in I_{to}$, $I(g^\mu(y)) \cap \Omega_i^\delta$ is either empty, or a singleton, or an interval of positive measure.*

When (S1)-(S3) are satisfied, we say that a stationary point (u, y) of $(\mathcal{P}^\mu)$ has a *neighboring structure* to that of $(\bar{u}, \bar{y})$.

Remark 3.5. We can actually state a "local" version of Th. 3.4. More precisely, if a stationary point $(\bar{u}, \bar{y})$ of $(\mathcal{P}^{\mu_0})$ satisfying (A3) has a boundary arc $[\bar{\tau}_{en}, \bar{\tau}_{ex}]$ (resp. a touch point $\bar{\tau}_{to}$) and if assumptions (A2) and (A4)-(A6) hold locally over $(\bar{\tau}_{en}-\delta, \bar{\tau}_{ex}+\delta)$ (resp. over $(\bar{\tau}_{to}-\delta, \bar{\tau}_{to}+\delta)$) for some $\delta > 0$, then all stationary points (u, y) of $(\mathcal{P}^\mu)$ with (u, μ) in the neighborhood of $(\bar{u}, \mu_0)$ have exactly one boundary arc on $(\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta)$ (resp. have at most either one touch point or one boundary arc on $(\bar{\tau}_{to} - \delta, \bar{\tau}_{to} + \delta)$).

The proof of Theorem 3.4 is given in section 3.10 and will use two lemmas below. Note that by continuity of the mapping $(u, \mu) \mapsto g^\mu(y_u^\mu)$, it is immediate that all stationary points of a stable extension $(\mathcal{P}^\mu)$ with (u, μ) in the neighborhood of $(\bar{u}, \mu_0)$ satisfy (S1). Let us first define alternative multipliers needed in lemma 3.6 (see also [93, 88, 68, 53, 65] where these multipliers are used)

$$\eta^1(t) := \int_t^T \mathrm{d}\eta(s) = -\eta(t^+) \tag{3.32}$$

$$p^1(t) := p(t) - \eta^1(t) g_y^\mu(y(t)). \tag{3.33}$$

With this definition, and without any assumptions on the arc structure of the trajectory (i.e. without assuming a finite number of junction points), we have that

$$-\mathrm{d}p^1 = (H_y^\mu(u, y, p^1) + (g^\mu)_y^{(1)}(u, y)\eta^1)\mathrm{d}t,$$

and hence, the new alternative costate p^1 is absolutely continuous. Consequently, an equivalent form of (3.13)-(3.14) (when $\alpha = 1$) and (3.18) is, a.e. on $[0, T]$:

$$-\dot{p}^1(t) = H_y^\mu(u(t), y(t), p^1(t)) + (g^\mu)_y^{(1)}(u(t), y(t))\eta^1(t), \quad p^1(T) = \phi_y^\mu(y(T)) \tag{3.34}$$

$$0 = H_u^\mu(u(t), y(t), p^1(t)) + (g^\mu)_u^{(1)}(u(t), y(t))\eta^1(t). \tag{3.35}$$

In addition, (3.16) implies the following (weaker) relations, since η^1 is constant on interior arcs:

$$0 = (g^\mu)^{(1)}(u(t), y(t)) \quad \text{on boundary arcs,} \qquad 0 = \dot{\eta}^1(t) \quad \text{on interior arcs.} \tag{3.36}$$

Note that given a trajectory (u, y) of a stable extension $(\mathcal{P}^\mu)$, if (u, μ) is close enough to $(\bar{u}, \mu_0)$, Robinson's constraint qualification (3.17) still holds. This implies the uniqueness of the multipliers associated with a stationary point (u, y) of $(\mathcal{P}^\mu)$ with (u, μ) in the neighborhood of $(\bar{u}, \mu_0)$. The two lemmas below, used in the proof of Th. 3.4, are proved in section 3.10.

Lemma 3.6. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P}^{\mu_0})$ satisfying (A2)-(A3) with multipliers $(\bar{p}, \bar{\eta})$, and let the associated alternative multipliers $(\bar{p}^1, \bar{\eta}^1)$ be given by* (3.32)-(3.33)*. Consider a stable extension $(\mathcal{P}^\mu)$, and let $(u_n, y_n = y_{u_n}^{\mu_n})$ be a stationary point of $(\mathcal{P}^{\mu_n})$, such that $u_n \to \bar{u}$ in L^∞ and $\mu_n \to \mu_0$. Denote by p_n, η_n the (unique) multipliers associated with (u_n, y_n), and let p_n^1, η_n^1 be given by* (3.32)-(3.33)*. Then:*

1. *The sequence $(\mathrm{d}\eta_n)$ is bounded in $\mathcal{M}[0, T]$;*
2. *$\|\mathrm{d}\eta_n - \mathrm{d}\bar{\eta}\|_{1,\infty *} \to 0$, where $\|\cdot\|_{1,\infty *}$ denote the norm of the dual of $W^{1,\infty}$ for the strong topology;*

3. $p_n^1 \to \bar{p}^1$ *uniformly over* $[0,T]$*;*

4. $\eta_n^1 \to \bar{\eta}^1$ *uniformly over* $[0,T]$*.*

Remark 3.7. Note that under the assumptions of Lemma 3.6, by (3.33) and (3.32), we deduce the uniform convergence of (p_n, η_n) towards $(\bar{p}, \bar{\eta})$.

The key tool for deriving the structural stability result of Th. 3.4 is the following lemma.

Lemma 3.8. *Let* $(\bar{u}, \bar{y})$ *be a stationary point of* $(\mathcal{P}^{\mu_0})$ *satisfying (A2)-(A6), and let* $\bar{\delta}$ *be defined as in Th. 3.4. Then for all* $0 < \delta < \bar{\delta}$ *and all stable extensions* $(\mathcal{P}^{\mu})$ *of* $(\mathcal{P}^{\mu_0})$*, there exists a neighborhood* $V_u \times V_\mu$ *of* $(\bar{u}, \mu_0)$ *in* $\mathcal{U} \times M$*, such that if* (u,y) *is a stationary point of* $(\mathcal{P}^{\mu})$ *with* $(u,\mu) \in V_u \times V_\mu$*, then* (u,y) *has* no interior arc *contained in* Ω_i^δ*, for all* $i = 1, \ldots, N$*.*

3.4 Statement of the main result

Let us first recall the second-order conditions of [18, 21]. Let the linearized control and state spaces be respectively $\mathcal{V} := L^2(0,T)$ and $\mathcal{Z} := H^1(0,T;\mathbb{R}^n)$, where $H^1(0,T) = W^{1,2}(0,T)$. The quadratic function over $\mathcal{V} \times \mathcal{Z}$ involved in the second-order conditions is:

$$\begin{aligned} \mathcal{J}(v,z) \;:=\; & \int_0^T H_{(u,y),(u,y)}(\bar{u},\bar{y},\bar{p})((v,z),(v,z))\mathrm{d}t \;+\; z(T)^* \phi_{yy}(\bar{y}(T))z(T) \\ & + \int_0^T z(t)^* g_{yy}(\bar{y}(t))z(t)\mathrm{d}\bar{\eta}(t) \end{aligned} \tag{3.37}$$

and the set of constraints (defining the critical cone):

$$\dot{z} = f_u(\bar{u},\bar{y})v + f_y(\bar{u},\bar{y})z \quad \text{on } [0,T], \qquad z(0) = 0 \tag{3.38}$$

$$g_y(\bar{y}(t))z(t) = 0 \qquad t \in \bar{\mathcal{I}}_b \tag{3.39}$$

$$g_y(\bar{y}(\tau))z(\tau) \le 0 \qquad \tau \in \bar{\mathcal{T}}_{to}, \tag{3.40}$$

where $\bar{\mathcal{I}}_b$ and $\bar{\mathcal{T}}_{to}$ denote respectively the union of boundary arcs and the set of touch points of $(\bar{u}, \bar{y})$.

Theorem 3.9 ([18, 21]). (i) *Let* $(\bar{u}, \bar{y})$ *be a local solution of* $(\mathcal{P})$ *satisfying (A2)-(A5). Then*

$$\mathcal{J}(v,z) \ge 0, \quad \textit{for all } (v,z) \in \mathcal{V} \times \mathcal{Z} \textit{ satisfying } (3.38)\text{-}(3.40). \tag{3.41}$$

(ii) *Let* $(\bar{u}, \bar{y})$ *be a stationary point of* $(\mathcal{P})$ *satisfying (A2)-(A5). Then*

$$\mathcal{J}(v,z) > 0, \quad \textit{for all } (v,z) \in \mathcal{V} \times \mathcal{Z},\ (v,z) \neq 0, \textit{ satisfying } (3.38)\text{-}(3.40), \tag{3.42}$$

iff $(\bar{u}, \bar{y})$ *is a local solution of* $(\mathcal{P})$ *satisfying the quadratic growth condition:*

$$\exists\, c, \rho > 0, \qquad J(u) \ge J(\bar{u}) + c\|u - \bar{u}\|_2^2, \quad \forall\, u \in \mathcal{U}\ ;\ G(u) \in K,\ \|u - \bar{u}\|_\infty \le \rho. \tag{3.43}$$

Let us recall that a quadratic form Q on an Hilbert space $\mathcal{H}$ is a *Legendre form*, if Q is weakly lower semicontinuous and if for all weakly convergent subsequence $(v_n) \in \mathcal{H}^{\mathbb{N}}$, say $v_n \rightharpoonup v$, we have that $v_n \to v$ strongly if $Q(v_n) \to Q(v)$. Using (A2) we can show that the quadratic form $\mathcal{J}$ is a Legendre form (see [74, 24]). This plays a role to obtain the no-gap second-order conditions of Th. 3.9.

In the stability and sensitivity analysis, we will use the condition below, stronger than (3.42):

$$\mathcal{J}(v,z) > 0, \quad \text{for all } (v,z) \in \mathcal{V} \times \mathcal{Z},\ (v,z) \neq 0, \text{ satisfying } (3.38)\text{-}(3.39). \tag{3.44}$$

Definition 3.10. Let $(\bar{u},\bar{y}) = (u^{\mu_0}, y^{\mu_0})$ be a stationary point of $(\mathcal{P}^{\mu_0})$. We say that $(\bar{u},\bar{y})$ satisfies the *uniform quadratic growth condition*, if for all stable extensions $(\mathcal{P}^{\mu})$ of $(\mathcal{P}^{\mu_0})$ satisfying (3.4), there exist $c,\rho > 0$ and an open neighborhood V_0 of μ_0, such that for all $\mu \in V_0$, there exists a *unique stationary point* (u^μ, y^μ) of $(\mathcal{P}^\mu)$ with $\|u^\mu - \bar{u}\|_\infty \leq \rho$, and this point satisfies

$$J^\mu(u) \geq J^\mu(u^\mu) + c\|u-u^\mu\|_2^2, \quad \forall\, u \ \in \mathcal{U}\ ;\ G^\mu(u) \in K,\ \|u-\bar{u}\|_\infty \leq \rho, \qquad \forall\, \mu \in V_0. \tag{3.45}$$

Of course (3.45) implies that (u^μ, y^μ) is a local solution of $(\mathcal{P}^\mu)$. Note that the constants c and ρ in the uniform growth condition (3.45) does not depend on μ.

The arc structure of the trajectory (in the sense of number and order of boundary arcs and touch points) *is not necessarily stable* under a small perturbation. However, by (A5), boundary arcs are locally preserved, and by (A6), the only three possibilities for a touch point is to become a boundary arc, remain a touch point or become inactive at a local solution of the perturbed problem, i.e. the solutions of the perturbed problems have a *neighboring arc structure* of active constraints to that of $(\bar{u},\bar{y})$ (see Th. 3.4). Below is our main result (together with Theorems 3.4 and 3.30), that will be proved later in section 3.6.

Theorem 3.11. *Let $(\bar{u},\bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)-(A6). Then assertions* (i) *and* (ii) *below are equivalent:*
(i) *The uniform quadratic growth (Def. 3.10) holds.*
(ii) *The strong second-order sufficient condition* (3.44) *holds.*

If either point (i) or (ii) is satisfied, for $\mu \in V_0$ denote by (u^μ, y^μ) the unique local solution of $(\mathcal{P}^\mu)$ with $\|u^\mu - \bar{u}\| \leq \rho$, and by (p^μ, η^μ) the (unique) associated multipliers. Then (u^μ, y^μ) has a neighboring structure to that of $(\bar{u},\bar{y})$, and the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu) \in C^0[0,T] \times C^1([0,T];\mathbb{R}^n) \times C^0([0,T];\mathbb{R}^{n*}) \times C^0[0,T]$ is Lipschitz continuous on V_0.

The above result implies that the solutions of the perturbed problems satisfy the quadratic growth condition (3.45), and hence the no-gap sufficient condition (3.42) by Th. 3.9(ii). The lemma below (proved at the end of section 3.6) shows that the strong second-order sufficient condition (3.44) remains satisfied as well for the perturbed problems (this will be useful for the analysis of the homotopy algorithm in section 3.9).

Lemma 3.12. *Under assumptions (A2)-(A6), if either point (i) or (ii) of Th. 3.11 is satisfied, then the locally unique stationary point (u^μ, y^μ) of $(\mathcal{P}^\mu)$ satisfies the strong second-order sufficient condition* (3.44)*, for μ close enough to μ_0.*

Remark 3.13. We show more precisely (see Lemma 3.26) that under assumptions (A2)-(A6) and point (i) or (ii) of Th. 3.11, then the *shooting parameters* associated with (u^μ, y^μ) (initial costate, jump parameters at entry times and all junction times, see the next section) are Lipschitz continuous functions of μ.

Related results to Theorem 3.11, based on a shooting approach (see the next section) too, are [93, Th. 8.3], where the existence of a locally unique local solution of $(\mathcal{P}^\mu)$ having the same structure as $(\bar{u},\bar{y})$ was shown (but the uniqueness of the stationary point or the converse implication "(i) $\Rightarrow$ (ii)" are not discussed), and [19, Th. 4.3][2], where only the uniqueness of stationary points satisfying some restrictions on the arc structure is argued. In addition, both results assume the absence of touch points for state constraints of first-order. Here we are able to show that (u^μ, y^μ) is locally the unique stationary point of $(\mathcal{P}^\mu)$ (see Lemma 3.29)

[2]Theorem 2.34 of this thesis.

thanks to the analysis done in section 3.3. As mentioned in the Introduction, this is difficult to compare to [88, 89, 53] where an infinite dimensional approach was used, which required weaker assumptions, e.g. (A4)-(A6) are not needed, so their results are more general than Th. 3.11, but the conclusions obtained are also weaker than those of Th. 3.11.

In section 3.7, we will provide the first-order expansion of the local optimal solution and associated multipliers of the perturbed problem (see Theorem 3.30).

3.5 Alternative and Shooting Formulations

3.5.1 Alternative formulation of optimality conditions

In presence of pure state constraints, a reformulation of the optimality conditions is needed to apply shooting methods. Our results are based on the following alternative formulation of optimality conditions, see e.g. [29, 75, 68, 98, 19]. We use in this alternative formulation another set of alternative multipliers, that we denote by (p_1, η_1), different from the alternative multipliers (p^1, η^1) used in section 3.3. Whereas the latter are continuous, (p_1, η_1) have jumps at entry points. The jumps of p_1 at entry times τ_{en}, denoted by $\nu^1_{\tau_{en}}$, are part of the *shooting parameters* used in the shooting algorithm.

Definition 3.14. A trajectory (u, y) is solution of the *alternative formulation*, if it has finitely many junction times $\mathcal{T}$ and $g^\mu(y(T)) < 0$, if $(u, y) \in PC^0_{\mathcal{T}}[0,T] \times PC^{1,0}_{\mathcal{T}}([0,T];\mathbb{R}^n)$ and if there exist $p_1 \in PC^1_{\mathcal{T}}([0,T];\mathbb{R}^{n*})$, $\eta_1 \in PC^1_{\mathcal{T}}[0,T]$, and alternative jump parameters $\nu^1_{\mathcal{T}_{en}}$ and $\nu_{\mathcal{T}_{to}}$, such that the following relations are satisfied, with the augmented Hamiltonian (3.8) (time dependence is omitted):

$$\begin{aligned}
\dot{y} &= f^\mu(u,y) \qquad \text{on } [0,T], \qquad y(0) = y_0^\mu && (3.46)\\
-\dot{p}_1 &= \tilde{H}^\mu_y(u,y,p_1,\eta_1) \qquad \text{on } [0,T]\setminus\mathcal{T} && (3.47)\\
0 &= \tilde{H}^\mu_u(u,y,p_1,\eta_1) \qquad \text{on } [0,T]\setminus\mathcal{T} && (3.48)\\
(g^\mu)^{(1)}(u,y) &= 0 \qquad \text{on } \mathcal{I}_b && (3.49)\\
\eta_1(t) &= 0 \qquad \text{on } [0,T]\setminus\mathcal{I}_b && (3.50)\\
p_1(T) &= \phi^\mu_y(y(T)) && (3.51)\\
g^\mu(y(\tau_{en})) &= 0, \qquad \tau_{en} \in \mathcal{T}_{en} && (3.52)\\
g^\mu(y(\tau_{to})) &= 0, \qquad \tau_{to} \in \mathcal{T}_{to} && (3.53)\\
[p_1(\tau_{en})] &= -\nu^1_{\tau_{en}} g^\mu_y(y(\tau_{en})), \qquad \tau_{en} \in \mathcal{T}_{en} && (3.54)\\
[p_1(\tau_{ex})] &= 0, \qquad \tau_{ex} \in \mathcal{T}_{ex} && (3.55)\\
[p_1(\tau_{to})] &= -\nu_{\tau_{to}} g^\mu_y(y(\tau_{to})), \qquad \tau_{to} \in \mathcal{T}_{to}. && (3.56)
\end{aligned}$$

A solution of the alternative formulation satisfies the *additional conditions*, if the conditions below hold:

$$\begin{aligned}
g^\mu(y(t)) &< 0 \qquad \text{on } [0,T]\setminus(\mathcal{I}_b \cup \mathcal{T}_{to}) && (3.57)\\
\dot{\eta}_1(t) &\le 0 \qquad \text{on } \operatorname{int}\mathcal{I}_b && (3.58)\\
\nu^1_{\tau_{en}} &= \eta_1(\tau_{en}^+), \quad \tau_{en} \in \mathcal{T}_{en}\ ; \qquad \eta_1(\tau_{ex}^-) = 0, \quad \tau_{ex} \in \mathcal{T}_{ex} && (3.59)\\
\nu_{\tau_{to}} &= 0 \qquad \tau \in \mathcal{T}_{to}. && (3.60)
\end{aligned}$$

Proposition 3.15 (See e.g. [116, 75, 68]). *Let $(\bar{u}, \bar{y})$ be a local solution of $(\mathcal{P})$, satisfying (A2)-(A4). Then $(\bar{u}, \bar{y})$ is solution of alternative formulation* (3.46)-(3.56)*, and satisfies additional conditions* (3.57)-(3.60)*.*

The following remarks comment on those optimality conditions and on the relations existing between the different sets of multipliers.

Remark 3.16. It can be shown (see [19, Prop. 2.10][3]) that under assumptions (3.22) (resp. (A2)) and (A3)-(A4), relations (3.46)-(3.60) characterize regular ($\alpha = 1$) Pontryagin extremals (resp. stationary points), and the (unique) *classical* multipliers $\mathrm{d}\eta \in \mathcal{M}_+[0,T]$ and $p \in BV([0,T];\mathbb{R}^{n*})$ of Def. 3.1 are given by (recall that we adopted the convention $\eta(T^+) = 0$):

$$\eta(t) = -\sum_{\tau\in\mathcal{T}_{en}} \nu^1_\tau \mathbf{1}_{[0,\tau)}(t) - \eta_1(t^+), \qquad p(t) = p_1(t) + \eta_1(t) g^\mu_y(y(t)), \tag{3.61}$$

with $\mathbf{1}_{[0,\tau)}(t) = 1$ if $0 \le t < \tau$ and zero otherwise. Equivalently, η is given by $\mathrm{d}\eta(t) = -\dot\eta_1(t)\mathrm{d}t$.

The classical multipliers (p,η) and alternative ones (p_1,η_1) can be recovered from each other by (3.61) and (3.59). By (3.54)-(3.56) and additional conditions (3.59)-(3.60), we have $(p,\eta) \in PC^{1,0}_{\mathcal{T}}([0,T];\mathbb{R}^{n*}) \times PC^{1,0}_{\mathcal{T}}[0,T]$. It is also easy to see that, when (3.61) holds, $\tilde H^\mu(\cdot,y,p_1,\eta_1) = H^\mu(\cdot,y,p)$, and hence, (3.20) is equivalent (with $\bar p_1$ and $\bar\eta_1$ the alternative multipliers associated with $\bar u$) to:

$$\tilde H_{uu}(\bar u(t),\bar y(t),\bar p_1(t^\pm),\bar\eta_1(t^\pm)) \ge \alpha, \quad \text{for all } t\in[0,T]. \tag{3.62}$$

Remark 3.17. On $[0,T]\setminus\mathcal{T}$, the multipliers η^1 and p^1 in section 3.3 are related to p_1 and η_1 by the following relations:

$$\eta^1(t) = \sum_{\tau\in\mathcal{T}_{en}} \nu^1_\tau \mathbf{1}_{[0,\tau)}(t) + \eta_1(t), \qquad p^1(t) = p_1(t) - \sum_{\tau\in\mathcal{T}_{en}} \nu^1_\tau \mathbf{1}_{[0,\tau)}(t) g^\mu_y(y(t)). \tag{3.63}$$

Remark 3.18. By (3.58)-(3.59), the following necessary condition holds:

$$\nu^1_{\tau_{en}} \;\ge\; 0, \qquad \tau_{en}\in\mathcal{T}_{en}. \tag{3.64}$$

Lemma 3.19. *Let (u,y) be a trajectory of $(\mathcal{P}^\mu)$ satisfying the alternative formulation. Assume that there exist $\alpha,\beta,\varepsilon > 0$ such that (we denote here $[u(t^-),u(t^+)] := \{(1-\sigma)u(t^-) + \sigma u(t^+)\ ;\ \sigma\in[0,1]\}$)*

$$\alpha \;\le\; \tilde H^\mu_{uu}(\hat u, y(t), p_1(t^\pm), \eta_1(t^\pm)) \qquad \textit{for all } \hat u\in[u(t^-),u(t^+)] \textit{ and all } t\in[0,T] \tag{3.65}$$

$$\beta \;\le\; |(g^\mu)^{(1)}_u(\hat u, y(t))| \qquad \textit{for all } \hat u\in[u(t^-),u(t^+)] \textit{ and all } t : \mathrm{dist}\{t; I(g^\mu(y))\} \le \varepsilon. \tag{3.66}$$

Then (3.59) *is equivalent to the condition below*

$$(g^\mu)^{(1)}(u(\tau^-_{en}),y(\tau_{en})) = 0, \quad \tau_{en}\in\mathcal{T}_{en}\ ; \qquad (g^\mu)^{(1)}(u(\tau^+_{ex}),y(\tau_{ex})) = 0, \quad \tau_{ex}\in\mathcal{T}_{ex}. \tag{3.67}$$

Also (3.59) *or* (3.67) *is equivalent to the continuity of the control at entry/exit points.*

Proof. We recall here the proof (see [93] and [19, Prop. 2.15][4].) since the arguments will be used later in Lemma 3.27. Since $(g^\mu)^{(1)}(u(\tau^+_{en}),y(\tau_{en})) = 0 = (g^\mu)^{(1)}(u(\tau^-_{ex}),y(\tau_{ex}))$, by (3.66), (3.67) is equivalent to the continuity of the control at entry and exit times. Now let $\tau\in\mathcal{T}_{en}$. By (3.48) and (3.54),

$$\begin{aligned}\tilde H^\mu_u(u(\tau^-),y(\tau),p_1(\tau^-),\eta_1(\tau^-)) &= \tilde H^\mu_u(u(\tau^+),y(\tau),p_1(\tau^+),\eta_1(\tau^+))\\ &= \tilde H^\mu_u(u(\tau^+),y(\tau),p_1(\tau^-),\eta_1(\tau^+)-\nu^1_\tau).\end{aligned}$$

[3]Proposition 2.10 of this thesis.
[4]Proposition 2.15 of this thesis.

If (3.59) holds, then we obtain (since $\eta_1(\tau^-) = 0$)

$$\tilde{H}^\mu_u(u(\tau^-), y(\tau), p_1(\tau^-), \eta_1(\tau^-)) \;=\; \tilde{H}^\mu_u(u(\tau^+), y(\tau), p_1(\tau^-), \eta_1(\tau^-)),$$

which implies by (3.65) that $u(\tau^-) = u(\tau^+)$. Conversely, if (3.67) holds, i.e. if u is continuous at τ, then we obtain

$$(\eta_1(\tau^+) - \nu^1_\tau)(g^\mu)^{(1)}_u(u(\tau), y(\tau)) = 0.$$

Since by (3.66), $(g^\mu)^{(1)}_u(u, y) \neq 0$, we obtain the result. Similar arguments hold at exit points. □

Remark 3.20. By (3.56) and (3.60), (3.48) and hypothesis (3.65), we can show similarly that a solution (u, y) of the alternative formulation and additional conditions satisfying (3.65)-(3.66) is such that u is also continuous at touch points, and hence $(u, y) \in PC^{1,0}_{\mathcal{T}}[0, T] \times PC^{2,1}_{\mathcal{T}}([0, T]; \mathbb{R}^n)$.

Remark 3.21. At a touch point τ_{to}, the function $t \mapsto g^\mu(y(t))$ has a local isolated maximum, and a continuous derivative at τ_{to} (due to the continuity of u), hence the condition below is satisfied (compare to (3.67)):

$$(g^\mu)^{(1)}(u(\tau_{to}), y(\tau_{to})) = 0, \quad \tau \in \mathcal{T}_{to}. \tag{3.68}$$

The next lemma provides in particular a proof for Lemma 3.3.

Lemma 3.22. *Let (u, y) be a trajectory of $(\mathcal{P}^\mu)$ solution of the alternative formulation and additional conditions. Assume that there exist $\alpha, \beta, \varepsilon > 0$ such that* (3.65) *and* (3.66) *holds. Then, for all $\tau_{en} \in \mathcal{T}_{en}$ and $\tau_{ex} \in \mathcal{T}_{ex}$,*

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g^\mu(y(t))|_{t=\tau^-_{en}} < 0 \quad \textit{iff} \quad \dot\eta_1(\tau^+_{en}) < 0 \;; \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2} g^\mu(y(t))|_{t=\tau^+_{ex}} < 0 \quad \textit{iff} \quad \dot\eta_1(\tau^-_{ex}) < 0. \tag{3.69}$$

Proof. Let $\tau_{en} \in \mathcal{T}_{en}$. We omit in the proof the superscript $^\mu$ on $\tilde{H}$, g and f. Derivation w.r.t. time of the relation (3.48) on the left and right neighborhood of τ_{en} yields (omitting the dependence in t and arguments (u, y, p_1, η_1) of $\tilde{H}$):

$$\tilde{H}_{uu}\dot{u} + \tilde{H}_{uy} f(u, y) - \tilde{H}_y f_u(u, y) + g^{(1)}_u(u, y)\dot\eta_1 \;=\; 0. \tag{3.70}$$

Recall that $g^{(1)}(u, y) = g_y(y)f(u, y)$. By Lemma 3.19 and (3.59), u is continuous, so it follows that, taking the jumps at time τ_{en} (omitting again arguments and setting $\nu^1 := \nu^1_{\tau_{en}}$):

$$\begin{aligned}
[\tilde{H}_{uu}] \;&=\; [p_1]f_{uu} + [\eta_1]g^{(1)}_{uu} \;=\; -\nu^1 g_y f_{uu} + \nu^1 g^{(1)}_{uu} \;=\; 0,\\
[\tilde{H}_{uy}]f - [\tilde{H}_y]f_u \;&=\; ([p_1]f_{uy} + [\eta_1]g^{(1)}_{uy})f - ([p_1]f_y + [\eta_1]g^{(1)}_y)f_u\\
&=\; (-\nu^1 g_y f_{uy} + \nu^1 g^{(1)}_{uy})f - (-\nu^1 g_y f_y + \nu^1 g^{(1)}_y)f_u \;=\; 0.
\end{aligned}$$

Taking then the jump in (3.70) at time τ_{en}, the above relations imply that

$$\tilde{H}_{uu}[\dot{u}] + g^{(1)}_u[\dot\eta_1] \;=\; 0. \tag{3.71}$$

Since u, y, p_1 and η_1 are all continuous at exit times by Lemma 3.19, (3.71) holds as well at exit times. Since the function $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g(y(t)) = g^{(2)}(\dot{u}, u, y)$, with $g^{(2)}$ given by (3.5), vanishes on (τ_{en}, τ_{ex}), and (u, y) is continuous, we have by (3.66) that $g^{(2)}(\dot{u}, u, y)$ is discontinuous at τ iff $\dot{u}$ is, and hence by (3.71) and (3.65)-(3.66) iff $\dot\eta_1$ is. Since $\dot\eta_1 = 0$ locally outside (τ_{en}, τ_{ex}), and $\dot\eta_1 \leq 0$ on (τ_{en}, τ_{ex}) by (3.58), the result follows. □

Remark 3.23. We know by [19, Lemma 3.6][5] that we can express the quadratic cost $\mathcal{J}$, using $(\bar{p}_1, \bar{\eta}_1)$ defined by (3.61) instead of $(\bar{p}, \bar{\eta})$, over the space of linearized trajectories (v, z) satisfying (3.38), by $\mathcal{J}(v, z) = \mathcal{J}_1(v, z)$, with

$$\begin{aligned} \mathcal{J}_1(v,z) \;:=\; & \int_0^T \tilde{H}_{(u,y),(u,y)}(\bar{u},\bar{y},\bar{p}_1,\bar{\eta}_1)((v,z),(v,z))\mathrm{d}t \\ & + z(T)^* \phi_{yy}(\bar{y}(T))z(T) \;+\; \sum_{\tau \in \bar{\mathcal{T}}_{en}} \bar{\nu}^1_\tau z(\tau)^* g_{yy}(\bar{y}(\tau))z(\tau), \end{aligned} \tag{3.72}$$

where $\tilde{H}$ is the augmented Hamiltonian (3.8), and the constraint (3.39) is equivalent to

$$g_y(\bar{y}(\tau))z(\tau) = 0 \qquad \tau \in \bar{\mathcal{T}}_{en} \tag{3.73}$$

$$g^{(1)}_{(u,y)}(\bar{u}(t),\bar{y}(t))(v(t),z(t)) = 0 \qquad t \in \bar{\mathcal{I}}_b. \tag{3.74}$$

Remark 3.24. The second-order sufficient condition (3.44) used in the stability and sensitivity analysis, is equivalent by Rem. 3.23 to

$$\mathcal{J}_1(v,z) > 0, \quad \text{for all } (v,z) \in \mathcal{V} \times \mathcal{Z},\ (v,z) \neq 0, \text{ satisfying (3.38) and (3.73)-(3.74).} \tag{3.75}$$

This condition is weaker than the one in [93], where the entry-point constraint (3.73) is omitted. The authors present a numerical method, based on Riccati equations, allowing to check the coercivity of the quadratic form $\mathcal{J}_1$ over the subspace defined by (3.38) and (3.74), which is of interest in applications, while the verification of (3.44) or (3.75) in practice remains open.

3.5.2 Shooting formulation with nonessential touch points

By (A2)-(A4), applying the implicit function theorem to (3.48)-(3.50), we may express the algebraic variables (u, η_1) on each arc as C^1 functions of the differential variables (y, p_1). Denote by $\mathcal{F}^\mu_b$ and $\mathcal{F}^\mu_i$ the flows on (y, p_1) obtained respectively on boundary and interior arcs, by eliminating the algebraic variables, and write $(y, p_1)(t) = (y(t), p_1(t))$. On each arc (t_1, t_2), we have that

$$(y,p_1)(t_2^-) = \mathcal{F}^\mu_a((y,p_1)(t_1^+), t_2 - t_1) \tag{3.76}$$

where $\mathcal{F}^\mu_a$ equals $\mathcal{F}^\mu_b$ for a boundary arc, and $\mathcal{F}^\mu_i$ for an interior arc. So we can (and this is precisely the idea of shooting methods) describe the alternative optimality system (3.46)-(3.56) as a sequence of applications of mappings $\mathcal{F}^\mu_b$ and $\mathcal{F}^\mu_i$, combined with junction conditions. Note that the mappings $(x, t_1, t_2) \to \mathcal{F}^\mu_a(x, t_2 - t_1)$, $a = i, b$, are (locally) C^1 w.r.t. all arguments, and allow in particular $t_2 - t_1$ to be nonpositive.

Now let us view a touch point as a boundary arc of zero length. This makes sense since, as we will see later, under a small perturbation, a touch point may switch into a boundary arc. So we have an entry point and an exit point, τ_{en} and τ_{ex}, whose common value is the one of the touch point. The jump $\nu^1_{\tau_{en}}$ at entry point τ_{en} equals $\nu_{\tau_{to}}$ (i.e., zero). There is a zero jump of p_1 at the entry (and exit) time τ_{en}.

Assume that we have N_{ba} boundary arcs and N_{to} touch points. Let $N := N_{ba} + N_{to}$. We have now N entry and N exit points. Denote by t^{en} (resp. t^{ex}) the N dimensional vector of entry (resp. exit) points, taken in the chronological order, and $\nu^1_i := \nu^1_{t^{en}_i}$. We use the notation

[5] Lemma 2.26 of this thesis.

$t_0^{ex} := 0$ and $t_{N+1}^{en} := T$. We may rewrite the alternative formulation as follows, taking into account the continuity of state and of costate at exit points:

$$
\begin{aligned}
(y,p_1)(0) &= (y_0^\mu, p_0) && (3.77)\\
(y,p_1)(t_i^{en-}) &= \mathcal{F}_i^\mu((y,p_1)(t_{i-1}^{ex}), t_i^{en} - t_{i-1}^{ex}), \quad i = 1,\dots,N+1, && (3.78)\\
(y,p_1)(t_i^{ex}) &= \mathcal{F}_b^\mu((y,p_1)(t_i^{en+}), t_i^{ex} - t_i^{en}), \quad i = 1,\dots,N, && (3.79)\\
[p_1(t_i^{en})] &= -\nu_i^1 g_y^\mu(y(t_i^{en})), \quad i = 1,\dots,N, && (3.80)\\
p_1(T) &= \phi_y^\mu(y(T)) && (3.81)\\
g^\mu(y(t_i^{en})) &= 0, \quad i = 1,\dots,N, && (3.82)
\end{aligned}
$$

where $p_0 \in \mathbb{R}^{n*}$ denotes the initial value of the costate.

We come now to the definition of the shooting mapping. Let $\Theta := \mathbb{R}^n \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N$ be the space of shooting parameters, of dimension $\bar{N} := n + 3N$. A vector of shooting parameters is denoted by

$$\theta = (p_0^*, \nu^1, t^{en}, t^{ex}) \in \Theta. \tag{3.83}$$

The shooting mapping F is defined over a neighborhood $V_\theta \times V_\mu$ of (θ_0, μ_0) in $\mathbb{R}^{\bar{N}} \times M_0$ into $\mathbb{R}^{\bar{N}}$, by

$$F(\theta,\mu) = \begin{pmatrix} p_1(T) - \phi_y^\mu(y(T)) \\ g^\mu(y(t^{en})) \\ (g^\mu)^{(1)}(u(t^{en-}), y(t^{en})) \\ (g^\mu)^{(1)}(u(t^{ex+}), y(t^{ex})) \end{pmatrix}, \tag{3.84}$$

where the values of (y, p_1, u) at times $t_i^{en\pm}, t_i^{ex\pm}, T$ are given by (3.77)-(3.80), and where we used e.g. the notation

$$(g^\mu)^{(1)}(u(t^{en-}), y(t^{en})) := \Big((g^\mu)^{(1)}(u(t_i^{en-}), y(t_i^{en}))\Big)_{1\le i\le N} \in \mathbb{R}^N.$$

Being a composition of C^1 mappings, the shooting mapping is itself locally of class C^1.

Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$, satisfying (A2)-(A4), with finite set of junction times $\bar{\mathcal{T}}$. Let I_{ba} and I_{to} denote the (disjoint) sets of index in $\{1,\dots,N\}$ corresponding respectively to boundary arcs and touch points of the trajectory $(\bar{u}, \bar{y})$. Split F into two components:

$$F(\theta,\mu) = (\Phi(\theta,\mu)^*, \Psi(\theta,\mu)^*)^*,$$

where Ψ corresponds to the components $g^\mu(y(t_i^{en}))$ for $i \in I_{to}$, denoted by the vector $g^\mu(y(t_{to}^{en})) \in \mathbb{R}^{N_{to}}$. Denote similarly by ν_{to}^1 the vector of components ν_i^1, for $i \in I_{to}$. Consider the following nonlinear complementarity problem, for μ close to μ_0:

$$\text{Find} \quad \theta \in \Theta \quad \text{such that} \qquad \Phi(\theta,\mu) = 0 \quad \text{and} \quad \Psi(\theta,\mu) \in N(\theta), \tag{3.85}$$

where

$$N(\theta) := \begin{cases} \mathbb{R}_-^{N_{to}} \cap (\nu_{to}^1)^\perp & \text{if } \nu_{to}^1 \in \mathbb{R}_+^{N_{to}}, \\ \emptyset & \text{otherwise.} \end{cases} \tag{3.86}$$

Note that by (3.81)-(3.82) and (3.67)-(3.68), $\theta_0 := (\bar{p}_1(0)^*, \bar{\nu}^1, \bar{t}^{en}, \bar{t}^{ex})$ is solution of (3.85) for $\mu = \mu_0$, with $\bar{t}^{en}$ and $\bar{t}^{ex}$ the vectors of times in $\bar{\mathcal{T}}_{en} \cup \bar{\mathcal{T}}_{to}$ and $\bar{\mathcal{T}}_{ex} \cup \bar{\mathcal{T}}_{to}$ respectively, in increasing order, $\bar{\nu}_i^1 = \bar{\nu}_{\bar{t}_i^{en}}^1$ if $i \in I_{ba}$, and $\bar{\nu}_i^1 = 0$ if $i \in I_{to}$.

It should be underlined that we allow, in formulation of problem (3.85), entry times to be greater than exit times. However, we will check in the next section, after having shown that (3.85) has a locally unique solution, that the constraint $\nu^1_{to} \geq 0$ in (3.85) (compare with (3.64)) is sufficient, with assumption (A6), to ensure locally for μ in the neighborhood of μ_0 that the solution of (3.85) is such that $t_i^{en} \leq t_i^{ex}$ for all $i \in I_{to}$. In addition, we will show that by (3.26), strict complementarity $\dot{\eta}_1 < 0$ holds on the boundary arc (t_i^{en}, t_i^{ex}) whenever $t_i^{en} < t_i^{ex}$.

As we will see, the formulation (3.85) is strongly related with the associated linear-quadratic tangent problem $\min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \mathcal{J}_1(v,z)$ subject to the equality constraints (3.38) and (3.73)-(3.74), *and the inequality constraint* (3.40).

Remark 3.25. When the state constraint is of higher order, under small perturbations, a nonessential touch point satisfying (3.26) cannot switch into a boundary arc, i.e. it either becomes inactive, remains nonessential, or becomes an essential touch point (with a nonzero jump of the costate), see [19].

3.6 Stability Analysis

In problem (3.85), there are inequality constraints that cannot be reduced to equality ones since strict complementarity does not hold at touch points, and those inequality constraints introduce nonsmoothness. Therefore we cannot apply the classical implicit function Theorem as it is done in [93]. Our stability analysis uses the notion of strong regularity, introduced by Robinson in [121], applied to the complementarity problem (3.85).

The point θ_0 solution of (3.85) for $\mu = \mu_0$ is *strongly regular*, if there exist neighborhoods (V'_θ, V_δ) in $\mathbb{R}^{\bar{N}} \times \mathbb{R}^{\bar{N}}$ of $(\theta_0, 0)$, such that, for all $\delta \in V_\delta$, $\delta = (\delta_1, \delta_2) \in \mathbb{R}^{\bar{N}-N_{to}} \times \mathbb{R}^{N_{to}}$, there exists in V'_θ a unique solution θ of:

$$\begin{cases} D_\theta \Phi(\theta_0, \mu_0)(\theta - \theta_0) - \delta_1 = 0 \\ D_\theta \Psi(\theta_0, \mu_0)(\theta - \theta_0) - \delta_2 \in N(\theta) \end{cases} \tag{3.87}$$

and the mapping $\Xi : \delta \to \theta(\delta)$ is Lipschitz continuous over V_δ.

If θ_0 is strongly regular, then by [121], there exist neighborhoods (V_θ, V_μ) of (θ_0, μ_0), such that for each $\mu \in V_\mu$, (3.85) has in V_θ a unique solution θ^μ,

$$\theta^\mu = (p_0^{\mu *}, \nu^{\mu,1}, t^{\mu,en}, t^{\mu,ex}), \tag{3.88}$$

and there exists $\kappa > 0$ such that for all $\mu, \mu' \in V_\mu$,

$$|\theta^\mu - \theta^{\mu'}| \leq \kappa \|\mu - \mu'\|. \tag{3.89}$$

In addition, the following expansion of θ^μ holds (see e.g. [24], p.413 equation (5.41)):

$$\theta^\mu = \Xi(-D_\mu F(\theta_0, \mu_0)(\mu - \mu_0)) + o(\|\mu - \mu_0\|). \tag{3.90}$$

Lemma 3.26. *Under assumptions (A2)-(A6),* (3.44) *implies that* θ_0 *is a strongly regular solution of* (3.85) *for* $\mu = \mu_0$. *More precisely, given* $\delta = (\delta_1, \delta_2) \in \mathbb{R}^{\bar{N}-N_{to}} \times \mathbb{R}^{N_{to}}$, $\delta_1 = (a_T, b_{ba}, c^{en}, c^{ex}) \in \mathbb{R}^n \times \mathbb{R}^{N_{ba}} \times \mathbb{R}^N \times \mathbb{R}^N$, $\delta_2 = b_{to}$, *there exists a unique* $\omega \in \Theta$, $\omega = (\pi_0^*, \gamma^1, \sigma^{en}, \sigma^{ex})$, *solution of the following relation, equivalent to* (3.87) *with* $\omega = \theta - \theta_0$*:*

$$\begin{cases} D_\theta \Phi(\theta_0, \mu_0)\omega - \delta_1 = 0 \\ D_\theta \Psi(\theta_0, \mu_0)\omega - \delta_2 \in N(\omega), \end{cases} \tag{3.91}$$

and ω is given as follows. Let $(v_\delta, z_\delta, \pi_\delta, \zeta_\delta, \lambda^1_\delta)$ be the unique solution and associated multipliers of the following linear-quadratic problem (recall that $\mathcal{J}_1$ is given by (3.72)*)*

$$
\begin{aligned}
(\mathcal{P}^\delta) \qquad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \quad & \tfrac{1}{2}\mathcal{J}_1(v,z) + a_T^* z(T) && (3.92)\\
\text{subject to} \quad & \text{(3.38), (3.74)},\\
& g_y(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) = b_i, \quad i \in I_{ba} && (3.93)\\
& g_y(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) \le b_i, \quad i \in I_{to}, && (3.94)
\end{aligned}
$$

where the multipliers π_δ, ζ_δ and λ^1_δ are associated, respectively, with constraint (3.38)*,* (3.74) *and* (3.93)-(3.94)*. Then ω is given by: $\pi_0 = \pi_\delta(0)$, $\gamma^1 = \lambda^1_\delta$, and*

$$
\sigma_i^{en} = \frac{c_i^{en} - g^{(1)}_{(u,y)}(\bar u(\bar t_i^{en}), \bar y(\bar t_i^{en}))(v_\delta(\bar t_i^{en-}), z_\delta(\bar t_i^{en}))}{\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(\bar u, \bar y)|_{t=\bar t_i^{en-}}}, \quad i = 1, \dots, N, \tag{3.95}
$$

$$
\sigma_i^{ex} = \frac{c_i^{ex} - g^{(1)}_{(u,y)}(\bar u(\bar t_i^{ex}), \bar y(\bar t_i^{ex}))(v_\delta(\bar t_i^{ex+}), z_\delta(\bar t_i^{ex}))}{\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(\bar u, \bar y)|_{t=\bar t_i^{ex+}}}, \quad i = 1, \dots, N. \tag{3.96}
$$

Proof. The proof uses the block-decoupling property of the Jacobian of the shooting mapping w.r.t. junction times for first-order state constraints established in [93, Lemma 4.2]. See also [19, Lemma 4.5][6]. Let us first explicit the relation (3.91). Let (v, z, π, ζ) be the linearized control, state, costate and state constraint multiplier solution of the linearized shooting equations (3.77)-(3.80):

$$
\begin{aligned}
(z,\pi_1)(0) &= (0,\pi_0) && (3.97)\\
(z,\pi_1)(\bar t_i^{en-}) &= D\mathcal{F}_i^{\mu_0}((\bar y,\bar p_1)(\bar t_{i-1}^{ex}), \bar t_i^{en} - \bar t_{i-1}^{ex})(z,\pi_1)(\bar t_{i-1}^{ex}), \quad i = 1,\dots,N+1, && (3.98)\\
(z,\pi_1)(\bar t_i^{ex}) &= D\mathcal{F}_b^{\mu_0}((\bar y,\bar p_1)(\bar t_i^{en+}), \bar t_i^{ex} - \bar t_i^{en})(z,\pi_1)(\bar t_i^{en+}), \quad i = 1,\dots,N, && (3.99)\\
[\pi_1(\bar t_i^{en})] &= -\bar\nu_i^1 g_{yy}(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) - \gamma_i^1 g_y(\bar y(\bar t_i^{en})), \qquad i = 1,\dots,N. && (3.100)
\end{aligned}
$$

Then (3.91) writes

$$
\begin{aligned}
&\pi_1(T) = \phi_{yy}(\bar y(T))z(T) + a_T && (3.101)\\
&g_y(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) = b_i, \quad i \in I_{ba} && (3.102)\\
&g_y(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) \le b_i, \quad \gamma_i^1 \ge 0, \quad (g_y(\bar y(\bar t_i^{en}))z(\bar t_i^{en}) - b_i)\gamma_i^1 = 0, \quad i \in I_{to} && (3.103)\\
&Dg^{(1)}(\bar u(\bar t_i^{en}), \bar y(\bar t_i^{en}))(v(\bar t_i^{en-}), z(\bar t_i^{en})) + \sigma_i^{en}\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(\bar u,\bar y)|_{t=\bar t_i^{en-}} = 0, \quad i = 1,\dots,N && (3.104)\\
&Dg^{(1)}(\bar u(\bar t_i^{ex}), \bar y(\bar t_i^{ex}))(v(\bar t_i^{ex+}), z(\bar t_i^{ex})) + \sigma_i^{ex}\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(\bar u,\bar y)|_{t=\bar t_i^{ex+}} = 0, \quad i = 1,\dots,N. && (3.105)
\end{aligned}
$$

We recognize that (3.97)-(3.103) is the first-order optimality condition of problem $(\mathcal{P}_\delta)$, with γ_i^1 the multipliers associated with the constraints (3.93) and (3.94) for i in respectively I_{ba} and I_{to}. By (A2), we can show that the quadratic form $\mathcal{J}_1$ is a *Legendre form* over the space of linearized trajectories (v, z) satisfying (3.38). Therefore, (3.44), equivalent to (3.75) by Rem. 3.24, implies that $\mathcal{J}_1$ is *uniformly positive* over the linear space of $(v, z) \in \mathcal{V} \times \mathcal{Z}$ satisfying (3.38) and (3.73)-(3.74) (i.e. there exists $\alpha > 0$ such that $\mathcal{J}_1(v,z) \ge \alpha(\|v\|^2_{\mathcal{V}} + \|z\|^2_{\mathcal{Z}})$ for all $(v, z) \in \mathcal{V} \times \mathcal{Z}$ satisfying (3.38) and (3.73)-(3.74)). It follows then that problem $(\mathcal{P}^\delta)$ has, for all $\delta \in \mathbb{R}^{\bar N}$, a unique solution and multipliers $(v_\delta, z_\delta, \pi_\delta, \zeta_\delta, \lambda^1_\delta)$ that are Lipschitz continuous w.r.t. δ. Thus (3.91) has a unique solution, and by (3.104)-(3.105) and (A6) and (3.27), the variations of junction times σ_i^{en} and σ_i^{ex} are given by (3.95)-(3.96). □

[6] Lemma 2.36 of this thesis.

Lemma 3.27. *Under assumptions (A2)-(A6) and* (3.44)*, there exists a neighborhood* V_μ *of* μ_0*, such that the locally unique solution* θ^μ *of* (3.85) *given by* (3.88) *satisfies:*

$$t_i^{\mu,ex} \geq t_i^{\mu,en}, \qquad \text{for all } i \in I_{to} \tag{3.106}$$

and

$$t_i^{\mu,ex} = t_i^{\mu,en} \;\Leftrightarrow\; \nu_i^{\mu,1} = 0, \qquad i \in I_{to}. \tag{3.107}$$

In particular, the solution $(u^\mu, y^\mu, p_1^\mu, \eta_1^\mu)$ *of* (3.77)-(3.80) *with* $\theta = \theta^\mu$ *is well-defined over* $[0,T]$*, and there exists a constant* $\gamma > 0$*, such that for all* $i \in I_{to}$ *and all* $\mu \in V_\mu$*:*

$$\dot\eta_1^\mu(t) < -\gamma \quad \text{on} \quad [t_i^{\mu,en}, t_i^{\mu,ex}] \quad \text{whenever } t_i^{\mu,ex} > t_i^{\mu,en}. \tag{3.108}$$

Proof. Let $i \in I_{to}$. By strong regularity (Lemma 3.26), we have that

$$t_i^{\mu,ex} - t_i^{\mu,en} = \mathcal{O}(\|\mu - \mu_0\|), \qquad \nu_i^{\mu,1} = \mathcal{O}(\|\mu - \mu_0\|). \tag{3.109}$$

Denote by (u, y, p_1, η_1) the solution of (3.77)-(3.80) for $\theta = \theta^\mu$. Note that this is well-defined *on each arc*, but not a priori as function of time, since it may take several values for $t \in (\!(t_i^{\mu,en}, t_i^{\mu,ex})\!)$ if $t_i^{\mu,en} > t_i^{\mu,ex}$ (where $(\!(a,b)\!)$ stands for (a,b) if $a \leq b$ and (b,a) otherwise). We will see that this last case cannot occur, i.e. (3.106) holds (and clearly also holds by continuity with a strict inequality for $i \in I_{ba}$), and is satisfied with equality iff $\nu_i^{\mu,1} = 0$.

Note first that by (A2)-(A3) and the strong regularity property, for $\|\mu - \mu_0\|$ small enough, (3.65)-(3.66) are satisfied on each arc. Suppose first that $t_i^{\mu,ex} = t_i^{\mu,en}$. Then (u, y, η_1, p_1) is defined as function of time without ambiguity in the neighborhood of $t_i^{\mu,en}$ (the algebraic variables are given by the dynamics on interior arcs). By (3.77)-(3.80), there is a jump of p_1 at entry time and no jump at exit time, and thus $(y,p_1)(t_i^{\mu,en+}) = (y,p_1)(t_i^{\mu,ex-}) = (y,p_1)(t_i^{\mu,ex+})$. By definition of the problem (3.85), we have

$$(g^\mu)^{(1)}(u(t_i^{\mu,en-}), y(t_i^{\mu,en})) = (g^\mu)^{(1)}(u(t_i^{\mu,ex+}), y(t_i^{\mu,ex})) = 0,$$

and hence, since $t_i^{\mu,ex} = t_i^{\mu,en}$, (3.66) implies that u is continuous at time $t_i^{\mu,en}$. We deduce that:

$$0 = [H_u^\mu(u(t_i^{\mu,en}), y(t_i^{\mu,en}), p_1(t_i^{\mu,en}))] = -\nu_i^{\mu,1}(g^\mu)_u^{(1)}(u(t_i^{\mu,en}), y(t_i^{\mu,en})).$$

Since $(g^\mu)_u^{(1)}(u(t_i^{\mu,en}), y(t_i^{\mu,en})) \neq 0$ by (3.66), it follows that $\nu_i^{\mu,1} = 0$. This proves the "$\Rightarrow$" implication in (3.107).

Suppose now that $t_i^{\mu,ex} \neq t_i^{\mu,en}$. In order to avoid any confusion, denote the solution of (3.77)-(3.80) for $\theta = \theta^\mu$ by $(u^-, y^-, p_1^-, \eta_1^-)$ on the boundary arc $(\!(t_i^{\mu,en}, t_i^{\mu,ex})\!)$, and by $(u^+, y^+, p_1^+, \eta_1^+)$ on the succeeding interior arc $(t_i^{\mu,ex}, t_{i+1}^{\mu,en})$. Note that the limits of these functions and of their time derivative at endpoints of the interval where they are defined do exist, and are continuous w.r.t. μ (this follows from the implicit function Theorem applied by (3.65)-(3.66) on each arc of the trajectory). This holds in particular for $\dot u^\mu$. Here the jump has the following signification, for instance $[u(t_i^{\mu,ex})] := u^+(t_i^{\mu,ex}) - u^-(t_i^{\mu,ex})$.

Since (3.65)-(3.66) are satisfied, we can show using the same local arguments as in Lemma 3.19 that

$$(u^+, y^+, p_1^+, \eta_1^+)(t_i^{\mu,ex}) = (u^-, y^-, p_1^-, \eta_1^-)(t_i^{\mu,ex}), \tag{3.110}$$

and we denote this common value by $(u(t_i^{\mu,ex}), y(t_i^{\mu,ex}), p_1(t_i^{\mu,ex}), \eta_1(t_i^{\mu,ex}))$. By (A6), there exists by continuity a constant $c > 0$ such that, for μ close enough to μ_0,

$$\lim_{t \to t_i^{\mu,ex+}} \frac{\mathrm{d}}{\mathrm{d}t}(g^\mu)^{(1)}(u^+(t), y^+(t)) \;<\; -c. \tag{3.111}$$

On the other hand, we have on the boundary arc $((t_i^{\mu,en}, t_i^{\mu,ex}))$:

$$\lim_{t \to t_i^{\mu,ex}} \frac{\mathrm{d}}{\mathrm{d}t}(g^\mu)^{(1)}(u^-(t), y^-(t)) \;=\; 0. \tag{3.112}$$

Since $\frac{\mathrm{d}}{\mathrm{d}t}(g^\mu)^{(1)}(u^\pm(t), y^\pm(t)) = (g^\mu)^{(2)}(\dot{u}^\pm, u^\pm, y^\pm)$ with $(g^\mu)^{(2)}$ given by (3.5), the jump of $\dot{u}$ at $t_i^{\mu,ex}$ satisfies

$$(g^\mu)_u^{(1)}(u(t_i^{\mu,ex}), y(t_i^{\mu,ex}))[\dot{u}(t_i^{\mu,ex})] = [\frac{\mathrm{d}}{\mathrm{d}t}(g^\mu)^{(1)}(u(t), y(t))|_{t=t_i^{\mu,ex}}] < -c, \tag{3.113}$$

and hence, $\dot{u}^-(t_i^{\mu,ex}) \neq \dot{u}^+(t_i^{\mu,ex})$. By time-derivation of (3.48) on the boundary arc $((t_i^{\mu,en}, t_i^{\mu,ex}))$ of nonzero length and on the interior arc $(t_i^{\mu,ex}, t_{i+1}^{\mu,en})$, we obtain (omitting the arguments $(u^\pm(t), y^\pm(t), p_1^\pm(t), \eta_1^\pm(t))$):

$$\tilde{H}_{uu}^\mu \dot{u}^\pm + \tilde{H}_{yu}^\mu f^\mu - \tilde{H}_y^\mu f_u^\mu + (g^\mu)_u^{(1)} \dot{\eta}_1^\pm = 0. \tag{3.114}$$

Hence, taking the jump at time $t_i^{\mu,ex}$ gives, since (u, y, p_1, η_1) is continuous at $t_i^{\mu,ex}$ by (3.110):

$$\tilde{H}_{uu}^\mu(u, y, p_1, \eta_1)(t_i^{\mu,ex})[\dot{u}(t_i^{\mu,ex})] + (g^\mu)_u^{(1)}(u, y)(t_i^{\mu,ex})[\dot{\eta}_1(t_i^{\mu,ex})] = 0.$$

Since $\dot{\eta}_1^+(t_i^{\mu,ex}) = 0$, by (3.113) and (3.65)-(3.66) there exists by continuity a constant $C > 0$ such that, for $\|\mu - \mu_0\|$ small enough,

$$\dot{\eta}_1^-(t_i^{\mu,ex}) = -[\dot{\eta}_1(t_i^{\mu,ex})] = \frac{\tilde{H}_{uu}^\mu(u, y, p_1, \eta_1)(t_i^{\mu,ex})}{((g^\mu)_u^{(1)}(u, y)(t_i^{\mu,ex}))^2}(g^\mu)_u^{(1)}(u, y)(t_i^{\mu,ex})[\dot{u}(t_i^{\mu,ex})] < -C. \tag{3.115}$$

By (3.114) and time derivation of (3.49), we see that $\dot{\eta}_1^-(t)$ is given by a Lipschitz continuous function of time on $((t_i^{\mu,en}, t_i^{\mu,ex}))$, uniformly w.r.t. μ, so there exists $m > 0$ independent of μ, such that

$$\dot{\eta}_1^-(t) \;\leq\; -C + m|t_i^{\mu,ex} - t_i^{\mu,en}|, \qquad t \in ((t_i^{\mu,en}, t_i^{\mu,ex})). \tag{3.116}$$

In view of (3.109), this implies that $\dot{\eta}_1^-$ is negative on $((t_i^{\mu,en}, t_i^{\mu,ex}))$ for sufficiently small $\|\mu - \mu_0\|$, and consequently, $\eta_1^-(t_i^{\mu,en}) = \eta_1^-(t_i^{\mu,en}) - \eta_1^-(t_i^{\mu,ex})$ is nonzero and has the sign of $t_i^{\mu,ex} - t_i^{\mu,en}$. By similar arguments to Lemma 3.19, we can show that $\eta_1^-(t_i^{\mu,en}) = \nu_i^{\mu,1}$, and since $\nu_i^{\mu,1} \geq 0$ by definition of the problem (3.85), it follows that $t_i^{\mu,ex} > t_i^{\mu,en}$ necessarily holds whenever $t_i^{\mu,en} \neq t_i^{\mu,ex}$, which proves (3.106). In addition, (3.116) implies that $\nu_i^{\mu,1} = \eta_1(t_i^{\mu,en+}) > 0$ for μ close enough to μ_0, which show by contraposition the "$\Leftarrow$" implication in (3.107). Finally, relation (3.108) follows from (3.115) and (3.109), which completes the proof. □

Lemma 3.28. *Under assumptions (A2)-(A6) and* (3.44)*, the solution* $(u^\mu, y^\mu, p_1^\mu, \eta_1^\mu)$ *of* (3.77)-(3.80) *for* $\theta = \theta^\mu$*, where* θ^μ *is solution of* (3.85)*, is, for* $\|\mu - \mu_0\|$ *small enough, such that* (u^μ, y^μ) *is a stationary point of* $(\mathcal{P}^\mu)$*, with classical multipliers* (p^μ, η^μ) *given by* (3.61)*, and the mapping* $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu) \in C^0[0,T] \times C^1([0,T]; \mathbb{R}^n) \times C^0([0,T]; \mathbb{R}^{n*}) \times C^0[0,T]$ *is Lipschitz continuous on a neighborhood of* μ_0*.*

Proof. By Lemma 3.27, we see that $(u^\mu, y^\mu, p_1^\mu, \eta_1^\mu)$ is well-defined over $[0,T]$, and by definition of the problem (3.85), satisfies the alternative formulation (3.46)-(3.56). By (A2)-(A3), (3.65)-(3.66) hold for $\|\mu - \mu_0\|$ small enough, so Lemma 3.19 implies that the additional condition (3.59) is satisfied, and that u^μ is continuous on $[0,T]$, as well as η^μ and p^μ given by (3.61). In

view of Rem. 3.16, in order to show that (u^μ, y^μ) is a stationary point of $(\mathcal{P}^\mu)$ it remains to show that the additional conditions (3.57), (3.58) and (3.60) are satisfied. By (3.65)-(3.66), the implicit function Theorem applied on each arc shows that $\dot{u}^\mu(t_i^{\mu,en-})$ and $\dot{u}^\mu(t_i^{\mu,ex+})$ are continuous w.r.t. μ, for all $i = 1, \ldots, N$, as well as $\dot{\eta}_1^\mu(t_i^{\mu,en+})$ and $\dot{\eta}_1^\mu(t_i^{\mu,ex-})$ for $i \in I_{ba}$. So let $\|\mu - \mu_0\|$ be so small that, by (3.25)-(3.27) and (3.6),

$$\text{(i) } \frac{\mathrm{d}^2}{\mathrm{d}t^2} g^\mu(y^\mu(t))|_{t=t_i^{\mu,en-}, t_i^{\mu,ex+}} < 0,\ i = 1, \ldots, N, \qquad \text{(ii) } \dot{\eta}_1^\mu(t) \geq \frac{\beta}{2} \text{ on } (t_i^{\mu,en}, t_i^{\mu,ex}),\ i \in I_{ba}. \tag{3.117}$$

Let $i \in N_{to}$. If $\nu_i^{\mu,1} = 0$, then by Lemma 3.27, $t_i^{\mu,en} = t_i^{\mu,ex}$, u^μ and its time derivative are continuous at $t_i^{\mu,en}$, and $(g^\mu)^{(1)}(u^\mu(t_i^{\mu,en}), y^\mu(t_i^{\mu,en})) = 0$. By (A6) and standard continuity arguments, there exists $\varepsilon > 0$ such that $g^\mu(y^\mu(\cdot))$ attains its maximum over $(\bar{t}_i^{en} - \varepsilon, \bar{t}_i^{en} + \varepsilon)$ at the unique point $t_i^{\mu,en}$. Therefore if $g^\mu(y^\mu(t_i^{\mu,en})) < 0$, the state constraint is locally not active. If $g^\mu(y^\mu(t_i^{\mu,en})) = 0$, then $t_i^{\mu,en}$ is a touch point of the perturbed problem, and (3.60) holds by (3.107). If $\nu_i^{\mu,1} > 0$, then by Lemma 3.27, $t_i^{\mu,en} < t_i^{\mu,ex}$ and we have a boundary arc. By (3.108), additional condition (3.58) holds on this boundary arc. If $i \in I_{ba}$, then (3.58) holds on the boundary arc $(t_i^{\mu,en}, t_i^{\mu,ex})$ by (3.117)(ii). Finally, (3.57) holds near the junction points by (3.117)(i), and outside a small neighborhood of contact points, we obtain $g^\mu(y^\mu) < 0$ by a standard compactness argument. Hence (u^μ, y^μ) is a stationary point, with classical multipliers (p^μ, η^μ) given by (3.61).

Lipschitz continuity of the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu)$ follows from Lipschitz continuity of the mapping $\mu \mapsto \theta^\mu$ by strong regularity (Lemma 3.26), Lipschitz continuity of $(\theta, \mu) \mapsto (u, y, p, \eta)|_k$, where $(u, y, p, \eta)|_k$ denotes the restriction of the solution of (3.77)-(3.80) and (3.61) to "arc" k (possibly a singleton), for all $k = 1, \ldots, 2N + 1$, and continuity of u^μ, $\dot{y}^\mu$, p^μ and η^μ on $[0, T]$. □

Thanks to Th. 3.4, we can show that (u^μ, y^μ) is the *locally unique stationary point* of $(\mathcal{P}^\mu)$.

Lemma 3.29. *Under assumptions (A2)-(A6) and* (3.44)*, there exist a L^∞ neighborhood V_u of $\bar{u}$ and a neighborhood V_μ of μ_0, such that for all $\mu \in V_\mu$, (u^μ, y^μ) is the locally unique stationary point of $(\mathcal{P}^\mu)$ with $u \in V_u$.*

Proof. Let (u, y) be a stationary point of $(\mathcal{P}^\mu)$ with (u, μ) in the neighborhood of $(\bar{u}, \mu_0)$. By Th. 3.4, (u, y) satisfies (S1)-(S3), and therefore has finitely many junction times, so it makes sense to speak of the finite-dimensional vector of "shooting parameters" θ (initial costate, jump parameters at entry times, and junction times) such that (u, y) is solution of the alternative formulation (Def. 3.14). Now construct its *augmented set of shooting parameters* $\hat{\theta}$ as follows. For all $i \in I_{to}$, if the state constraint is not active on Ω_i^δ, add to the set of shooting parameters θ the (unique by (A6)) time in Ω_i^δ where $g^\mu(y)$ attains its maximum over Ω_i^δ, duplicate all such times as well as touch points, add a zero jump parameter for each of them, and obtain then a $\hat{\theta} \in \Theta$ such that $\hat{\theta}$ is solution of (3.85), and (u, y) is the trajectory associated with $\hat{\theta}$.

Let us show that this augmented set of shooting parameters $\hat{\theta}$ is arbitrarily close to θ_0 when $\|\mu - \mu_0\|$ and $\|u - \bar{u}\|_\infty$ are small enough. Indeed, the convergence of the initial costate is a consequence of Rem. 3.7. For $i \in I_{ba}$, since we know by Th. 3.4 that $\Omega_i^\delta \cap I(g^\mu(y))$ is an interval $[\tau_{en,i}^\mu, \tau_{ex,i}^\mu]$, letting $\delta \to 0$, we obtain that $\bar{t}_i^{en} \leq \liminf_{\mu \to \mu_0} \tau_{en,i}^\mu$ and $\bar{t}_i^{ex} \geq \limsup_{\mu \to \mu_0} \tau_{ex,i}^\mu$. The converse inequalities $\bar{t}_i^{en} \geq \limsup_{\mu \to \mu_0} \tau_{en,i}^\mu$ and $\bar{t}_i^{ex} \leq \liminf_{\mu \to \mu_0} \tau_{ex,i}^\mu$ are obtained as follows. Assume e.g. by contradiction that $\bar{t}_i^{en} < \limsup_{\mu \to \mu_0} \tau_{en,i}^\mu$. Then there exist $\delta > 0$, a stable extension $(\mathcal{P}^\mu)$, a sequence $\mu_n \to \mu_0$, and a stationary point (u_n, y_n) of $(\mathcal{P}^{\mu_n})$, with multipliers (p_n, η_n), such that $u_n \to \bar{u}$ in L^∞ and $\tau_{en,i}^{\mu_n} \geq \bar{t}_i^{en} + \delta$ for all n. Let φ be a C^∞

function with support in $[\bar{t}_i^{en}, \bar{t}_i^{en} + \delta]$ and positive on $(\bar{t}_i^{en}, \bar{t}_i^{en} + \delta)$. Then $\int_0^T \varphi(t)\mathrm{d}\eta_n(t) = 0$, for all n. But by (A5), $\int_0^T \varphi(t)\mathrm{d}\bar{\eta}(t) > 0$, which contradicts the second assertion in Lemma 3.6. This achieves to show the convergence of entry/exit points for $i \in I_{ba}$. Letting $\delta \to 0$ in (S3), we obtain similarly the convergence of touch points and entry/exit points of boundary arcs to the common value $\bar{t}_i^{en}$, for all $i \in I_{to}$. The convergence of nonactive local isolated maxima of $g^\mu(y)$ in Ω_i^δ when $i \in I_{to}$, is obtained by classical arguments, since (3.26) holds and locally on Ω_i^δ, the second-order derivative (3.6) is continuous on interior arcs since u and $\dot{u}$ are (indeed, for $\|u - \bar{u}\|_\infty$ and $\|\mu - \mu_0\|$ small enough, $H_{uu}^\mu(u, y, p) \geq \alpha/2 > 0$ by (A2) and Rem. 3.7), so that $g^\mu(y)$ belongs to a C^2 neighborhood of $g^{\mu_0}(\bar{y})$. Finally, the convergence of jump parameters at entry times follows from assertion (4) in Lemma 3.6, since η^1 and η_1 are related by (3.63), and η_1 satisfies (3.50) and (3.59).

Hence if (μ, u) is close enough to $(\mu_0, \bar{u})$, the augmented set of shooting parameters $\hat{\theta}$ belongs to the neighborhood V_θ of θ_0, on which (3.85) has a unique solution θ^μ by Lemma 3.26, and (u, y) is the (unique) trajectory associated with $\hat{\theta}$. Consequently, $\hat{\theta} = \theta^\mu$ and $(u, y) = (u^\mu, y^\mu)$ is the unique stationary point of $(\mathcal{P}^\mu)$ with (u, μ) in the neighborhood of $(\bar{u}, \mu_0)$. □

Now we can prove the main result. Under assumptions (A2)-(A6) and point (ii) of Th. 3.11, for μ in the neighborhood of μ_0 and $v \in L^2$, denote by z_v^μ the unique solution in $\mathcal{Z}$ of the linearized state equation

$$\dot{z}_v^\mu \;=\; f_u^\mu(u^\mu, y^\mu)v + f_y^\mu(u^\mu, y^\mu)z_v^\mu \qquad \text{a.e. on } [0,T], \qquad z_v^\mu(0) = 0 \tag{3.118}$$

and by Q^μ the quadratic form over L^2 defined by

$$Q^\mu(v) \;=\; \mathcal{J}^\mu(v, z_v^\mu) \tag{3.119}$$

where $\mathcal{J}^\mu$ is defined by (3.37) for $(\mathcal{P}^\mu)$ and its stationary point and multipliers $(u^\mu, y^\mu, p^\mu, \eta^\mu)$.

Proof of Theorem 3.11. By Lemmas 3.26-3.29, to achieve the proof of (ii) ⇒ (i), it remains to show that u^μ satisfies the uniform quadratic growth condition. The arguments used are similar to those in the proof of [19, Th. 4.3][7]. We argue by contradiction. Assume that the uniform quadratic growth does not hold. Then there exist a sequence μ_n converging to μ_0 and a sequence $u_n \to \bar{u}$ in L^∞ such that for all n, $G^{\mu_n}(u_n) \in K$ and

$$J^{\mu_n}(u_n) \;\leq\; J^{\mu_n}(u^{\mu_n}) + o(\|u_n - u^{\mu_n}\|_2^2). \tag{3.120}$$

Introducing the *Lagrangian* of (3.11) defined by $L^\mu(u, \eta) := J^\mu(u) + \langle \eta, G^\mu(u)\rangle$, with $\langle \cdot, \cdot \rangle$ the duality product in $\mathcal{M}[0,T] \times C^0[0,T]$ defined by $\langle \eta, x\rangle = \int_0^T x(t)\mathrm{d}\eta(t)$, we obtain that

$$L^{\mu_n}(u_n, \eta^{\mu_n}) \;\leq\; L^{\mu_n}(u^{\mu_n}, \eta^{\mu_n}) + o(\|u_n - u^{\mu_n}\|_2^2).$$

Set $\varepsilon_n := \|u_n - u^{\mu_n}\|_2 \to 0$ and $v_n := (u_n - u^{\mu_n})/\varepsilon_n$. A second-order expansion of the Lagrangian shows that

$$L^{\mu_n}(u_n, \eta^{\mu_n}) = L^{\mu_n}(u^{\mu_n}, \eta^{\mu_n}) + \varepsilon_n^2 Q^{\mu_n}(v_n) + o(\varepsilon_n^2),$$

where Q^{μ_n} is defined by (3.119). It follows then that $Q^{\mu_n}(v_n) \;\leq\; o(1)$. Since (v_n) is bounded in $\mathcal{V} = L^2$, we may assume that it converges weakly to some $\bar{v} \in L^2$. In view of the compact

[7] Theorem 2.34 of this thesis.

inclusion of $H^1(0,T)$ in $C^0[0,T]$, the associated linearized state $z_n := z_{v_n}^{\mu_n}$ defined by (3.118) converges uniformly to $\bar{z} := z_{\bar{v}}^{\mu_0}$. We may write that

$$Q^{\mu_n}(v_n) = Q^{\mu_0}(v_n) + Q^{\mu_n}(v_n) - Q^{\mu_0}(v_n),$$

and using that $\|v_n\|_2$ is bounded it is not difficult to check that $Q^{\mu_n}(v_n) - Q^{\mu_0}(v_n) \to 0$. Therefore by weak lower-semicontinuity of the Legendre form $Q = Q^{\mu_0}$ by (3.20), we obtain that

$$\mathcal{J}(\bar{v},\bar{z}) = Q(\bar{v}) \;\le\; \liminf_{n\to+\infty} Q(v_n) \;\le\; \limsup_{n\to+\infty} Q(v_n) \;\le\; 0. \tag{3.121}$$

Moreover, $\bar{v}$ and $\bar{z}$ satisfy (3.39). Indeed, since $G^{\mu_n}(u_n) \in K$, we have that $g_y^{\mu_n}(y^{\mu_n})z_n + r_n \le 0$ on $I(g^{\mu_n}(y^{\mu_n}))$, where r_n satisfies $\|r_n\|_\infty = \mathcal{O}(\varepsilon_n)$. Since $\frac{\mathrm{d}}{\mathrm{d}t} g_y^{\mu_n}(y^{\mu_n}(t))z_n(t) = (g_u^{\mu_n})^{(1)}(u^{\mu_n},y^{\mu_n})v_n + (g_y^{\mu_n})^{(1)}(u^{\mu_n},y^{\mu_n})z_n$, it follows from Cauchy-Schwarz inequality that the functions (of time) $g_y^{\mu_n}(y^{\mu_n})z_n$ are uniformly Hölder continuous. Therefore, there exists a constant $C > 0$ such that, for all large enough n, using Lemma 3.26,

$$\sup_{t\in\cup_{i=1}^N [\bar{t}_i^{en},\bar{t}_i^{ex}]} g_y^{\mu_n}(y^{\mu_n}(t))z_n(t) \;\le\; \mathcal{O}(\varepsilon_n) + C\sqrt{\max_{i=1,\dots,N}\{|t_i^{\mu_n,en} - \bar{t}_i^{en}|, |t_i^{\mu_n,ex} - \bar{t}_i^{ex}|\}} \;=\; o(1). \tag{3.122}$$

Since $g_y^{\mu_n}(y^{\mu_n})z_n \to g_y(\bar{y})\bar{z}$ uniformly, it follows that $g_y(\bar{y})\bar{z} \le 0$ on $\cup_{i=1}^N[\bar{t}_i^{en},\bar{t}_i^{ex}]$. In addition, by (3.120), we have that $\langle \eta^{\mu_n}, g_y^{\mu_n}(y^{\mu_n})z_n\rangle = -DJ^{\mu_n}(u^{\mu_n})v_n \ge \mathcal{O}(\varepsilon_n)$. Therefore, $\langle \bar{\eta}, g_y(\bar{y})\bar{z}\rangle \ge 0$, which implies finally by (A5) that $g_y(\bar{y})\bar{z} = 0$ on $\cup_{i=1}^N[\bar{t}_i^{en},\bar{t}_i^{ex}]$, i.e. (3.39) holds. Thus (3.44) and (3.121) imply that $\bar{v} = 0$. But then $Q(v_n) \to Q(\bar{v})$, and hence, by the property of Legendre forms, $v_n \to \bar{v}$ strongly, contradicting that $\|v_n\|_2 = 1$ for all n.

To prove the converse implication, we construct a perturbation of the constraint g^μ, so that (nonessential) touch points becomes inactive on the perturbed problem $(\mathcal{P}^\mu)$, and $(\bar{u},\bar{y})$ is a stationary point of $(\mathcal{P}^\mu)$. This is where we need nonautonomous perturbations. Let φ be a C^∞ function with support in $[-1,1]$ and positive on $(-1,1)$. Set $\mu_0 = 0$ and $g^\mu(y) := g(y) - \sum_{\tau\in\bar{\mathcal{T}}_{to}} \mu^5\varphi((y_n-\tau)/\mu)$ for $\mu \ne 0$ (recall that we assume (3.4)). Then (ℓ,ϕ,f,g^μ,y_0) is a stable extension of $(\mathcal{P})$, $(\bar{u},\bar{y})$ is a stationary point of $(\mathcal{P}^\mu)$ for all $|\mu|$ small enough, and $g^\mu(\bar{y}(\tau)) < 0$ for all nonessential touch point τ. By the definition of the uniform growth condition, $(\bar{u},\bar{y})$ is a local solution of $(\mathcal{P}^\mu)$ satisfying (3.45), so it follows from Th. 3.9(ii) that the strong second-order sufficient condition (3.44) holds. □

We end this section by the proof of Lemma 3.12.

Proof of Lemma 3.12. Denote by Q^μ the quadratic form (3.119) and $\tilde{C}^\mu$ the set of $v \in \mathcal{V}$ satisfying the constraints (3.38)-(3.39) for $(\mathcal{P}^\mu)$ and its stationary point (u^μ,y^μ), i.e. such that

$$g_y^\mu(y^\mu)z_v^\mu \;=\; 0 \qquad \text{on } [t_i^{\mu,en},t_i^{\mu,ex}], \qquad \text{for all } i = 1,\dots,N. \tag{3.123}$$

Let us show that there exists $\alpha' > 0$ such that for all μ close enough to μ_0 and all $v \in \tilde{C}^\mu(u^\mu)$, we have $Q^\mu(v) \ge \alpha'\|v\|_2^2$, which will give the result.

We argue by contradiction, as in the proof of the uniform growth condition in Th. 3.11. Assume this is not the case. Then there exist sequences $(\mu_n)_{n\in\mathbb{N}^*}$ and $(v_n)_{n\in\mathbb{N}^*}$, such that $\mu_n \to \mu_0$, $v_n \in \tilde{C}^{\mu_n}$ for all n, and

$$Q^{\mu_n}(v_n) \;\le\; o(1)\|v_n\|_2^2. \tag{3.124}$$

Since $\tilde{C}^{\mu_n}$ is a cone (in fact, here, a linear subspace of $\mathcal{V}$), and Q^{μ_n} is a quadratic form, assume w.l.o.g. that $\|v_n\|_2 = 1$, and taking a subsequence if necessary, that the sequence

(v_n) converges weakly to some $\bar{v} \in \mathcal{V}$. Then the associated state $z_n := z_{v_n}^{\mu_n}$ given by (3.118) is weakly convergent to $\bar{z} := z_{\bar{v}}^{\mu_0}$ in H^1, and hence $z_n \to \bar{z}$ uniformly. By the same argument as in the proof of Th. 3.11 (see (3.122)), since $v_n \in \tilde{C}^{\mu_n}$, we deduce that $\sup_{t\in\cup_{i=1}^N[\bar{t}_i^{en},\bar{t}_i^{ex}]} |g_y^{\mu_n}(y^{\mu_n}(t))z_n(t)| \leq C\sqrt{\max_{i=1,\ldots,N}\{|t_i^{\mu_n,en} - \bar{t}_i^{en}|, |t_i^{\mu_n,ex} - \bar{t}_i^{ex}|\}} = o(1)$. It follows then that $\bar{v} \in \tilde{C}^{\mu_0}$. But (3.124) implies that $Q^{\mu_0}(\bar{v}) \leq 0$, therefore $\bar{v} = 0$ by (3.44), and then $Q^{\mu_0}(v_n) \to Q^{\mu_0}(\bar{v})$. Since Q^{μ_0} is a Legendre form, it follows that $v_n \to \bar{v}$ strongly, contradicting that $\|v_n\|_2 = 1$ for all n. This achieves the proof. □

3.7 Sensitivity Analysis

Under assumptions (A2)-(A6) and point (i) or (ii) of Th. 3.11, we investigate in this section directional differentiability of solutions. Given a stable extension $(\mathcal{P}^\mu)$, by Lemma 3.26, strong regularity holds, and the mapping $\Xi : V_\delta \to V_\theta'$, $\delta \mapsto \theta(\delta)$ solution of (3.87) is given by $\Xi(\delta) = \theta_0 + \omega(\delta)$, where $\omega(\delta)$ is the solution of (3.91). It is easy to see that the mapping $\delta \mapsto \omega(\delta)$ is positively homogeneous of degree one, and it follows then from (3.90) that the mapping $\mu \mapsto \theta^\mu$ is Fréchet directionally differentiable. The directional derivatives in direction $d \in M$ are obtained by substituting into (3.91) δ by $-D_\mu F(\theta_0, \mu_0)d$. Therefore,

$$\theta^{\mu_0+d} = \theta_0 + \omega_d + o(\|d\|), \tag{3.125}$$

where

$$\omega_d \;=\; (\pi_{d,0}^*, \gamma_d^1, \sigma_d^{en}, \sigma_d^{ex}) \in \mathbb{R}^n \times \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^N \tag{3.126}$$

is as follows. Denote by

$$(v_d, z_d, \pi_{1,d}, \zeta_{1,d}, \lambda_d^1) \tag{3.127}$$

the (unique) optimal solution, costate and multipliers of the linear-quadratic problem below:

$$\begin{aligned}(\mathcal{P}_d) \qquad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \; & \tfrac{1}{2}\int_0^T D^2_{(u,y,\mu),(u,y,\mu)}\tilde{H}(\bar{u},\bar{y},\bar{p}_1,\bar{\eta}_1,\mu_0)((v,z,d),(v,z,d))\mathrm{d}t \\ & + \tfrac{1}{2}D^2\hat{\phi}(\bar{y}(T),\mu_0)((z(T),d),(z(T),d)) \\ & + \tfrac{1}{2}\sum_{i\in I_{ba}} \bar{\nu}_i^1 D^2\hat{g}(\bar{y}(\bar{t}_i^{en}),\mu_0)((z(\bar{t}_i^{en}),d),(z(\bar{t}_i^{en}),d))\end{aligned}$$

$$\text{subject to:} \qquad \dot{z} = D\hat{f}(\bar{u},\bar{y},\mu_0)(v,z,d) \quad \text{on } [0,T], \quad z(0) = D\hat{y}_0(\mu_0)d \tag{3.128}$$

$$D\hat{g}^{(1)}(\bar{u},\bar{y},\mu_0)(v,z,d) = 0 \quad \text{on } \bar{\mathcal{I}}_b \tag{3.129}$$

$$D\hat{g}(\bar{y}(\bar{t}_i^{en}),\mu_0)(z(\bar{t}_i^{en}),d) = 0, \quad i \in I_{ba} \tag{3.130}$$

$$D\hat{g}(\bar{y}(\bar{t}_i^{en}),\mu_0)(z(\bar{t}_i^{en}),d) \leq 0, \quad i \in I_{to}, \tag{3.131}$$

with $\pi_{1,d}$ associated with the constraint (3.128), $\zeta_{1,d}$ with (3.129), and λ_d^1 with (3.130)-(3.131). Then we have

$$\pi_{d,0} \;=\; \pi_{1,d}(0) \tag{3.132}$$

$$\gamma_d^1 \;=\; \lambda_d^1 \tag{3.133}$$

$$\sigma_{d,i}^{en} \;=\; -\frac{D\hat{g}^{(1)}(\bar{u}(\bar{t}_i^{en}),\bar{y}(\bar{t}_i^{en}),\mu_0)(v_d(\bar{t}_i^{en-}),z_d(\bar{t}_i^{en}),d)}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(\bar{u},\bar{y})|_{t=\bar{t}_i^{en-}}}, \; i = 1,\ldots,N, \tag{3.134}$$

$$\sigma_{d,i}^{ex} \;=\; -\frac{D\hat{g}^{(1)}(\bar{u}(\bar{t}_i^{ex}),\bar{y}(\bar{t}_i^{ex}),\mu_0)(v_d(\bar{t}_i^{ex+}),z_d(\bar{t}_i^{ex}),d)}{\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(\bar{u},\bar{y})|_{t=\bar{t}_i^{ex+}}}, \; i = 1,\ldots,N. \tag{3.135}$$

Since the mapping $\mu \mapsto \theta^\mu$ is Fréchet directionally differentiable and the solution $(u^\mu, y^\mu, p_1^\mu, \eta_1^\mu)$ of (3.77)-(3.81) is, on each arc, a C^1 function of (θ^μ, μ), combining with the continuity of u^μ and of the classical multipliers p^μ and η^μ given by (3.61) (which follows from Lemma 3.28), we obtain the following result.

Theorem 3.30. *Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)-(A6). If either point (i) or (ii) of Th. 3.11 is satisfied, then there exists a neighborhood V_μ of μ, such that the mapping $\mu \mapsto (u^\mu, y^\mu, p^\mu, \eta^\mu)$ is Fréchet directionally differentiable in the space*

$$L^r(0,T) \times W^{1,r}(0,T;\mathbb{R}^n) \times L^r(0,T;\mathbb{R}^{n*}) \times L^r(0,T), \qquad \text{for all } 1 \le r < +\infty,$$

and the derivatives of the state and control in direction d are the optimal solution (v_d, z_d) of linear-quadratic problem $(\mathcal{P}_d)$, while those of the costate p^μ and state constraint multiplier η^μ are obtained, respectively, a.e. by

$$\pi_d(t) = \pi_{1,d}(t) + \zeta_{1,d}(t) g_y^{\mu_0}(\bar y(t)) + \bar\eta_1(t) D\hat g_y(\bar y(t), \mu_0)(z_d(t), d) \tag{3.136}$$

$$\zeta_d(t) = -\sum_{i=1}^{N} \gamma_{d,i}^1 \mathbf{1}_{[0,\bar t_i^{en})}(t) - \zeta_{1,d}(t). \tag{3.137}$$

In addition, all shooting parameters (initial costate, jump parameters and junction times) are Fréchet directionally differentiable w.r.t. μ, and their directional derivative in direction d are given by (3.132)-(3.135).

Remark 3.31. We can show that an equivalent formulation of $(\mathcal{P}_d)$ is (see Rem. 3.23) to minimize

$$\int_0^T D^2_{(u,y,\mu),(u,y,\mu)} H(\bar u, \bar y, \bar p, \mu_0)((v,z,d),(v,z,d))\mathrm{d}t \; + \; D^2\hat\phi(\bar y(T), \mu_0)((z(T),d),(z(T),d)) \\ + \int_0^T D^2 \hat g(\bar y(t), \mu_0)((z(t),d),(z(t),d))\mathrm{d}\bar\eta(t) \tag{3.138}$$

for $(v,z) \in \mathcal{V} \times \mathcal{Z}$ subject to the constraints (3.128), (3.131) and

$$D\hat g(\bar y, \mu_0)(z,d) = 0 \qquad \text{on } \bar{\mathcal{I}}_b. \tag{3.139}$$

This last constraint is equivalent to (3.129)-(3.130) since we have that $D\hat g^{(1)}(\bar u, \bar y, \mu_0)(v,z,d) = \frac{\mathrm{d}}{\mathrm{d}t} D\hat g(\bar y(t), \mu_0)(z(t), d)$. Then, using the relation (3.136), we can show that π_d, the directional derivative of p^μ w.r.t. μ, is the multiplier associated with (3.128) in formulation (3.138)-(3.139) of $(\mathcal{P}_d)$, and that the directional derivative of $\frac{\mathrm{d}\eta^\mu}{\mathrm{d}t}$ w.r.t. μ, equal by (3.137) to $\dot\zeta_d = -\dot\zeta_{1,d}$, is equal to the multiplier associated with the constraint (3.139).

Let us conclude this section by the following observation. For $i \in I_{to}$, since $\bar t_i^{en} = \bar t_i^{ex}$, the optimality system of $(\mathcal{P}_d)$, easily obtained, yields that $H_{uu} v_d + H_{uy} z_d + \pi_{1,d} f_u = 0$ at $\bar t_i^{en\pm}$, and that the jump of $\pi_{1,d}$ is given by $[\pi_{1,d}(\bar t_i^{en})] = -\gamma_{d,i}^1 g_y(\bar y(\bar t_i^{en}))$. Hence, the jump of v_d is given by

$$[v_d(\bar t_i^{en})] = \gamma_{d,i}^1 H_{uu}^{-1}(\bar u, \bar y, \bar p)(\bar t_i^{en}) g_y(\bar y(\bar t_i^{en})) f_u(\bar u, \bar y)(\bar t_i^{en}) = \gamma_{d,i}^1 H_{uu}^{-1}(\bar u, \bar y, \bar p)(\bar t_i^{en}) g_u^{(1)}(\bar u, \bar y)(\bar t_i^{en}),$$

and we obtain from (3.134)-(3.135)

$$\sigma_{d,i}^{ex} - \sigma_{d,i}^{en} = -\frac{g_u^{(1)}(\bar u, \bar y)(\bar t_i^{en})[v_d(\bar t_i^{en})]}{\frac{\mathrm{d}}{\mathrm{d}t} g^{(1)}(\bar u, \bar y)|_{t=\bar t_i^{en}}} = C_i \gamma_{d,i}^1 \tag{3.140}$$

with

$$C_i := \frac{H_{uu}^{-1}(\bar{u},\bar{y},\bar{p})(\bar{t}_i^{en})(g_u^{(1)}(\bar{u},\bar{y})(\bar{t}_i^{en}))^2}{-\frac{\mathrm{d}}{\mathrm{d}t}g^{(1)}(\bar{u},\bar{y})|_{t=\bar{t}_i^{en}}} > 0.$$

Since $\gamma_{d,i}^1 \geq 0$ for $i \in I_{to}$, we see that $\sigma_{d,i}^{ex} - \sigma_{d,i}^{en} \geq 0$, with equality iff $\gamma_{d,i}^1 = 0$. It follows that, for $\mu - \mu_0 = d$, the length of the boundary arc and the jump parameter are related, at first order, by

$$t_i^{\mu,ex} - t_i^{\mu,en} = C_i \nu_i^{\mu,1} + o(\|\mu - \mu_0\|). \tag{3.141}$$

Remark 3.32. It was quite expected that nonessential touch points generally turn into boundary arcs for constraints of first order (see e.g. [29]). However it was surprising to be able to describe this transition between touch points and boundary arcs by a shooting approach when the structure is not stable, and obtain the differentiability of the shooting mapping, and in particular of the entry and exit times of the appearing boundary arcs.

Note that those results are false for control constraints. Consider for example the problem below:

$$\min_{u\in\mathcal{U}} \int_0^2 (u(t) - (t-1)^2)^2 \mathrm{d}t.$$

Here we have no state, or more precisely, the state is equal to the time. Obviously the solution is $u(t) = (t-1)^2$. Add now a constraint $u(t) \geq \varepsilon$ for $\varepsilon > 0$. Then the optimal solution is $u(t) = \varepsilon$ on $[\tau_-^\varepsilon, \tau_+^\varepsilon]$ with $\tau_\pm^\varepsilon = 1 \pm \sqrt{\varepsilon}$, and $u(t) = (t-1)^2$ on $[0, \tau_-^\varepsilon) \cup (\tau_+^\varepsilon, 2]$. So for $\varepsilon > 0$ a boundary arc appear, whose end points τ_-^ε and τ_+^ε are not differentiable at the point $\varepsilon = 0$, and whose length is of order $\sqrt{\varepsilon}$ and not ε. A fortiori the shooting mapping is not differentiable at the point $\varepsilon = 0$, and the algorithm described in section 3.9 has no obvious extension to control constraints (or more generally to mixed control-state constraints).

3.8 Example of sensitivity analysis

We illustrate the results of this paper on a very basic example. We consider the problem of an elastic line of positive mass, fixed at its endpoints and submitted to a vertical uniform force (g). The problem is to find the equilibrium position, i.e. minimize the energy. Assuming the elastic potential to be quadratic with unit constant, this can be written as the optimal control problem (with t replaced by $x \in [0,1]$):

$$\min \int_0^1 \left(\frac{u(x)^2}{2} + g y(x)\right) \mathrm{d}x, \qquad \dot{y}(x) = u(x), \qquad y(0) = 0 = y(1). \tag{3.142}$$

We add a first-order state constraint, e.g. the level of the floor

$$y(x) \geq -h. \tag{3.143}$$

Here g and h denotes positive constants.

Remark 3.33. Our results can be extended with only slight adaptations to the case when there are also finitely many equality and inequality constraints on the final state, if we assume in addition a *controllability condition*. In the case of a fixed final state, $y(T) = y_T$ given in $\mathbb{R}^n$, this controllability condition is assumption (A1') below. Recall that given $\delta > 0$, we denote by $\Omega^\delta := \{t \in [0,T],\ \mathrm{dist}\{t; I(g(\bar{y}))\} < \delta\}$.

(A1') (i) The initial and final conditions satisfy $g(y_0) < 0$ and $g(y_T) < 0$;
(ii) There exists $\delta > 0$ such that the linear mapping $\mathcal{U} \to W^{1,\infty}(\Omega^\delta) \times \mathbb{R}^n$; $v \mapsto (g_y(\bar{y}(\cdot))z_v(\cdot)|_{\Omega^\delta}, z_v(T))$, where z_v is the solution of (3.24) and $|_{\Omega^\delta}$ denotes the restriction to the set Ω^δ, is onto (and therefore has a bounded right inverse by the open mapping Theorem).

This assumption (A1') plays the role of Lemma 3.2 in the proofs. Note that when the dynamics f is linear, i.e. $f(u, y) = Ay + Bu$, then (A1')(ii) is satisfied if the pair (A, B) is controllable, and if (A1')(i) and (A3) hold.

For the example considered here, (A1') is obviously satisfied so all the previous results are valid. The unconstrained optimal trajectory when $h/g \geq 1/8$ is given by:

$$y(x) = \tfrac{1}{2}gx^2 - \tfrac{1}{2}gx, \qquad u(x) = gx - \tfrac{1}{2}g. \tag{3.144}$$

The resolution of the constrained problem when $h/g \leq 1/8$ is as follows. The trajectory is:

$$u(x) = \begin{cases} g(x - x_{en}) & \text{on } [0, x_{en}] \\ 0 & \text{on } [x_{en}, x_{ex}] \\ g((x-1) - (x_{ex} - 1)) & \text{on } [x_{ex}, 1] \end{cases}$$

$$y(x) = \begin{cases} g(x^2/2 - x_{en}x) & \text{on } [0, x_{en}] \\ -h & \text{on } [x_{en}, x_{ex}] \\ g((x-1)^2/2 - (x_{ex} - 1)(x-1)) & \text{on } [x_{ex}, 1]. \end{cases}$$

Entry and exit positions x_{en} and x_{ex} are given by:

$$x_{en} = \sqrt{2h/g}, \qquad x_{ex} = 1 - \sqrt{2h/g}. \tag{3.145}$$

The alternative state constraint multiplier on $[x_{en}, x_{ex}]$ is given by:

$$\eta_1(x) = p_1(x) = -g(x - x_{ex}) \geq 0, \qquad \dot{\eta}_1(x) = -g < 0,$$

and hence, the jump parameter at entry time is:

$$\nu^1_{en} = \eta_1(x_{en}) = g(x_{ex} - x_{en}) = g\left(1 - 2\sqrt{2h/g}\right) \geq 0. \tag{3.146}$$

We consider perturbations w.r.t. nominal values of parameters $g = g_0 = 1$ and $h = h_0 = 1/8$, for which there is a touch point at $x = 1/2$. The strong sufficient second-order condition (3.44) clearly holds, since the linear-quadratic problem:

$$\min \int_0^1 \frac{v^2(x)}{2} \mathrm{d}x, \qquad \dot{z}(x) = v(x), \qquad z(0) = 0 = z(1)$$

having a strongly convex cost function, has $(v, z) = 0$ for unique solution. Let us then study the perturbed quadratic problem at (g_0, h_0) in direction $d := (\gamma, \eta)$:

$$\min \int_0^1 \left(\frac{v(x)^2}{2} - \gamma z(x)\right) \mathrm{d}x, \qquad \dot{z}(x) = v(x), \qquad z(0) = 0 = z(1),$$

subject to the interior point inequality constraint:

$$z(1/2) \geq -\eta. \tag{3.147}$$

The unconstrained trajectory is:

$$z_d(x) = \gamma\left(\frac{x^2}{2} - \frac{x}{2}\right), \qquad v_d(x) = \gamma\left(x - \frac{1}{2}\right). \tag{3.148}$$

Therefore, the constraint is active, iff $\eta \leq \gamma/8$. If $\eta > \gamma/8$, (3.148) corresponds to the directional derivative of the unconstrained trajectory (3.144). When $\eta \leq \gamma/8$, the constraint (3.147) is active, i.e. $z_d(1/2) = -\eta$, and therefore, the solution of the linear-quadratic problem is as follows:

$$v_d(x) = \begin{cases} \gamma x - (2\eta + \gamma/4) & \text{on } [0, 1/2] \\ \gamma(x-1) + (2\eta + \gamma/4) & \text{on } [1/2, 1]. \end{cases}$$

$$z_d(x) = \begin{cases} \gamma x^2/2 - (2\eta + \gamma/4)x & \text{on } [0, 1/2] \\ \gamma(x-1)^2/2 + (2\eta + \gamma/4)(x-1) & \text{on } [1/2, 1]. \end{cases}$$

The multiplier λ_d associated with the constraint (3.147) is, by (3.140):

$$\lambda_d = [\pi_d(1/2)] = -[v_d(1/2)] = -2(2\eta - \gamma/4) \geq 0, \tag{3.149}$$

and, by (3.134)-(3.135), the variations of entry and exit points $\sigma_{d,en}$ and $\sigma_{d,ex}$ are given by:

$$\sigma_{d,en} = -\frac{v(1/2^-)}{g_0} = -\gamma/4 + 2\eta, \qquad \sigma_{d,ex} = -\frac{v(1/2^+)}{g_0} = \gamma/4 - 2\eta. \tag{3.150}$$

By (3.146) and (3.145), we check that the above formula corresponds to the first-order variations, with $g = g_0 + \gamma$ and $h = h_0 + \eta$, $|\gamma|, |\eta|$ small, of:

$$\nu^1_{en} = (1+\gamma)\left(1 - 2\sqrt{\frac{1/4 + 2\eta}{1+\gamma}}\right), \qquad x_{en} = \sqrt{\frac{1/4 + 2\eta}{1+\gamma}}, \qquad x_{ex} = 1 - \sqrt{\frac{1/4 + 2\eta}{1+\gamma}}.$$

We consider perturbations in three directions $d = (\gamma, \eta)$:

Case (a) $(\gamma, \eta) = (0, -0.02)$

Case (b) $(\gamma, \eta) = (1, 0)$

Case (c) $(\gamma, \eta) = (1, -0.02)$.

Case (a) corresponds to an elevation of the ground level, case (b) corresponds to an increasing of the "gravitational" force g, both of them leading to the emergence of a boundary arc, and case (c) combines elevation of the ground and increasing of g. The perturbed trajectories and directional derivatives of the state in $W^{1,r}$, $1 \leq r < +\infty$, are presented for each case in Fig. 3.1. The unconstrained trajectory for (g_0, h_0) is a parabola. In Fig. 3.2, we focus on the appearance of the boundary arc in case (c), check that its length is of the order of the perturbation and compare with the directional derivatives of variation of junction times (3.150).

3.9 Homotopy method

We present in this section an algorithm that combines shooting and continuation (or homotopy) methods for solving optimal control problems with a scalar first-order state constraint,

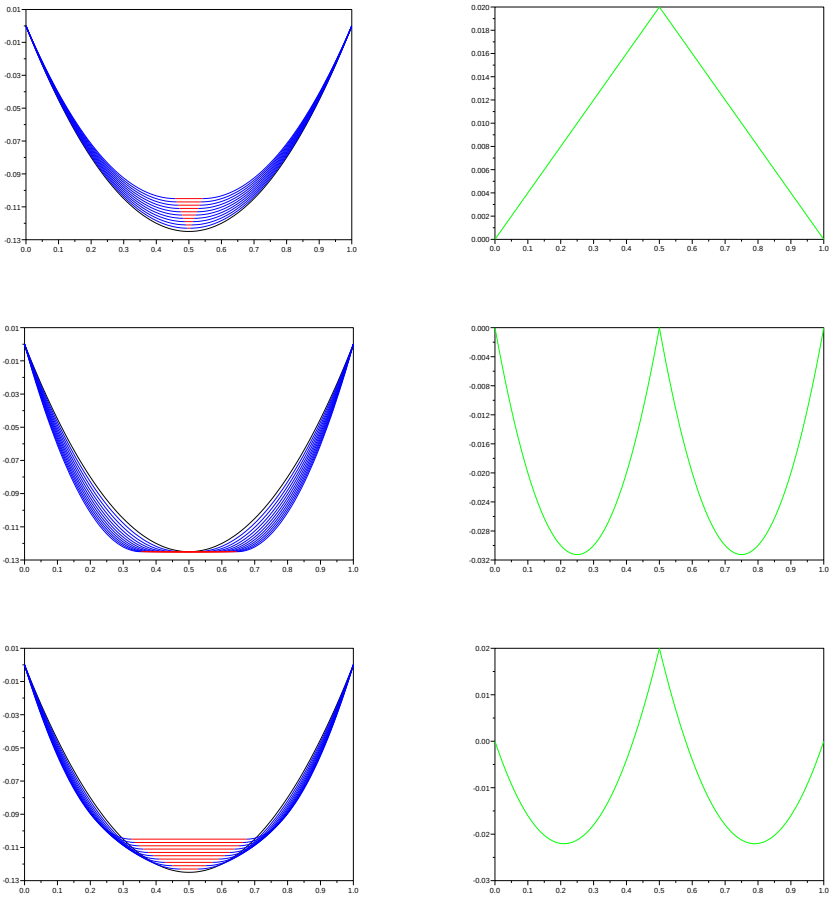


Figure 3.1: Perturbation of the state (left) and directional derivatives (right) in case (a) to (c) (from top to bottom)

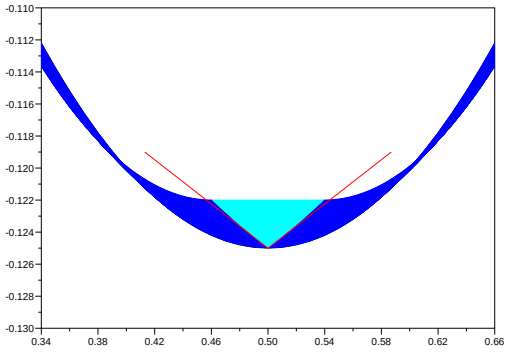


Figure 3.2: Variation of the length of the boundary arc in case (c).

when the structure of the trajectory is unknown. It keeps the advantages of shooting methods regarding to the (high) precision and the (low) complexity, and enables to get rid of the (sometimes) hard task to guess a priori the structure of the trajectory, and of the initialization of some of the shooting parameters (only the initialization of the initial costate is left to the user). The idea is to handle automatically the appearance (and disappearance) of boundary arcs, so that the algorithm finds itself the structure of the trajectory. The results of the previous sections are used.

General results on homotopy methods can be found in e.g. [1], [45, Chap. 5], and applications of homotopy methods to optimal control problems in e.g. [31, 63, 97].

3.9.1 Description of the algorithm

The problem to be solved is the following:

$$
\begin{aligned}
(\mathcal{P}) \qquad & \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} && \int_0^T \ell(u(t),y(t))\mathrm{d}t + \phi(y(T)) && (3.151)\\
& \text{subject to} && \dot{y}(t) = f(u(t),y(t)) \quad \text{a.e. on } [0,T], \quad y(0) = y_0, && (3.152)\\
& && g(y(t)) \leq 0 \quad \text{on } [0,T]. && (3.153)
\end{aligned}
$$

We assume that $(\mathcal{P})$ satisfies (A0)-(A1). In view of remark 3.33, we can more generally consider a fixed final state $y(T) = y_T$ and $\phi = 0$ if we assume in addition that the controllability condition (A1') holds.

We consider the natural homotopy on the state constraint $(\mathcal{P}^\mu)$, for $\mu \in [0,1]$, defined by $(\ell^\mu, \phi^\mu, f^\mu, y_0^\mu) := (\ell, \phi, f, y_0)$ and

$$
g^\mu(y) \;:=\; g(y) - (1-\mu)K, \tag{3.154}
$$

where the constant $K > 0$ is large enough, so that the state constraint of problem $(\mathcal{P}^0)$ is not active, except maybe at finitely many (isolated) touch points in $(0,T)$. We explain later how we choose K in the algorithm. We thus have $(\mathcal{P}^1) \equiv (\mathcal{P})$.

The shooting mapping (3.84) for $(\mathcal{P}^\mu)$ is denoted by $F(\theta,\mu)$, where θ is the vector of shooting parameters, of variable dimension depending on the structure of the trajectory, and μ is the (scalar) homotopy parameter. Since we only have here one state constraint of first order, note that the structure of the trajectory, and hence F, is entirely determined by the dimension of θ. More precisely, the number of boundary arcs of the trajectory N_{ba} is given by (assuming the state constraint inactive at initial and final times)

$$
N_{ba} \;=\; \frac{\dim(\theta) - n}{3} \;\in\; \mathbb{N}. \tag{3.155}
$$

The structure of the trajectory follows then from the alternation between interior and boundary arcs. We denote by $y^{\theta,\mu}$ the state solution of the alternative formulation for the shooting parameter θ and the value of the homotopy parameter μ. The algorithm is as follows (see Algorithm 3.34).

The algorithm is initialized by solving the unconstrained problem (without the state constraint) (3.151)-(3.152). We thus obtain a vector of shooting parameters θ_0 (reduced to the initial costate), associated with a stationary point of (3.151)-(3.152), which is a local solution of (3.151)-(3.152) if the second-order sufficient condition (3.42) holds. The constant K in (3.154) is taken equal to $K := \max_{t\in[0,T]} g(y^{\theta_0,0}(t))$. If $K \leq 0$, then θ_0 is a vector of shooting

parameters associated with a local solution of $(\mathcal{P})$. If $K > 0$, we start the homotopy from $\mu = \mu_0 := 0$ in (3.154) to $\mu = 1$.

The variable m_k denotes the maximum of $g^{\mu_k}(y^{\theta_k,\mu_k})$, attained at time τ_k. If m_k is positive, this means that the state constraint is violated so the structure is not correct and we have to add a boundary arc (step A). The variable i_k equals zero iff all entry and exit times of boundary arcs are such that entry times are lower than or equal to the corresponding exit times, and equals $i > 0$ if the entry time of the i-th boundary arc is greater than the corresponding exit time. If $i_k = i > 0$, the structure is not correct again so we have to delete the i-th boundary arc (step A). All this will be justified later in subsection 3.9.3 under some assumptions. If both $m_k \leq 0$ and $i_k = 0$, this means that the structure is correct, i.e. the current iterate θ_k is a vector of shooting parameters associated with a stationary point (u^{μ_k}, y^{μ_k}) of $(\mathcal{P}^{\mu_k})$. We thus increase the value of μ and do a simple predictor-corrector iteration (steps B-C), keeping the same structure for the shooting mapping. Then in step D we calculate the new values of m_{k+1} and i_{k+1} that say whether the structure is still correct or has to be updated in the next iteration. We do so until reaching the value $\mu = 1$.

If the Newton algorithm in step C fails, then we decrease the value of the step $\Delta\mu_k$, and go back to the last value $(\mu_{k-1}, \theta_{k-1})$ satisfying $F(\mu_{k-1}, \theta_{k-1}) = 0$ and $\max(m_{k-1}, i_{k-1}) = 0$.

Algorithm 3.34 (Homotopy Algorithm).

INITIALIZATION

Input $p_0 \in \mathbb{R}^{n*}$ *and* $\delta \in (0, 1]$.

- *Solve by the shooting algorithm (initialized by the value* p_0*) the unconstrained problem* (3.151)-(3.152)*, and obtain a vector of shooting parameters* θ_0*.*
- *Set* $K := \max g(y^{\theta_0,0}(t))$*. If* $K \leq 0$ *set* $\mu_0 := 1$*, else set* $\mu_0 := 0$*. Set* $m_0 := 0$*,* $i_0 := 0$*,* $k := 0$*,* $\Delta\mu_1 := \delta$*.*

While $\mu_k < 1$ **or** $\max(m_k, i_k) > 0$

If $\max(m_k, i_k) > 0$ **then** STEP A *(Update the structure)*

IF $m_k > 0$ THEN *(Addition of a boundary arc)*
Initialize the new shooting parameters $(\nu^1, \tau_{en}, \tau_{ex})$ *associated with this boundary arc by:*

$$\nu^1 = 0 \qquad \textit{and} \qquad \tau_{en} = \tau_{ex} = \tau_k. \tag{3.156}$$

Take the remaining shooting parameters equal to the previous value θ_k*, and obtain a vector of shooting parameters* $\bar{\theta}_k$ *of dimension* $\dim(\theta_k) + 3$*.*

END IF

IF $i_k > 0$ *(Suppression of a boundary arc)*
Remove the shooting parameters $(\nu^1, \tau_{en}, \tau_{ex})$ *corresponding to the* i_k*-th boundary arc from the vector of shooting parameters* θ_k*, and obtain a new vector of shooting parameters* $\bar{\theta}_k$ *of dimension* $\dim(\theta_k) - 3$*.*

END IF

Set $\bar{\mu}_k := \mu_k$ *(the value of* μ *is unchanged by this step).*

Else STEP B *(Prediction)*

Set $k := k + 1$
$\bar{\mu}_k := \min\{\mu_{k-1} + \Delta\mu_k; 1\}$
$\bar{\theta}_k := \theta_{k-1} - D_\theta \mathcal{F}(\theta_{k-1}, \mu_{k-1})^{-1} D_\mu \mathcal{F}(\theta_{k-1}, \mu_{k-1})(\bar{\mu}_k - \mu_{k-1})$.

End if

STEP C *(Correction) Try to solve, by a Newton method,* $\mathcal{F}(\theta, \bar{\mu}_k) = 0$. *The Newton algorithm is initialized by the value* $\bar{\theta}_k$.

If *the Newton algorithm fails* **then** *(go back to old values of* μ *and* θ *and decrease the step) Set* $\mu_k := \mu_{k-1}$, $\theta_k := \theta_{k-1}$, $m_k := m_{k-1}$, $i_k := i_{k-1}$, $\tau_k := \tau_{k-1}$, $\Delta\mu_k := \Delta\mu_k/2$ *and* $k := k - 1$.

Else *(success) obtain a solution* θ_k *such that* $F(\theta_k, \bar{\mu}_k) = 0$. *Set* $\mu_k := \bar{\mu}_k$.

STEP D *(Verify if the structure is correct)*

- *Set* $m_k := \max g^{\mu_k}(y^{\theta_k,\mu_k}(t))$ *and* $\tau_k \in \operatorname{argmax} g^{\mu_k}(y^{\theta_k,\mu_k}(t))$.
- *Set* $i_k := 0$. *For all* $i = 1, \ldots, N^k_{ba}$ *(*N^k_{ba} *given by* (3.155)*), if* θ_k *is such that the entry time corresponding to the* i*-th boundary arc is greater than the exit time corresponding to the* i*-th boundary arc, then* $i_k := i$.
- IF $\max(m_k, i_k) = 0$ THEN *set* $\Delta\mu_{k+1} := \delta$.

End if

End while

Remark 3.35. Note that the Newton algorithm converges quadratically, provided that the initial point is good enough. Therefore, we can see rapidly in step C whether the Newton algorithm converges or not and if we need to decrease the step $\Delta\mu_k$.

Remark 3.36. Clearly, the present algorithm does not take into account all possible events, since it principally assumes the stability of boundary arcs (which holds when uniform strict complementarity is satisfied, see assumption $(\mathcal{H}_2)$ below). If uniform strict complementarity does not hold along the homotopy path, then it may happen for example that a boundary arc splits into two boundary arcs, or on the contrary that two boundary arcs melt into one.

3.9.2 Existence of the homotopy path

Assume that the following holds:

$(\mathcal{H}_0)$ For $\mu_0 = 0$, the unconstrained problem $(\mathcal{P}^0)$ has a local solution $(\bar{u}, \bar{y})$ that satisfies (A0)-(A3), the contact set $I(g^0(\bar{y}))$ is composed of finitely many (nonessential) touch points in $(0, T)$, all of them satisfying (3.26), and the strong second-order sufficient condition (3.44) is satisfied.

By Th. 3.11, $(\mathcal{H}_0)$ implies that there exists $\tilde{\mu} > 0$ such that for all $\mu \in [0, \tilde{\mu})$, $(\mathcal{P}^\mu)$ has a locally unique local solution (u^μ, y^μ) with multipliers (p^μ, η^μ), that satisfies assumptions (A1)-(A3) for $(\mathcal{P}^\mu)$. In addition, this local solution (u^μ, y^μ) of $(\mathcal{P}^\mu)$ has a neighboring structure to that of $(\bar{u}, \bar{y})$, implying that if $(\bar{u}, \bar{y})$ has N touch points, then (u^μ, y^μ) has at most N boundary arcs or touch points, i.e. satisfies (A4). Further, strict complementarity holds on the boundary arcs of (u^μ, y^μ), and the touch points satisfy (3.26) by continuity, i.e. (A5)-(A6) are satisfied. Finally, (u^μ, y^μ) satisfies the strong second-order condition (3.44) for $(\mathcal{P}^\mu)$ by Lemma 3.12.

Consequently, assumption $(\mathcal{H}_0)$ ensures that the homotopy path is well-defined on an interval $[0,\tilde{\mu}) \subset [0,1]$, and that assumptions (A1)-(A6) as well as the strong second-order sufficient condition (3.44) remain satisfied on this neighborhood. Let

$$\mu_{max} := \sup\{\, \tilde{\mu} \in [0,1] : \text{for all } \mu \in [0,\tilde{\mu}],\ \text{the locally unique local solution } (u^\mu, y^\mu) \text{ of } (\mathcal{P}^\mu) \text{ satisfies (A1)-(A6) and (3.44).}\,\}.$$

The preceding discussion shows that assumption $(\mathcal{H}_0)$ implies that $\mu_{max} > 0$.

Lemma 3.37 (Existence of the homotopy path). *Assume that $(\mathcal{H}_0)$ holds, that there exists $L > 0$ such that for all $\mu \in [0,\mu_{max})$,*

$$\|\dot{u}^\mu\|_1 + \|u^\mu\|_\infty \ \leq\ L, \tag{3.157}$$

and that (A1) and (A3) are uniformly satisfied, i.e. there exist $\beta, \varepsilon, \zeta > 0$ such that for all $\mu \in [0,\mu_{max})$,

$$g^\mu(y_0^\mu) < -\zeta \quad \textit{and} \quad |(g^\mu)_u^{(1)}(u^\mu(t), y^\mu(t))| \geq \beta, \quad \text{for all } t,\ \operatorname{dist}\{t; I(g^\mu(y^\mu))\} \leq \varepsilon. \tag{3.158}$$

Then there exists a sequence $(\mu_n)_{n\in\mathbb{N}^*}$ such that $\mu_n \uparrow \mu_{max}$, $(u^{\mu_n}, y^{\mu_n}) \to (\tilde{u},\tilde{y})$ uniformly, $(p^{\mu_n}, \mathrm{d}\eta^{\mu_n})$ weakly- converges to $(\tilde{p}, \mathrm{d}\tilde{\eta})$ in $L^\infty(0,T;\mathbb{R}^{n*}) \times \mathcal{M}[0,T]$, and $(\tilde{u},\tilde{y},\tilde{p},\tilde{\eta})$ is a stationary point and its multipliers of $(\mathcal{P}^{\mu_{max}})$.*

Moreover, if $(\tilde{u},\tilde{y},\tilde{p},\tilde{\eta})$ satisfies assumptions (A1)-(A6) and the strong second-order sufficient condition (3.44), then $(u^\mu, y^\mu, p^\mu, \eta^\mu)$ converges when $\mu \uparrow \mu_{max}$ to a locally unique local solution of $(\mathcal{P}^{\mu_{max}})$ and its multipliers $(\tilde{u},\tilde{y},\tilde{p},\tilde{\eta}) =: (u^{\mu_{max}}, y^{\mu_{max}}, p^{\mu_{max}}, \eta^{\mu_{max}})$, and $\mu_{max} = 1$, i.e. the homotopy path is locally well-defined over $\mu \in [0,1]$.

Proof. Consider a sequence $(\mu_n)_{n\in\mathbb{N}^*} \subset [0,\mu_{max})$ such that $\mu_n \to \mu_{max}$ when $n \to +\infty$. Since $W^{1,1}(0,T)$ is compactly embedded in $C^0[0,T]$, (3.157) implies that there exists a subsequence, still denoted by (μ_n), such that the sequence (u^{μ_n}) converges uniformly to some $\tilde{u} \in \mathcal{U}$. By (3.157), we may pass to the limit in the state equation (3.2) and obtain that y^{μ_n} converges in $\mathcal{Y}$ to the state $\tilde{y} := y_{\tilde{u}}^{\mu_{max}}$ solution of (3.10).

By (3.158), Robinson's constraint qualification (3.17) is uniformly satisfied for all $\mu \in [0,\mu_{max})$, i.e. the positive constant γ in (3.17) does not depend on μ. It follows then from [24, Prop. 4.43] and (3.157) that $\|\mathrm{d}\eta^{\mu_n}\|_{\mathcal{M}[0,T]}$ is uniformly bounded. Therefore there exists a weakly-* convergent subsequence $\mathrm{d}\eta^{\mu_n} \overset{*}{\rightharpoonup} \mathrm{d}\tilde{\eta}$ in $\mathcal{M}[0,T]$. Since $\mathrm{d}\eta^\mu \in N_K(g^\mu(y^\mu))$ for all $\mu \in [0,\mu_{max})$, and $g^{\mu_n}(y^{\mu_n}) \to g^{\mu_{max}}(\tilde{y})$ strongly (i.e. uniformly), we deduce easily from the definition of the normal cone that $\mathrm{d}\tilde{\eta} \in N_K(g^{\mu_{max}}(\tilde{y}))$. By the costate equation (3.13) (with $\alpha = 1$), $\mathrm{d}p^\mu$ is uniformly bounded in $\mathcal{M}([0,T];\mathbb{R}^{n*})$. Therefore, there exists a weakly-* convergent subsequence $\mathrm{d}p^{\mu_n} \overset{*}{\rightharpoonup} \mathrm{d}\tilde{p} \in \mathcal{M}([0,T];\mathbb{R}^{n*})$. Due to the convergence of the final condition (3.14), we deduce easily from the integration by parts formula [58, p.154]

$$\int_0^T p(t)\varphi(t)\mathrm{d}t = -\int_0^T \mathrm{d}p(t)\Phi(t) + p(T)\Phi(T) \ \text{ for all } (p,\varphi) \in BV \times L^1 \text{ with } \Phi(t) := \int_0^t \varphi(s)\mathrm{d}s$$

that p^{μ_n} weakly-* converges in $L^\infty(0,T;\mathbb{R}^{n*})$ to a limit $\tilde{p}$ given by $\tilde{p}(t) := \int_T^t \mathrm{d}\tilde{p}(s) + \phi_y^{\mu_{max}}(\tilde{y}(T))$. Since (3.18) and (3.13) are linear in p and η, we may pass to the weak-* limit and obtain that $(\tilde{u},\tilde{y})$ is a stationary point of $(\mathcal{P}^{\mu_{max}})$ with multipliers $(\tilde{p},\tilde{\eta})$.

Now assume that this stationary point $(\tilde{u},\tilde{y})$ of $(\mathcal{P}^{\mu_{max}})$ satisfies assumptions (A1)-(A6) and the strong second-order sufficient condition (3.44). These assumptions imply by Th. 3.11

that $(\tilde{u}, \tilde{y})$ is an isolated stationary point of $(\mathcal{P}^{\mu_{max}})$, which shows the local uniqueness of the stationary point $(\tilde{u}, \tilde{y})$ of $(\mathcal{P}^{\mu_{max}})$ constructed above and of its multipliers. In addition $(\tilde{u}, \tilde{y})$ is a local solution of $(\mathcal{P}^{\mu_{max}})$, and by Th. 3.11, we obtain the existence of the homotopy path on the interval $[\mu_{max}, \mu_{max} + \varepsilon)$, for some $\varepsilon > 0$, and assumptions (A1)-(A6) hold on this interval by Th. 3.11, as well as the strong second-order condition (3.44) by Lemma 3.12. This implies that $\mu_{max} = 1$, otherwise this would contradict the definition of μ_{max}. Therefore the homotopy path is locally well-defined over $[0, 1]$. □

We thus make the assumptions below:

($\mathcal{H}_1$) For all $\mu \in [0,1]$, (u^μ, y^μ) satisfies (A2), there exist $L > 0$ and $\beta, \varepsilon, \zeta > 0$ such that (3.157) and (3.158) hold, and $g^\mu(y^\mu(T)) < 0$.

($\mathcal{H}_2$) For all $\mu \in [0,1]$, (u^μ, y^μ) has finitely many boundary arcs, and there exists $\beta > 0$ such that for all $\mu \in [0,1]$, $\dot{\eta}_1^\mu < -\beta$ on the boundary arcs of (u^μ, y^μ) (with η_1^μ the alternative state constraint multiplier associated with (u^μ, y^μ)).

($\mathcal{H}_3$) For all $\mu \in [0,1]$, (u^μ, y^μ) has finitely many (nonessential) touch points, all of them satisfying (3.26).

($\mathcal{H}_4$) For all $\mu \in [0,1]$, (u^μ, y^μ) satisfies the strong second-order sufficient condition (3.44) for $(\mathcal{P}^\mu)$.

Actually the algorithm 3.34 is correct only if we replace assumption ($\mathcal{H}_3$) by:

($\mathcal{H}_3'$) For all $\mu \in [0,1]$, (u^μ, y^μ) has *at most one* (nonessential) touch point, and the latter satisfies (3.26).

But the algorithm can be generalized to the more general case case when ($\mathcal{H}_3$) holds (see Rem. 3.47).

Remark 3.38. Assumptions ($\mathcal{H}_0$)-($\mathcal{H}_4$) needed to ensure the existence (and local uniqueness) of the homotopy path, and the convergence of the algorithm, are rather strong, but they also give some indications on why the algorithm fails, if it fails (for other reasons than numerical ones, see Rem. 3.46). Either (3.157) is not satisfied (i.e. u^μ is not uniformly Lipschitz continuous), or the problem becomes singular (i.e. (3.158) fails), or a solution with infinitely many boundary arcs or touch points is met during the homotopy, or strict complementarity on boundary arcs fails, or finally the strong second-order sufficient condition (3.44) fails.

3.9.3 Correctness of the algorithm

The existence of a locally unique local solution (u^μ, y^μ) of $(\mathcal{P}^\mu)$, for all $\mu \in [0,1]$, is guaranteed by assumptions ($\mathcal{H}_1$)-($\mathcal{H}_4$). In addition, for all $\mu \in [0,1]$, the locally unique local solution (u^μ, y^μ) of $(\mathcal{P}^\mu)$ has finitely many boundary arcs and touch points. So to prove the correctness of the algorithm, it suffices to show that the algorithm does find, in finitely many steps, these local solutions (u^μ, y^μ) for a finite increasing sequence of values of μ, until $\mu = 1$ (in fact, the algorithm gives the vector of shooting parameters θ^μ, of appropriate dimensions, associated with the trajectory (u^μ, y^μ)). For this Lemmas 3.39 to 3.43 given below will be useful.

Lemma 3.39. *Assume that* ($\mathcal{H}_0$)-($\mathcal{H}_4$) *hold. Then the trajectories* $(u^\mu, y^\mu)_{\mu \in [0,1]}$ *have finitely many different structures, and the mapping* $\mu \mapsto \theta^\mu$ *is globally Lipschitz continuous over* $[0, 1]$.

Here, since the dimension of θ^μ may vary, by "globally Lipschitz continuous" we mean that on any subinterval of $[0,1]$ where the trajectories (u^μ, y^μ) have "neighboring structures", then the mapping $\mu \mapsto \theta^\mu$ is Lipschitz continuous with a Lipschitz constant uniform on $[0,1]$.

Proof. By assumptions $(\mathcal{H}_1)$-$(\mathcal{H}_4)$ and Th. 3.11, for all $\mu \in [0,1]$, there exists an open neighborhood V_μ of μ such that for all $\mu' \in V_\mu$, the locally unique local solution $(u^{\mu'}, y^{\mu'})$ of $(\mathcal{P}^{\mu'})$ has a neighboring structure to that of (u^μ, y^μ), and the mapping $\mu' \mapsto \theta^{\mu'}$ is Lipschitz continuous over V_μ. We can thus extract from $(V_\mu)_{\mu\in[0,1]}$ a finite covering $(V_{\hat\mu_k})_{k=0,\dots,M}$ of $[0,1]$. Since for each $\hat\mu_k$, there exist finitely many possible neighboring structures to that of $(u^{\hat\mu_k}, y^{\hat\mu_k})$, and $\mu \mapsto \theta^\mu$ is Lipschitz continuous on each $V_{\hat\mu_k}$, the result follows. □

Although by Lemma 3.39 the trajectories $(u^\mu, y^\mu)_{\mu\in[0,1]}$ have finitely many different structures, assumptions $(\mathcal{H}_0)$-$(\mathcal{H}_4)$ do not imply that there are finitely many changes in the structure of the trajectory along the homotopy path (see Rem. 3.40 below). More precisely, we say that the *structure of the trajectory changes* at $\bar\mu \in [0,1)$, if $(u^{\bar\mu}, y^{\bar\mu})$ has a touch point that either disappears or turns into a boundary arc (of positive length) when $\mu \to \bar\mu^+$. We will therefore make the following assumption in the proof of correctness of the algorithm (Prop. 3.44), in addition to $(\mathcal{H}_0)$-$(\mathcal{H}_4)$ that ensure the existence of the homotopy path.

$(\mathcal{H}_5)$ There exist finitely many values of $\mu \in (0,1)$ for which the structure of the trajectory (u^μ, y^μ) changes.

Remark 3.40. Consider the problem (3.142), with $g = 1$, subject to the state constraint (3.143) where h depends on $\mu \in [0,1]$, i.e. $y \geq h^\mu$ with $h^\mu = -1/8 + \mu^5 \sin(1/\mu)$. For $\mu = 0$, there is a nonessential touch point at $\tau = 1/2$. When $\mu^5 \sin(1/\mu) > 0$, i.e. $\mu \in \cup_{n\in\mathbb{N}^*}(\frac{1}{(2n+1)\pi}, \frac{1}{2n\pi}) \cup (\frac{1}{\pi}, 1]$, then the latter turns into a boundary arc, and when $\mu^5 \sin(1/\mu) < 0$, i.e. $\mu \in \cup_{n\in\mathbb{N}^*}(\frac{1}{2n\pi}, \frac{1}{(2n-1)\pi})$, the boundary arc disappear (the state constraint is not active). Therefore, for any $\varepsilon > 0$ arbitrarily small, the structure changes for infinitely many values of μ in the interval $[0, \varepsilon]$. By Th. 3.30, the computation of the directional derivatives in direction $d = 1$ at point $\mu = 0$ shows that problem $(\mathcal{P}_d)$ has zero for unique solution, and therefore the directional derivatives of the entry/exit points and jump parameters at entry times are all zero in that case.

After this general description of the homotopy path, we will focus now on the changes in the structure, i.e. when there are nonessential touch points. So consider a value $\bar\mu \in [0,1]$ for which $(u^{\bar\mu}, y^{\bar\mu})$ has $N_{to} \geq 1$ (nonessential) touch points $\bar\tau_i$, $i = 1, \dots, N_{to}$. Denote by F_j, for $j = 1, \dots, 2^{N_{to}}$, the shooting mappings corresponding to *all* possible neighboring structures to that of $(u^{\bar\mu}, y^{\bar\mu})$, i.e. each touch point $\bar\tau_i$ *is or not* converted into a boundary arc like in subsection 3.5.2. Denote by $\bar\theta_j$ the appropriate vector of shooting parameters of $(u^{\bar\mu}, y^{\bar\mu})$ for F_j. Thus we have

$$F_j(\bar\theta_j, \bar\mu) \;=\; 0, \qquad \text{for all } j = 1, \dots, 2^{N_{to}}.$$

For μ in the neighborhood of $\bar\mu$, and all $j = 1, \dots, 2^{N_{to}}$, we consider the problem:

$$\text{Find } \theta \text{ of appropriate dimensions solution of:} \qquad F_j(\theta, \mu) \;=\; 0. \tag{3.159}$$

Lemma 3.41. *Assume that* $(\mathcal{H}_0)$-$(\mathcal{H}_4)$ *hold. Let* $\bar\mu \in [0,1]$ *be such that* $(u^{\bar\mu}, y^{\bar\mu})$ *has* $N_{to} \geq 1$ *(nonessential) touch points* $\bar\tau_i$, $i = 1, \dots, N_{to}$. *Then there exist an open neighborhood* $\bar V_\mu$ *of* $\bar\mu$ *and open neighborhoods* V_j *of* $\bar\theta_j$, $j = 1, \dots, 2^{N_{to}}$, *such that for all* $j = 1, \dots, 2^{N_{to}}$ *and for all* $\mu \in \bar V_\mu$, *the problem* (3.159) *has in* V_j *a unique solution* θ_j^μ, *and the mappings* $\bar V_\mu \to V_j$, $\mu \mapsto \theta_j^\mu$, *are of class* C^1.

Proof. By $(\mathcal{H}_3)$, the touch points $\bar{\tau}_i$ all satisfy (3.26). By $(\mathcal{H}_4)$ the strong second-order sufficient condition (3.44) is satisfied, and hence the Jacobians $D_\theta F_j(\bar{\theta}_j, \bar{\mu})$ are nonsingular, for all $j = 1, \ldots, 2^{N_{to}}$ (by the same arguments as in the proof of Lemma 3.26). So it follows from the classical implicit function Theorem that (3.159) has a locally a unique solution θ_j^μ, which is C^1 w.r.t. μ. □

Under the assumptions of Lemma 3.41, for $\mu \in \bar{V}_\mu$ and $j = 1, \ldots, 2^{N_{to}}$, denote by y_j^μ the state associated with θ_j^μ, i.e. solution of (3.77)-(3.80) for the arc structure of F_j. Note that y_j^μ is well-defined on each arc of the trajectory only (not on $[0, T]$), since some entry times in θ_j^μ may be greater than the corresponding exit times. Let $\hat{\theta}_j^\mu$ denote the *augmented vector of shooting parameters* obtained from θ_j^μ by adding, for each touch point $\bar{\tau}_i$ that was not converted into a boundary arc in F_j, a zero jump parameter for the costate and an entry and exit time both equal to the unique local maximum of $g^\mu(y_j^\mu(t))$ in the neighborhood of $\bar{\tau}_i$. Thus the augmented vectors of shooting parameters $\hat{\theta}_j^\mu$ have the same dimension for all j, which is also the dimension of the shooting mapping F in (3.84) for which *all* the N_{to} touch points are converted into boundary arcs. For $\mu = \bar{\mu}$, we denote the augmented vector of shooting parameters by $\bar{\theta} = \hat{\theta}_j^{\bar{\mu}}$, for all j.

Lemma 3.42. *Under the assumptions of Lemma 3.41, there exists an open neighborhood $\bar{\bar{V}}_\mu$ of $\bar{\mu}$ such that for all $j = 1, \ldots, 2^{N_{to}}$, the mapping $\mu \mapsto \hat{\theta}_j^\mu$ is C^1 over $\bar{\bar{V}}_\mu$, and for all $\mu \in \bar{\bar{V}}_\mu$, the augmented vector of shooting parameters $\hat{\theta}_j^\mu$ is solution of* (3.85)*, iff the two conditions below are satisfied:*

$$g^\mu(y_j^\mu(t)) \leq 0, \qquad \text{on each arc,} \tag{3.160}$$

$$\tau_{en,j}^\mu \leq \tau_{ex,j}^\mu, \qquad \text{for all boundary arcs,} \tag{3.161}$$

where for each boundary arc of F_j, $\tau_{en,j}^\mu$ and $\tau_{ex,j}^\mu$ denote the components of θ_j^μ corresponding respectively to the entry and exit point of the boundary arc.

Proof. In the neighborhood of a touch point $\bar{\tau}_i$ that was not converted into a boundary arc in F_j, for all $\mu \in \bar{V}_\mu$, the function $g^\mu(y_j^\mu(\cdot))$ is locally well-defined and C^2. Therefore, since $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g^{\bar{\mu}}(y^{\bar{\mu}})|_{t=\bar{\tau}_i} < 0$, the function that with $g^\mu(y_j^\mu)$ associates its (unique) local maximum time in the neighborhood of $\bar{\tau}_i$ is C^1, and hence, by Lemma 3.41, $\mu \mapsto \hat{\theta}_j^\mu$ is C^1. Now denote by t_{en}^i and ν_i^1 respectively the entry time and jump parameter of the boundary arc associated with the touch point $\bar{\tau}_i$ in $\hat{\theta}_j^\mu$, $j = 1, \ldots, 2^{N_{to}}$. By the arguments of the proof of Lemma 3.27, we have that (3.161) is equivalent to $\nu_i^1 \geq 0$ for all $i = 1, \ldots, N_{to}$, and for each i we have either $g^\mu(y_j^\mu(t_{en}^i)) = 0$ or $\nu_i^1 = 0$. Therefore (3.160)-(3.161) are equivalent to the condition $\Psi(\hat{\theta}_j^\mu, \mu) \in N(\hat{\theta}_j^\mu)$. The conclusion follows. □

Let $j_1, j_2 \in \{1, \ldots, 2^{N_{to}}\}$, $j_1 \neq j_2$, and $\mu \in \bar{\bar{V}}_\mu$. Given a solution $\theta_{j_1}^\mu$ of (3.159) for $j = j_1$, let us explain now how to initialize the Newton algorithm in order to find a solution of (3.159) for $j = j_2$. The initial point $\tilde{\theta}_{j_1,j_2}^\mu$ is obtained from $\theta_{j_1}^\mu$ as follows:

- For every touch point $\bar{\tau}_i$ that was converted into a boundary arc in F_{j_1} but not in F_{j_2}, remove from $\theta_{j_1}^\mu$ the shooting parameters associated with this boundary arc;
- For every touch point $\bar{\tau}_i$ that was converted into a boundary arc in F_{j_2} but not in F_{j_1}, add to $\theta_{j_1}^\mu$ the three shooting parameters associated with this boundary arc $(\nu^{1,i}, \tau_{en}^i, \tau_{ex}^i)$ as follows: $\nu^{1,i} = 0$, and τ_{en}^i and τ_{ex}^i are both equal to the unique point of local maximum of $g^\mu(y_{j_1}^\mu)$ in the neighborhood of $\bar{\tau}_i$.

Lemma 3.43. *Under the assumptions of Lemma 3.41, for all $j_1, j_2 \in \{1, \ldots, 2^{N_{to}}\}$, $j_1 \neq j_2$, there exists $\bar{\delta}_{j_1,j_2} > 0$, such that for all μ, $|\mu - \bar{\mu}| \leq \bar{\delta}_{j_1,j_2}$, the Newton method to solve the equation (3.159) for $j = j_2$ is convergent to $\theta^\mu_{j_2}$, whenever the initial point $\tilde{\theta}^\mu_{j_1,j_2}$ is obtained from the solution $\theta^\mu_{j_1}$ of (3.159) for $j = j_1$ as explained above.*

Proof. By Lemma 3.41, there exists $\rho^\mu_{j_2} > 0$ such that the Newton algorithm to solve (3.159) for $j = j_2$ converges to $\theta^\mu_{j_2}$ for all initial point θ_0 satisfying $|\theta_0 - \theta^\mu_{j_2}| < \rho^\mu_{j_2}$, and this constant is uniformly positive, i.e. $\rho^\mu_{j_2} \geq \rho > 0$ for all μ in a compact neighborhood of $\bar{\mu}$ [8]. Let $|\mu - \bar{\mu}| \leq \bar{\delta}_{j_1,j_2} := \min(\kappa^{-1}_{j_1}, \kappa^{-1}_{j_2})\rho/3$, with κ_j the Lipschitz constant of the mapping $\mu \mapsto \hat{\theta}^\mu_j$ (Lemma 3.42). Let $\bar{\theta} := \hat{\theta}^{\bar{\mu}}_{j_1} = \hat{\theta}^{\bar{\mu}}_{j_2}$ and note that we obviously have $\tilde{\theta}^{\bar{\mu}}_{j_1,j_2} = \bar{\theta}_{j_2}$. It follows then that $|\tilde{\theta}^\mu_{j_1,j_2} - \theta^\mu_{j_2}| \leq |\tilde{\theta}^\mu_{j_1,j_2} - \tilde{\theta}^{\bar{\mu}}_{j_1,j_2}| + |\bar{\theta}_{j_2} - \theta^\mu_{j_2}| \leq |\hat{\theta}^\mu_{j_1} - \bar{\theta}| + |\bar{\theta} - \hat{\theta}^\mu_{j_2}| \leq \frac{2}{3}\rho$, from which the result follows. □

We give now a theoretical proof of correctness of the algorithm.

Proposition 3.44. *Assume that $(\mathcal{H}_0)$-$(\mathcal{H}_2)$, $(\mathcal{H}'_3)$ and $(\mathcal{H}_4)$-$(\mathcal{H}_5)$ hold. Then there exists $\delta_0 > 0$ such that, whenever p_0 is close enough to $\bar{p}(0)$, for all $0 < \delta < \delta_0$ the algorithm 3.34 follows the homotopy path previously described, and ends with a vector of shooting parameters θ^1 of adapted dimension associated with a local solution (u^1, y^1) of $(\mathcal{P}^1) \equiv (\mathcal{P})$. In addition, if $0 < \delta < \delta_0$, the steps $\Delta\mu_k$ are not reduced by the algorithm (i.e. Newton's algorithm in step C do not fail).*

Proof. By $(\mathcal{H}_5)$, there exist finitely many values of $\mu \in (0, 1)$, $0 < \bar{\mu}_1 < \ldots < \bar{\mu}_m < 1$, for which the structure of the trajectory (u^μ, y^μ) changes. By $(\mathcal{H}'_3)$, this implies that for all $j = 1, \ldots, m$, the trajectory associated with $\bar{\mu}_j$ has exactly one touch point $\bar{\tau}^j_{to}$. Set $\bar{\mu}_0 := 0$ and $\bar{\mu}_{m+1} := 1$. For all $j = 0, \ldots, m$, denote by F_j the shooting mapping corresponding to the structure of (u^μ, y^μ) for $\mu \in (\bar{\mu}_j, \bar{\mu}_{j+1})$. We have $F_j \neq F_{j+1}$, for all $j = 0, \ldots, m$.

Let $j = 0, \ldots, m$. For all $\mu \in [\bar{\mu}_j, \bar{\mu}_{j+1}]$, by (3.44) and Lemma 3.41, there exists a constant $\rho_j > 0$ (uniform w.r.t. μ, see [8]) such that the shooting algorithm (i.e. Newton's algorithm to solve $F_j(\theta, \mu) = 0$) converges to θ^μ for all initial point θ_0 satisfying $|\theta_0 - \theta^\mu| < \rho_j$. For all $\mu, \mu' \in [\bar{\mu}_j, \bar{\mu}_{j+1}]$, with θ' the solution of the prediction step obtained from θ^μ by

$$D_\theta F_j(\theta^\mu, \mu)(\theta' - \theta^\mu) + D_\mu F_j(\theta^\mu, \mu)(\mu' - \mu) \;=\; 0,$$

it is easy to see that there exists a constant C_j [9] such that $|\theta' - \theta^{\mu'}| \;\leq\; C_j|\mu - \mu'|^2$. Therefore the convergence of the Newton algorithm to $\theta^{\mu'}$ with the initial point θ' is guaranteed if $|\mu - \mu'| < \hat{\delta} := \min_{j=0}^m(\rho_j/C_j)^{1/2}$. Now let $\delta_0 > 0$ be the minimum of $\hat{\delta}$ defined above, of all the finitely many constants $\bar{\delta}_{j_1,j_2} > 0$ of Lemma 3.43 involved at the changes of structure of the trajectory, and finally of $\bar{\mu}_{j+1} - \bar{\mu}_j > 0$, for $j = 0, \ldots, m$.

Let $\delta \in (0, \delta_0)$. The proof of the the algorithm is by finite induction on the property below, for $k \geq 0$:

($\mathcal{A}_k$) At each passage in the prediction step (step B), before k is increased, we have $\mu_k = \min(k\delta, 1)$, $m_k = 0$, $i_k = 0$ and

[8] From the proof of the Newton algorithm, it can be seen that this constant ρ^μ_j depends continuously on the Lipschitz constant of $D_\theta F_j(\cdot, \mu)$, on $\|D_\theta F_j(\theta^\mu_j, \mu)^{-1}\|^{-1}$ and on the modulus of continuity of $D_\theta F_j(\cdot, \mu)^{-1}$, and is therefore a continuous function of μ.

[9] This constant C_j depends on $\|D_\theta F_j(\theta^\mu, \mu)^{-1}\|$, on the Lipschitz constant of DF_j and on the Lipschitz constant of the mapping $\mu \mapsto \theta^\mu$ on $[\bar{\mu}_j, \bar{\mu}_{j+1}]$ (Lemma 3.39).

- if $\mu_k \notin \{\bar{\mu}_j\}_{j=0,\dots,m}$, $\theta_k = \theta^{\mu_k}$ is the (unique) vector of shooting parameters associated with (u^{μ_k}, y^{μ_k}),
- if $\mu_k = \bar{\mu}_j$ for some $j = 0, \dots, m$,
  - if either $k = 0$ or the touch point of $\bar{\mu}_j$ is either inactive or a (nonessential) touch point when $\mu \to \bar{\mu}_j^-$, then θ_k is the vector of shooting parameters associated with (u^{μ_k}, y^{μ_k}) that does not contain the touch point of $\bar{\mu}_j$ as a boundary arc of zero length;
  - if the touch point of $\bar{\mu}_j$ is a boundary arc for $\mu \to \bar{\mu}_j^-$, then θ_k is the vector of shooting parameters associated with (u^{μ_k}, y^{μ_k}) that contains the touch point of $\bar{\mu}_j$ as a boundary arc of zero length.

For p_0 sufficiently close to $\bar{p}(0)$, the initialization step of the algorithm succeeds in obtaining the initial vector of shooting parameters (reduced to the initial costate) $\theta^0 = \theta^{\mu_0}$ associated with the local solution $(\bar{u}, \bar{y})$ of $(\mathcal{P}^0)$. So $(\mathcal{A}_0)$ holds. Assume now that $(\mathcal{A}_{k-1})$ holds, and let $j \in \{0, \dots, m+1\}$ be such that

$$\bar{\mu}_j \;<\; \mu_{k-1} \;\leq\; \bar{\mu}_{j+1}.$$

We thus go through the prediction step B and then to step C. By $(\mathcal{A}_{k-1})$, we try to solve, by the Newton algorithm, the equation

$$F_j(\theta, \mu_k) = 0. \tag{3.162}$$

By construction of δ_0, the Newton algorithm succeeds and obtain a solution θ'_k of (3.162). So we go to step D. There are two cases to consider. Either (a) $\mu_k \leq \bar{\mu}_{j+1}$ or (b) $\mu_k > \bar{\mu}_{j+1}$.

In case (a), the structure of the trajectory does not change, so we obtain the vector of shooting parameters $\theta_k := \theta'_k = \theta^{\mu_k}$ associated with (u^{μ_k}, y^{μ_k}). Therefore $m_k \leq 0$ and $i_k = 0$, which shows $(\mathcal{A}_k)$.

In case (b), by construction of δ_0, we have $\mu_k \in (\bar{\mu}_{j+1}, \bar{\mu}_{j+2})$. Therefore θ^{μ_k} is the (locally unique) solution of

$$F_{j+1}(\theta, \mu_k) = 0. \tag{3.163}$$

By Lemma 3.42, among all the "augmented vectors of shooting parameters" associated with one of the (two) possible neighboring structures to $(u^{\bar{\mu}_{j+1}}, y^{\bar{\mu}_{j+1}})$, only θ^{μ_k} satisfies (3.160)-(3.161). Therefore we deduce that necessarily, the augmented vector of shooting parameters $\hat{\theta}_k$ obtained from θ'_k solution of (3.162) does not satisfy either (3.160) or (3.161), i.e. either $m_k > 0$ or $i_k > 0$.

Assume e.g. that $m_k > 0$, i.e. $g^{\mu_k}(y^{\theta'_k, \mu_k})$ has positive values. Using $(\mathcal{H}_2)$ and Lemma 3.22, this can only happen in the neighborhood of the touch point $\bar{\tau}_{to}^{j+1}$ of $\bar{\mu}_{j+1}$, i.e. τ_k is close to $\bar{\tau}_{to}^{j+1}$. Note that this is possible only if $\bar{\tau}_{to}^{j+1}$ was not converted in a boundary arc in F_j. So we go to step A and add a boundary arc. Here, $\bar{\mu}_j$ having a single touch point, there are only two possible neighboring structures to that of $(u^{\bar{\mu}_j}, y^{\bar{\mu}_j})$. Having eliminated F_j, it remains only one possible structure, i.e. with $\bar{\tau}_{to}^{j+1}$ as a boundary arc, which corresponds necessarily to F_{j+1}. The shooting parameters associated with this new boundary arc are initialized by (3.156), and hence we obtain an augmented vector of shooting parameters $\tilde{\theta}_k$, that by Lemma 3.43 belongs, by construction of δ_0, to the neighborhood of $\theta^{\bar{\mu}_{j+1}}$ for which the Newton algorithm solving (3.163) is convergent to θ^{μ_k}. We thus obtain $\theta_k = \theta^{\mu_k}$, which satisfies $m_k = 0$ and $i_k = 0$, and therefore $(\mathcal{A}_k)$ holds.

The case $i_k > 0$ is dealt with similarly, i.e. if it happens that for a boundary arc, the entry time is greater than the exit time, this can only happen in the neighborhood of the touch point $\bar{\tau}_{to}^{j+1}$ of $\bar{\mu}_{j+1}$, and this implies that this touch point is converted in a boundary arc in F_j. So we remove in step A this boundary arc, and conclude with the same arguments that $(\mathcal{A}_k)$ holds again. The result follows by finite induction on k, since the algorithms ends for the smaller integer $k \geq \frac{1}{\delta}$. □

Remark 3.45. The process of reduction of $\Delta\mu_k$ is not active if δ is small enough, as appears from Prop. 8.8. However, in practice we do not know what a correct value of $\Delta\mu_k$ is so that this reduction process is useful.

Of course when initialized with $\delta > \delta_0$ it may happen that Newton's method converges to a point that does not belong to the continuous path (u^μ, y^μ), i.e., it computes another critical point, say $(\hat{u}^\mu, \hat{y}^\mu)$. If the latter satisfies conditions of Th. 3.11, then the algorithm continues despite the jump to another branch of solutions.

Remark 3.46. We could theoretically give an explicit expression for the constant δ_0 that ensures the convergence in Prop. 3.44, but the latter depends on constants involving, among other, bounds on the hessian of the shooting mapping that are almost impossible to calculate. In case of ill-conditioning (δ_0 is very small), the convergence may be difficult, if not impossible, to achieve in practice, due to numerical errors.

Remark 3.47. Algorithm 3.34 and Prop. 3.44 can be extended to the case when $(\mathcal{H}_3)$ holds instead of $(\mathcal{H}_3')$. If $(\mathcal{H}_3')$ does not hold, but $(\mathcal{H}_3)$ do, this means that there exists $\bar{\mu} \in (0,1)$ such that $(u^{\bar{\mu}}, y^{\bar{\mu}})$ has N_{to} touch points, $N_{to} \geq 2$. If the structure of the shooting mapping changes at this point, there are a priori $2^{N_{to}}$ possibilities for the new structure when $\mu \to \hat{\mu}^+$. It is possible to enumerate all of them, i.e. solve (3.159), for all $j = 1, \ldots, 2^{N_{to}}$, for $\mu > \bar{\mu}$ close to $\bar{\mu}$. Lemma 3.42 ensures that if (3.160)-(3.161) are satisfied for some j, then we have found the new structure, and Lemma 3.26 ensures that (3.160)-(3.161) will be satisfied for at least one j.

A possibility that may reduces the enumeration is to use the directional differentiability of solutions in Th. 3.30. One can e.g. solve the problem $(\mathcal{P}_d)$, and whenever the variation $\sigma_{d,i}^{ex} - \sigma_{d,i}^{en}$ given by (3.132) is positive (resp. negative), this tells us that the touch point τ_{to}^i have to be converted into boundary arc (resp. removed from the shooting mapping). For touch points such that $\sigma_{d,i}^{ex} - \sigma_{d,i}^{en} = 0$, this gives no information on τ_{to}^i so it possibly remains different possibilities to enumerate.

3.9.4 Numerical Implementation

The convergence of the algorithm presented in the previous subsections is illustrated on the academic problem below:

$$(\mathcal{P}) \qquad \min \int_0^1 \left(\frac{u^2(t)}{2} + g(t)y(t) \right) \mathrm{d}t$$
$$\text{s.t.} \qquad \dot{y}(t) = u(t), \qquad y(0) = y(1) = 0, \qquad y(t) \geq h$$

with

$$g(t) := g_0(c - \sin(\alpha t)), \qquad c, \alpha > 0.$$

The time is introduced as a state variable, and let $\mu = (h - h_0)/(h_1 - h_0)$ be the homotopy parameter, with $h_0 = \min \bar{y}(t)$, for $\bar{y}$ the solution of the problem without the state constraint,

and $h_1 = h$ the desired value of the state constraint. Numerical values of constants are taken equal to

$$g_0 := 10, \qquad \alpha = 10\pi, \qquad c = 0.1, \qquad h_1 = -0.001.$$

The algorithm is initialized with the value $p_0 = 0$, and $\delta = 1/5$ to initialize the steps $\Delta\mu_k$. Let us comment Figure 3.3 where the results of the algorithm are presented. The algorithm reduces the step $\Delta\mu_k$ once, in the next to last iteration, since the Newton algorithm was not converging, meaning here that it was not converging quadratically. Thus the solution was computed for the values $\mu_0 = 0$, $\mu_k = k\delta = k/5$ for $k = 1, \ldots, 4$, $\mu_5 = 9/10$ and $\mu_6 = 1$. We plotted in dark blue the state y_k solution of $(\mathcal{P}^{\mu_k})$ obtained at the exit of the WHILE loop when $m_k = i_k = 0$, for $k = 0, \ldots, 6$. In light blue we plotted the previous iterations, including the states obtained when $m_k > 0$ at the exit of the WHILE loop (so we can see the algorithm add a boundary arc at the following iteration when this happens).

For $k = 0$, we just have the solution of the unconstrained problem. For $k = 1$, the algorithm adds a single boundary arc around time $t = 0.55$. At each iteration $k = 2, 3, 4$, the algorithm detects that the state constraint is violated so it adds a boundary arc. So for $k = 4$ we have $\mu_k = 0.8$ and four boundary arcs. Then the algorithm tries to pursue the homotopy with $\mu = 1$. It detects that it has to add a boundary arc but Newton algorithm fails. Therefore it decreases the step and obtained the solution for μ_5 (see the figure for $k = 5$) that has a fifth boundary arc. It then increases μ to $\mu_6 = 1$ and obtain the solution of $(\mathcal{P})$ which exhibits five boundary arcs.

At each passage in the Newton algorithm (step C), the latter converges very rapidly in 2 or 3 iterations (for the tolerance $|F(\theta_k, \mu_k)|_\infty \leq 10^{-10}$) excepted of course the time it failed because $\Delta\mu_k$ was too large, and at the very last passage (which requires 5 iterations).

Finally, let us check that the uniform strict complementarity hypothesis $(\mathcal{H}_2)$ is satisfied. On a boundary arc, (3.48) gives

$$u_b + p_1 - \eta_1 = 0 \qquad \text{with } u_b = 0,$$

i.e. $p_1 = \eta_1$. Hence, $\dot\eta_1 \leq \beta < 0$ on boundary arcs iff p_1 is (uniformly) decreasing. This is the case, see the figure bottom right in Fig. 3.3 on which we plotted p_1 for the final solution for $\mu_6 = 1$ (the portions corresponding to boundary arcs are plotted in red). We can also check similarly that this uniform strict complementarity assumption is satisfied as well for all other values of μ_k, $k = 1, \ldots, 5$.

3.10 Proof of Theorem 3.4

We start by the proof of Lemma 3.6, then give that of Lemma 3.8, and finally that of Th. 3.4.

Proof of Lemma 3.6. Let $\delta > 0$. By continuity of the mapping $(u, \mu) \mapsto g^\mu(y_u^\mu)$, there exists $\delta > 0$, such that for n large enough (this is precisely assertion (S1)),

$$I(g^{\mu_n}(y_n)) \subset \Omega^\delta := \cup_{i=1}^N \Omega_i^\delta. \tag{3.164}$$

The first assertion of the lemma is a classical consequence of Robinson's constraint qualification (3.17) (see e.g. [24, Prop. 4.43]). By Lemma 3.2, reducing δ if necessary, the mapping (3.23) is onto. Since $\mathrm{supp}(\mathrm{d}\eta_n) \subset I(g^{\mu_n}(y_n)) \subset \Omega^\delta$ by (3.164), the second assertion follows from [24,

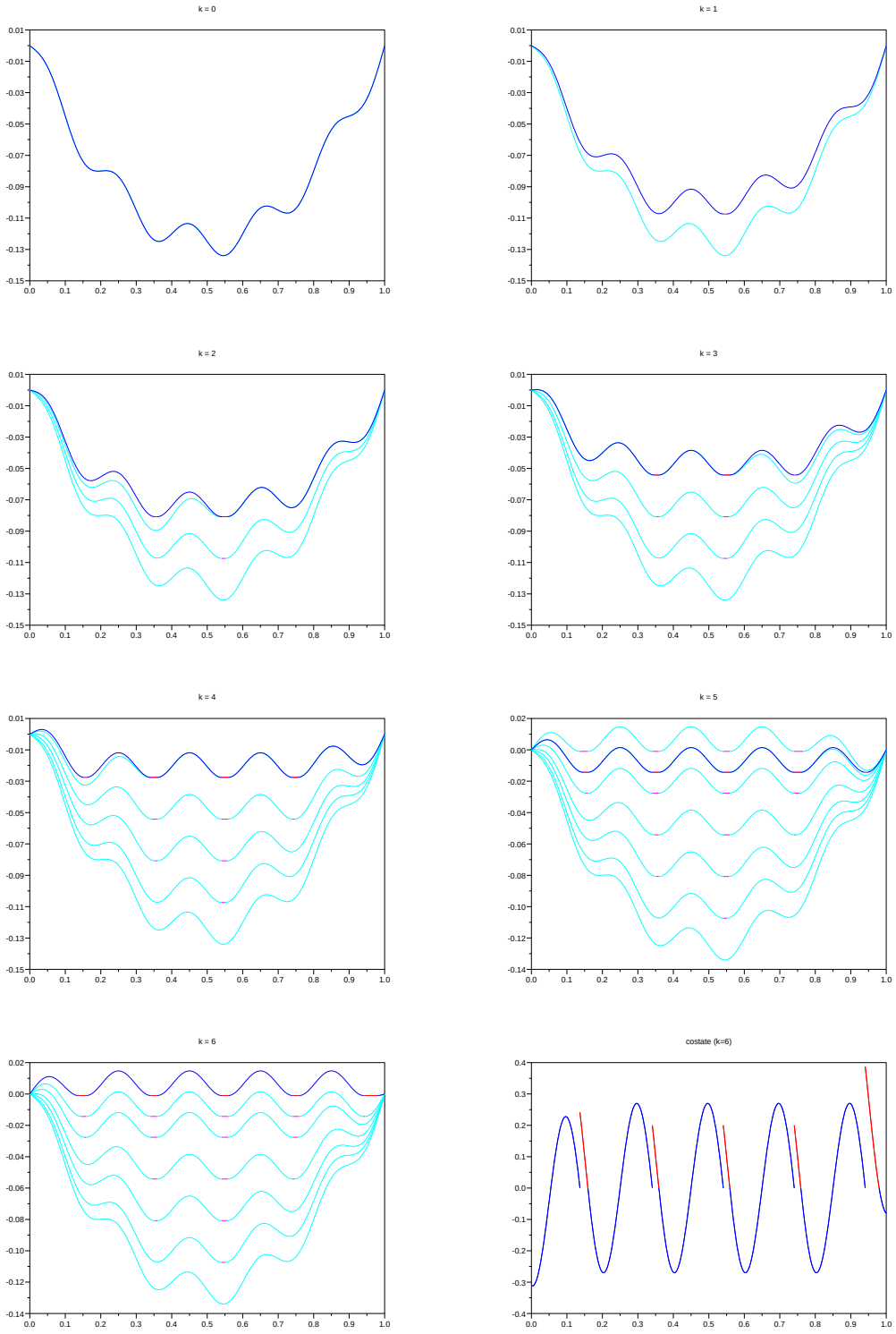


Figure 3.3: Iterations of the homotopy algorithm and costate p_1 (for $k = 6$)

Prop. 4.44 and Rem. 4.45(i)], meaning that

$$\sup_{\Phi\in W^{1,\infty}(0,T),\ \Phi\not\equiv 0} \frac{\left|\int_0^T \Phi(t)(\mathrm{d}\eta_n-\mathrm{d}\eta)(t)\mathrm{d}t\right|}{\|\Phi\|_{1,\infty}} \underset{n\to+\infty}{\longrightarrow} 0. \tag{3.165}$$

Let p_n^1 and η_n^1 be the multipliers associated with the stationary point (u_n, y_n) of $(\mathcal{P}^{\mu_n})$ by (3.32)-(3.33). By (3.34),

$$\frac{\mathrm{d}}{\mathrm{d}t}(p_n^1-\bar p^1) \;=\; (p_n^1-\bar p^1)f_y(\bar u,\bar y)+g_y^{(1)}(\bar u,\bar y)(\eta_n^1-\bar\eta^1)+r_n(t) \qquad \text{a.e. on } [0,T],$$

with $\|r_n\|_\infty\to 0$ when $n\to+\infty$. By Gronwall's Lemma, there exists a constant $C>0$ such that

$$\begin{aligned} |p_n^1(t)-\bar p^1(t)| &\le C|\phi_y^{\mu_n}(y_n(T))-\phi_y^{\mu_0}(\bar y(T))|+C\int_t^T|\eta_n^1(s)-\bar\eta^1(s)|\mathrm{d}s+o_\infty(1)\\ &\le C\|\eta_n^1-\bar\eta^1\|_1+o_\infty(1), \end{aligned} \tag{3.166}$$

where $o_\infty(1)$ denotes a function that goes to zero in L^∞ when $n\to+\infty$. Let us show that $\eta_n^1\to\bar\eta^1$ in L^1. The sequence $(\mathrm{d}\eta_n)_{n\in\mathbb{N}^*}$ being bounded in $\mathcal{M}[0,T]$ by the first assertion (1), it follows that $(\eta_n^1)_{n\in\mathbb{N}^*}$ is bounded in BV, for the norm $\|\eta\|_{BV}=\|\eta\|_1+\|\mathrm{d}\eta\|_{\mathcal{M}}$. By the compactness Theorem in BV [2, Th. 3.23], there exists a subsequence $(\eta^1_{\psi(n)})_{n\in\mathbb{N}^*}$ converging in L^1 to some $\tilde\eta\in BV(0,T)$, and such that $\mathrm{d}\eta_{\psi(n)}\overset{*}{\rightharpoonup}-\mathrm{d}\tilde\eta$ in $\mathcal{M}[0,T]$. It suffices then to show that necessarily, $-\mathrm{d}\tilde\eta=\mathrm{d}\bar\eta$ and $\tilde\eta=\bar\eta^1$ in order to obtain the convergence of the whole sequence $(\eta_n^1)_{n\in\mathbb{N}^*}$ to $\bar\eta^1$ in L^1. So let us do that. The space $W^{1,\infty}(0,T)$ being dense in $C^0[0,T]$, it follows easily from (3.165) that $\mathrm{d}\eta_n\overset{*}{\rightharpoonup}\mathrm{d}\bar\eta$, and hence $-\mathrm{d}\tilde\eta=\mathrm{d}\bar\eta$. Thus $\tilde\eta$ equals $\bar\eta$ up to a constant. Using Fubini's Theorem and (3.165), we obtain

$$\int_0^T\eta_n^1(t)\mathrm{d}t \;=\; \int_0^T s\mathrm{d}\eta_n(s) \underset{n\to+\infty}{\longrightarrow} \int_0^T s\mathrm{d}\bar\eta(s) \;=\; \int_0^T\bar\eta^1(t)\mathrm{d}t,$$

implying finally that $\tilde\eta=\bar\eta$, and consequently, that $\eta_n^1\to\bar\eta^1$ in L^1. By (3.166), we deduce then that $p_n^1\to\bar p^1$ uniformly over $[0,T]$.

Finally, for $\|u_n-\bar u\|_\infty$ small enough, $|(g^{\mu_n})_u^{(1)}(u_n,y_n)|\ge\beta/2>0$ on Ω^δ, so by (3.35) we have on Ω^δ:

$$\eta_n^1 \;=\; -\frac{H_u^{\mu_n}(u_n,y_n,p_n^1)}{(g^{\mu_n})_u^{(1)}(u_n,y_n)} \;\to\; -\frac{H_u^{\mu_0}(\bar u,\bar y,\bar p^1)}{(g^{\mu_0})_u^{(1)}(\bar u,\bar y)} \;=\; \bar\eta^1 \quad \text{ uniformly on } \Omega^\delta,$$

and η_n^1 is piecewise constant on $[0,T]\setminus\Omega^\delta$, which shows the last assertion. □

Proof of Lemma 3.8. Let (u,y) be a stationary point of $(\mathcal{P}^\mu)$ with multipliers (p^1,η^1) given by (3.32)-(3.33). By time derivation of (3.35), we have, using the augmented Hamiltonian (3.8),

$$\tilde H_{uu}^\mu(u,y,p^1,\eta^1)\dot u+\tilde H_{uy}^\mu(u,y,p^1,\eta^1)f^\mu(u,y)-\tilde H_y^\mu(u,y,p^1,\eta^1)f_u^\mu(u,y)+(g^\mu)_u^{(1)}(u,y)\dot\eta^1 \;=\; 0. \tag{3.167}$$

For $\|\mu-\mu_0\|$ and $\|u-\bar u\|_\infty$ small enough, then $\|y-\bar y\|_\infty$ is arbitrarily small, as well as $\|p^1-\bar p^1\|_\infty$ and $\|\eta^1-\bar\eta^1\|_\infty$ by Lemma 3.6. Consequently, for (u,μ) close enough to $(\bar u,\mu_0)$, we have by

(3.20) that $\tilde{H}^{\mu}_{uu}(u,y,p^1,\eta^1) \geq \alpha/2$ on $[0,T]$. Multiplying (3.167) by $(g^\mu)^{(1)}_u(u,y)/\tilde{H}^{\mu}_{uu}(u,y,p^1,\eta^1)$, we obtain that

$$(g^\mu)^{(1)}_u(u,y)\dot{u} + \frac{(g^\mu)^{(1)}_u(u,y)^2}{\tilde{H}^{\mu}_{uu}(u,y,p^1,\eta^1)}\dot{\eta}^1 \;\rightarrow\; g^{(1)}_u(\bar{u},\bar{y})\dot{\bar{u}} + \frac{g^{(1)}_u(\bar{u},\bar{y})^2}{\tilde{H}_{uu}(\bar{u},\bar{y},\bar{p}^1,\bar{\eta}^1)}\dot{\bar{\eta}}^1 \tag{3.168}$$

uniformly over $[0,T]$. In view of (3.5)-(3.6), it follows that

$$(g^\mu)^{(2)}(\dot{u},u,y) + \frac{(g^\mu)^{(1)}_u(u,y)^2}{\tilde{H}^{\mu}_{uu}(u,y,p^1,\eta^1)}\dot{\eta}^1 \;\rightarrow\; g^{(2)}(\dot{\bar{u}},\bar{u},\bar{y}) + \frac{g^{(1)}_u(\bar{u},\bar{y})^2}{\tilde{H}_{uu}(\bar{u},\bar{y},\bar{p}^1,\bar{\eta}^1)}\dot{\bar{\eta}}^1 \tag{3.169}$$

again, uniformly over $[0,T]$.

Now on every Ω^δ_i, for small enough $\delta > 0$, we have by (A5)-(A6), (3.27) and (3.21) the existence of a constant $\kappa_1 > 0$ such that either $g^{(2)}(\dot{\bar{u}},\bar{u},\bar{y}) < -\kappa_1$ and $\dot{\bar{\eta}}^1 = 0$, or $g^{(2)}(\dot{\bar{u}},\bar{u},\bar{y}) = 0$, $|g^{(1)}_u(\bar{u},\bar{y})| \geq \kappa_1$ and $\dot{\bar{\eta}}^1 \leq -\kappa_1$. It follows that, for some $\kappa_2 > 0$, δ small enough and (μ,u) close to $(\mu_0,\bar{u})$,

$$(g^\mu)^{(2)}(\dot{u},u,y) + \frac{(g^\mu)^{(1)}_u(u,y)^2}{\tilde{H}^{\mu}_{uu}(u,y,p^1,\eta^1)}\dot{\eta}^1 \;\leq\; -\kappa_2 \qquad \text{on } \Omega^\delta_i. \tag{3.170}$$

If $g^\mu(y(t)) < 0$, then $\dot{\eta}^1(t) = 0$, and hence, $(g^\mu)^{(2)}(\dot{u},u,y)(t) < -\kappa_2/2$. But on an interior arc included in Ω^δ_i, $g^\mu(y)$ would attain its minimum at some point t where $(g^\mu)^{(2)}(\dot{u},u,y)(t) \geq 0$, which gives the desired contradiction. □

Remark 3.48. It follows from (3.170) that the property of uniform strict complementarity is stable, in the sense that if the state constraint is active, then $\dot{\eta}^1$ remains uniformly far from zero (uniformly over $[0,T]$).

Now we are ready to give the proof of Th. 3.4.

Proof of Th. 3.4. Assertion (S1) is immediate, and (S3) follows directly from Lemma 3.8 since there is no interior arc of (u,y) in Ω^δ_i. In view of Lemma 3.8, to complete the proof of (S2), it remains to show that $\Omega^\delta_i \cap I(g^\mu(y))$ is an interval of positive measure, i.e. a boundary arc. Assume that this is false. Then there exist a stable extension $(\mathcal{P}^\mu)$, sequences $u_n \rightarrow \bar{u}$ in L^∞, $\mu_n \rightarrow \mu_0$, and (u_n,y_n) a stationary point of $(\mathcal{P}^{\mu_n})$, such that for all n, $\Omega^\delta_i \cap I(g^{\mu_n}(y_n))$ is either empty or a singleton by Lemma 3.8. Taking if necessary a subsequence, this implies that there exists an interval of positive measure $(t_1,t_2) \subset [\bar{t}^{en}_i, \bar{t}^{ex}_i]$, such that $(t_1,t_2) \cap I(g^{\mu_n}(y_n)) = \emptyset$ for all n, and hence, $(t_1,t_2) \cap \mathrm{supp}(\mathrm{d}\eta_n) = \emptyset$. Let φ be a C^∞ function with support in $[t_1,t_2]$ which is positive on (t_1,t_2). Then we have $\int_0^T \varphi(t)\mathrm{d}\eta_n(t) = 0$, for all n. But by (A5), $\bar{\eta}$ has a positive density over (t_1,t_2), and hence, $\int_0^T \varphi(t)\mathrm{d}\bar{\eta}(t) > 0$, which contradicts the second assertion in Lemma 3.6. This achieves the proof of assertion (S2). □

Acknowledgement The authors thank an anonymous referee for his remarks that helped to improve the paper.

Chapitre 4

Le cas de plusieurs contraintes*

Abstract This paper deals with the optimal control problem of an ordinary differential equation with several pure state constraints, of arbitrary orders, as well as mixed control-state constraints. We assume (i) the control to be continuous and the strengthened Legendre-Clebsch condition to hold, and (ii) a linear independence condition of the active constraints at their respective order to hold. We give a complete analysis of the smoothness and junction conditions of the control and of the constraints multipliers. This allows us to obtain, when there are finitely many nontangential junction points, a theory of no-gap second-order optimality conditions and a characterization of the well-posedness of the shooting algorithm. These results generalize those obtained in the case of a scalar-valued state constraint and a scalar-valued control.

Résumé Dans cet article on s'intéresse au problème de commande optimale d'une équation différentielle ordinaire avec plusieurs contraintes pures sur l'état, d'ordres quelconques, et des contraintes mixtes sur la commande et sur l'état. On suppose que (i) la commande est continue et la condition forte de Legendre-Clebsch satisfaite, et (ii) une condition d'indépendance linéaire des contraintes actives est satisfaite. Des résultats de régularité des solutions et multiplicateurs et des conditions de jonction sont donnés. Lorsqu'il y a un nombre fini de points de jonction, on obtient des conditions d'optimalité du second ordre nécessaires ou suffisantes, ainsi qu'une caractérisation du caractère bien posé de l'algorithme de tir. Ces résultats généralisent les résultats obtenus dans le cas d'une contrainte sur l'état et d'une commande scalaires.

4.1 Introduction

This paper deals with optimal control problems with a vector-valued state constraint. Mixed control-state constraints (state constraints of order zero) are included in the analysis. It is assumed that the control is continuous and the strengthened Legendre-Clebsch condition holds, and that each component of the state constraint is of arbitrary (but finite) order q_i.

Second-order optimality conditions for state-constrained optimal control problems were recently studied in [80, 112, 113, 20]. The presence of pure state constraints introduces an additional curvature term in the second-order necessary condition, in contrast with mixed

*Joint work with J.F. Bonnans. Accepted for publication in Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire, under the title *Second-order analysis for optimal control problems with pure state constraints and mixed control-state constraints.*

control-state constraints, see [108, 105]. An analysis of the junction conditions may help to narrow the gap with the second-order sufficient condition. There are, to our knowledge, relatively few papers dealing with optimal control problems with several state constraints *of order greater than one*. One of them is an unpublished paper by Maurer [98]. In e.g. [65, 88, 53, 54, 93, 95], several constraints of *first-order* were considered, but when dealing with constraints of higher order, then often only one constraint (and sometimes also a scalar control) is considered, see e.g. [75, 68, 94]. When there are several constraints of different orders, and more control variables than active constraints, then even the regularity of the control and of the state constraint multipliers on the interior of the arcs of the trajectory is not an obvious question. In [98, Lemma 4.1], it is shown that the control u is $C^{q_{max}}$ (where q_{max} is the bigger order of the active constraints), under the assumption that there are as many active state constraints as control variables. In [98, Th. 4.2], it is shown that the state constraints multipliers are smooth on the interior of arcs, but with the extra assumption that the control u is $C^{q_{max}}$.

The motivation of this paper is to extend the no-gap second-order optimality conditions and the characterization of the well-posedness of the shooting algorithm, obtained in [18, 21] and [19], respectively, for an optimal control problem with a scalar-valued state constraint and control, to the case of a vector-valued state constraint and control. The critical step is the extension of the junctions conditions obtained in the scalar case (i.e., with a scalar-valued state constraint and control) by Jacobson, Lele and Speyer [75]. This result says that some of the time derivatives of the control are continuous at a junction point until an order that depend on the *order* of the (scalar) state constraint, and on the nature of the junction point (entry/exit of boundary arcs versus touch points). This result has an important role when deriving the second-order necessary condition, since, with this regularity result and under suitable assumptions, it can be shown that boundary arcs have typically no contribution to the curvature term. This enables to derive a second-order sufficient condition as close as possible to the necessary one (no-gap), and to obtain a characterization of the well-posedness of the shooting algorithm. We show in particular that the shooting algorithm is ill-posed if a component of the state constraint of order $q_i \geq 3$ has a boundary arc.

In this paper, the focus is on the proofs that are not directly obtained from the scalar case, and in particular the (nontrivial) extension of the junction condition result of [75]. Our main assumption is the simplest one that the gradients w.r.t. the control variable of the time derivatives of the active constraints at their respective order are linearly independent. This enables to write locally the system under a "normal form", where the dynamics corresponding to the state constraints is linearized, and the different components of the constraints are decoupled.

The paper is organized as follows. In section 4.2, we present the problem, notation, basic definitions and assumptions. In section 4.3, we give sufficient conditions implying the continuity of the control, and we show local higher regularity of the control and constraints multipliers on the interior of arcs. In section 4.4, we give some technical lemmas needed to put the system under a "normal form". This will be used in section 4.5, where we give the junction conditions results. In section 4.6, the no-gap second-order optimality conditions is stated. In section 4.7, we recall the shooting formulation and state a characterization of the well-posedness of the shooting algorithm, under the additional assumption that the junction times of the different components of the state constraint do not coincide.

4.2 Framework

Let n, m, r, s be positive integers. If r and/or s is equal to zero, then the statements of this paper remain correct if the corresponding terms are removed. Denote by $\mathcal{U} := L^\infty(0,T;\mathbb{R}^m)$ (resp. $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$) the control (resp. state) space. We consider the following optimal control problem:

$$(\mathcal{P}) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell(u(t),y(t))\mathrm{d}t + \phi(y(T)) \tag{4.1}$$

$$\text{subject to} \qquad \dot{y}(t) = f(u(t),y(t)) \ \text{ for a.a. } t\in[0,T]; \quad y(0) = y_0 \tag{4.2}$$

$$g_i(y(t)) \le 0 \quad \text{for all } t\in[0,T], \quad i = 1,\dots,r \tag{4.3}$$

$$c_i(u(t),y(t)) \le 0 \quad \text{for a.a. } t\in[0,T], \quad i = r+1,\dots,r+s. \tag{4.4}$$

The data of the problem are the distributed cost $\ell : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, final cost $\phi : \mathbb{R}^n \to \mathbb{R}$, dynamics $f : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, pure state constraint $g : \mathbb{R}^n \to \mathbb{R}^r$, mixed control-state constraint $c : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^s$, (fixed) final time $T > 0$, and (fixed) initial condition $y_0 \in \mathbb{R}^n$. We make the following assumptions on the data:

(A0) The mappings ℓ, ϕ, f, g and c are (at least) of class C^2 with locally Lipschitz continuous second-order derivatives, and the dynamics f is Lipschitz continuous.

(A1) The initial condition satisfies $g_i(y_0) < 0$ for all $i = 1,\dots,r$.

Throughout the paper it is assumed that assumption (A0) holds.

Notations The space of row vectors is denoted by $\mathbb{R}^{n*}$. We denote by $A^\top$ the adjoint operator of a linear operator A or the transpose operator in $\mathbb{R}^{n\times m}$. Given a measurable set $\mathcal{I} \subset (0,T)$, we denote by $L^s(\mathcal{I})$ the Lebesgue space of measurable functions such that $\|u\|_s := (\int_{\mathcal{I}} |u(t)|^s \mathrm{d}t)^{1/s}$ (resp. $\|u\|_\infty := \operatorname{supess}_{t\in\mathcal{I}} |u(t)|$) for $1 \le s < +\infty$ (resp. $s = +\infty$) is finite. Given an open set $\mathcal{I} \subset (0,T)$, $k \in \mathbb{N}^*$ and $1 \le s \le +\infty$, the space $W^{k,s}(\mathcal{I})$ denotes the Sobolev space of functions having their weak derivatives until order k in $L^s(\mathcal{I})$. The standard norm of $W^{k,s}$ is denoted by $\|\cdot\|_{k,s}$. We say that a function is nonpositive, if it takes values in $\mathbb{R}_-$.

The Banach space of vector-valued continuous functions is denoted by $C([0,T];\mathbb{R}^r)$ and supplied with the product norm $\|x\|_\infty := \sum_{i=1}^r \|x_i\|_\infty$. The space of vector-valued Radon measures, dual space to $C([0,T];\mathbb{R}^r)$, is denoted by $\mathcal{M}([0,T];\mathbb{R}^{r*})$ and identified with vector-valued functions of bounded variation (BV) vanishing at T. The duality product between $C([0,T];\mathbb{R}^r)$ and $\mathcal{M}([0,T];\mathbb{R}^{r*})$ is denoted by $\langle \eta, x\rangle = \sum_{i=1}^r \int_0^T x_i \mathrm{d}\eta_i$. The cones of nonpositive continuous functions and nonnegative Radon measures over $[0,T]$ are denoted respectively by $K := C_-([0,T];\mathbb{R}^r)$ and $\mathcal{M}_+([0,T];\mathbb{R}^{r*})$.

The dual space to $L^\infty(0,T)$, denoted by $(L^\infty)^*(0,T)$, is the space of finitely additive set functions (see [58, p.258]) letting invariant the sets of zero Lebesgue's measure. The duality product over $(L^\infty)^*$ and L^∞ is denoted by $\langle \lambda, x\rangle$, and when $\lambda \in L^1$, we have $\langle \lambda, x\rangle = \int_0^T \lambda(t)x(t)\mathrm{d}t$. The set of vector-valued essentially bounded functions $L^\infty(0,T;\mathbb{R}^s)$ is supplied with the product topology. The set of essentially bounded functions with value in $\mathbb{R}^s_-$ almost everywhere is denoted by $\mathcal{K} := L^\infty_-(0,T;\mathbb{R}^s)$, and the set of elements λ in $(L^\infty)^*(0,T;\mathbb{R}^s)$ such that $\langle \lambda, x\rangle$ is nonpositive for all $x \in L^\infty_-(0,T;\mathbb{R}^s)$ is denoted by $(L^\infty)^*_+(0,T;\mathbb{R}^s)$.

We denote by B_X the unit (open) ball of the Banach space X. By $\operatorname{cl} S$, $\operatorname{int} S$ and ∂S we denote respectively the closure, interior and boundary of the set S. The cardinal of a finite

set J is denoted by $|J|$. The restriction of a function φ defined over $[0,T]$ to a set $A \subset [0,T]$ is denoted by $\varphi|_A$. The indicator function of a set A is denoted by $\mathbf{1}_A$. Given a Banach space X and $A \subset X^*$ the dual space to X, we denote by $A^\perp$ the space of $x \in X$ such that $\langle \xi, x\rangle = 0$ for all $\xi \in A$. If A is a singleton, then $\xi^\perp := \{\xi\}^\perp$. The left and right limits of a function of bounded variation φ over $[0,T]$ are denoted by $\varphi(\tau^\pm) := \lim_{t\to\tau^\pm} \varphi(t)$ and jumps are denoted by $[\varphi(\tau)] := \varphi(\tau^+) - \varphi(\tau^-)$. Fréchet derivatives of f, g_i, etc. w.r.t. arguments $u \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, etc. are denoted by a subscript, for instance $f_u(u,y) = D_u f(u,y)$, $g_{i,y}(y) = D_y g_i(y)$. An exception to this rule is that given $u \in \mathcal{U}$, we denote by y_u the (unique) solution in $\mathcal{Y}$ of the state equation (4.2).

Abstract formulation We denote by $J : \mathcal{U} \to \mathbb{R}$, $G : \mathcal{U} \to C([0,T];\mathbb{R}^r)$ and $\mathcal{G} : \mathcal{U} \to L^\infty(0,T;\mathbb{R}^s)$ the cost function $J(u) := \int_0^T \ell(u(t), y_u(t))\mathrm{d}t + \phi(y_u(T))$ and the constraints mappings defined by $G(u) := g(y_u)$ and $\mathcal{G}(u) := c(u, y_u)$. Recall that the constraints cones are defined by $K = C_-([0,T];\mathbb{R}^r)$ and $\mathcal{K} = L^\infty_-(0,T;\mathbb{R}^s)$. The abstract formulation of $(\mathcal{P})$ (used in section 4.6 and in the Appendix) is the following:

$$(\mathcal{P}) \qquad \min_{u\in\mathcal{U}} J(u), \quad \text{subject to } G(u) \in K,\ \mathcal{G}(u) \in \mathcal{K}. \tag{4.5}$$

The choice of the functional space for the pure state constraints (here, the space of continuous functions) is discussed later in Remark 4.4.

A *trajectory* (u,y) is an element of $\mathcal{U}\times\mathcal{Y}$ satisfying the state equation (4.2). A *feasible* trajectory is one that satisfies the constraints (4.3) and (4.4). We say that a feasible trajectory $(u,y) = (u,y_u)$ is a *local solution* (weak minimum) of $(\mathcal{P})$, if it minimizes (4.1) over the set of feasible trajectories $(\tilde u, \tilde y)$ satisfying $\|\tilde u - u\|_\infty \le \delta$, for some $\delta > 0$.

4.2.1 Constraint qualification condition

Given a measurable (nonpositive) function x, we denote the *contact set* by

$$\Delta(x) \ := \ \{t \in [0,T] \ : \ x(t) = 0\} \tag{4.6}$$

and, for $n \in \mathbb{N}^*$,

$$\Delta_n(x) \ := \ \{t \in [0,T] \ : \ x(t) \ge -\frac{1}{n}\}. \tag{4.7}$$

Given a feasible trajectory (u,y), define the sets of *active state constraints* and *active mixed constraints* at a.a. time $t \in [0,T]$ respectively by:

$$I^g(t) \ := \ \{i \in \{1,\dots,r\} \ : \ g_i(y(t)) = 0\} \tag{4.8}$$

$$I^c(t) \ := \ \{i \in \{r+1,\dots,r+s\} \ : \ t \in \Delta(c_i(u,y))\}, \tag{4.9}$$

and let

$$I(t) \ := \ I^g(t) \cup I^c(t). \tag{4.10}$$

An *arc* of the trajectory (u,y) is a maximal *open* interval of *positive measure* $\mathcal{I} = (\tau_1,\tau_2)$, such that $I(t)$ is constant, for all $t \in (\tau_1,\tau_2)$.

For $\varepsilon > 0$, $n \in \mathbb{N}^*$ and a.a. $t \in [0,T]$, define the set of *nearly active state constraints* and *nearly active mixed constraints* respectively by:

$$I^g_\varepsilon(t) \ := \ \cup\{\ I(\sigma)\ ;\ \sigma \in (t-\varepsilon, t+\varepsilon) \cap [0,T]\} \tag{4.11}$$

$$I^c_n(t) \ := \ \{i \in \{r+1,\dots,r+s\} \ : \ t \in \Delta_n(c_i(u,y))\} \tag{4.12}$$

and the set of nearly active constraints by

$$I_{\varepsilon,n}(t) \ := \ I^g_\varepsilon(t) \cup I^c_n(t). \tag{4.13}$$

The contact sets of the constraints are denoted by

$$\Delta_i \ := \ \Delta(g_i(y)) \qquad \text{for } i = 1, \dots, r, \tag{4.14}$$

$$\Delta_i \ := \ \Delta(c_i(u,y)) \qquad \text{for } i = r+1, \dots, r+s \tag{4.15}$$

and, for $\delta > 0$ and $n \in \mathbb{N}^*$,

$$\Delta^\delta_i \ := \ \{t \in (0,T) \ : \ \mathrm{dist}\{t, \Delta(g_i(y))\} \ < \ \delta\}, \qquad i = 1, \dots, r \tag{4.16}$$

$$\Delta^n_i \ := \ \Delta_n(c_i(u,y)), \qquad i = r+1, \dots, r+s. \tag{4.17}$$

Orders of the state constraints Let $i = 1, \dots, r$. If f and g_i are C^{q_i} mappings, we may define inductively the functions $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$, $g^{(j)}_i(u,y) := g^{(j-1)}_{i,y}(y) f(u,y)$ for $j = 1, \dots, q_i$, with $g^{(0)}_i := g_i$, if we have $g^{(j)}_{i,u} \equiv 0$ for all $j = 0, \dots, q_i - 1$, i.e. $g^{(j)}_{i,u}(u,y) = 0$ for all $(u,y) \in \mathbb{R}^m \times \mathbb{R}^n$. Then $\frac{\mathrm{d}^j}{\mathrm{d}t^j} g_i(y(t)) = g^{(j)}_i(u(t), y(t))$, and for all $j < q_i$, we have that $g^{(j)}_i(u,y) = g^{(j)}_i(y)$. Let q_i be the smallest number of derivations, so that a dependence w.r.t. u appears, i.e. such that $g^{(q_i)}_{i,u}$ is not identically zero over $\mathbb{R}^m \times \mathbb{R}^n$ (this intrinsic definition of the order does not depend on a given trajectory $(u,y) \in \mathcal{U} \times \mathcal{Y}$ nor on the time). If q_i is finite, we say that q_i is the *order* of the component g_i. If q_i is finite, for all i, we define the *highest order* $q_{max} := \max_{i=1}^r q_i$, and the *orders vector* $q := (q_1, \dots, q_r) \in \mathbb{N}^r$ is the vector of orders of the constraint $g = (g_1, \dots, g_r)$. In all the paper, it is assumed in addition to (A0) that

(A0$_q$) Each component of the state constraint g_i, $i = 1, \dots, r$, is of finite order q_i, and f and g are (at least) $C^{q_{max}+1}$.

Remark 4.1. When performing the analysis in the L^∞-vicinity of a given trajectory $(u,y) \in \mathcal{U} \times \mathcal{Y}$, it is sufficient, for the results of this paper, to restrict the variable $y \in \mathbb{R}^n$ in the above definition of the mappings $g^{(j)}_i$ and of the order q_i to an open neighborhood in $\mathbb{R}^n$ of $\{y(t) \ ; \ t \in \Delta_i\}$ for each $i = 1, \dots, r$. Likewise, the order of the constraint q_i needs only to be defined in the neighborhood of each connected component of the contact set Δ_i and may differ over two distinct connected components.

Note that when the state constraint g_i is of order q_i, relations such as

$$g^{(j)}_{i,y}(u,y) \ = \ g^{(j-1)}_{i,yy}(y) f(u,y) + g^{(j-1)}_{i,y}(y) f_y(u,y), \tag{4.18}$$

are satisfied, for all $j = 1, \dots, q_i$. This will be useful in some of the proofs.

We assume w.l.o.g. in this paper that $u \to c_{i,u}(u,y)$ is not identically zero, for all $i = r+1, \dots, r+s$, since otherwise $c_i(u,y)$ is a pure state constraint. We may interpret mixed control-state constraints as state constraint of order zero, setting

$$q_i \ := \ 0 \quad \text{and} \quad g^{(0)}_i(u,y) \ := \ c_i(u,y), \qquad \text{for all } i = r+1, \dots, r+s. \tag{4.19}$$

Given a subset $J \subset \{1, \dots, r+s\}$, say $J = \{i_1 < \dots < i_k\}$, define the mapping $G^{(q)}_J : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^{|J|}$ by:

$$G^{(q)}_J(u,y) := \begin{pmatrix} g^{(q_{i_1})}_{i_1}(u,y) \\ \vdots \\ g^{(q_{i_k})}_{i_k}(u,y) \end{pmatrix}, \quad \text{for all } (u,y) \in \mathbb{R}^m \times \mathbb{R}^n. \tag{4.20}$$

By (4.19), mixed control-state constraints are taken into account in this definition. When $J = \{1, \dots, r+s\}$, we denote just (4.20) by $G^{(q)}(u, y)$.

The controllability lemma For $\kappa \in [1, +\infty]$, let

$$\mathcal{V}_\kappa \ := \ L^\kappa(0, T; \mathbb{R}^m), \qquad \mathcal{Z}_\kappa \ := \ W^{1,\kappa}(0, T; \mathbb{R}^n). \tag{4.21}$$

Given a trajectory (u, y) and $v \in \mathcal{V}_\kappa$, we denote by z_v the (unique) solution in $\mathcal{Z}_\kappa$ of the *linearized state equation*

$$\dot{z}(t) = f_u(u(t), y(t))v(t) + f_y(u(t), y(t))z(t) \quad \text{a.e. on } [0, T], \quad z(0) = 0. \tag{4.22}$$

Lemma 4.2. *Let (u, y) be a trajectory, and let $\kappa \in [1, +\infty]$. For all $v \in \mathcal{V}_\kappa$ and all $i = 1, \dots, r$, we have that $g_{i,y}(y(\cdot))z_v(\cdot) \in W^{q_i,\kappa}(0, T)$ and:*

$$\frac{\mathrm{d}^j}{\mathrm{d}t^j}(g_{i,y}(y(t))z_v(t)) \ = \ g_{i,y}^{(j)}(y(t))z_v(t), \quad \textit{for all } j = 1, \dots, q_i - 1, \tag{4.23}$$

$$\frac{\mathrm{d}^{q_i}}{\mathrm{d}t^{q_i}}(g_{i,y}(y(t))z_v(t)) \ = \ g_{i,u}^{(q_i)}(u(t), y(t))v(t) + g_{i,y}^{(q_i)}(u(t), y(t))z_v(t). \tag{4.24}$$

Proof. It suffices to use the linearized state equation (4.22), the relation (4.18), and that $g_{i,y}^{(j-1)} f_u = g_{i,u}^{(j)} \equiv 0$ for all $j = 1, \dots, q_i - 1$ to obtain (4.23)-(4.24) by induction on j. □

Consider the following constraint qualification condition:

$$\begin{array}{l} \text{there exist } \gamma, \varepsilon > 0 \text{ and } n \in \mathbb{N}^* \text{ such that} \\ \gamma\,|\xi| \ \leq \ \left| G^{(q)}_{I_{\varepsilon,n}(t),u}(u(t), y(t))^\top \xi \right|, \text{ for all } \xi \in \mathbb{R}^{|I_{\varepsilon,n}(t)|} \text{ and a.a. } t \in [0, T]. \end{array} \tag{4.25}$$

Lemma 4.3. *Let (u, y) be a trajectory satisfying (A1) and (4.25). Then for all $\kappa \in [1, +\infty]$ and all $\delta \in (0, \varepsilon)$, where ε is given in (4.25), the linear mapping*

$$\begin{array}{rcl} \mathcal{V}_\kappa & \to & \prod_{i=1}^r W^{q_i,\kappa}(\Delta_i^\delta) \times \prod_{i=r+1}^{r+s} L^\kappa(\Delta_i^n) \\ v & \mapsto & \begin{pmatrix} \left((g_{i,y}(y(\cdot))z_v(\cdot))|_{\Delta_i^\delta}\right)_{1 \leq i \leq r} \\ \left((c_{i,u}(u(\cdot), y(\cdot))v(\cdot) + c_{i,y}(u(\cdot), y(\cdot))z_v(\cdot))|_{\Delta_i^n}\right)_{r+1 \leq i \leq r+s} \end{pmatrix} \end{array} \tag{4.26}$$

where z_v is the unique solution in $\mathcal{Z}_k$ of the linearized state equation (4.22), is onto, and hence has a bounded right inverse by the open mapping Theorem.

Recall that $\varphi|_{\mathcal{I}}$ denotes the restriction of the function φ to the set $\mathcal{I} \subset [0, T]$.

Proof. Let $\psi = (\psi_i)_{1 \leq i \leq r+s} \in \prod_{i=1}^r W^{q_i,\kappa}(\Delta_i^\delta) \times \prod_{i=r+1}^{r+s} L^\kappa(\Delta_i^n)$. In order to have $\psi_i = g_{i,y}(y)z_v$ on Δ_i^δ for all $i = 1, \dots, r$, it is necessary and sufficient by Lemma 4.2 that, a.e. on Δ_i^δ,

$$g_{i,u}^{(q_i)}(u, y)v + g_{i,y}^{(q_i)}(u, y)z_v \ = \ \psi_i^{(q_i)} \tag{4.27}$$

and that, for every point τ in the left boundary of Δ_i^δ (note that there exist finitely many such points),

$$g_{i,y}^{(j)}(y(\tau))z_v(\tau) \ = \ \psi_i^{(j)}(\tau), \qquad \text{for all } j = 0, \dots, q_i - 1. \tag{4.28}$$

The relation (4.27) with $q_i = 0$, $g_i^{(0)} = c_i$ and $\psi_i^{(0)} := \psi_i$ must be satisfied as well a.e. on Δ_i^n for all $i = r+1, \dots, r+s$. Set $M(t) := G^{(q)}_{I_{\varepsilon,n}(t),u}(u(t), y(t))$. By (4.25), the matrix $M(t)M(t)^\top$ is invertible at a.a. t, so we may take a.e., if $I_{\varepsilon,n}(t) \neq \emptyset$ (take $v(t) = 0$ if $I_{\varepsilon,n}(t) = \emptyset$):

$$v(t) \;=\; M(t)^\top (M(t)M(t)^\top)^{-1} \{\varphi(t) - G^{(q)}_{I_{\varepsilon,n}(t),y}(u(t), y(t)) z_v(t)\}, \tag{4.29}$$

where z_v is the solution of (4.22) with v given by (4.29), and the right-hand side $\varphi = (\varphi_i)_{i \in I_{\varepsilon,n}(t)}$ is as follows. We have $\varphi_i(t) = \psi_i(t)$ if $i = r+1, \dots, r+s$ and $t \in \Delta_i^n$, and $\varphi_i(t) = \psi_i^{(q_i)}(t)$ if $i = 1, \dots, r$ and $t \in \Delta_i^\delta$. On $\Delta_i^\varepsilon \setminus \Delta_i^\delta$, φ_i can be chosen equal e.g. to a polynomial function of order $2q_i - 1$, in order to match, in arbitrary small time $\varepsilon - \delta > 0$, the first $q_i - 1$ time derivatives of $g_{i,y}(y) z_v$ with those of ψ_i, i.e. so that (4.28) holds for all left endpoints τ of Δ_i^δ. □

If the control u is continuous (see Prop. 4.8 and assumption (A2)), (4.25) is always satisfied if the *linear independence condition* below holds:

$$\begin{array}{l} \text{there exists } \gamma > 0 \text{ such that} \\ \gamma\,|\xi| \;\leq\; \left| G^{(q)}_{I(t),u}(u(t), y(t))^\top \xi \right|, \quad \text{for all } \xi \in \mathbb{R}^{|I(t)|} \quad \text{and a.a. } t \in [0,T], \end{array} \tag{4.30}$$

i.e. $G^{(q)}_{I(t),u}(u(t), y(t))$ is uniformly onto, for all $t \in [0,T]$. This assumption (without the mixed control-state constraints) was already used in [98].

For $J = \{i_1 < \dots < i_k\} \subset \{r+1, \dots, r+s\}$, let us denote

$$c_J(u,y) \;:=\; (c_{i_1}(u,y), \dots, c_{i_k}(u,y))^\top.$$

We will also use in Proposition 4.8 the constraint qualification (4.31) below, weaker than (4.25), involving only the mixed control-state constraints:

$$\begin{array}{l} \text{there exist } n \in \mathbb{N}^* \text{ and } \gamma > 0 \text{ such that} \\ \gamma|\xi| \leq |c_{I_n^c(t),u}(u(t), y(t))^\top \xi| \quad \text{for all } \xi \in \mathbb{R}^{|I_n^c(t)|} \quad \text{and a.a. } t \in [0,T]. \end{array} \tag{4.31}$$

Remark 4.4. There are two possible natural choices for the functional space of the pure state constraints: either the space of continuous functions $C^0 := C([0,T]; \mathbb{R}^r)$, or the space $W^{q,\infty} := \prod_{i=1}^r W^{q_i,\infty}(0,T)$, where q_i denotes the order of the i-th component of the constraint, in which the constraint is "onto" by Lemma 4.3. Considering the state constraints in C^0 instead of $W^{q,\infty}$, we have multipliers in $\mathcal{M}([0,T]; \mathbb{R}^{r*})$ rather than in the dual space of $W^{q,\infty}$. Existence of multipliers in $\mathcal{M}([0,T]; \mathbb{R}^{r*})$ is ensured under natural hypotheses (see below). Moreover, since the inclusion of $W^{q,\infty}$ in C^0 is dense and continuous, by surjectivity of the constraint in $W^{q,\infty}$ we obtain that the multipliers associated in both formulations are one to one, and we inherit nice properties such as uniqueness of the multiplier in $\mathcal{M}([0,T]; \mathbb{R}^{r*})$.

4.2.2 First-order Optimality Condition

Define the classical *Hamiltonian* and *Lagrangian* functions of $(\mathcal{P})$, $H : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$ and $L : \mathcal{U} \times M([0,T]; \mathbb{R}^{r*}) \times (L^\infty)^*(0,T; \mathbb{R}^{s*}) \to \mathbb{R}$ by:

$$H(u,y,p) \;:=\; \ell(u,y) + pf(u,y) \tag{4.32}$$

$$L(u; \eta, \lambda) \;:=\; J(u) + \langle \eta, G(u) \rangle + \langle \lambda, \mathcal{G}(u) \rangle, \tag{4.33}$$

for the duality products in the appropriate spaces.

Robinson's constraint qualification for the abstract problem (4.5) is as follows:

$$\exists\, \varepsilon > 0, \qquad \varepsilon B_{C\times L^\infty} \subset (G(u), \mathcal{G}(u)) + (DG(u), D\mathcal{G}(u))\mathcal{U} - K \times \mathcal{K}. \tag{4.34}$$

It is easy to see that under the assumptions of Lemma 4.3, (4.34) holds. Some elements of proof of the next theorem are recalled in the Appendix (subsection 4.9.2). The existence and uniqueness of the multipliers are a consequence of Lemma 4.3.

Theorem 4.5. *Let $(u, y) \in \mathcal{U} \times \mathcal{Y}$ be a a local solution of $(\mathcal{P})$, satisfying (A1),* (4.34) *and* (4.31)*. Then there exist $p \in BV([0,T];\mathbb{R}^{n*})$, $\eta \in \mathcal{M}([0,T];\mathbb{R}^{r*})$ and $\lambda \in L^\infty(0,T;\mathbb{R}^{s*})$ such that*

$$\begin{aligned}
\dot y(t) &= f(u(t), y(t)) \quad \textit{for a.a. } t \in [0,T]; \qquad y(0) = y_0 && (4.35)\\
-\mathrm{d}p(t) &= \{H_y(u(t), y(t), p(t)) + \lambda(t) c_y(u(t), y(t))\}\mathrm{d}t + \mathrm{d}\eta(t) g_y(y(t)) && (4.36)\\
p(T^+) &= \phi_y(y(T)) && (4.37)\\
0 &= H_u(u(t), y(t), p(t)) + \lambda(t) c_u(u(t), y(t)) \quad \textit{for a.a. } t \in [0,T] && (4.38)\\
0 &\geq g_i(y(t)), \quad \mathrm{d}\eta_i \geq 0, \quad \int_0^T g_i(y(t))\mathrm{d}\eta_i(t) = 0, \qquad i = 1, \dots, r && (4.39)\\
0 &\geq c_i(u(t), y(t)), \quad \lambda_i(t) \geq 0 \quad \textit{a.e.}, \int_0^T c_i(u(t), y(t))\lambda_i(t)\mathrm{d}t = 0, && (4.40)\\
&\qquad i = r+1, \dots, r+s.
\end{aligned}$$

We say that (u, y) is a *stationary point* of $(\mathcal{P})$, if there exist $p \in BV([0,T];\mathbb{R}^{n*})$, $\eta \in \mathcal{M}([0,T];\mathbb{R}^{r*})$ and $\lambda \in L^\infty(0,T;\mathbb{R}^{s*})$ such that (4.35)-(4.40) hold.

When the Hamiltonian and the mixed control-state constraints are convex w.r.t. the control variable (and in particular when assumption (4.44) below holds), then (4.38) and (4.40) are equivalent to

$$u(t) \in \operatorname*{argmin}_{w\in\mathbb{R}^m,\ c(w,y(t))\leq 0} H(w, y(t), p(t)) \qquad \text{for a.a. } t \in [0,T]. \tag{4.41}$$

Here $\lambda(t)$ is the multiplier associated with the constraint (in $\mathbb{R}^m$) $c(w, y(t)) \leq 0$. We thus recover in this particular case Pontryagin's Minimum Principle, see [57, 50, 104].

Assumptions Let the *augmented Hamiltonian of order zero* $H^0 : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \times \mathbb{R}^{s*} \to \mathbb{R}$ be defined by

$$H^0(u, y, p, \lambda) \;:=\; H(u, y, p) + \lambda c(u, y). \tag{4.42}$$

Given (u, y) a stationary point of $(\mathcal{P})$, we will make the assumptions below:

(A2) The control u is *continuous* on [0,T], and (strengthened Legendre-Clebsch condition)

$$\begin{aligned}
&\text{there exists } \alpha > 0 \text{ such that for all } t \in [0,T],\\
&\qquad \alpha|\upsilon|^2 \;\leq\; H^0_{uu}(u(t), y(t), p(t), \lambda(t))(\upsilon, \upsilon) \qquad \text{for all } \upsilon \in \mathbb{R}^m.
\end{aligned} \tag{4.43}$$

(A3) The data of the problem are (at least) $C^{2q_{max}}$, and the linear independence condition (4.30) is satisfied.

Remark 4.6. The only condition (4.43) is not enough to ensure the continuity of the control, as shows the following example:

$$\min_{u\in L^\infty(0,T)} \int_0^2 \{u(t)^4 - 2u(t)^2 + (y(t)-1)u(t)\}\mathrm{d}t, \quad \dot{y}(t) = 1, \quad y(0) = 0,$$

where the minimizer u jumps from the minimum close to 1 for $t = y(t) < 1$ to the minimum close to -1 for $t = y(t) > 1$, although (4.43) holds.

We will see in Prop. 4.8 that if (u, y) is a stationary point such that the Hamiltonian $H(\cdot, y(t), p(t))$ is uniformly strongly convex and the mixed control-state constraints are convex w.r.t. the control along the trajectory, which is equivalent to the condition below (stronger than (4.43))

$$\begin{aligned} &\text{there exists } \alpha > 0 \text{ such that for all } t \in [0,T] \text{ and all } (\hat{u}, \hat{\lambda}) \in \mathbb{R}^m \times \mathbb{R}^{s*}_+, \\ &\quad \alpha|v|^2 \le H^0_{uu}(\hat{u}, y(t), p(t), \hat{\lambda})(v, v) \qquad \text{for all } v \in \mathbb{R}^m, \end{aligned} \tag{4.44}$$

and if (4.31) holds, then u is continuous on $[0,T]$. Therefore (4.44) and (4.31) imply that (A2) holds.

Remark 4.7. In some of the results of section 4.3 and 4.5, assumption (4.43) in (A2) can be weakened by assuming the uniform positivity of H^0_{uu} only on a subspace of $\mathbb{R}^m$ depending on the active constraints, namely

$$\begin{aligned} &\text{there exists } \alpha > 0 \text{ such that for a.a. } t \in [0,T], \\ &\quad \alpha|v|^2 \;\le\; H^0_{uu}(u(t), y(t), p(t), \lambda(t))(v, v) \quad \text{for all } v \in \mathbb{R}^m \text{ satisfying} \\ &\quad g^{(q_i)}_{i,u}(u(t), y(t))v = 0 \quad \text{for all } i = 1, \dots, r+s \text{ such that } t \in \operatorname{int} \Delta_i. \end{aligned} \tag{4.45}$$

4.3 First regularity results

In the scalar case (when both the state constraint $g(y)$ and the control are scalar-valued, i.e. $m = r = 1$), and when there is no constraint on the control, the regularity of the control on the interior of arcs follows from the implicit function Theorem, applied by (A2) to the relation $H_u(u(t), y(t), p(t)) = 0$ on the interior of unconstrained arcs (when $g(y(t)) < 0$), and by (A3) to $g^{(q)}(u(t), y(t)) = 0$ on the interior of boundary arcs (when $g(y(t)) = 0$). Knowing that u (and y) are smooth on boundary arcs, we can then differentiate w.r.t. t (in the measure sense) the relation $H_u(u(t), y(t), p(t))$ on boundary arcs, as many times as necessary, until we express, using (A3), the measure $\mathrm{d}\eta$ as $\eta_0(t)\mathrm{d}t$, with $\eta_0(t)$ a smooth function of $(u(t), y(t), p(t))$. Therefore we obtain that the state constraint multiplier η is continuously differentiable on the interior of boundary arcs.

Maurer in [98] extended this approach to the particular case when $r = m$ (and $s = 0$) (as many control as active state constraints), but this proof has no direct extension to the case $1 \le r < m$.

In subsection 4.3.1, we show that assumptions (4.44) and (4.31) imply the continuity of the control over $[0,T]$ (Prop. 4.8), and therefore also (A2) (no constraint regularity for the state constraint is needed). Moreover, (A2)-(A3) imply that the multipliers associated with mixed control-state constraints and with state constraints of first-order are continuous. In subsection 4.3.2 we show higher regularity of the control and of the constraints multipliers on the interior of the arcs of the trajectory (Prop. 4.13). Our proof is based on the use of *alternative multipliers* (Def. 4.10).

4.3.1 Continuity of the control

Proposition 4.8. *Let (u, y) be a stationary point of $(\mathcal{P})$.*
(i) *Assume that* (4.44) *and* (4.31) *hold. Then the control u is continuous on $[0, T]$.*
(ii) *Assume that (A2) and* (4.30) *hold. Then the multiplier λ associated with the mixed control-state constraints and the multipliers η_i associated with components g_i of the state constraint of* first order *($q_i = 1$) are continuous on $[0, T]$.*

In the absence of constraints of order greater than one, point (ii) is well-known, see e.g. [65, 68].

Proof of Prop. 4.8. Assumption (4.44) implies that for each $t \in [0, T]$, the problem (4.41) has a strongly convex cost function and convex constraints, therefore the control $u(t)$ is the unique solution of (4.41). In view of (4.31), $\lambda(t)$ is the unique associated multiplier. By (4.31) and (4.44), classical results on stability analysis in nonlinear programming (e.g. an easy application of Robinson's strong regularity theory [121], see also [76]) show that there exists a Lipschitz continuous function $\Upsilon : \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}^m \times \mathbb{R}^{s*}$ such that $(u(t), \lambda(t)) = \Upsilon(y(t), p(t))$, for a.a. $t \in [0, T]$. Since the composition of a Lipschitz continuous function with a function of bounded variation is a function of bounded variation, it follows that u and λ are of bounded variation, and hence have a right- and a left limit everywhere.

Fix $t \in [0, T]$. We sometimes omit the time argument t. Denote respectively by u^+ and u^- the right- and left limits of u at time t. Set $[u] := u^+ - u^-$ and for $\sigma \in [0, 1]$, $u^\sigma := \sigma u^+ + (1 - \sigma)u^-$. We use similar notations for λ and p. By the costate equation (4.36), p has at most countably many jumps, of type

$$[p] = p^+ - p^- = -\sum_{i=1}^{r} \nu_i g_{i,y}(y(t)), \qquad \text{with } \nu_i := [\eta_i(t)] \geq 0. \tag{4.46}$$

Recall that H^0 denotes the augmented Hamiltonian of order zero (4.42). It follows from (4.38) that

$$\begin{aligned} 0 &= H^0_u(u^+, y, p^+, \lambda^+) - H^0_u(u^-, y, p^-, \lambda^-) \\ &= \int_0^1 \{H^0_{uu}(u^\sigma, y, p^\sigma, \lambda^\sigma)[u] + [p] f_u(u^\sigma, y) + [\lambda] c_u(u^\sigma, y)\} \mathrm{d}\sigma. \end{aligned}$$

Using (4.46) and observing that, by definition of the order of the state constraint, $g_{i,y} f_u = g^{(1)}_{i,u}$ equals zero if $q_i > 1$, we obtain that

$$\int_0^1 H^0_{uu}(u^\sigma, y, p^\sigma, \lambda^\sigma)[u] \mathrm{d}\sigma \;=\; \int_0^1 \sum_{i:q_i=1} \nu_i g^{(1)}_{i,u}(u^\sigma, y) \mathrm{d}\sigma - \int_0^1 [\lambda] c_u(u^\sigma, y) \mathrm{d}\sigma. \tag{4.47}$$

Noticing that $H^0_{uu}(u^\sigma, y, p^\sigma, \lambda^\sigma) = \sigma H^0_{uu}(u^\sigma, y, p^+, \lambda^+) + (1 - \sigma) H^0_{uu}(u^\sigma, y, p^-, \lambda^-)$ and taking the scalar product of both sides of (4.47) by $[u]$, we get using hypothesis (4.44) that

$$\alpha |[u]|^2 \;\leq\; \sum_{i:q_i=1} \nu_i [g^{(1)}_i(u, y)] - [\lambda][c(u, y)]. \tag{4.48}$$

If $\nu_i > 0$, then $g_i(y(t)) = 0$, and hence $[g^{(1)}_i(u, y)] \leq 0$ since t is a local maximum of $g_i(y)$. By (4.40), $\lambda^\pm(t)$ belongs to the normal cone to $\mathbb{R}^s_-$ at point $c(u^\pm(t), y(t))$. By monotonicity

of the normal cone, we obtain that $[\lambda][c(u,y)] \geq 0$. Therefore, the right-hand side in (4.48) is nonpositive, implying that $[u] = 0$, i.e. u is continuous at t. This shows (i).

Since $[u] = 0$, the right-hand side of (4.47) equals zero. By (4.30), the vectors $(g^{(1)}_{i,u}(u,y))$ for $i \in I^g(t) \cap \{i : q_i = 1\}$ and $c_{i,u}(u,y)$ for $i \in I^c(t)$ are jointly linearly independent. It follows that $[\lambda] = 0$ and $\nu_i = 0$, for all i corresponding to first-order state-constraint components. This achieves the proof of (ii). □

Remark 4.9. For point (ii) in Prop. 4.8, it is sufficient to have the linear independence condition (4.30) for mixed control-state constraints and *first-order* components of the state constraint only.

4.3.2 Higher Regularity on interior of arcs

We recall that an arc of the trajectory (u,y) is a maximal open interval of positive measure with a constant set of active constraints (4.10), and that mixed control-state constraints are considered as state constraint of order zero by (4.19).

Definition 4.10. Let (u,y) be a stationary point of $(\mathcal{P})$, and (τ_1,τ_2) an arc of the trajectory, with constant set of active constraints $I(t) = J \subset \{1,\dots,r+s\}$, for all $t \in (\tau_1,\tau_2)$. The *alternative multipliers* on (τ_1,τ_2) are as follows. Define the functions η_i^j for $i = 1,\dots,r+s$ and $j = 1,\dots,q_i$ if $i \leq r$, $j = 0$ if $i > r$, by

$$\begin{array}{rcll} \eta_i^1(t) & := & -\displaystyle\int \mathrm{d}\eta_i(\sigma) = Cst - \eta_i(t), & i \in J, \quad i \leq r, \\ \eta_i^j(t) & := & -\displaystyle\int \eta_i^{j-1}(\sigma)\mathrm{d}\sigma & j = 2,\dots,q_i, \ i \in J, \quad i \leq r \\ \eta_i^j(t) & := & 0, & j = 1,\dots,q_i, \ i \in \{1,\dots,r\} \setminus J \\ \eta_i^0(t) & := & \lambda_i(t), & i \in J, \quad i > r. \end{array} \tag{4.49}$$

We denote here by Cst an arbitrary integration constant. The alternative multipliers (p^q,η^q) are defined by $\eta^q := (\eta_1^{q_1},\dots,\eta_{r+s}^{q_{r+s}})$ and

$$p^q(t) \ := \ p(t) - \sum_{i=1}^{r}\sum_{j=1}^{q_i} \eta_i^j(t) g_{i,y}^{(j-1)}(y(t)). \tag{4.50}$$

The *alternative Hamiltonian of order q* $H^q : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \times \mathbb{R}^{(r+s)*} \to \mathbb{R}$ is defined by:

$$H^q(u,y,p^q,\eta^q) := H(u,y,p^q) + \eta^q G^{(q)}(u,y) = H(u,y,p^q) + \sum_{i=1}^{r+s} \eta_i^{q_i} g_i^{(q_i)}(u,y), \tag{4.51}$$

with H the classical Hamiltonian (4.32).

Lemma 4.11. *Let (u,y) be a stationary point of $(\mathcal{P})$, with multipliers (p,η,λ). Then on the interior of each arc (τ_1,τ_2) of the trajectory, with a constant set of active constraints $I(t) = J \subset \{1,\dots,r+s\}$ on (τ_1,τ_2), the following holds, with the alternative multipliers of Def. 4.10, for all $t \in (\tau_1,\tau_2)$: p_q is absolutely continuous on (τ_1,τ_2) and*

$$-\dot{p}^q(t) \ = \ H^q_y(u(t),y(t),p^q(t),\eta^q(t)), \tag{4.52}$$

$$H^q(\cdot,y(t),p^q(t),\eta^q(t)) \ = \ H^0(\cdot,y(t),p(t),\lambda(t)), \tag{4.53}$$

and for all $i = 1, \dots, r+s$*:*

$$g_i^{(q_i)}(u(t), y(t)) = 0, \qquad i \in J \tag{4.54}$$
$$\eta_i^{q_i}(t) = 0, \qquad i \notin J. \tag{4.55}$$

Remark 4.12. An obvious consequence of (4.53) is that u minimizes $H^0(\cdot, y(t), p(t), \lambda(t))$ iff it minimizes $H^q(\cdot, y(t), p^q(t), \eta^q(t))$, and in particular, by (4.38), a stationary point satisfies

$$0 = H_u^q(u(t), y(t), p^q(t), \eta^q(t)). \tag{4.56}$$

Proof. For the sake of completeness of the paper, let us recall the proof, due to Maurer in [98] when there are no mixed control-state constraints. Relation (4.54) follows from differentiation w.r.t. $t \in (\tau_1, \tau_2)$ of the relation $g_i(y(t)) = 0$, for $i \in J$, $i \leq r$ and (4.55) follows from definition (4.49). By definition of the constraint order q_i, the function $g_i^{(j)}(u, y)$ does not depend on u, for all $j = 1, \dots, q_i - 1$ and $i = 1, \dots, r$, and hence, for all $\hat{u} \in \mathbb{R}^m$, we have:

$$\begin{aligned} H^0(\hat{u}, y, p, \lambda) &= H^0(\hat{u}, y, p^q, \lambda) + (p - p^q) f(\hat{u}, y) \\ &= H^0(\hat{u}, y, p^q, \lambda) + \textstyle\sum_{i=1}^r \sum_{j=1}^{q_i} \eta_i^j g_i^{(j)}(\hat{u}, y) \\ &= H^q(\hat{u}, y, p^q, \eta^q) + F(t), \end{aligned}$$

where

$$F(t) := \sum_{i=1}^{r} \sum_{j=1}^{q_i - 1} \eta_i^j(t) g_i^{(j)}(y(t))$$

does not depend on $\hat{u}$. For all $i = 1, \dots, r$, if $i \in J$, then $g_i^{(j)}(y(t)) = 0$, and if $i \notin J$, then $\eta_i^j(t) = 0$ by (4.49). Consequently, $F(t) = 0$, which proves (4.53).

We show now (4.52). Using (4.50) and that $\dot{\eta}_i^j = -\eta_i^{j-1}$, for $j = 2, \dots, q_i$, $i \leq r$, we have:

$$-\mathrm{d}p^q = -\mathrm{d}p + \sum_{i=1}^{r} \{ \sum_{j=1}^{q_i} \eta_i^j g_{i,yy}^{(j-1)}(y) f(u, y) \mathrm{d}t - \sum_{j=2}^{q_i} \eta_i^{j-1} g_{i,y}^{(j-1)}(y) \mathrm{d}t - \mathrm{d}\eta_i g_{i,y}(y) \}. \tag{4.57}$$

Since

$$-\mathrm{d}p \;=\; H_y(u, y, p^q)\mathrm{d}t + (p - p^q) f_y(u, y)\mathrm{d}t + \sum_{i=1}^{r} \mathrm{d}\eta_i g_{i,y}(y) + \sum_{i=r+1}^{r+s} \lambda_i c_{i,y}(u, y)\mathrm{d}t,$$

substituting $p - p^q$ into (4.57) using (4.50), we obtain:

$$\begin{aligned} -\mathrm{d}p^q \;=\; & H_y(u, y, p^q)\mathrm{d}t + \sum_{i=r+1}^{r+s} \eta_i^0 g_{i,y}^{(0)}(u, y)\mathrm{d}t \\ & + \sum_{i=1}^{r} \{ \sum_{j=1}^{q_i} \eta_i^j (g_{i,y}^{(j-1)}(y) f_y(u, y) + g_{i,yy}^{(j-1)}(y) f(u, y)) - \sum_{j=2}^{q_i} \eta_i^{j-1} g_{i,y}^{(j-1)}(y) \} \mathrm{d}t. \end{aligned}$$

Using (4.18), it follows that

$$-\mathrm{d}p^q \;=\; H_y(u, y, p^q)\mathrm{d}t + \sum_{i=1}^{r+s} \eta_i^{q_i} g_{i,y}^{(q_i)}(u, y)\mathrm{d}t,$$

which shows (4.52) and achieves the proof. □

Proposition 4.13. *Assume that the data are (at least) $C^{2q_{max}}$. Let (u, y) be a stationary point of $(\mathcal{P})$, with multipliers (p, η, λ), and let $(\tau_1, \tau_2) \subset [0, T]$ be such that $I(t)$ is constant on (τ_1, τ_2), u is continuous on (τ_1, τ_2), and (4.45) and (4.30) are satisfied on (τ_1, τ_2). Then on (τ_1, τ_2), u is $C^{q_{max}}$, y is $C^{q_{max}+1}$, p is C^1, λ is $C^{q_{max}}$ and the state constraint multiplier η_i is $C^{q_{max}-q_i+1}$, for all $i = 1, \ldots, r$.*

Proof. Denote by $J \subset \{1, \ldots, r+s\}$ the constant set of active constraints $I(t)$ for $t \in (\tau_1, \tau_2)$. The Jacobian w.r.t. u and $(\eta_i^{q_i})_{i\in J}$ of the equations (4.56) and (4.54), the latter being rewritten as $G_J^{(q)}(u(t), y(t)) = 0$, is given by

$$\begin{pmatrix} H_{uu}(u, y, p^q) + \sum_{i\in J} \eta_i^{q_i} g_{i,uu}^{(q_i)}(u, y) & G_{J,u}^{(q)}(u, y)^\top \\ G_{J,u}^{(q)}(u, y) & 0 \end{pmatrix}. \tag{4.58}$$

By (4.53),

$$H_{uu}(u, y, p^q) + \sum_{i\in J} \eta_i^{q_i} g_{i,uu}^{(q_i)}(u, y) = H_{uu}^q(u, y, p^q, \eta^q) = H_{uu}^0(u, y, p, \lambda)$$

is positive definite on $\operatorname{Ker} G_{J,u}^{(q)}(u, y)$ by (4.45), and by (4.30), $G_{J,u}^{(q)}(u, y)$ is onto. Since by assumption u is continuous, by (4.30) and (4.56), we deduce that $(\eta_i^{q_i})_{i\in J}$ is also continuous. Thus we can apply the implicit function Theorem to express u and $(\eta_i^{q_i})_{i\in J}$ as $C^{q_{max}}$ implicit functions of (y, p^q). Since (y, p^q) is solution of a $C^{q_{max}-1}$ differential equation system (4.2) and (4.52), we deduce that $(y, p^q, u, \eta_i^{q_i})$, $i \in J$, are $C^{q_{max}}$ on (τ_1, τ_2). By (4.55), the components $\eta_i^{q_i}$ for $i \notin J$ being equal to zero on (τ_1, τ_2) are also trivially $C^{q_{max}}$ on (τ_1, τ_2). Finally, recall that the classical multipliers η_i and p are related to the alternative ones by (4.49), i.e. $\eta_i(t) = (-1)^{q_i} \frac{\mathrm{d}^{q_i-1}}{\mathrm{d}t^{q_i-1}} \eta_i^{q_i}(t)$, and (4.50). It follows that each component η_i is $C^{q_{max}-q_i+1}$ for $i \leq r$, $\lambda_i = \eta_i^0$ is $C^{q_{max}}$, for all $i = r+1, \ldots, r+s$, and p is C^1, locally on (τ_1, τ_2). □

4.4 Local exact linearization of the "constraint dynamics"

We first give in subsection 4.4.1 a result of "local invariance" of stationary points by a local change of coordinates and nonlinear feedback (Lemma 4.15). We use this result in subsection 4.4.3 to show that, assuming (A3) and the continuity of u, we can locally "linearize the constraints dynamics" (Lemma 4.19), and we will use this "normal form" of the system in the proof of the junctions conditions results in Prop. 4.22. For that, a technical lemma (Lemma 4.17) given in subsection 4.4.2 is needed, which will also be used in the proofs of Prop. 4.29 and Th. 4.33.

4.4.1 Local invariance of stationary points by change of coordinates

Definition 4.14. Let (u, y) be a trajectory, and $t_0 \in (0, T)$. A couple of mappings (ϕ, ψ) is a *C^k local change of state variables and nonlinear feedback* at time t_0, $k \geq 1$, if there exist $\delta > 0$ and an open neighborhood $V_u \times V_y$ in $\mathbb{R}^m \times \mathbb{R}^n$ of $\{(u(t), y(t)) \ ; \ t \in (t_0 - \delta, t_0 + \delta)\}$, such that $\phi : V_y \to \phi(V_y) =: V_z$, $\psi : V_u \times V_y \to \psi(V_u \times V_y) =: V_v$ and there exist $\bar{\phi} : V_z \to V_y$ and $\bar{\psi} : V_v \times V_z \to V_u$ such that for all $(u, y, v, z) \in V_u \times V_y \times V_v \times V_z$, we have

$$z = \phi(y) \ \Leftrightarrow \ y = \bar{\phi}(z); \qquad v = \psi(u, y) \ \Leftrightarrow \ u = \bar{\psi}(v, z)$$

and the inverse mappings $\bar{\phi}$ and $\bar{\psi}$ are C^k over V_z and $V_v \times V_z$, respectively.

Lemma 4.15 (Invariance of stationarity equations). *Let (u, y) be a trajectory, and $t_0 \in (0, T)$. Let (ϕ, ψ) be a local change of state variable and nonlinear feedback at time t_0, with $\delta > 0$ as in Def. 4.14. Then (u, y) satisfies with multipliers (p, η, λ) the stationarity equations (4.35)-(4.36) and (4.38)-(4.40) locally on $(t_0 - \delta, t_0 + \delta)$, iff (v, z, π) defined on $(t_0 - \delta, t_0 + \delta)$ by*

$$z(t) := \phi(y(t)); \qquad v(t) := \psi(u(t), y(t)); \qquad \pi(t) := p(t)\phi_y^{-1}(y(t)) \tag{4.59}$$

satisfies on $(t_0 - \delta, t_0 + \delta)$:

$$\begin{aligned}
\dot{z}(t) &= \hat{f}(v(t), z(t)) && (4.60)\\
-\mathrm{d}\pi(t) &= \hat{H}_z(v(t), z(t), \pi(t))\mathrm{d}t + \mathrm{d}\eta(t)\hat{g}_z(z(t)) + \lambda(t)\hat{c}_z(v(t), z(t))\mathrm{d}t && (4.61)\\
0 &= \hat{H}_v(v(t), z(t), \pi(t)) + \lambda(t)\hat{c}_v(v(t), z(t)) \quad a.e. && (4.62)\\
\hat{g}(z(t)) &\le 0; \qquad \mathrm{d}\eta \ge 0; \qquad \textstyle\int_{t_0-\delta}^{to+\delta} \mathrm{d}\eta(t)\hat{g}(z(t)) = 0; && (4.63)\\
\hat{c}(v(t), z(t)) &\le 0; \quad \lambda(t) \ge 0 \ \ a.e.; \quad \textstyle\int_{t_0-\delta}^{to+\delta} \lambda(t)\hat{c}(v(t), z(t))\mathrm{d}t = 0; && (4.64)
\end{aligned}$$

with the new dynamics, integral cost function, Hamiltonian, and state and mixed constraints given by

$$\begin{aligned}
\hat{f}(v, z) &:= \phi_y(\bar{\phi}(z))f(\bar{\psi}(v, z), \bar{\phi}(z)) && (4.65)\\
\hat{\ell}(v, z) &:= \ell(\bar{\psi}(v, z), \bar{\phi}(z)) && (4.66)\\
\hat{H}(v, z, \pi) &:= \hat{\ell}(v, z) + \pi\hat{f}(v, z) && (4.67)\\
\hat{g}(z) &:= g(\bar{\phi}(z)) && (4.68)\\
\hat{c}(v, z) &:= c(\bar{\psi}(v, z), \bar{\phi}(z)). && (4.69)
\end{aligned}$$

In addition, the augmented Hamiltonian of order 0 and the time derivatives of the state constraint (all components supposed to be of finite order q_i, $i = 1, \ldots, r$), are invariant, i.e., on $V_z \times V_v$:

$$\begin{aligned}
\hat{H}^0(v, z, \pi, \lambda) &:= \hat{H}(v, z, \pi) + \lambda\hat{c}(v, z) = H^0(\bar{\psi}(v, z), \bar{\phi}(z), \pi\phi_y(\bar{\phi}(z)), \lambda); && (4.70)\\
\hat{g}_i^{(j)}(z) &= g_i^{(j)}(\bar{\phi}(z)), \quad \text{for all } j = 1, \ldots, q_i - 1, \ i = 1, \ldots, r; && (4.71)\\
\hat{g}_i^{(q_i)}(v, z) &= g_i^{(q_i)}(\bar{\psi}(v, z), \bar{\phi}(z)), \quad i = 1, \ldots, r. && (4.72)
\end{aligned}$$

Proof. Assume that (u, y, p, η, λ) satisfies (4.35)-(4.36) and (4.38)-(4.40) for $t \in (t_0 - \delta, t_0 + \delta)$, and let us show that $(v, z, \pi, \eta, \lambda)$ satisfies (4.60)-(4.64) on $(t_0 - \delta, t_0 + \delta)$. The converse is proved similarly by symmetry. By (4.59), (4.65) and (4.68)-(4.69), it is obvious that (4.60), (4.63) and (4.64) follow from (4.35) and (4.39)-(4.40). Moreover, we have

$$\begin{aligned}
\hat{H}_v^0(v, z, \pi, \lambda) &= D_v\{\ell(\bar{\psi}(v, z), \bar{\phi}(z)) + \pi\phi_y(\bar{\phi}(z))f(\bar{\psi}(v, z), \bar{\phi}(z)) + \lambda c(\bar{\psi}(v, z), \bar{\phi}(z))\}\\
&= H_u^0(\bar{\psi}(v, z), \bar{\phi}(z), p, \lambda)\ \bar{\psi}_v(v, z).
\end{aligned}$$

Since $\bar{\psi}_v$ is invertible, this gives (4.62). It remains to check the costate equation. We have

$$\begin{aligned}
\hat{H}_z^0(v, z, \pi, \lambda) = H_u^0(\bar{\psi}(v, z), \bar{\phi}(z), p, \lambda)\bar{\psi}_z(v, z) + H_y^0(\bar{\psi}(v, z), \bar{\phi}(z), p, \lambda)\bar{\phi}_z(z)\\
+ \pi\phi_{yy}(\bar{\phi}(z))(\bar{\phi}_z(z), f(\bar{\psi}(v, z), \bar{\phi}(z))).
\end{aligned} \tag{4.73}$$

By definition of π in (4.59), we have

$$\begin{aligned}
\mathrm{d}p(t) &= \mathrm{d}\{\pi(t)\phi_y(\bar{\phi}(z(t)))\}\\
&= \mathrm{d}\pi(t)\phi_y(\bar{\phi}(z(t))) + \pi(t)\phi_{yy}(\bar{\phi}(z))f(\bar{\psi}(v, z), \bar{\phi}(z))\mathrm{d}t.
\end{aligned}$$

Since $\phi_y(\bar\phi(z))\bar\phi_z(z) \equiv I_d$, using (4.36), (4.73) and (4.38) on $(t_0-\delta, t_0+\delta)$, we obtain

$$\begin{aligned} -\mathrm{d}\pi(t) &= -\mathrm{d}p(t)\bar\phi_z(z) + \pi(t)\phi_{yy}(\bar\phi(z))(f(\bar\psi(v,z),\bar\phi(z)),\bar\phi_z(z))\mathrm{d}t \\ &= \hat H^0_z(v,z,\pi,\lambda)\mathrm{d}t + \mathrm{d}\eta g_y(\bar\phi(z))\bar\phi_z(z) \;=\; \hat H^0_z(v,z,\pi,\lambda)\mathrm{d}t + \mathrm{d}\eta\hat g_z(z), \end{aligned}$$

which gives (4.61). From (4.65) and (4.68), by induction for $j = 1,\dots,q_i$, we obtain

$$\begin{aligned} \hat g_i^{(j)}(v,z) &= \hat g_{i,z}^{(j-1)}(z)\hat f(v,z) \\ &= g_{i,y}^{(j-1)}(\bar\phi(z))\bar\phi_z(z)\phi_y(\bar\phi(z))f(\bar\psi(v,z),\bar\phi(z)) \\ &= g_{i,y}^{(j-1)}(\bar\phi(z))f(\bar\psi(v,z),\bar\phi(z)) \;=\; g_i^{(j)}(\bar\psi(v,z),\bar\phi(z)), \end{aligned}$$

which shows (4.71)-(4.72) and achieves the proof. □

Remark 4.16. With the notations and assumptions of Lemma 4.15, we have

$$\hat H^0_{vv}(v,z,\pi,\lambda) = H^0_{uu}(u,y,p,\lambda)(\bar\psi_v(v,z),\bar\psi_v(v,z)) + H^0_u(u,y,p,\lambda)\,\bar\psi_{vv}(v,z) \tag{4.74}$$

and, for $J \subset \{1,\dots,r+s\}$, defining $\hat G^{(q)}_J(v,z) := \left(\hat g_i^{(q_i)}(v,z)\right)_{i\in J}$, with still $q_i := 0$ and $\hat g_i^{(0)} := \hat c_i$ for $i = r+1,\dots,r+s$, we obtain by (4.72) and (4.69):

$$\hat G^{(q)}_{J,v}(v(t),z(t)) \;=\; G^{(q_i)}_{J,u}(u(t),y(t))\bar\psi_v(v(t),z(t)).$$

Since $H^0_u(u,y,p,\lambda) = 0$ at a stationary point, and $\bar\psi_v(v,z)$ is invertible over $V_v \times V_z$, we obtain that if (u,y) is a stationary point, then assumptions (4.43) (or (4.45)) and (4.30) are locally invariant by local change of coordinate and nonlinear feedback (but of course, with possibly different positive constants α and γ).

4.4.2 The Linear Independence Lemma

Given $J \subset \{1,\dots,r\}$, we denote by $|q_J| := \sum_{i\in J} q_i$ and $|q| := \sum_{i=1}^r q_i$. Define the mapping $\Gamma_J : \mathbb{R}^n \to \mathbb{R}^{|q_J|}$ that with y associates the "J" state constraints and their time derivatives depending on y only, by:

$$\Gamma_J(y) \;:=\; \begin{pmatrix} g_{i_1}(y) \\ \vdots \\ g_{i_1}^{(q_{i_1}-1)}(y) \\ \vdots \\ g_{i_s}(y) \\ \vdots \\ g_{i_s}^{(q_{i_s}-1)}(y) \end{pmatrix}, \qquad J = \{i_1,\dots,i_s\}, \quad i_1 < \cdots < i_s. \tag{4.75}$$

Lemma 4.17. *Let $\hat y \in \mathbb{R}^n$ and $J \subset \{1,\dots,r\}$. Assume that there exists $\hat w \in \mathbb{R}^m$ such that $G^{(q)}_{J,u}(\hat w,\hat y)$ has full rank $|J|$. Then the matrix $\Gamma_{J,y}(\hat y)$ has full rank, equals to $|q_J|$.*

The above result is well-known in the case when the dynamics and the constraints are linear, but since we were not able to find a reference for it in the general nonlinear case, we give a proof below, which uses the relations (4.77) established in [98].

Proof. For $\tau \in (0,T)$ and small $\delta > 0$, consider the solution y of the state equation $\dot{y}(t) = f(u(t),y(t))$ over $(\tau-\delta,\tau+\delta)$, with $y(\tau) = \hat{y}$ and $u : (\tau-\delta,\tau+\delta) \to \mathbb{R}^m$ is here any $C^{q_{max}}$ function such that $u(\tau) = \hat{w}$. For $k = 1,\dots,q_{max}-1$, define the mappings $A_k : (\tau-\delta,\tau+\delta) \to \mathbb{R}^{n\times m}$ by:

$$\left\{ \begin{array}{l} A_0(t) := f_u(u(t),y(t)) \\ A_k(t) := f_y(u(t),y(t))A_{k-1}(t) - \dot{A}_{k-1}(t) \qquad 1 \leq k \leq q_{max}-1. \end{array} \right. \tag{4.76}$$

The proof of the lemma is based on the following relations, due to [98]. For all $t \in (\tau-\delta,\tau+\delta)$ and $i = 1,\dots,r$, we have:

$$\left\{ \begin{array}{l} g_{i,y}^{(j)}(y(t))A_k(t) = 0 \qquad \text{for } k,j \geq 0, \quad k+j \leq q_i-2, \\ g_{i,y}^{(j)}(y(t))A_{q_i-j-1}(t) = g_{i,u}^{(q_i)}(u(t),y(t)) \qquad \text{for } 0 \leq j \leq q_i-1. \end{array} \right. \tag{4.77}$$

For the sake of completeness of the paper, let us recall how to prove (4.77). We first show that for all $j = 0,\dots,q_i-1$, the following assertion

$$g_{i,y}^{(j)}(y(t))A_k(t) = 0 \qquad \forall\, t \in (\tau-\delta,\tau+\delta) \tag{4.78}$$

implies that

$$g_{i,y}^{(j+1)}(u(t),y(t))A_k(t) = g_{i,y}^{(j)}(y(t))A_{k+1}(t) \qquad \forall\, t \in (\tau-\delta,\tau+\delta). \tag{4.79}$$

Indeed, by derivation of (4.78) w.r.t. time, we get using (4.18)

$$\begin{array}{rcl} 0 & = & g_{i,yy}^{(j)}(y)f(u,y)A_k + g_{i,y}^{(j)}(y)\dot{A}_k \\ & = & g_{i,yy}^{(j)}(y)f(u,y)A_k + g_{i,y}^{(j)}(f_y(u,y)A_k - A_{k+1}) \\ & = & g_{i,y}^{(j+1)}(u,y)A_k - g_{i,y}^{(j)}(y)A_{k+1}. \end{array}$$

This gives (4.79). We also have that $g_{i,u}^{(j)}(u,y) = g_{i,y}^{(j-1)}(y)f_u(u,y) = g_{i,y}^{(j-1)}(y)A_0$ for $j = 1,\dots,q_i$. Since $g_{i,u}^{(j)} = 0$ for $j \leq q_i-1$, it follows that $g_{i,y}^{(j)}A_0 = 0$ for $j = 0,\dots,q_i-2$. By (4.79), we deduce that $g_{i,y}^{(j)}A_1 = 0$ for $j = 0,\dots,q_i-3$. By induction, this proves the first equation in (4.77). Since $g_{i,y}^{(q_i-2)}A_0 = 0 = g_{i,y}^{(q_i-3)}A_1 = \cdots = g_{i,y}A_{q_i-2}$, by (4.79) we obtain $g_{i,u}^{(q_i)} = g_{i,y}^{(q_i-1)}A_0 = g_{i,y}^{(q_i-2)}A_1 = \cdots = g_{i,y}A_{q_i-1}$, which proves the second equation in (4.77).

Assume w.l.o.g. that $J = \{1,\dots,r'\}$, with $r' \leq r$, and that $q_1 \geq q_2 \geq \cdots \geq q_{r'} \geq 1$. Consider the matrix

$$K(t) \;:=\; \left(\begin{array}{cccc} A_{q_1-1}(t) & \dots & A_1(t) & A_0(t) \end{array} \right) \in \mathbb{R}^{n\times mq_1}, \tag{4.80}$$

and form the product matrix

$$P(t) \;:=\; \Gamma_{J,y}(y(t))K(t) \in \mathbb{R}^{|q_J|\times mq_1}. \tag{4.81}$$

Let $\tilde{q}_i := \sum_{l=1}^{i} q_l$, and for $i = 1,\dots,r'$, denote by $P_i(t) \in \mathbb{R}^{q_i\times mq_1}$ the submatrix formed by the rows $\tilde{q}_{i-1}+1$ to $\tilde{q}_i$ of $P(t)$. By (4.77), we have

$$P_i(t) \;=\; \left(\begin{array}{cccc} * & g_{i,u}^{(q_i)}(u(t),y(t)) & \dots & 0 \\ * & \vdots & \ddots & \vdots \\ \underbrace{*}_{m(q_1-q_i)} & * & \cdots & g_{i,u}^{(q_i)}(u(t),y(t)) \end{array} \right). \tag{4.82}$$

Let us show that $P(\tau)$ has full rank $|q_J|$. For that consider a linear combination of the rows ℓ_j of $P(\tau)$, $\sum_{j=1}^{|q_J|} \beta_j \ell_j = 0$. By (4.82), only the rows of $P(\tau)$ for $j = \tilde{q}_i$, $i = 1, \dots, r'$, have a contribution to the last m components of $\sum_{j=1}^{|q_J|} \beta_j \ell_j$. It is easily seen that these last m components are a linear combination of the rows of $G^{(q)}_{J,u}(u(\tau), y(\tau))$, with coefficients $\beta_{\tilde{q}_i}$. Since $u(\tau) = \hat{w}$ and $G^{(q)}_{J,u}(\hat{w}, y(\tau))$ has full rank by hypothesis, it follows that $\beta_{\tilde{q}_i} = 0$ for all $i = 1, \dots, r'$. Repeating the same argument, we obtain that $\beta_j = 0$ for all $j = 1, \dots, |q_J|$, i.e. the product matrix $P(t)$ has rank $|q_J|$. Therefore, the matrix $\Gamma_{J,y}(y(\tau))$ has rank $|q_J|$. □

Corollary 4.18. *Let a trajectory (u, y) satisfy (4.30). Then the matrix $\Gamma_{I^g(t),y}(y(t))$ has full rank, equals to $|q_{I^g(t)}|$, for all $t \in [0, T]$ (and consequently, $\sum_{i \in I^g(t)} q_i \leq n$).*

4.4.3 Locally Normal form of the state equation

Lemma 4.19. *Let (u, y) be a trajectory and $t_0 \in (0, T)$ such that u is continuous at t_0. Assume that f, g are (at least) $C^{2q_{max}}$, that (4.30) holds at $t = t_0$, and w.l.o.g. that $I(t_0) = \{1, \dots, r'\} \cup \{r+1, \dots, r+s'\} =: J$. Then there exists a $C^{q_{max}}$ local change of variable and nonlinear feedback (ϕ, ψ), defined over a neighborhood of $(u(t_0), y(t_0))$, such that, with the notations of Lemma 4.15, the new dynamics $\hat{f}$ writes on $(t_0 - \delta, t_0 + \delta)$, with $\tilde{q}_i := \sum_{l=1}^{i} q_l$ (and $\tilde{q}_0 = 0$):*

$$\left\{ \begin{array}{rcll} \dot{z}_{\tilde{q}_{i-1}+1}(t) & = & z_{\tilde{q}_{i-1}+2}(t) & \\ & \vdots & & i = 1, \dots, r' \\ \dot{z}_{\tilde{q}_i - 1}(t) & = & z_{\tilde{q}_i}(t) & \\ \dot{z}_{\tilde{q}_i}(t) & = & v_i(t) & \end{array} \right. \qquad (4.83)$$

$$\dot{z}_N(t) = \hat{f}_N(v(t), z(t)),$$

where z_N and $\hat{f}_N$ denote components $|q_J| + 1, \dots, n$ of z and $\hat{f}$, and the state and mixed constraints $\hat{g}$ and $\hat{c}$ are given by:

$$\hat{g}_i(z(t)) = z_{\tilde{q}_{i-1}+1}(t) \leq 0, \qquad i = 1, \dots, r' \qquad (4.84)$$

$$\hat{c}_i(v(t), z(t)) = v_{i-r+r'}(t) \leq 0, \qquad i = r+1, \dots, r+s'. \qquad (4.85)$$

Under this change of coordinates, the active state constraints $\hat{g}_i$ and their time derivatives until order q_i are linear, and the active mixed control-state constraints $\hat{c}_i$ are linear as well, and depend only on the control.

Proof. By Coro. 4.18, the Jacobian $\Gamma_{J,y}(y(t_0))$ has full-rank, equal to $|q_J|$, and since y is continuous at t_0, there exist $\delta > 0$ and a diffeomorphism ϕ defined over an open neighborhood V_y in $\mathbb{R}^n$ of $\{y(t)\,;\, t \in (t_0 - \delta, t_0 + \delta)\}$, such that $\phi_k(y) = \Gamma_J(y)|_k$, for all $k = 1, \dots, |q_J|$.

By (4.30), there exists then an open neighborhood V_u of $u(t_0)$ in $\mathbb{R}^m$, such that all $u \in V_u$ can be partitioned in $u = (u_G, u_N) \in \mathbb{R}^{r'+s'} \times \mathbb{R}^{m-r'-s'}$, and $G^{(q)}_{J,u_G}(u(t_0), y(t_0))$ is invertible

(note that $|J| = r' + s'$). Consequently, reducing V_u and V_y if necessary, the mapping

$$\psi(\cdot, y) \ : \ u \mapsto \begin{pmatrix} g_1^{(q_1)}(u,y) \\ \vdots \\ g_{r'}^{(q_{r'})}(u,y) \\ c_{r+1}(u,y) \\ \vdots \\ c_{r+s'}(u,y) \\ u_N \end{pmatrix} \tag{4.86}$$

has an invertible Jacobian $\psi_u(u,y)$, for all $(u,y) \in V_u \times V_y$. Since by assumption, u is continuous at t_0, reducing δ if necessary, V_u is a neighborhood of $\{u(t);\ t \in (t_0 - \delta, t_0 + \delta)\}$.

Therefore, (ϕ, ψ) is a $C^{q_{max}}$ local change of state variables and nonlinear feedback, so Lemma 4.15 applies, and formulae (4.65) and (4.68)-(4.69) give the expressions (4.83) and (4.84)-(4.85). □

4.5 Junctions Conditions Analysis

In Prop. 4.13, it was shown that when assumptions (A2) and (A3) hold, the control and multipliers are smooth on the interior of the arcs of the trajectory. In this section we study the regularity of the control and multipliers at the junction between two arcs. The main result of this section is Prop. 4.22 which generalizes the result obtained by Jacobson, Lele, and Speyer [75] in the particular case of a scalar control and scalar state constraint.

4.5.1 Junction points

The set of *junction points* (or junction times) of constraint $i = 1, \ldots, r+s$, is defined as the endpoints in $(0,T)$ of the contact set Δ_i and is denoted by $\mathcal{T}^i := \partial\Delta_i$.

A *boundary* (resp. *interior*) *arc* of component g_i is a maximal *open* interval of positive measure $\mathcal{I}_i \subset [0,T]$, such that $g_i(y(t)) = 0$ (resp. $g_i(y(t)) < 0$) for all $t \in \mathcal{I}_i$. If $(\tau_{en}^i, \tau_{ex}^i)$ is a boundary arc of g_i, then τ_{en}^i and τ_{ex}^i are called respectively *entry* and *exit* point (or time) of the constraint g_i. A *touch* point τ_{to}^i in $(0,T)$ is an isolated contact point for constraint g_i (endpoint of two interior arcs). Similar definitions of boundary and interior arcs, entry, exit and touch points for the mixed control-state constraints c_i, $i = r+1, \ldots, r+s$, hold. Thus entry, exit and touch points are by definition junction points.

Definition 4.20. We say that a junction point τ is *regular*, if it is endpoint of two arcs.

By the above definition, a cluster point of junction times is not a regular junction time. The (disjoint and possibly empty) sets of *regular* entry, exit and touch points of constraint g_i and c_i will be respectively denoted by $\mathcal{T}_{en}^i$, $\mathcal{T}_{ex}^i$, and $\mathcal{T}_{to}^i$. Thus $\mathcal{T}^i \supset \mathcal{T}_{en}^i \cup \mathcal{T}_{ex}^i \cup \mathcal{T}_{to}^i$ with equality for all $i = 1, \ldots, r+s$ iff all the junction points are regular (equivalently, iff $\mathcal{T}^i$ is *finite* for all $i = 1, \ldots, r+s$). The set of all junctions times of the trajectory (u,y) will be denoted by $\mathcal{T}$, with

$$\mathcal{T} := \bigcup_{i=1}^{r+s} \mathcal{T}^i. \tag{4.87}$$

Definition 4.21. A touch point $\tau_{to}^i \in \mathcal{T}_{to}^i$ of the state constraint g_i, for $i = 1, \ldots, r$, is said to be *essential*, if it belongs to the support of the multiplier η_i, that is if $[\eta_i(\tau_{to}^i)] > 0$.

In other words, a touch point is essential, if strict complementarity locally holds at that touch point. Otherwise, it is said *nonessential*. The set of essential (resp. nonessential) touch points for constraint i will be denoted by $\mathcal{T}_{to}^{i,ess}$ (resp. $\mathcal{T}_{to}^{i,nes}$). For mixed control-state constraints, since $\lambda \in L^\infty$, we will say by extension that touch points of mixed control-state constraints are always nonessential. The regularity of u, η, λ given in Prop. 4.13 is not affected by the presence of nonessential touch points.

Recall now the alternative multipliers in subsection 4.3.2. Let τ be a *regular* junction time, i.e. τ is the right and left endpoint of two arcs, (τ_1, τ) and (τ, τ_2), with constant set of active constraints J_1 and J_2, respectively. Note that $J_1 \cup J_2 \subset I(\tau)$, the inclusion being strict iff τ is a touch point for at least one of the constraint. The multipliers η_i^j for $j = 1, \ldots, q_i$ and $i = 1, \ldots, r$ being defined in (4.49) up to a polynomial function of order $j-1$ on each arc (τ_1, τ) and (τ, τ_2), their jump at τ are well-defined. According to (4.50) and (4.36), it holds, with $\nu_\tau^i := [\eta_i(\tau)] \geq 0$:

$$\begin{aligned} [p^q(\tau)] &= [p(\tau)] - \sum_{i \in I(\tau)} \sum_{j=1}^{q_i} [\eta_i^j(\tau)] g_{i,y}^{(j-1)}(y(\tau)) \\ &= -\sum_{i \in I(\tau)} \{(\nu_\tau^i + [\eta_i^1(\tau)]) g_{i,y}(y(\tau)) + \sum_{j=2}^{q_i} [\eta_i^j(\tau)] g_{i,y}^{(j-1)}(y(\tau))\}. \end{aligned} \tag{4.88}$$

4.5.2 Junction conditions

We say that a function $u \in L^\infty(0,T;\mathbb{R}^m)$ is *continuous until order* $k \geq 0$ at point $\tau \in (0,T)$, if u and its time derivatives $\dot{u}, \ldots, u^{(k)}$ are continuous at τ. We say that u is *discontinuous at order* $k' \geq 1$ at point τ, if u is continuous until order $k'-1$ and if the time derivative $u^{(k')}$ of order k' is discontinuous at τ. This integer k' will be called the *order of discontinuity* of the control. If u is not continuous at τ (resp. if u is C^∞ at τ), we say that u has order of discontinuity 0 (resp. ∞).

The next theorem is an extension of the junction conditions results of Jacobson, Lele and Speyer [75] to the case of a vector-valued state constraint and control. Let us recall their result. Given an optimal control problem with a scalar control $u(t) \in \mathbb{R}$ and a scalar state constraint $g(y(t)) \leq 0$, if (u, y) is a stationary point satisfying assumptions (A2)-(A3), then the time derivatives of u are continuous at a regular junction point until an order that depends on the *order* q of the (scalar) state constraint, and on the *nature* of the junction point (regular entry/exit points versus essential touch points). More precisely, for constraints of first order, u is continuous at entry/exit points, and essential touch points cannot occur (see Prop. 4.8(ii)). For constraints of *even* order $q \geq 2$, u is continuous until order $q-2$ at regular entry/exit points and essential touch points. For constraints of *odd* order $q \geq 3$, u is continuous until order $q-1$ at regular entry/exit points and until order $q-2$ at essential touch points. The result is illustrated in figure 4.1 below. The junction condition results for mixed control-constraints ($q = 0$) were added.

When studying the second-order necessary condition (see section 4.6), we have to compute the expression (4.120) at junction points τ. To this end, we use Taylor expansions of the nominator and denominator in the neighborhood of τ, and for this we need to know the order of discontinuity of the function $g_i(y(t))$ at regular entry/exit points. Since $\frac{\mathrm{d}^{q_i}}{\mathrm{d}t^{q_i}} g_i(y(t)) = g_i^{(q_i)}(u(t), y(t))$, we see that the order of discontinuity of $g_i(y(t))$ is at least q_i plus the order of discontinuity of the control.

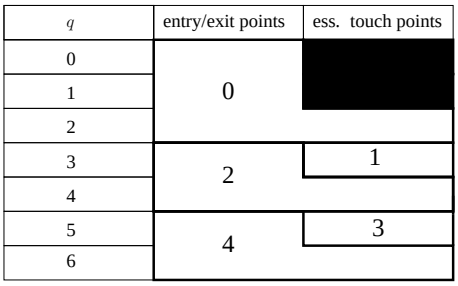


Figure 4.1: Order of continuity of the control at a regular junction point, in function of the order of the constraint q and the nature of the junction point (in the scalar case).

Proposition 4.22. *Assume that the data are (at least) $C^{2q_{max}}$. Let (u, y) be a stationary point of $(\mathcal{P})$, and let $\tau \in (0, T)$ be a regular junction point. Assume that u is continuous at τ and that* (4.45) *and* (4.30) *are satisfied at $t = \tau$. Let*

$$q_\tau := \min\{\, q_i \;;\; \tau \in \mathcal{T}^i_{en} \cup \mathcal{T}^i_{ex} \cup \mathcal{T}^{i,ess}_{to},\ i \in I(\tau)\}. \tag{4.89}$$

(i) *If $q_\tau \geq 3$, then the control is continuous at τ until order $q_\tau - 2$.*
(ii) *If in addition, the following holds:*

$$\begin{array}{c} q_\tau \text{ is odd, and for all } i \text{ such that } q_i = q_\tau \text{ and } \tau \in \mathcal{T}^i \setminus \mathcal{T}^{i,nes}_{to} \\ \tau \text{ is an entry or exit point, i.e. } \tau \in \mathcal{T}^i_{en} \cup \mathcal{T}^i_{ex}, \end{array} \tag{4.90}$$

then the control is continuous at τ until order $q_\tau - 1$.

The alternative multipliers $\eta_i^{q_i}$ for all $i = 1, \ldots, r + s$ such that $\tau \in \operatorname{int} \Delta_i$ are continuous at τ until the same order as the control. In particular,

$$\text{(i}') \text{ If } q_\tau \geq 3, \qquad \nu^i_\tau = [\eta_i(\tau)] = 0 \quad \text{for all } i \in I(\tau) \text{ such that } q_i < q_\tau, \tag{4.91}$$

$$\text{(ii}') \text{ If (4.90) holds}, \quad \nu^i_\tau = [\eta_i(\tau)] = 0 \quad \text{for all } i \in I(\tau) \text{ such that } q_i \leq q_\tau. \tag{4.92}$$

Remark 4.23. If $q_\tau = 1$, then (4.90) always holds since components of first order of the state constraint have no essential touch points by Prop. 4.8(ii). It follows then from Prop. 4.8 that point (i') (resp. (ii')) of Prop. 4.22 holds true when $q_\tau = 2$ (resp. $q_\tau = 1$).

Proof. Let $\tau \in \mathcal{T}$ be such that $q_\tau > 2$. Assume w.l.o.g. that

$$I(\tau) = \{1, \ldots, r'\} \cup \{r + 1, \ldots, r + s'\} =: J, \qquad 1 \leq q_1 \leq \ldots \leq q_{r'}. \tag{4.93}$$

We will use the local invariance of stationary points of Lemma 4.15 for the particular choice of (ϕ, ψ) given in Lemma 4.19, and write the optimality conditions in these variables (v, z). Since $u(t) = \bar\psi(v(t), z(t))$, $\bar\psi$ is $C^{q_{max}}$, and $\bar\psi_v(v(t), z(t))$ is invertible in the neighborhood of τ, the continuity of $u, \ldots, u^{(j)}$ for $j \leq q_{max}$ is equivalent to the continuity of $v, \ldots, v^{(j)}$. Assume w.l.o.g. that $\delta > 0$ is so small that $\mathcal{T} \cap (\tau - \delta, \tau + \delta) = \{\tau\}$. Define

$$r_k := \operatorname{Card}\{i \in I(\tau) \;;\; 1 \leq q_i \leq k\}, \qquad 0 \leq k \leq q_{max}, \qquad r_0 := 0.$$

Then $r_{q_{max}} = r'$, and the useful relation below holds, for all $1 \leq i \leq r'$ and $1 \leq k \leq q_{max}$:

$$r_{k-1} < i \leq r_k \quad \text{iff} \quad q_i = k. \tag{4.94}$$

Denote the nonlinear part of the Hamiltonian by:

$$\hat{L}(v,z,\pi_N) \;:=\; \hat{\ell}(v,z) + \sum_{k=|q_J|+1}^{n} \pi_k \hat{f}_k(v,z) \;=\; \hat{\ell}(v,z) + \pi_N \hat{f}_N(v,z),$$

where, similarly to y_N and $\hat{f}_N$, we denote by π_N the last $n-|q_J|$ components of π, and still denote $\tilde{q}_i := \sum_{l=1}^{i} q_l$ for $i = 0,\dots,r'$. Then (v,z) is solution on $(\tau-\delta,\tau+\delta)$ of the state equation (4.83), and, since

$$\hat{G}_J^{(q)}(v,z) \;=\; (v_1,\dots,v_{r'},v_{r'+1},\dots,v_{r'+s'})^\top,$$

the *alternative* costate and control equations (recall Lemma 4.11 and Rem. 4.12) satisfied on $(\tau-\delta,\tau)\cup(\tau,\tau+\delta)$ are respectively given by:

$$\left\{\begin{array}{rcll} -\dot{\pi}^q_{\tilde{q}_{i-1}+1}(t) &=& \hat{L}_{z_{\tilde{q}_{i-1}+1}}(v(t),z(t),\pi^q_N(t)) & \\ -\dot{\pi}^q_{\tilde{q}_{i-1}+2}(t) &=& \hat{L}_{z_{\tilde{q}_{i-1}+2}}(v(t),z(t),\pi^q_N(t)) + \pi^q_{\tilde{q}_{i-1}+1}(t) & \\ \qquad\vdots & & & i=1,\dots,r' \\ -\dot{\pi}^q_{\tilde{q}_i}(t) &=& \hat{L}_{z_{\tilde{q}_i}}(v(t),z(t),\pi^q_N(t)) + \pi^q_{\tilde{q}_i-1}(t) & \end{array}\right. \tag{4.95}$$

$$-\dot{\pi}^q_N(t) \;=\; \hat{L}_{z_N}(v(t),z(t),\pi^q_N(t)); \tag{4.96}$$

$$0 \;=\; \hat{L}_{v_i}(v(t),z(t),\pi^q_N(t)) \;+\; \pi^q_{\tilde{q}_i}(t) \;+\; \eta^{q_i}_i(t), \quad i=1,\dots,r' \tag{4.97}$$

$$0 \;=\; \hat{L}_{v_i}(v(t),z(t),\pi^q_N(t)) \;+\; \eta^0_{i-r'+r}(t), \quad i=r'+1,\dots,r'+s' \tag{4.98}$$

$$0 \;=\; \hat{L}_{v_N}(v(t),z(t),\pi^q_N(t)), \tag{4.99}$$

where v_N denotes the remaining $m-r'-s'$ components of the control. Since $\hat{g}^{(j-1)}_{i,y}(z)$ is the $(\tilde{q}_{i-1}+j)$-th basis vector, by (4.88), the jump of each component of π^q satisfies, using that $\tilde{q}_{i-1}+1=i$ if $i\le r_1$ (recall that here, $\nu^i_\tau=[\eta_i(\tau)]\ge 0$ and by Prop. 4.8(ii), $\nu^i_\tau=0$ if $q_i=1$, i.e. if $i\le r_1$ by (4.94)):

$$\begin{array}{rcll} [\pi^q_i(\tau)]+[\eta^1_i(\tau)] &=& -\nu^i_\tau \;=\; 0 & i=1,\dots,r_1 \\ [\pi^q_{\tilde{q}_{i-1}+1}(\tau)]+[\eta^1_i(\tau)] &=& -\nu^i_\tau \;\le\; 0 & i=r_1+1,\dots,r' \\ [\pi^q_{\tilde{q}_{i-1}+j}(\tau)]+[\eta^j_i(\tau)] &=& 0, \qquad j=2,\dots,q_i, & i=r_1+1,\dots,r' \\ [\pi^q_N(\tau)] &=& 0. & \end{array} \tag{4.100}$$

For future reference, we rewrite the above relations as

$$\begin{array}{rcll} [\pi^q_{\tilde{q}_i}(\tau)]+[\eta^{q_i}_i(\tau)] &=& -\nu^i_\tau \;=\; 0 & i=1,\dots,r_1 \\ [\pi^q_{\tilde{q}_i-q_i+1}(\tau)]+[\eta^1_i(\tau)] &=& -\nu^i_\tau \;\le\; 0 & i=r_1+1,\dots,r' \\ [\pi^q_{\tilde{q}_i-j}(\tau)]+[\eta^{q_i-j}_i(\tau)] &=& 0, \qquad j=0,\dots,q_i-2, & i=r_1+1,\dots,r' \\ [\pi^q_N(\tau)] &=& 0. & \end{array} \tag{4.101}$$

By Prop. 4.13, the control and state constraint alternative multiplier η^q are $C^{q_{max}}$ on interiors of arcs, therefore we may define over $(\tau-\delta,\tau)\cup(\tau,\tau+\delta)$ the functions a^j_i for

$i = 1, \dots, r' + s'$ and $j = 0, \dots, q_{max}$ by:

$$\left\{ \begin{array}{rcll} a_i^0(t) & := & \hat{L}_{v_i}(v(t), z(t), \pi_N^q(t)), & \\ a_i^{j+1}(t) & := & -\frac{\mathrm{d}}{\mathrm{d}t} a_i^j(t) + \hat{L}_{z_{\tilde{q}_i - j}}(v(t), z(t), \pi_N^q(t)), & 0 \le j \le q_i - 1 \\ a_i^{j+1}(t) & := & -\frac{\mathrm{d}}{\mathrm{d}t} a_i^j(t), & q_i \le j \le q_{max}. \end{array} \right.$$

After j derivations of row i of (4.97) and (4.98), $1 \le j \le q_{max}$, we obtain using (4.95) that the following holds, on $(\tau - \delta, \tau) \cup (\tau, \tau + \delta)$:

$$0 = a_i^j(t) + \pi^q_{\tilde{q}_i - j}(t) + \eta_i^{q_i - j}(t), \quad 1 \le j \le q_i - 1, \ i = 1, \dots, r', \tag{4.102}$$

$$0 = a_i^j(t) + (-1)^{q_i - j} \eta_i^{q_i - j}(t), \quad q_i \le j \le q_{max}, \ i = 1, \dots, r', \tag{4.103}$$

$$0 = a_i^j(t) + (-1)^{-j} \eta^{-j}_{i - r' + r}(t), \quad 1 \le j \le q_{max}, \ i = r' + 1, \dots, r' + s'. \tag{4.104}$$

Here, for all $i \in J$, we define for $q_i - j \le 0$, $\eta_i^{q_i - j} := (-1)^{q_i} \frac{\mathrm{d}^j}{\mathrm{d}t^j} \eta_i^{q_i}(t)$. We have, by definition of the functions a_i^j, for all $1 \le j \le q_{max}$ and $i = 1, \dots, r' + s'$, with (4.95)-(4.96),

$$\begin{array}{rcl} a_i^j(t) & = & (-1)^j \hat{L}_{v_i v}(v(t), z(t), \pi_N^q(t)) v^{(j)}(t) \\ & & + \text{ a continuous function of } (v^{(j-1)}(t), \dots, v(t), z(t), \pi_N^q(t)). \end{array} \tag{4.105}$$

This implies in particular that if $v, \dots, v^{(j-1)}$ are continuous at τ, then the jump of a_i^j at time τ is given by

$$[a_i^j(\tau)] = (-1)^j \hat{L}_{v_i v}(v(t), z(t), \pi_N^q(\tau))[v^{(j)}(\tau)].$$

Similarly, by derivations of (4.99), we obtain, for all $1 \le j \le q_{max}$:

$$\begin{array}{rcl} 0 & = & (-1)^j \hat{L}_{v_N v}(v(t), z(t), \pi_N^q(t)) v^{(j)}(t) \\ & & + \text{ a continuous function of } (v^{(j-1)}(t), \dots, v(t), z(t), \pi_N^q(t)). \end{array} \tag{4.106}$$

Let us show now that the time derivatives of the control v are continuous until order $q_\tau - 2$. By assumption, v is continuous at τ. By induction, assume that $v, \dots, v^{(j-1)}$ are continuous at τ, for $j < q_\tau - 2$. Taking the jump at τ in (4.102)-(4.103) and (4.106), we obtain, for $i = 1, \dots, r' + s'$ (recall that by (4.94), $i \le r_j$ iff $1 \le q_i \le j$):

$$\begin{array}{l} 0 = (-1)^j \hat{L}_{v_i v}(v(\tau), z(\tau), \pi_N^q(\tau))[v^{(j)}(\tau)] + (-1)^{q_i - j}[\eta_i^{q_i - j}(\tau)], \quad i \le r_j \\ 0 = (-1)^j \hat{L}_{v_i v}(v(\tau), z(\tau), \pi_N^q(\tau))[v^{(j)}(\tau)] + [\pi^q_{\tilde{q}_i - j}(\tau)] + [\eta_i^{q_i - j}(\tau)], \quad r_j < i \le r' \\ 0 = (-1)^j \hat{L}_{v_i v}(v(\tau), z(\tau), \pi_N^q(\tau))[v^{(j)}(\tau)] + (-1)^{-j}[\eta^{-j}_{i - r' + r}(\tau)], \quad i > r' \\ 0 = (-1)^j \hat{L}_{v_N v}(v(\tau), z(\tau), \pi_N^q(\tau))[v^{(j)}(\tau)]. \end{array} \tag{4.107}$$

We denote in the sequel by $v_{k+1:l}$ the subvector of components $k+1, \dots, l$ of v. Similarly, $\nu_\tau^{k+1:l}$ denotes the column vector of components ν_τ^i for $i = k+1, \dots, l$. Recall that by (4.94), $q_i - j = 1$ iff $r_j < i \le r_{j+1}$, and $q_i - j > 1$ iff $i > r_{j+1}$. Since $\hat{H}^0_{vv} = \hat{L}_{vv}$ depends only on $(v, z, \pi_N^q = \pi_N)$, we write in what follows $\hat{H}^0_{vv}(v, z, \pi_N^q)$ instead of $\hat{H}^0_{vv}(v, z, \pi, \lambda)$, and using (4.101), equations (4.107) become:

$$\hat{H}^0_{vv}(v(\tau), z(\tau), \pi_N^q(\tau)) \begin{pmatrix} \left[v^{(j)}_{1:r_j}(\tau)\right] \\ \left[v^{(j)}_{r_j+1:r_{j+1}}(\tau)\right] \\ \left[v^{(j)}_{r_{j+1}+1:r'}(\tau)\right] \\ \left[v^{(j)}_{r'+1:r'+s'}(\tau)\right] \\ \left[v^{(j)}_{r'+s'+1:m}(\tau)\right] \end{pmatrix} = \begin{pmatrix} (-1)^{q_i+1}[\eta_i^{q_i - j}(\tau)] \\ (-1)^j \nu_\tau^{r_j+1:r_{j+1}} \\ 0 \\ -[\eta^{-j}_{i-r'+r}(\tau)] \\ 0 \end{pmatrix}. \tag{4.108}$$

By remark 4.16, $\hat{H}^0_{vv}(v(\tau), z(\tau), \pi^q_N(\tau))$ satisfies (4.45) for some positive constant α'. Since $[v^{(j)}(\tau)]$ is such that $\hat{g}^{(q_i)}_{i,v}(v(\tau), z(\tau))[v^{(j)}(\tau)] = [v_i^{(j)}(\tau)] = 0$ for all $i = 1, \dots, r'$ such that $\tau \in \operatorname{int} \Delta_i$, and $\hat{g}^{(q_i)}_{i,v}(v(\tau), z(\tau))[v^{(j)}(\tau)] = [v^{(j)}_{i+r'-r}(\tau)] = 0$ for all $i = r+1, \dots, r+s'$ such that $\tau \in \operatorname{int} \Delta_i$, it follows that

$$\alpha' |[v^{(j)}(\tau)]|^2 \ \leq \ [v^{(j)}(\tau)]^\top \hat{H}^0_{vv}(v(\tau), z(\tau), \pi^q_N(\tau))[v^{(j)}(\tau)]. \tag{4.109}$$

For all $j \leq q_\tau - 1$, by definition of q_τ, we have $\tau \in \operatorname{int} \Delta_i$, for all $i = 1, \dots, r_j$ and hence, $[v_i^{(j)}(\tau)] = 0$ for all $i = 1, \dots, r_j$. Since $q_\tau > 0$, we have for the same reason $[v_i^{(j)}(\tau)] = 0$ for all $i = r' + 1, \dots, r' + s'$. Therefore, (4.108) writes

$$\hat{H}^0_{vv}(v(\tau), z(\tau), \pi^q_N(\tau)) \begin{pmatrix} 0 \\ \left[v^{(j)}_{r_j+1:r_{j+1}}(\tau)\right] \\ \left[v^{(j)}_{r_{j+1}+1:r'}(\tau)\right] \\ 0 \\ \left[v^{(j)}_{r'+s'+1:m}(\tau)\right] \end{pmatrix} = \begin{pmatrix} (-1)^{q_i+1}[\eta_i^{q_i-j}(\tau)] \\ (-1)^j \nu_\tau^{r_j+1:r_{j+1}} \\ 0 \\ -[\eta^{-j}_{i-r'+r}(\tau)] \\ 0 \end{pmatrix}. \tag{4.110}$$

For $j \leq q_\tau - 2$, we also have $\tau \in \operatorname{int} \Delta_i$, for all $i \leq r_{j+1}$, and hence $[v^{(j)}_{r_j+1:r_{j+1}}(\tau)] = 0$. Multiplying on the left (4.110) by $[v^{(j)}(\tau)]^\top$, we obtain that the product with the right-hand side is zero, and therefore $[v^{(j)}(\tau)]^\top \hat{H}^0_{vv}(v(\tau), z(\tau), \pi^q_N(\tau))[v^{(j)}(\tau)] = 0$. From (4.109) it follows that $v^{(j)}$ is continuous at τ, and the right-hand side in (4.110) is equal to zero. This implies that the alternative multipliers $\eta_i^{q_i}$ are C^j at τ, and the second row of (4.100) is satisfied with equality, that is $\nu^i_\tau = 0$, for all $i = 1, \dots, r_{j+1}$, i.e. such that $q_i \leq j+1 \leq q_\tau - 1$ and $\tau \in \operatorname{int} \Delta_i$. By induction, we proved that $v, \dots, v^{(q_\tau - 2)}$ are continuous. This shows (i) and (i').

Let now $j = q_\tau - 1$. Assume that (4.90) holds, i.e. q_τ is odd, and attained at entry/exit points. Then we have, near the boundary arc, due to the continuity of $v_i, \dots, v_i^{(q_\tau-2)}$ vanishing at entry/exit of boundary arc, for all $i = r_{q_\tau - 1} + 1, \dots, r_{q_\tau}$ (and hence $q_i = q_\tau$):

$$z_{\tilde{q}_{i-1}+1}(t) \ = \ \frac{(t-\tau)^{(2q_\tau-1)}}{(2q_\tau - 1)!} v_i^{(q_\tau-1)}(\tau^\pm) + \mathcal{O}((t-\tau)^{2q_\tau}) \ \leq \ 0,$$

from which we deduce that $[v_i^{(q_\tau-1)}(\tau)] \leq 0$ at both entry and exit times. We still have $[v_i^{(q_\tau-1)}(\tau)] = 0$ for $i \leq r_{q_\tau - 1}$ and for $i = r' + 1, \dots, r' + s'$, since $q_i \leq q_\tau - 1$ implies that we are on the interior of a boundary arc for constraint i. Since $v, \dots, v^{(q_\tau-2)}$ are continuous, (4.110) holds for $j = q_\tau - 1$, hence we obtain by (4.109) and (4.100), since $\nu_\tau^{r_{q_\tau-1}+1:r_{q_\tau}} \geq 0$:

$$\begin{aligned} \alpha' |[v^{(q_\tau-1)}(\tau)]|^2 \ &\leq \ [v^{(q_\tau-1)}(\tau)]^\top \hat{H}^0_{vv}(v(\tau), z(\tau), \pi^q_N(\tau))[v^{(q_\tau-1)}(\tau)] \\ &= \ (-1)^{q_\tau - 1}[v^{(q_\tau-1)}_{r_{q_\tau-1}+1:r_{q_\tau}}(\tau)]^\top \nu_\tau^{r_{q_\tau-1}+1:r_{q_\tau}} \ \leq \ 0, \end{aligned}$$

which implies that $v^{(q_\tau-1)}$ is also continuous, and $\nu^i_\tau = 0$ for all $i \in I(\tau)$ such that $q_i = q_\tau$. This shows (ii) and (ii') and achieves the proof. □

4.6 No-Gap Second-order Optimality Conditions

In this section, we extend the no-gap second-order optimality conditions of [21] given in the scalar case, to several state constraints, and include mixed control-state constraints. The main results of the section are Theorem 4.24 and Corollary 4.25.

4.6.1 Abstract Optimization Framework and Main result

We consider here the abstract formulation (4.5) of $(\mathcal{P})$. We say that a local solution u of (4.5) satisfies the *quadratic growth condition*, if there exist $c, \rho > 0$ such that

$$J(u') \geq J(u) + c\|u' - u\|_2^2, \quad \text{for all } u' : \left\|u' - u\right\|_\infty < \rho,\ G(u') \in K,\ \mathcal{G}(u') \in \mathcal{K}. \tag{4.111}$$

Recall that the Lagrangian is given by (4.33). Let $(u, y = y_u)$ be a local solution of $(\mathcal{P})$ satisfying the assumptions of Th. 4.5, with (unique) multipliers p, η and λ. A second-order necessary condition for (4.5) due to Kawasaki [77] is as follows:

$$D^2_{uu}L(u;\eta,\lambda)(v,v) - \sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) - \sigma(\lambda, T_{\mathcal{K}}^{2,i}(\mathcal{G}(u), D\mathcal{G}(u)v)) \geq 0, \tag{4.112}$$

for all directions v in the critical cone $C(u)$ defined by

$$C(u) := \{v \in \mathcal{U} : DJ(u)v \leq 0,\ DG(u)v \in T_K(G(u)),\ D\mathcal{G}(u)v \in T_{\mathcal{K}}(\mathcal{G}(u))\}. \tag{4.113}$$

Here $T_P(x)$ (for $P = K$ or $\mathcal{K}$) denotes the tangent cone (in the sense of convex analysis) to the set P at point $x \in P$, $T_P^{2,i}(x,h)$ is the *inner second-order tangent set* to P at $x \in P$ in direction h,

$$T_P^{2,i}(x,h) := \{w \ :\ \mathrm{dist}(x + \varepsilon h + \frac{\varepsilon^2}{2}w, P) = o(\varepsilon^2),\ \forall \varepsilon > 0\},$$

and $\sigma(\cdot, S)$ denotes the *support function* of the set S, defined for $\xi \in X^*$ by $\sigma(\xi, S) = \sup_{x\in S}\langle \xi, x\rangle$. The critical cone can be characterized as follows:

$$C(u) = \{v \in \mathcal{U} \ :\ DG(u)v \in T_K(G(u)) \cap \eta^\perp,\ D\mathcal{G}(u)v \in T_{\mathcal{K}}(\mathcal{G}(u)) \cap \lambda^\perp\}. \tag{4.114}$$

The term

$$\Sigma(u,v) := \sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) + \sigma(\lambda, T_{\mathcal{K}}^{2,i}(\mathcal{G}(u), D\mathcal{G}(u)v)) \tag{4.115}$$

in (4.112) is called the *curvature term*. It is nonpositive, for all $v \in C(u)$. Note that the component i of $DG(u)v$ (resp. $D\mathcal{G}(u)v$) is the function $g_{i,y}(y(\cdot))z_v(\cdot)$ (resp. $c_{i,u}(u(\cdot), y(\cdot))v(\cdot) + c_{i,y}(u(\cdot), y(\cdot))z_v(\cdot)$), where z_v is the solution of the linearized state equation (4.22).

When there are only mixed control-state constraints, it is known that the latter have no contribution in the curvature term (4.115). This follows from the extended polyhedricity framework, see [24, Propositions 3.53 and 3.54] (the cone $\mathcal{K}$ is a polyhedric subset of L^∞ and $D\mathcal{G}(u)$ is "onto" by (4.31)). On the contrary, pure state constraints may have a non zero contribution in the curvature term (4.115).

Since K has a product form, $K \equiv (K_0)^r$ with $K_0 := C_-[0,T]$, the *inner* second-order tangent set is also given under a product expression. This would be false, however, for the *outer* second-order tangent-set, see e.g. [24, p.168]. Therefore we have, for $x = (x_i)_{1\leq i\leq r} \in K$ and $h = (h_i)_{1\leq i\leq r} \in T_K(x)$:

$$T_K^{2,i}(x,h) = \prod_{i=1}^r T_{K_0}^{2,i}(x_i, h_i). \tag{4.116}$$

Since the support function of a cartesian product of sets is the sum of the support function for each set, the expression of pure state constraints in the curvature term can be deduced from the result by Kawasaki [79] for $K_0 = C_-[0,T]$. Recall that Δ_i is given by (4.14), and the *second-order contact set* is defined, for $v \in \mathcal{V}$, by

$$\Delta_i^2 \ := \ \{t \in \Delta_i \ ;\ g_{i,y}(y(t))z_v(t) = 0\}, \qquad i = 1, \ldots, r. \tag{4.117}$$

Then, by [79], we have

$$\sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) = \sum_{i=1}^{r} \sigma(\eta_i, T_{K_0}^{2,i}(g_i(y), g_{i,y}(y)z_v)) = \sum_{i=1}^{r} \int_0^T \varsigma_i(t)\mathrm{d}\eta_i(t),$$

where, for all $i = 1, \ldots, r$:

$$\varsigma_i(t) = \begin{cases} 0 & \text{if } t \in (\operatorname{int} \Delta_i) \cap \Delta_i^2 \\ \liminf\limits_{t' \to t;\ g_i(y(t'))<0} \dfrac{(\{g_{i,y}(y(t'))z_v(t')\}_+)^2}{2g_i(y(t'))} & \text{if } t \in (\partial\Delta_i) \cap \Delta_i^2 \\ +\infty & \text{otherwise} \end{cases} \tag{4.118}$$

where $h_+(t) := \max(0, h(t))$. We denote in the sequel by $\operatorname{supp}(\mathrm{d}\eta_i)$ the *support* of the measure η_i. We make the following assumption:

(A4) **(i)** Each component of the state constraint g_i, $i = 1, \ldots, r$, has *finitely many* junctions times, and the state constraint is not active at final time, $g_i(y(T)) < 0$, $i = 1, \ldots, r$.

This assumption implies that all entry and exit times of state constraints are regular. Using (4.118), and the fact that $\operatorname{supp}(\mathrm{d}\eta_i) \subset \Delta_i^2$ for all critical directions v, the curvature term has the expression below, for $v \in C(u)$ (see [79]), with $\nu_\tau^i = [\eta_i(\tau)]$

$$\sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) = \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}_i \cap \Delta_i^2} \nu_\tau^i \varsigma_i(\tau). \tag{4.119}$$

We thus need to compute, for junction times $\tau \in \mathcal{T}_i \cap \Delta_i^2$,

$$\varsigma_i(\tau) = \liminf_{t \to \tau;\ g_i(y(t))<0} \frac{(\{g_{i,y}(y(t))z_v(t)\}_+)^2}{2g_i(y(t))}. \tag{4.120}$$

The tangentiality conditions (see assumption (A5)(i) below), under which boundary arcs with regular entry/exit points of state constraints have no contribution to the curvature term, are more delicate to state than in the scalar case, due to the possibility of having coinciding junction times of different components of the state constraints. Let $i = 1, \ldots, r$ and $\tau \in \mathcal{T}_{en}^i \cup \mathcal{T}_{ex}^i$. Denote by k_i^τ the order of discontinuity at point τ of the function (of time) $g_i^{(q_i)}(u(t), y(t))$. By Prop. 4.22, we necessarily have $k_i^\tau \geq q_\tau - 1$. A Taylor expansion of the denominator in (4.120) gives then, in the neighborhood of τ on the interior arc-side

$$g_i(y(t)) = g_i^{(q_i+k_i^\tau)}(\tau^\pm)\frac{(t-\tau)^{q_i+k_i^\tau}}{(q_i+k_i^\tau)!} + o((t-\tau)^{q_i+k_i^\tau}), \tag{4.121}$$

with $\tau^\pm = \tau^-$ (resp. τ^+) if $\tau \in \mathcal{T}_{en}^i$ (resp. $\tau \in \mathcal{T}_{ex}^i$), and $g_i^{(q_i+k_i^\tau)}(\tau^\pm) := \frac{\mathrm{d}^{q_i+k_i^\tau}}{\mathrm{d}t^{q_i+k_i^\tau}} g_i(y(t))|_{t=\tau^\pm}$ is nonzero by definition of k_i^τ.

Assume now that strict complementarity holds near τ on the boundary arc, in the sense that there exists $\varepsilon > 0$ small such that

$$[\tau, \tau+\varepsilon] \subset \operatorname{supp}(\mathrm{d}\eta_i) \text{ if } \tau \in \mathcal{T}_{en}^i \quad (\text{resp. } [\tau-\varepsilon, \tau] \subset \operatorname{supp}(\mathrm{d}\eta_i) \text{ if } \tau \in \mathcal{T}_{ex}^i). \tag{4.122}$$

Since $g_{i,y}(y)z_v \in W^{q_i,\infty}(0,T)$ by Lemma 4.2, for all critical directions $v \in C(u)$, the first $q_i - 1$ time derivatives of $g_{i,y}(y)z_v$ being continuous vanish at entry/exit of boundary arcs, and hence

the following expansion holds, for t in the neighborhood of τ on the side of the interior arc of g_i:

$$g_{i,y}(y(t))z_v(t) = \mathcal{O}((t-\tau)^{q_i}). \tag{4.123}$$

We thus obtain with (4.121) and (4.123) that there exists a constant $C > 0$ such that

$$|\varsigma_i(\tau)| \ \leq \ \lim_{t\to\tau} C|t-\tau|^{q_i-k_i^\tau}. \tag{4.124}$$

It follows that

$$\varsigma_i(\tau) > -\infty \quad \text{if} \ \ k_i^\tau \leq q_i \qquad \text{and} \qquad \varsigma_i(\tau) = 0 \quad \text{if} \ \ k_i^\tau < q_i. \tag{4.125}$$

Since $k_i^\tau \geq q_\tau - 1$ by Prop. 4.22, and $q_i \geq q_\tau$ whenever τ is an entry or exit point of constraint g_i, it makes sense to assume that $q_\tau - 1 \leq k_i^\tau \leq q_i$. In addition, the continuity of u implies that $k_i^\tau \geq 1$. By (4.125), we see that whenever

$$\max(1, q_\tau - 1) \ \leq \ k_i^\tau \ < \ q_i \tag{4.126}$$

then $\varsigma_i(\tau) = 0$, and hence $\nu_\tau^i \varsigma_i(\tau) = 0$.

Clearly, (4.126) requires that $q_i > 1$. In addition, when (4.90) holds and $q_i = q_\tau$, then it is necessary by Prop. 4.22(ii) that $k_i^\tau \geq q_\tau = q_i$, which is incompatible with (4.126). Therefore, we cannot assume that (4.126) holds when either $q_i = 1$ or (4.90) holds and $q_i = q_\tau$, and will rather assume in that case that

$$k_i^\tau \ = \ q_i. \tag{4.127}$$

By (4.125), assumption (4.127) ensures that $\varsigma_i(\tau)$ is finite. Moreover, if $q_i = 1$, then $\nu_\tau^i = 0$ by Prop. 4.8(ii), implying that $\nu_\tau^i \varsigma_i(\tau) = 0$. If (4.90) holds and $q_i = q_\tau$, then by Prop. 4.22(ii'), we have $\nu_\tau^i = 0$, i.e. $\nu_\tau^i \varsigma_i(\tau) = 0$ again. This shows that boundary arcs have no contribution to the curvature term (4.119) when assumptions (4.122) and (A5)(i) below hold:

(A5) (i) For all junction point $\tau \in \mathcal{T}_i$, $i = 1, \ldots, r$, if τ is an *entry or exit time* of constraint g_i, the function of time $g_i(y(t))$ has order of discontinuity $q_i + k_i^\tau$, and k_i^τ satisfies

$$\begin{cases} (4.127) & \text{if } q_i = 1 \text{ or if (4.90) holds and } q_i = q_\tau, \\ (4.126) & \text{otherwise.} \end{cases}$$

In the case when the junction times of the different components of the state constraints do not coincide (see assumption (A7) in section 4.7), then assumption (A5)(i) has the simpler form (4.202) (see Remark 4.32).

The contribution of touch points to the curvature term (4.119) is classical, when the touch points are reducible, in the following sense. A touch point τ of a component g_i of the state constraint of order $q_i \geq 2$ is said to be *reducible*, if $t \mapsto \frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))$ is continuous at τ, and if

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau} \ < \ 0. \tag{4.128}$$

We will make the assumption that

(A5) (ii) All *essential* touch points of constraint g_i, for all $i = 1, \ldots, r$, are *reducible*, i.e. satisfy (4.128).

Finally, we will also need the following assumption, implying (4.122):

(A6) (i) (Strict complementarity on interior of boundary arcs)

$$\frac{\mathrm{d}\eta_i}{\mathrm{d}t}(t) \ > \ 0, \qquad \text{for a.a. } t \in \operatorname{int}\Delta_i, \qquad \text{for all } i = 1, \ldots, r. \tag{4.129}$$

Let $\mathcal{V} := \mathcal{V}_2 = L^2(0,T;\mathbb{R}^m)$ and $\mathcal{Z} := \mathcal{Z}_2 = H^1(0,T;\mathbb{R}^n)$. Let

$$\hat{T}_{\mathcal{K}}(\mathcal{G}(u)) := \{\omega \in L^2(0,T;\mathbb{R}^s) \ : \ \omega_i \leq 0 \text{ a.e. on } \Delta_i, \ i = r+1, \ldots, r+s\}. \tag{4.130}$$

This is the extension of the tangent cone $T_{\mathcal{K}}(\mathcal{G}(u))$ over L^2. Since $\lambda \in L^\infty(0,T;\mathbb{R}^{r*})$, λ can be extended to a continuous linear form over $L^2(0,T;\mathbb{R}^r)$. We may then consider the extension of the critical cone over L^2 as follows:

$$\hat{C}_{L^2}(u) := \{v \in \mathcal{V} \ : \ DG(u)v \in T_K(G(u)) \cap \eta^\perp, \ D\mathcal{G}(u)v \in \hat{T}_{\mathcal{K}}(\mathcal{G}(u)) \cap \lambda^\perp\}. \tag{4.131}$$

We can now state the no-gap second-order conditions, that do not assume strict complementarity at touch points for the state constraints, and make no additional assumptions for the mixed control-state constraints.

Theorem 4.24. (i) *(Necessary condition) Let (u,y) be a local solution of $(\mathcal{P})$ and (p,η,λ) its (unique) associated multipliers, satisfying (A1)-(A3), (A4)(i), (A5)(i)(ii) and (A6)(i), and $\nu^i_\tau = [\eta_i(\tau)]$. Then*

$$D^2_{uu}L(u;\eta,\lambda)(v,v) - \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}^{i,ess}_{to}} \nu^i_\tau \frac{(g^{(1)}_{i,y}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}} \ \geq \ 0 \quad \forall v \in \hat{C}_{L^2}(u). \tag{4.132}$$

(ii) *(Sufficient condition) Let (u,y) be a stationary point of $(\mathcal{P})$ with multipliers (p,η,λ), satisfying* (4.43)*, and $\nu^i_\tau = [\eta_i(\tau)]$. For $i = 1, \ldots, r$ such that $q_i \geq 2$, let $\mathcal{T}^i_{red}$ denote a* finite *set (possibly empty) of reducible touch points of constraint g_i. If*

$$D^2_{uu}L(u;\eta,\lambda)(v,v) - \sum_{i \,:\, q_i \geq 2} \sum_{\tau \in \mathcal{T}^i_{red}} \nu^i_\tau \frac{(g^{(1)}_{i,y}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}} \ > \ 0 \quad \forall v \in \hat{C}_{L^2}(u) \setminus \{0\}, \tag{4.133}$$

then (u,y) is a local solution of $(\mathcal{P})$ satisfying the quadratic growth condition (4.111)*.*

Note that under (A2)-(A3), $\mathcal{T}^{i,ess}_{to} = \emptyset$ if $q_i \leq 1$. It is easy to obtain from the above theorem a characterization of the quadratic growth.

Corollary 4.25. *Let (u,y) be a stationary point of $(\mathcal{P})$ with multipliers (p,η,λ), satisfying (A1)-(A3), (A4)(i), (A5)(i)(ii) and (A6)(i), and $\nu^i_\tau = [\eta_i(\tau)]$. Then (u,y) is a local solution of $(\mathcal{P})$ satisfying the quadratic growth condition* (4.111) *iff*

$$D^2_{uu}L(u;\eta,\lambda)(v,v) - \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}^{i,ess}_{to}} \nu^i_\tau \frac{(g^{(1)}_{i,y}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}} \ > \ 0 \quad \forall v \in \hat{C}_{L^2}(u) \setminus \{0\}. \tag{4.134}$$

Denote by $Q(v)$ the left-hand side of (4.132) and (4.134). An explicit computation of the Hessian of the Lagrangian $D^2_{uu}L(u;\eta,\lambda)(v,v)$ shows that

$$\begin{aligned} Q(v) \ = \ & \int_0^T H^0_{(u,y),(u,y)}(u,y,p,\lambda)((v,z_v),(v,z_v))\mathrm{d}t + \ \phi_{yy}(y(T))(z_v(T),z_v(T)) \\ & + \sum_{i=1}^{r} \left(\int_0^T g_{i,yy}(y(t))(z_v(t),z_v(t))\mathrm{d}\eta_i(t) - \sum_{\tau \in \mathcal{T}^{i,ess}_{to}} \nu^i_\tau \frac{(g^{(1)}_{i,y}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}} \right). \end{aligned} \tag{4.135}$$

Let us recall that a *Legendre form* Q (see [74]) is a weakly lower semi-continuous quadratic form defined over an Hilbert space, that satisfies the following property: for all weakly convergent sequences (v_n), $(v_n) \rightharpoonup \bar{v}$, we have that $v_n \to \bar{v}$ strongly if $Q(v_n) \to Q(\bar{v})$. An example of a Legendre form is $v \mapsto \|v\|^2$, with $\|\cdot\|$ the norm of the Hilbert space. Under assumption (4.43), it is not difficult to show that (4.135) is a Legendre form (see e.g. [21, Lemma 21][1]). This is no more true if (4.43) is replaced by the weaker hypothesis (4.45).

4.6.2 Proof of Th. 4.24

Denote the *radial cone* to $\mathcal{K}$ at point $x \in \mathcal{K}$ by:

$$\mathcal{R}_{\mathcal{K}}(x) = \{h \in L^\infty \, ; \exists\, \varepsilon_0 > 0,\ x + \varepsilon h \in \mathcal{K},\ \text{for all } 0 < \varepsilon < \varepsilon_0\}. \tag{4.136}$$

Since $\mathcal{K}$ is a closed convex set, $T_{\mathcal{K}}(x) = \mathrm{cl}(\mathcal{R}_{\mathcal{K}}(x))$. Let

$$\begin{array}{rcl} C_0(u) & := & \{v \in C(u),\ DG(u)v|_i(\tau) < 0,\ \text{for all } \tau \in \mathcal{T}_{to}^{nes,i},\ i = 1, \ldots, r, \\ & & \quad D\mathcal{G}(u)v \in \mathcal{R}_{\mathcal{K}}(\mathcal{G}(u))\}. \end{array} \tag{4.137}$$

This subset of the critical cone contains the critical directions that "avoid" nonessential touch points of the state constraint, and such that the derivatives of the mixed constraints belong to the radial cone $\mathcal{R}_{\mathcal{K}}(\mathcal{G}(u))$.

Lemma 4.26. *Under the assumptions of Th. 4.24(i), for all $v \in C_0(u)$, the term* (4.115) *has the expression*

$$\Sigma(u, v) = \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}_{to}^{i,ess}} \nu_\tau^i \frac{(g_{i,y}^{(1)}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}}. \tag{4.138}$$

Proof. It is easy to see that if $D\mathcal{G}(u)v \in \mathcal{R}_{\mathcal{K}}(\mathcal{G}(u))$, then $0 \in T_{\mathcal{K}}^{2,i}(\mathcal{G}(u), D\mathcal{G}(u)v)$. Hence $\sigma(\lambda, T_{\mathcal{K}}^{2,i}(\mathcal{G}(u), D\mathcal{G}(u)v)) = 0$. It remains then in (4.115) the contribution of state constraints. As shown in the previous subsection, when assumptions (A5)(i) and (A6)(i) hold, entry and exit points of boundary arcs of the state constraints have a zero contribution to the curvature term. The term (4.120) for the contribution of essential touch points satisfying (4.128) is computed explicitly, in the same manner as in the scalar case (see [21, Prop. 14][2].). Finally, nonessential touch points do not belong to Δ_i^2 for $v \in C_0(u)$, and hence have no contribution in the sum (4.119). The results follows. □

Lemma 4.27. *Under the assumptions of Th. 4.24(i):*
(i) *The set $C_0(u)$ is dense in $C(u)$.*
(ii) *The set $C(u)$ is dense in the set $\hat{C}_{L^2}(u)$.*

The key point in the proof below is the controllability Lemma 4.3, that enables to handle separately the arguments for the state constraints and for the mixed control-state constraints, in the following way. Under the assumptions of Lemma 4.3, with n_0 the n of (4.25), for all $\kappa \in [1, +\infty]$, there exists a constant $C = C(\kappa) > 0$ such that for all $(w, \omega) \in \mathcal{W}_\kappa \times L^\kappa(0, T; \mathbb{R}^s)$, with

$$\mathcal{W}_\kappa \ := \ \prod_{i=1}^{r} W^{q_i,\kappa}(0, T), \tag{4.139}$$

[1] Lemma 1.21 of this thesis.
[2] Proposition 1.14 of this thesis

there exists $v \in \mathcal{V}_\kappa$ such that

$$\begin{aligned} g_{i,y}(y)z_v &= w_i \quad \text{on } \Delta_i, \quad \forall\, i = 1, \dots, r, && (4.140)\\ c_{i,u}(u,y)v + c_{i,y}(u,y)z_v &= \omega_i \quad \text{a.e. on } \Delta_i^{n_0}, \quad \forall\, i = r+1, \dots, r+s, && (4.141)\\ \|v\|_\kappa &\le C(\|w\|_{\mathcal{W}_\kappa} + \|\omega\|_\kappa). && (4.142) \end{aligned}$$

Proof. (i) Let $v \in C(u)$, and set $w := DG(u)v$ and $\omega := D\mathcal{G}(u)v$. Let φ be a C^∞ function with support in $[-1,1]$ and which is positive on $(-1,1)$. Set $w_{n,i} := w_i - \sum_{\tau \in \mathcal{T}_{to}^{i,nes}} \frac{1}{n^{q_i+1}} \varphi(n(\cdot - \tau))$ for $i = 1, \dots, r$. Then, for n large enough, $w_{n,i}(\tau) < 0$ for all $\tau \in \mathcal{T}_{to}^{i,nes}$, $w_{n,i} = w_i$ outside a neighborhood of $\mathcal{T}_{to}^{i,nes}$, and $\|w_{n,i} - w_i\|_{q_i,\infty} \to 0$ when $n \to +\infty$. Further, since $\mathcal{R}_\mathcal{K}(\mathcal{G}(u)) \cap \lambda^\perp$ in dense in $T_\mathcal{K}(\mathcal{G}(u)) \cap \lambda^\perp$ (see Lemma 4.36 in the Appendix), there exists a sequence $(\omega_n) \subset \mathcal{R}_\mathcal{K}(\mathcal{G}(u)) \cap \lambda^\perp$ such that $\|\omega_n - \omega\|_\infty \to 0$. By the controllability lemma 4.3, there exists $v_n \in \mathcal{U}$ that satisfies (4.140)-(4.141) with (w_n, ω_n), and $\|v_n - v\|_\infty \le C(\|w_n - w\|_{\mathcal{W}_\infty} + \|\omega_n - \omega\|_\infty)$. By construction it follows that $v_n \in C_0(u)$, and $v_n \to v$ in L^∞.
(ii) Let $v \in \hat{C}_{L^2}(u)$, and again let $w := DG(u)v$ and $\omega := D\mathcal{G}(u)v$. By Lemmas 16-17 in [21][3] (this is where assumption (A6)(i) is used), we can construct a sequence $(w_n) \subset \prod_{i=1}^r W^{q_i,\infty}(0,T)$ such that $w_{n,i} = 0 = w_i$ on each boundary arc of g_i, $i = 1, \dots, r$, $w_{n,i}(\tau) = w_i(\tau)$ at each touch point $\tau \in \mathcal{T}_i$, and $\|w_{n,i} - w_i\|_{q_i,2} \to 0$. So $w_n \in T_K(G(u)) \cap \eta^\perp$. Now by Lemma 4.37 in the Appendix, there exists a sequence $(\omega_n) \subset T_\mathcal{K}(\mathcal{G}(u)) \cap \lambda^\perp$ such that $\|\omega_n - \omega\|_2 \to 0$. By Lemma 4.3 again, there exists $v_n \in \mathcal{U}$ that satisfies (4.140)-(4.141) with (w_n, ω_n) and $\|v_n - v\|_2 \le C(\|w_n - w\|_{\mathcal{W}_2} + \|\omega_n - \omega\|_2)$. By construction we have $v_n \in C(u)$, and $v_n \to v$ in L^2. □

Proof of Th. 4.24. For the necessary condition, we use the abstract condition (4.112) and compute the curvature term (4.115). By Lemma 4.26, we have the expression of the curvature term for all $v \in C_0(u)$. Since the right-hand side of (4.138) is continuous for the norm of L^2, we obtain the result by a density argument in view of Lemma 4.27.

For the sufficient condition, we follow [21, Th. 18 and 27][4]. The idea is to use a reduction approach, i.e. to reformulate the state constraint around finitely many reducible touch points of the components g_i of the state constraint of order $q_i \ge 2$. More precisely, for $\mathcal{T}_{red}^i := \{\tau_1^i, \dots, \tau_{N_i}^i\}$, $\varepsilon, \delta > 0$ small enough, and $\Omega_i := [0,T] \setminus \cup_{k=1}^{N_i} (\tau_k^i - \varepsilon, \tau_k^i + \varepsilon)$, the constraint $G(u') \in K$ in (4.5) can be equivalently replaced, for all $\|u' - u\|_\infty \le \delta$, by

$$g_i(y_{u'}(t)) \le 0 \ \text{ for all } t \in \Omega_i \ \text{ and } \ g_i(y_{u'}(t_k^i(u'))) \le 0, \quad k = 1, \dots, N_i, \ \forall i : q_i \ge 2 \qquad (4.143)$$

where $t_k^i(u')$ is the unique point of maximum of the function $g_i(y_{u'}(\cdot))$ over the set $(\tau_k^i - \varepsilon, \tau_k^i + \varepsilon)$. The Hessian of the Lagrangian of the reduced problem is equal to the quadratic form $Q(v)$, i.e. has an additional term that matches the curvature term. Now assume that (4.111) does not hold. Then there exists a sequence (u_n), $u_n \to u$ in L^∞, satisfying the constraints (4.143) and $\mathcal{G}(u_n) \in \mathcal{K}$, and such that

$$J(u_n) \le J(u) + o(\|u_n - u\|_2^2). \qquad (4.144)$$

Set $\varepsilon_n := \|u_n - u\|_2$ and $v_n := \varepsilon_n^{-1}(u_n - u)$. Being bounded in L^2, assume that $v_n \rightharpoonup v$ weakly in L^2. By (4.144), a second-order expansion of the Lagrangian of the reduced problem shows that

$$Q(v_n) \le o(1). \qquad (4.145)$$

[3] Lemmas 1.16 and 1.17 of this thesis.
[4] Theorems 1.18 and 1.27 of this thesis.

Moreover, since

$$\mathcal{K} \ni \mathcal{G}(u_n) = \mathcal{G}(u) + \varepsilon_n D\mathcal{G}(u)v_n + \varepsilon_n r_n$$

with $\|r_n\|_2 \to 0$, we deduce that $D\mathcal{G}(u)v_n + r_n \in \hat{T}_{\mathcal{K}}(\mathcal{G}(u))$. Taking the weak limit in L^2, we obtain that $D\mathcal{G}(u)v \in \hat{T}_{\mathcal{K}}(\mathcal{G}(u))$. Proceeding similarly for the state constraints, and since as a consequence of (4.144), we have $DJ(u)v \leq 0$, we deduce that $v \in \hat{C}_{L^2}(u)$. It follows then from (4.133) and (4.145), since Q is weakly lower semi-continuous, that $Q(v) = 0$, and hence, $Q(v_n) \to Q(v)$. Since Q is a Legendre form by hypothesis (4.43), this implies that $v_n \to v$ strongly, contradicting that $\|v_n\|_2 = 1$ for all n. This completes the proof. □

4.7 The shooting algorithm

In presence of state constraints, a reformulation of the optimality conditions is needed to apply so-called shooting methods. For an overview of the different formulations of optimality conditions existing in the literature, see the survey by Hartl et al. [68]. The shooting algorithm takes only into account a part of the optimality conditions, and the remainder conditions, referred as "additional conditions", have to be checked afterwards. In this section, we first recall the alternative formulation used in the shooting algorithm (Def. 4.28). Additional conditions are given, under which the alternative formulation is equivalent to the first-order optimality condition of $(\mathcal{P})$ (Prop. 4.29). It is shown that some of those additional conditions are automatically satisfied (Lemma 4.30). Finally we give a characterization of the well-posedness of the shooting algorithm (Th. 4.33), which is the main result of this section.

Given a finite subset $\mathcal{S}$ of $(0,T)$, we denote by $PC^k_{\mathcal{S}}[0,T]$ the set of functions over $[0,T]$ that are of class C^k outside $\mathcal{S}$ and have, as well as their first k derivatives, a left and a right limit over $\mathcal{S}$ and a left (resp. right) limit at T (resp. 0).

4.7.1 Shooting Formulation

The formulation for the shooting algorithm presented in this section was introduced by Bryson et al. [29]. The presence of additional conditions was first underlined by Jacobson, Lele and Speyer [75], see also Kreindler [83]. See an example of implementation in e.g. [107] and numerical applications in e.g. [30, 26].

Recall that H^q denotes the alternative Hamiltonian (4.51). We assume in the sequel that assumptions (A2)-(A4)(i) hold, and that first-order components of the state constraint do not have touch points (which is typically satisfied in view of Prop. 4.8(ii), since first-order components of the state constraint only have nonessential touch points). We assume in addition that

(A4) (ii) Each component of the mixed control-state constraint $c_i(u,y)$, $i = r+1,\dots,r+s$, has *finitely many* boundary arcs, and no touch points.

Under (A4) (which stands for (A4)(i)(ii)), we denote by $\mathcal{I}^i_b$ the closure of the union of boundary arcs of each constraint $i = 1,\dots,r+s$, i.e. $\mathcal{I}^i_b := \cup_{k=1}^{N^i_b}[\tau^{i,k}_{en}, \tau^{i,k}_{ex}]$ for $\mathcal{T}^i_{en} := \{\tau^{i,1}_{en} < \dots < \tau^{i,N^i_b}_{en}\}$ and a similar definition of $\mathcal{T}^i_{ex}$.

In the alternative formulation presented in Def. 4.10, the integration constants in (4.49) on a boundary arc of g_i are arbitrary. In the sequel, we will choose like in [98] these constants, on each boundary arc $(\tau^i_{en}, \tau^i_{ex})$ of g_i, such that the functions η^j_i for $i = 1,\dots,r$ and $j = 1,\dots,q_i$

satisfy, for $t \in (\tau^i_{en}, \tau^i_{ex})$,

$$\eta^1_i(t) \;:=\; \eta_i(\tau^{i+}_{ex}) - \eta_i(t), \qquad \eta^j_i(t) \;:=\; \int_t^{\tau^i_{ex}} \eta^{j-1}_i(\sigma)\mathrm{d}\sigma, \quad j = 2, \ldots, q_i,$$

and we still have $\eta^j_i = 0$ outside boundary arcs of g_i and $\eta^0_i = \lambda_i$ for $i = r+1, \ldots, r+s$. With this formulation, the alternative costate p_q is continuous at exit points and discontinuous at entry and touch points, which allows to take the jump parameters $\nu^{i,j}_\tau$ and ν^i_τ involved in the jump condition (4.154) as *shooting parameters* in the shooting algorithm.

Definition 4.28. A trajectory (u, y) having a finite set of junction times $\mathcal{T} = \cup_{i=1}^{r+s}\mathcal{T}_i$ satisfies the *alternative formulation*, if there exist $p^q \in PC^{q_{max}}_{\mathcal{T}}([0,T];\mathbb{R}^{n*})$, $\eta^q \in PC^{q_{max}}_{\mathcal{T}}([0,T];\mathbb{R}^{(r+s)*})$, and, for each $i = 1, \ldots, r$, for each entry time τ of g_i, there exist q_i jump parameters $(\nu^{i,j}_\tau)_{1\le j\le q_i}$ and for each touch point τ of g_i with $q_i \ge 2$, there exists a jump parameter ν^i_τ, such that the following relations are satisfied (dependence in time is omitted):

$$\begin{aligned}
\dot{y} &= f(u,y) \quad \text{on } [0,T]; \qquad y(0) = y_0 && (4.146)\\
-\dot{p}^q &= H^q_y(u,y,p^q,\eta^q) \;\text{ on } [0,T]\setminus\mathcal{T} && (4.147)\\
0 &= H^q_u(u,y,p^q,\eta^q) \;\text{ on } [0,T]\setminus\mathcal{T} && (4.148)\\
g_i^{(q_i)}(u(t),y(t)) &= 0 \qquad \text{on } \mathcal{I}^i_b, \quad i = 1,\ldots,r+s && (4.149)\\
\eta^{q_i}_i(t) &= 0 \qquad \text{on } [0,T]\setminus\mathcal{I}^i_b, \quad i = 1,\ldots,r+s && (4.150)\\
p^q(T) &= \phi_y(y(T)), && (4.151)
\end{aligned}$$

and, for all $i = 1, \ldots, r$ and each junction point $\tau \in \mathcal{T}^i$ of g_i:

$$\begin{aligned}
g_i^{(j)}(y(\tau)) &= 0 \qquad \text{if } \tau \in \mathcal{T}^i_{en}, \quad j = 0,\ldots,q_i-1, && (4.152)\\
g_i(y(\tau)) &= 0 \qquad \text{if } \tau \in \mathcal{T}^i_{to}, && (4.153)
\end{aligned}$$

and for each junction time $\tau \in \mathcal{T}$:

$$[p^q(\tau)] \;=\; -\sum_{i\le r\,:\,\tau\in\mathcal{T}^i_{en}} \sum_{j=1}^{q_i} \nu^{i,j}_\tau g^{(j-1)}_{i,y}(y(\tau)) - \sum_{i\le r\,:\,\tau\in\mathcal{T}^i_{to}} \nu^i_\tau g_{i,y}(y(\tau)). \qquad (4.154)$$

The shooting algorithm consists in finding a zero of a finite-dimensional shooting mapping, using e.g. a Newton method. The *structure* of active constraints of the optimal trajectory, i.e. the number and order of boundary arcs and touch points of each component of the constraint, is assumed to be known (or guessed). The arguments of the shooting mapping are called the *shooting parameters*, and are composed of the initial value of costate $p^0 \in \mathbb{R}^{n*}$, all the junction times (with the exception of nonessential touch points) of the pure state constraints and mixed control-state constraints, and all the jump parameters $\nu^{i,j}_\tau$ at entry times τ of g_i, $i = 1, \ldots, r$, $j = 1, \ldots, q_i$ and ν^i_τ at touch points τ of g_i, $i = 1, \ldots, r$, $q_i \ge 2$, that are involved in the jump condition of the costate (4.154).

By assumptions (A2)-(A3), the algebraic variable $(u(t), \eta^q(t)) \in \mathbb{R}^m \times \mathbb{R}^{(r+s)*}$ satisfying (4.148)-(4.150) can be expressed as implicit function of the differential variables $(y(t), p^q(t)) \in \mathbb{R}^n \times \mathbb{R}^{n*}$ on the interior of each arc of the trajectory (see the proof of Prop. 4.13). With a given set of shooting parameters is therefore associated at most a unique solution (u, y, p^q, η^q) of the Cauchy problem (4.146)-(4.147) with initial condition of the costate $p^q(0) = p^0$, the algebraic variable (u, η^q) satisfying (4.148)-(4.150) and the jump of p^q at junction times of pure state constraints being given by (4.154).

The shooting mapping is then defined as follows. With a given set of shooting parameters are associated the following conditions: the final condition (4.151), the interior point conditions (4.152)-(4.153), and the optimality conditions for junction times below, for all $\tau \in \mathcal{T}$ and all $i = 1, \ldots, r+s$:

$$
\begin{aligned}
g_i^{(q_i)}(u(\tau^-), y(\tau)) &= 0, \qquad \text{if } \tau \in \mathcal{T}_{en}^i && (4.155)\\
g_i^{(q_i)}(u(\tau^+), y(\tau)) &= 0, \qquad \text{if } \tau \in \mathcal{T}_{ex}^i && (4.156)\\
g_i^{(1)}(y(\tau)) &= 0, \qquad \text{if } \tau \in \mathcal{T}_{to}^i \text{ and if } q_i \geq 2. && (4.157)
\end{aligned}
$$

This is a mapping defined on a subset of $\mathbb{R}^{\bar{N}}$ to $\mathbb{R}^{\bar{N}}$, where $\bar{N}$ the dimension of the shooting mapping is as follows. Let N_{ba}^i be the total number of boundary arcs of constraints g_i for $i = 1, \ldots, r$ and c_i for $i = r+1, \ldots, r+s$, and N_{to} the total number of touch points of state constraints of order $q_i \geq 2$. Then

$$
\bar{N} = n + \sum_{i=1}^{r+s} (q_i + 2) N_{ba}^i + 2N_{to}. \tag{4.158}
$$

4.7.2 Additional Conditions

It is of importance to check whether solutions of the shooting algorithm (i.e. trajectory associated with a zero of the shooting function) are stationary points of $(\mathcal{P})$. For this, we need to make explicit the relation between the multipliers in the alternative formulation (Def. 4.28) and in Th. 4.5.

Given alternative multipliers (p^q, η^q) and jump parameters $(\nu_\tau^{i,j})$ at entry times and (ν_τ^i) at touch times, the related multipliers (p, η, λ) in Th. 4.5 are given by the following relations. Define first

$$
\eta_i^j(t) = (-1)^{q_i - j} \frac{\mathrm{d}^{q_i - j}}{\mathrm{d}t^{q_i - j}} \eta_i^{q_i}(t), \ j = 0, \ldots, q_i - 1, \ i = 1, \ldots, r, \ t \notin \mathcal{T}, \tag{4.159}
$$

then

$$
\begin{aligned}
\lambda_i(t) &= \eta_i^0(t), \qquad i = r+1, \ldots, r+s, \qquad t \notin \mathcal{T} && (4.160)\\
p(t) &= p^q(t) + \sum_{i=1}^{r} \sum_{j=1}^{q_i} \eta_i^j(t) g_{i,y}^{(j-1)}(y(t)), \qquad t \notin \mathcal{T}. && (4.161)
\end{aligned}
$$

Finally, let

$$
\mathrm{d}\eta_i(t) = \eta_i^0(t)\mathrm{d}t + \sum_{\tau \in \mathcal{T}} \nu_\tau^i \delta_\tau(t), \qquad i = 1, \ldots, r, \tag{4.162}
$$

where $\delta_\tau(t)$ denotes the Dirac measure at time τ, and the jumps parameters ν_τ^i at junction points $\tau \in \mathcal{T}$, for all $i = 1, \ldots, r$, are the ones in the alternative formulation if $\tau \in \mathcal{T}_{to}^i$, $\nu_\tau^i = 0$ if $i \notin I(\tau)$, and, if $\tau \in \mathcal{I}_b^i$, they are given by, in view of (4.88) and (4.154),

$$
\begin{aligned}
\nu_\tau^i &= \nu_\tau^{i,1} - \eta_i^1(\tau^+) \qquad \text{if } \tau \in \mathcal{T}_{en}^i, && (4.163)\\
\nu_\tau^i &= \eta_i^1(\tau^-) \qquad \text{if } \tau \in \mathcal{T}_{ex}^i, && (4.164)\\
\nu_\tau^i &= -[\eta_i^1(\tau)] \qquad \text{if } \tau \in \operatorname{int} \mathcal{I}_b^i. && (4.165)
\end{aligned}
$$

Conversely, Prop. 4.13 ensures, whenever assumptions (A2)-(A4) are satisfied, that each component η_i of η admits a (unique) decomposition under the form (4.162). Therefore, classical

multipliers (p, η, λ) of Th. 4.5 uniquely determine the alternative multipliers and alternative jump parameters so that (4.159)-(4.165) as well as (4.170), (4.172), (4.174) below hold, these three last conditions being needed in order to fix the integration constants in (4.49) and the jumps parameters at entry times $(\nu_\tau^{i,j})$ for $j \geq 2$.

The *additional conditions* needed to obtain the equivalence between the alternative formulation (4.146)-(4.154) and the first-order optimality condition (4.35)-(4.40) are the following:

$$g_i(y(t)) < 0 \quad \text{on } [0,T] \setminus (\mathcal{I}_b^i \cup \mathcal{T}_{to}^i), \quad \text{for all } i = 1, \dots, r, \tag{4.166}$$

$$c_i(u(t), y(t)) < 0 \quad \text{a.e. on } [0,T] \setminus \mathcal{I}_b^i, \quad \text{for all } i = r+1, \dots, r+s, \tag{4.167}$$

$$(-1)^{q_i} \frac{\mathrm{d}^{q_i}}{\mathrm{d}t^{q_i}} \eta_i^{q_i}(t) \geq 0 \quad \text{on } \operatorname{int} \mathcal{I}_b^i, \quad \text{for all } i = 1, \dots, r+s, \tag{4.168}$$

and, for all $\tau \in \mathcal{T}$ and all $i = 1, \dots, r$:

$$\nu_\tau^{i,1} - \eta_i^1(\tau^+) \geq 0, \qquad \text{if } \tau \in \mathcal{T}_{en}^i \tag{4.169}$$

$$\nu_\tau^{i,j} - \eta_i^j(\tau^+) = 0, \qquad \text{if } \tau \in \mathcal{T}_{en}^i, \ j = 2, \dots, q_i \tag{4.170}$$

$$\eta_i^1(\tau^-) \geq 0, \qquad \text{if } \tau \in \mathcal{T}_{ex}^i \tag{4.171}$$

$$\eta_i^j(\tau^-) = 0, \qquad \text{if } \tau \in \mathcal{T}_{ex}^i, \ j = 2, \dots, q_i \tag{4.172}$$

$$[\eta_i^1(\tau)] \leq 0, \qquad \text{if } \tau \in \operatorname{int} \mathcal{I}_b^i, \tag{4.173}$$

$$[\eta_i^j(\tau)] = 0, \qquad \text{if } \tau \in \operatorname{int} \mathcal{I}_b^i, \ j = 2, \dots, q_i \tag{4.174}$$

$$\nu_\tau^i \geq 0, \qquad \text{if } \tau \in \mathcal{T}_{to}^i, \tag{4.175}$$

$$\begin{gathered}\text{For all } i \text{ such that } q_i = 1, \text{ the inequalities (4.169), (4.171), (4.173)} \\ \text{and (4.175) are equalities.}\end{gathered} \tag{4.176}$$

Proposition 4.29. *Let (u, y) be a trajectory satisfying (A2)-(A4). Then (u, y) is a stationary point, with multipliers (p, η, λ), iff (u, y) satisfies both the alternative formulation (Def. 4.28) and the additional conditions* (4.166)-(4.176)*. Relations* (4.159)-(4.165) *and* (4.170), (4.172), (4.174) *are a one to one mapping between the multipliers (p, η, λ) involved in the first-order optimality condition of Th. 4.5, and the alternative multipliers (p^q, η^q) and alternative jumps parameters $(\nu_\tau^{i,j})$ and (ν_τ^i) at respectively entry and touch points in the alternative formulation and additional conditions.*

The higher the order q_i of the constraint is, the more additional conditions have to be checked at regular entry/exit points of boundary arcs. Those conditions are analogous to the known conditions in the scalar case, with in addition the conditions (4.173)-(4.174), that were not apparent in the scalar case, and to our knowledge not known in the literature. Thus, when assumptions (A2)-(A3) hold, we are led to think that, like in the scalar case, boundary arcs with regular entry/exit times for components of the state constraint of order $q_i \geq 3$ may occur only in degenerate situations. We underline that this was not, however, an immediate result, since now we allow more control variables (more than one) and hence, more degrees of freedom.

Proof of Prop. 4.29. Let us show the equivalence between, on the one hand, the first-order optimality system of $(\mathcal{P})$ (4.35)-(4.40), and on the other hand, the alternative formulation (4.146)-(4.154) and the additional conditions (4.166)-(4.176).

First, $g_i(y(t)) \leq 0$ in (4.39) is equivalent to $g_i(y(t)) = 0$ on $\mathcal{I}_b^i$, (4.153) at touch points and (4.166) outside the contact set, and then $g_i(y(t)) = 0$ on $\mathcal{I}_b^i$ is equivalent to (4.149) for

$i = 1, \dots, r$ with the q_i entry-point conditions (4.152). By Prop. 4.13, the state constraint multipliers η_i, $i = 1, \dots, r$ are regular on interiors of arcs, therefore, each component η_i can be put into the form (4.162), where jumps can occur only at junctions points, and the density of each component η_i^0 is continuous on the interior of arcs. It follows that η_i is a nonnegative measure ($\mathrm{d}\eta_i \geq 0$ in (4.39)), iff its density $\frac{\mathrm{d}\eta_i}{\mathrm{d}t}(t) = \eta_i^0(t) = (-1)^{q_i} \frac{\mathrm{d}^{q_i}}{\mathrm{d}t^{q_i}} \eta_i^{q_i}(t)$ is nonnegative, i.e. iff (4.168) holds for $i = 1, \dots, r$, and the jumps at junction times are nonnegative, i.e.

$$\nu_\tau^i \;=\; [\eta_i(\tau)] \;\geq\; 0, \quad \text{for all } i = 1, \dots, r \;\text{ and all }\; \tau \in \mathcal{T} = \cup_{i=1}^{r+s} \mathcal{T}^i. \tag{4.177}$$

The complementarity condition $\int_0^T g_i(y(t))\mathrm{d}\eta_i(t) = 0$ in (4.39) is then equivalent to (4.150) for $i = 1, \dots, r$ (the measure $\mathrm{d}\eta_i$ has support on the contact set of $g_i(y)$). Similarly, for mixed control-state constraints, since $\lambda \in L^\infty$, (4.40) is equivalent to (4.149)-(4.150) and (4.167)-(4.168) for $i = r+1, \dots, r+s$.

The state equations (4.35) and (4.146) are of course identical, and so are the final conditions of the costate (4.37) and (4.151) in view of (A4)(i). By Lemma 4.11, the costate and control equations (4.147) and (4.148) are equivalent, respectively, to the costate and control equations (4.36) and (4.38) *on the interior of arcs*. Now let us show the equivalence, at junction times, between on the one hand the costate equation (4.36) and (4.177), and on the other hand the jump condition (4.154) and the additional conditions (4.169)-(4.175). By (4.88) (recall that $[p(\tau)] = -\sum_{i \in I(\tau)} \nu_\tau^i g_{i,y}(y(\tau))$ with $\nu_\tau^i = [\eta_i(\tau)]$) and by (4.154), it holds respectively

$$[p^q(\tau)] \;=\; -\sum_{i \in I(\tau)} \{(\nu_\tau^i + [\eta_i^1(\tau)]) g_{i,y}(y(\tau)) + \sum_{j=2}^{q_i} [\eta_i^j(\tau)] g_{i,y}^{(j-1)}(y(\tau))\} \tag{4.178}$$

$$[p^q(\tau)] \;=\; -\sum_{i \leq r \,:\, \tau \in \mathcal{T}_{en}^i} \sum_{j=1}^{q_i} \nu_\tau^{i,j} g_{i,y}^{(j-1)}(y(\tau)) - \sum_{i \leq r \,:\, \tau \in \mathcal{T}_{to}^i} \nu_\tau^i g_{i,y}(y(\tau)). \tag{4.179}$$

By Corollary 4.18, the vectors $g_{i,y}^{(j-1)}(y(\tau))$ are linearly independent, for all $i \in I(\tau)$ and $j = 1, \dots, q_i$, hence the relations (4.178)-(4.179) are equal, iff the coefficients of $g_{i,y}^{(j-1)}(y(\tau))$ are equal. We thus obtain, for all $\tau \in \mathcal{T}$ and $i \in I(\tau)$, if $\tau \in \mathcal{T}_{en}^i$:

$$\nu_\tau^i + [\eta_i^1(\tau)] \;=\; \nu_\tau^{i,1} \qquad \text{and} \qquad [\eta_i^j(\tau)] \;=\; \nu_\tau^{i,j}, \quad j = 2, \dots, q_i$$

which, with (4.177), is equivalent to (4.169)-(4.170), using that $\eta_i^j(\tau^-) = 0$ at entry point. If now $\tau \in \mathcal{T}_{to}^i$, we obtain, since the multipliers η_i^j are equal to zero in the neighborhood of τ:

$$[\eta_i(\tau)] \;=\; \nu_\tau^i,$$

which, with (4.177), is equivalent to (4.175). Finally, if $\tau \in \operatorname{int} \mathcal{T}_b^i$ or if $\tau \in \mathcal{T}_{ex}^i$, then we have

$$[\eta_i(\tau)] + [\eta_i^1(\tau)] \;=\; 0 \qquad \text{and} \qquad [\eta_i^j(\tau)] \;=\; 0, \quad j = 2, \dots, q_i$$

which, with (4.177) again, is equivalent to (4.173)-(4.174) on interior of boundary arcs and to (4.171)-(4.172) at exit points, since $\eta_i^j(\tau^+) = 0$. Finally, whenever $q_i = 1$, then we know by Prop. 4.8 that η_i is continuous, i.e. $[\eta_i(\tau)] = 0$, and therefore all inequalities in (4.169)-(4.175) are in fact equalities. □

Like in the scalar case, the conditions (4.155)-(4.156) imposed in the shooting algorithm, related to the continuity of u, imply that some of the additional conditions are automatically satisfied by a solution of the shooting algorithm.

Lemma 4.30. *Let (u,y) satisfy the alternative formulation* (4.146)-(4.154)*, the strong assumption* (4.44) *and (A3)-(A4), and assume that $\mathcal{T}^i_{to}=\emptyset$, for all i such that $q_i=1$. Then the following assertions are equivalent:*
(i) *For all $i=1,\dots,r$ and all junction point $\tau\in\mathcal{T}$, if $q_i=1$ the additional conditions* (4.169)*,* (4.171) *and* (4.173) *are satisfied with equality and if $q_i\geq 2$, the additional conditions in* (4.170)*,* (4.172) *and* (4.174) *are satisfied for $j=q_i$, i.e.*

$$\begin{aligned}
\nu^{i,q_i}_\tau &= \eta^{q_i}_i(\tau^+), && \text{if } \tau\in\mathcal{T}^i_{en}, && (4.180)\\
\eta^{q_i}_i(\tau^-) &= 0, && \text{if } \tau\in\mathcal{T}^i_{ex}, && (4.181)\\
[\eta^{q_i}_i(\tau)] &= 0, && \text{if } \tau\in \operatorname{int}\mathcal{I}^i_b, && (4.182)
\end{aligned}$$

and for all $i=r+1,\dots,r+s$, $\eta^{q_i}_i=\lambda_i$ is continuous over $[0,T]$.
(ii) *The conditions* (4.155)-(4.156) *are satisfied, for all $\tau\in\mathcal{T}$ and all $i=1,\dots,r+s$.*
(iii) *The control u is continuous over $[0,T]$.*

Proof. Let $\tau\in\mathcal{T}$, and let $J:=I(\tau)\setminus\{i=1,\dots,r;\ \tau\in\mathcal{T}^i_{to}\}$. Set $u^\pm:=u(\tau^\pm)$, $[u]:=u^+-u^-$, and, for $\sigma\in[0,1]$, $u^\sigma:=u^-+\sigma(u^+-u^-)$. Similar notations for p^q, η^q are used. Denote by $\tilde\nu^q=(\tilde\nu^{q_i}_i)_{i\in J}$ the augmented (row) vector of jump parameters, satisfying $\tilde\nu^{q_i}_i=\nu^{i,q_i}_\tau$ for all $i\in J$ such that $\tau\in\mathcal{T}^i_{en}$ and $q_i\geq 1$, and $\tilde\nu^{q_i}_i=0$ for all $i\in J$ such that $\tau\in\operatorname{int}\mathcal{I}^i_b\cup\mathcal{T}^i_{ex}$ or $q_i=0$. By (4.148),

$$H^q_u(u^+,y(\tau),p^{q+},\eta^{q+}) \;=\; 0 \;=\; H^q_u(u^-,y(\tau),p^{q-},\eta^{q-}).$$

The alternative Hamiltonian H^q being affine in the variables p^q and η^q, we have

$$\begin{aligned}
0 \;=\; &\int_0^1 \{\sigma H^q_{uu}(u^\sigma,y(\tau),p^{q+},\eta^{q+})+(1-\sigma)H^q_{uu}(u^\sigma,y(\tau),p^{q-},\eta^{q-})\}[u]\mathrm{d}\sigma\\
&+\int_0^1 \{[p^q]f_u(u^\sigma,y(\tau))+[\eta^q]G^{(q)}_{J,u}(u^\sigma,y(\tau))\}\mathrm{d}\sigma.
\end{aligned} \tag{4.183}$$

Using the jump of p^q given by (4.154), and the fact that by hypothesis, first-order components of the state constraint do not have touch points, we easily get that

$$[p^q]f_u(u^\sigma,y(\tau))+[\eta^q]G^{(q)}_{J,u}(u^\sigma,y(\tau)) \;=\; ([\eta^q]-\tilde\nu^q)G^{(q)}_{J,u}(u^\sigma,y(\tau)). \tag{4.184}$$

In addition, (4.44) and (4.53) imply that $H^q_{uu}(u^\sigma,y,p^{q\pm},\eta^{q\pm})$ is uniformly positive definite, for all $\sigma\in[0,1]$, therefore, multiplying on the right (4.183) by $[u]$, and using (4.184), we obtain that

$$\alpha|[u]|^2 \;\leq\; (\tilde\nu^q-[\eta^q])\int_0^1 G^{(q)}_{J,u}(u^\sigma,y(\tau))[u]\mathrm{d}\sigma. \tag{4.185}$$

Note that point (i) is equivalent to the condition $[\eta^{q_i}_i]-\tilde\nu^{q_i}_i=0$ for all $i=1,\dots,r+s$. Therefore, the implication (i) $\Rightarrow$ (iii) follows from (4.185). Conversely, if (iii) holds, i.e. $[u]=0$, then (4.183)-(4.184) yields

$$([\eta^q]-\tilde\nu^q)G^{(q)}_{J,u}(u(\tau),y(\tau))=0,$$

implying (i) by (4.30). This shows the equivalence (iii) $\Leftrightarrow$ (i). Let us show now (iii) $\Leftrightarrow$ (ii). The implication (iii) $\Rightarrow$ (ii) is trivial. If (ii) holds, then

$$0 \;=\; G^{(q)}_J(u^+,y(\tau))-G^{(q)}_J(u^-,y(\tau)) \;=\; \int_0^1 G^{(q)}_{J,u}(u^\sigma,y(\tau))[u]\mathrm{d}\sigma. \tag{4.186}$$

By (4.185), it follows that $[u]=0$, i.e. (iii) holds, which completes the proof. □

4.7.3 Well-posedness of the shooting algorithm

We say that the shooting algorithm is (locally) well-posed in the neighborhood of a local solution, if the Jacobian of the shooting mapping is invertible. This allows us to apply locally a Newton method in order to find a zero of the shooting mapping with a very high precision, and low cost. If the additional conditions (4.166)-(4.176) are satisfied, we obtain a stationary point of $(\mathcal{P})$, and if the second-order sufficient condition (4.134) holds, we obtain a local solution of $(\mathcal{P})$.

The first step to study the well-posedness of the shooting algorithm is to compute the Jacobian of the shooting mapping. We denote by π^0 the variation of p^0, σ_τ^i the variation of τ for each $\tau \in \mathcal{T}^i$, $i = 1, \dots, r+s$, $\gamma_\tau^{i,j}$ the variations of alternative jump parameters at entry times $\nu_\tau^{i,j}$ for $\tau \in \mathcal{T}_{en}^i$, $i = 1, \dots, r$, $j = 1, \dots, q_i$, and γ_τ^i the variations of jump parameters at touch times ν_τ^i for $\tau \in \mathcal{T}_i^{to}$, $i = 1, \dots, r$ and $q_i \geq 2$. All of them will be called *variations of shooting parameters*.

Given a vector $\zeta \in \mathbb{R}^{(r+s)*}$ and $J := \{i_1 < \cdots < i_s\} \subset \{1, \dots, r+s\}$, the vector ζ_J denotes the row vector of component $(\zeta_{i_1}, \dots, \zeta_{i_s})$. We denote by $\bar{I}(t)$ the complement of $I(t)$ in $\{1, \dots, r+s\}$. With a set of variation of shooting parameters is associated a (unique by (A2)-(A3)) linearized trajectory and multipliers (z, v, π^q, ζ^q) solution of (arguments (u, y, p^q, η^q) and time are omitted):

$$
\begin{aligned}
\dot{z} &= f_y z + f_u v \quad \text{on } [0,T] \text{ a.e.;} \qquad z(0) = 0 && (4.187)\\
\dot{\pi}^q &= -(H_{yy}^q z + H_{yu}^q v + \pi^q f_y + \zeta^q G_y^{(q)}) \quad \text{on } [0,T] \setminus \mathcal{T} \text{ a.e.} && (4.188)\\
\pi^q(0) &= \pi^0 && (4.189)\\
0 &= H_{uy}^q z + H_{uu}^q v + \pi^q f_u + \zeta^q G_u^{(q)} \quad \text{on } [0,T] \setminus \mathcal{T} \text{ a.e.} && (4.190)\\
0 &= G_{I(t),u}^{(q)} v + G_{I(t),y}^{(q)} z \quad \text{on } [0,T] \setminus \mathcal{T} \text{ a.e.} && (4.191)\\
0 &= \zeta_{\bar{I}(t)}^q \qquad \text{on } [0,T] \setminus \mathcal{T} \text{ a.e.} && (4.192)
\end{aligned}
$$

and, for all $\tau \in \cup_{i=1}^r \mathcal{T}^i$, setting $\nu_\tau^{i,0} := 0$ for $\tau \in \mathcal{T}_{en}^i$:

$$
\begin{aligned}
[\pi^q(\tau)] \;=\; & -\sum_{i \leq r \,:\, \tau \in \mathcal{T}_{en}^i} \sum_{j=1}^{q_i} \{\nu_\tau^{i,j} g_{i,yy}^{(j-1)}(y(\tau))z(\tau) + (\gamma_\tau^{i,j} + \sigma_\tau^i \nu_\tau^{i,j-1}) g_{i,y}^{(j-1)}(y(\tau))\} \\
& - \sum_{i \leq r \,:\, \tau \in \mathcal{T}_{to}^i} \{\nu_\tau^i g_{i,yy}(y(\tau))z(\tau) + \gamma_\tau^i g_{i,y}(y(\tau)) + \sigma_\tau^i \nu_\tau^i g_{i,y}^{(1)}(y(\tau))\}.
\end{aligned} \tag{4.193}
$$

Lemma 4.31. *Let (u, y, p^q, η^q) be the trajectory associated with a zero of the shooting mapping, and assume that (A2)-(A4) hold and that $\mathcal{T}_{to}^i = \emptyset$ for all i such that $q_i = 1$. Let π^0, (σ_τ^i), $(\gamma_\tau^{i,j})$, and (γ_τ^i) be a set of variations of shooting parameters and denote by (z, v, π^q, ζ^q) the linearized trajectory and multipliers solution of* (4.187)-(4.193)*. Then this set of variations of shooting parameters belongs to the kernel of the Jacobian of the shooting mapping, iff:*

$$
\pi^q(T) \;=\; \phi_{yy}(y(T))z(T), \tag{4.194}
$$

and, for all junction time $\tau \in \mathcal{T}$ *and all* $i = 1, \dots, r+s$*:*

$$0 = g^{(j)}_{i,y}(y(\tau))z(\tau) \qquad \text{if } \tau \in \mathcal{T}^i_{en} \text{ and } q_i \geq 1,\ j = 0, \dots, q_i - 1 \tag{4.195}$$

$$0 = g_{i,y}(y(\tau))z(\tau) \qquad \text{if } \tau \in \mathcal{T}^i_{to} \text{ and } q_i \geq 2 \tag{4.196}$$

$$0 = g^{(q_i)}_{i,(u,y)}(u(\tau), y(\tau))(v(\tau^-), z(\tau)) + \sigma^i_\tau \frac{\mathrm{d}}{\mathrm{d}t} g^{(q_i)}_i(u, y)|_{t=\tau^-} \quad \text{if } \tau \in \mathcal{T}^i_{en} \tag{4.197}$$

$$0 = g^{(q_i)}_{i,(u,y)}(u(\tau), y(\tau))(v(\tau^+), z(\tau)) + \sigma^i_\tau \frac{\mathrm{d}}{\mathrm{d}t} g^{(q_i)}_i(u, y)|_{t=\tau^+} \quad \text{if } \tau \in \mathcal{T}^i_{ex} \tag{4.198}$$

$$0 = g^{(1)}_{i,y}(y(\tau))z(\tau) + \sigma^i_\tau g^{(2)}_i(u(\tau), y(\tau)) \qquad \text{if } \tau \in \mathcal{T}^i_{to} \text{ and } q_i \geq 2. \tag{4.199}$$

The proof of this result follows from the linearization of the shooting equations (for the jump of π^q at entry times, see [19, Lemma 3.7][5]).

In addition to the tangentiality conditions (A5)(i), reducibility condition (A5)(ii) and strict complementarity assumption on boundary arcs (A6)(i) made for pure state constraints in section 4.6, we will need the following assumptions, also for the mixed control-state constraints:

(A5) (iii) (Nontangentiality conditions for mixed control-state constraints)
For all $i = r+1, \dots, r+s$ and all $\tau^i_{en} \in \mathcal{T}^i_{en}$ and $\tau^i_{ex} \in \mathcal{T}^i_{ex}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} c_i(u(t), y(t))|_{t=\tau^{i-}_{en}} > 0, \qquad \frac{\mathrm{d}}{\mathrm{d}t} c_i(u(t), y(t))|_{t=\tau^{i+}_{ex}} < 0. \tag{4.200}$$

(A6) (ii) (Strict complementarity at touch points)

$$\mathcal{T}^{i,nes}_{to} = \emptyset, \qquad \text{for all } i = 1, \dots, r+s.$$

(iii) (Strict complementarity for mixed constraints)

$$\lambda_i(t) > 0, \quad \text{for a.a. } t \in \operatorname{int} \Delta_i, \quad \text{for all } i = r+1, \dots, r+s. \tag{4.201}$$

Assumption (A6)(ii) implies that constraints of order $q_i = 0, 1$ have no touch points.

We will finally make the assumption below:

(A7) The junctions times of different components of the constraint do not coincide (i.e. $i, j \in \{1, \dots, r+s\}$ and $i \neq j$ implies that $\mathcal{T}^i \cap \mathcal{T}^j = \emptyset$).

Remark 4.32. When (A7) holds, for all entry and exit points of state constraints $\tau \in \mathcal{T}^i_{en} \cup \mathcal{T}^i_{ex}$, $i = 1, \dots, r$, we have that $q_\tau = q_i$, and assumption (A5)(i) simply says that

$$\begin{aligned} \frac{\mathrm{d}^{q_i}}{\mathrm{d}t^{q_i}} g^{(q_i)}_i(u, y)|_{t=\tau^\pm} &\neq 0 \qquad \text{if } q_i \text{ is odd,} \\ \frac{\mathrm{d}^{q_i-1}}{\mathrm{d}t^{q_i-1}} g^{(q_i)}_i(u, y)|_{t=\tau^\pm} &\neq 0 \qquad \text{if } q_i \text{ is even,} \end{aligned} \tag{4.202}$$

where $\tau^\pm$ denotes τ^- (resp. τ^+) if τ is an entry point (resp. exit point).

Under (A4) and the strict complementarity assumption (A6), using Lemma 4.2, the critical cone $\hat{C}_{L_2}(u)$ defined by (4.131) is the set of $v \in \mathcal{V}$ satisfying (recall that $z_v \in \mathcal{Z}$ is the solution of the linearized state equation (4.22))

$$0 = g^{(q_i)}_{i,u}(u, y)v + g^{(q_i)}_{i,y}(u, y)z_v \quad \text{a.e. on } \mathcal{I}^i_b,\ i = 1, \dots, r+s, \tag{4.203}$$

$$0 = g^{(j)}_{i,y}(y(\tau))z_v(\tau), \quad \tau \in \mathcal{T}^i_{en},\ i = 1, \dots, r,\ j = 0, \dots, q_i - 1, \tag{4.204}$$

$$0 = g_{i,y}(y(\tau))z_v(\tau), \quad \tau \in \mathcal{T}^i_{to},\ i = 1, \dots, r. \tag{4.205}$$

[5] Lemma 2.27 of this thesis.

Theorem 4.33 (Well-posedness of the shooting algorithm). *Let (u, y) be a local solution of $(\mathcal{P})$ satisfying (A1)-(A7). Then the shooting algorithm is well-posed in the neighborhood of the trajectory (u, y), iff the two conditions below are satisfied:*
(i) *components of the state constraint of order $q_i \geq 3$ have no boundary arc;*
(ii) *the no-gap sufficient condition* (4.134) *holds, i.e. $Q(v) > 0$ for all $v \in \mathcal{V}$ satisfying* (4.203)-(4.205) *with the associated linearized state $z_v \in \mathcal{Z}$ solution of* (4.22) *and $Q(v)$ defined by* (4.135).

Once the junction conditions and the no-gap second-order optimality conditions have been established, and with assumption (A7), Th. 4.33 is an easy extension of [19, Th. 3.3][6] obtained in the scalar case. The next lemma relates the second-order conditions established in section 4.6 and the alternative multipliers used in the shooting algorithm.

Lemma 4.34. *Let (u, y) be a stationary point of $(\mathcal{P})$, satisfying (A2)-(A4) and (A5)(ii). Then an equivalent expression using the alternative Hamiltonian and multipliers for the quadratic form $Q(v)$ defined in* (4.135) *over $\mathcal{V}$ is:*

$$\begin{aligned} Q(v) \;=\; & \int_0^T H^q_{(u,y),(u,y)}(u, y, p^q, \eta^q)((v, z_v), (v, z_v))\mathrm{d}t + \phi_{yy}(y(T))(z_v(T), z_v(T)) \\ & + \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}^i_{en}} \sum_{j=1}^{q_i} \nu^{i,j}_\tau g^{(j-1)}_{i,yy}(y(\tau))(z_v(\tau), z_v(\tau)) \\ & + \sum_{i=1}^{r} \sum_{\tau \in \mathcal{T}^{i,ess}_{to}} \nu^i_\tau \left(g_{i,yy}(y(\tau))(z_v(\tau), z_v(\tau)) - \frac{(g^{(1)}_{i,y}(y(t))z_v(t))^2}{\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_i(y(t))|_{t=\tau}} \right). \end{aligned} \tag{4.206}$$

Proof. The contribution of mixed control-state constraints in both (4.135) and (4.206) is equal to $\int_0^T \lambda c_{(u,y),(u,y)}(u, y)((v, z_v), (v, z_v))\mathrm{d}t$, therefore, summing over the finitely many state constraints g_i, the proof is identical to [19, Lemma 3.6][7]. □

Proof of Th. 4.33. We first prove that if (i) does not hold, the Jacobian of the shooting mapping is singular. So assume that a constraint g_i of order $q_i \geq 3$ has a boundary arc $(\tau^i_{en}, \tau^i_{ex})$. By assumption (A7) and (4.89), we have that $q_{\tau^i_{en}} = q_{\tau^i_{ex}} = q_i$, and hence, by Prop. 4.22, u is continuous until order $q_i - 2 \geq 1$. Therefore $\dot{u}$ is continuous at τ^i_{en} and τ^i_{ex}, and consequently, $\frac{\mathrm{d}}{\mathrm{d}t} g^{(q_i)}(u(t), y(t))$ is also continuous, and vanishes at τ^{i-}_{en} and τ^{i+}_{ex}. Taking all variations of jump parameters equal to zero, except $\sigma^i_{\tau^i_{ex}} \neq 0$, we find by Lemma 4.31 a nonzero element in the kernel of the Jacobian of the shooting mapping. Therefore the shooting algorithm is ill-posed.

We assume now that (i) holds. We will prove that the Jacobian of the shooting mapping is invertible iff (ii) holds. The Jacobian of the shooting mapping is invertible, iff it is one-to-one, i.e. iff the only solution of equations (4.194)-(4.199), where (z, v, π^q, ζ^q) is the solution of (4.187)-(4.193), is $\pi^0 = 0$, $(\sigma^i_\tau) = 0$, $(\gamma^{i,j}_\tau) = 0$, $(\gamma^i_\tau) = 0$. We recognize that (4.187)-(4.193) and (4.194)-(4.196) and (4.199) (which enables, by (A5)(ii), to substitute $-g^{(1)}_{i,y}(y(\tau))z(\tau)/g^{(2)}_i(u(\tau), y(\tau))$ for σ^i_τ in (4.193) for all touch point τ), constitutes the first-order optimality condition for the problem

$$(PQ) \qquad \min_{v \in \mathcal{V}} \frac{1}{2} Q(v), \qquad v \in \hat{C}_{L_2}(u)$$

[6] Theorem 2.23 of this thesis.
[7] Lemma 2.26 of this thesis.

with $Q(v)$ given by (4.206) and $\hat{C}_{L_2}(u)$ by (4.203)-(4.205). Here (γ_τ^i) are the multipliers associated with the constraints (4.205), and those associated with the constraints (4.204) are equal to $\gamma_\tau^{i,j}$ if $j = 1$ and $\gamma_\tau^{i,j} + \sigma_\tau^i \nu_\tau^{i,j-1}$ if $j > 1$.

If (ii) holds, i.e. if the second-order sufficient condition (4.134) holds, then by Lemma 4.34 the unique solution of (PQ) is zero. By (A2), the cost function of (PQ) is a *Legendre form* over $\mathcal{V}$, and hence, the strict positivity of $Q(v)$ over the closed linear space $\hat{C}_{L_2}(u)$ implies its uniform positivity (i.e. there exists $\alpha > 0$ such that $Q(v) \geq \alpha \|v\|_2^2$ for all $v \in \hat{C}_{L_2}(u)$). In addition, the set $\hat{C}_{L_2}(u)$ is convex and the linear constraints (4.203)-(4.205) defining $\hat{C}_{L_2}(u)$ are *onto* by Lemma 4.3. Therefore the first-order optimality condition of (PQ) is necessary and sufficient for optimality, so (ii) implies that zero is the unique solution of the first-order optimality condition of (PQ). Therefore we have $(z, v, \pi^q, \zeta^q) = 0$, and all of π^0, (γ_τ^i), $(\gamma_\tau^{i,j})$ for $j = 1$ also equal zero by Corollary 4.18 since $[\pi^q(\tau)] = 0$, and we have as well

$$\gamma_\tau^{i,j} + \sigma_\tau^i \nu_\tau^{i,j-1} = 0, \qquad \text{for all } j = 2, \ldots, q_i,\ i = 1, \ldots, r,\ \tau \in \mathcal{T}_{en}^i. \tag{4.207}$$

Now whenever (i) holds, it holds for all entry/exit times that $q_\tau \leq q_i \leq 2$, and from assumptions (A5)(i) and (A5)(iii), it follows that $\frac{\mathrm{d}}{\mathrm{d}t} g_i^{(q_i)}(u, y)|_{t=\tau^-}$ is nonzero for all entry points $\tau \in \mathcal{T}_{en}^i$, for all $i = 1, \ldots, r+s$. Therefore, equations (4.197) with $(v, z) = 0$ and (4.207) imply that $\sigma_\tau^i = 0$, for all entry points $\tau \in \mathcal{T}_{en}^i$, $i = 1, \ldots, r+s$, and that $\gamma_\tau^{i,j} = 0$ for all $j = 2, \ldots, q_i$, $i = 1, \ldots, r$, $\tau \in \mathcal{T}_{en}^i$. Similarly, we obtain that (4.198) and (4.199) imply that $\sigma_\tau^i = 0$ for all exit and touch points. Therefore, whenever (i)-(ii) holds, the Jacobian of the shooting mapping is one-to-one, hence invertible, and thus the shooting algorithm is well-posed locally around the local solution (u, y).

Assume now that (ii) does not hold. By Th. 4.24(i), the second-order necessary condition (4.132) holds at the local solution (u, y), implying that $Q(v)$ is nonnegative over $\hat{C}_{L^2}(u)$. Therefore, if (4.134) is not satisfied, this implies that there exists a nonzero optimal solution of (PQ), and hence there exists a nonzero solution of its first-order optimality condition. It is then easy to see that the variations of shooting parameters associated as above with this nonzero solution of (PQ) are not all zero, and belong to the kernel of the Jacobian of the shooting mapping. This proves that the shooting algorithm is ill-posed. □

4.8 Final remark: Extension to constraints on the initial and final state

Let us comment on the extension of the results when there are additional equality and/or inequality constraints on the initial and final state:

$$\Psi_i(y(0), y(T)) = 0, \quad i = 1, \ldots, \varrho', \quad \Psi_i(y(0), y(T)) \leq 0, \quad i = \varrho' + 1, \ldots, \varrho \tag{4.208}$$

with $\Psi : \mathbb{R}^{2n} \to \mathbb{R}^\varrho$ a C^2 mapping ($0 \leq \varrho' \leq \varrho \leq n$). The results of this paper can easily be generalized, under an additional (strong) controllability assumption (A1') below, having the role of Lemma 4.3 in the proofs, and, for the second-order optimality conditions and the well-posedness of the shooting algorithm, also under an additional assumption that *strict complementarity* holds for the inequality constraints in (4.208). Denote by $\hat{\Psi}$ the mapping composed of the equality and active inequality constraints in (4.208), of dimension $\hat{\varrho}$. Given $\kappa \in [1, +\infty]$ and $(v, x) \in \mathcal{V}_\kappa \times \mathbb{R}^n$, let $z_{v,x}$ denote the (unique) solution in $\mathcal{Z}_\kappa$ of:

$$\dot{z}_{v,x} = f_u(u, y)v + f_y(u, y)z_{v,x}, \qquad z_{v,x}(0) = x.$$

(A1') For $\kappa = 2, \infty$, there exists $\delta > 0$ and $n \in \mathbb{N}^*$ such that the linear mapping $\mathcal{V}_\kappa \times \mathbb{R}^n \to \prod_{i=1}^r W^{q_i,\kappa}(\Delta_i^\delta) \times \prod_{i=r+1}^{r+s} L^\kappa(\Delta_i^n) \times \mathbb{R}^{\hat{\varrho}}$,

$$(v,x) \to \begin{pmatrix} \Big(g_{i,y}(y(\cdot))z_{v,x}(\cdot)|_{\Delta_i^\delta}\Big)_{1\le i\le r} \\ \big((c_{i,y}(u(\cdot),y(\cdot))z_{v,x}(\cdot) + c_{i,u}(u(\cdot),y(\cdot))v(\cdot))|_{\Delta_i^n}\big)_{r+1\le i\le r+s} \\ D_{y_0}\hat{\Psi}(y(0),y(T))x + D_{y_T}\hat{\Psi}(y(0),y(T))z_{v,x}(T) \end{pmatrix}$$

is onto, and therefore has a bounded right inverse by the open mapping Theorem.

Note that in the absence of mixed control-state constraints, this assumption (A1') is satisfied e.g. in the case of a linear system, i.e. $f(u,y) = Ay + Bu$, if the pair (A,B) is controllable, the initial and final conditions are fixed $y(0) = y_0$ and $y(T) = y_T$ and satisfy $g_i(y_0) < 0$ and $g_i(y_T) < 0$ for all $i = 1,\dots,r$, and (4.25) holds.

4.9 Appendix

4.9.1 Tangent and Normal cones in L^∞

Let us recall the characterization of the tangent and normal cones (in the sense of convex analysis) to $\mathcal{K} := L^\infty_-(0,T)$ at point $x \in \mathcal{K}$. The characterization of the tangent cone was obtained by Cominetti and Penot [42]:

$$T_{\mathcal{K}}(x) = \{h \in L^\infty \ : \ \|\mathbf{1}_{\Delta_n(x)}h_+\|_\infty \to 0 \text{ when } n \to +\infty\}, \tag{4.209}$$

with $\mathbf{1}_{\Delta_n(x)}$ the indicator function of the set $\Delta_n(x)$ defined by (4.7), and $h_+ := \max(h;0)$ a.e.

Since $\mathcal{K}$ is a cone, the normal cone satisfies $N_{\mathcal{K}}(x) = \{\lambda \in (L^\infty)^*_+,\ \langle \lambda, x\rangle = 0\}$. Define

$$\mathcal{N}_n(x) := \{y \in L^\infty(0,T) \ ; \ y(t) = 0 \text{ for a.a. } t \in \Delta_n(x)\}, \qquad n \in \mathbb{N}^*.$$

Then we have the following characterization of $N_{\mathcal{K}}(x)$.

Lemma 4.35. *Let $x \in \mathcal{K}$. Then*

$$N_{\mathcal{K}}(x) \ = \ \{\lambda \in (L^\infty)^*_+ \ ; \ \langle \lambda, y\rangle = 0, \ \forall\, y \in \cup_{n\in\mathbb{N}^*}\mathcal{N}_n(x)\}. \tag{4.210}$$

Proof. "$\subset$" Let $\lambda \in N_{\mathcal{K}}(x)$, $n \in \mathbb{N}^*$ and $y \in \mathcal{N}_n(x)$. Then the function $x \pm \frac{1}{n\|y\|_\infty}y$ is nonpositive a.e. on $[0,T]$, and hence, since $\lambda \ge 0$,

$$\langle \lambda, x \pm \frac{1}{n\|y\|_\infty}y\rangle \le 0.$$

Using then that $\langle \lambda, x\rangle = 0$, we obtain that $\pm\langle \lambda, y\rangle \le 0$, i.e. $\langle \lambda, y\rangle = 0$.
"$\supset$" Assume that $\lambda \in (L^\infty)^*_+$ and $\lambda \in \cap_{n\in\mathbb{N}^*}(\mathcal{N}_n(x))^\perp$. Then we have, for all $n \in \mathbb{N}^*$,

$$\langle \lambda, x\rangle \ \ = \ \ \langle \lambda, \mathbf{1}_{\Delta_n(x)}x\rangle$$

and hence, since $0 \ge x(t) \ge -\frac{1}{n}$ a.e. on $\Delta_n(x)$,

$$|\langle \lambda, x\rangle| \ \ \le \ \ \|\lambda\|_{\infty *}\|\mathbf{1}_{\Delta_n(x)}x\|_\infty \ \le \ \|\lambda\|_{\infty *}\frac{1}{n},$$

Letting $n \to +\infty$, we thus obtain that $\langle \lambda, x\rangle = 0$, which achieves the proof. □

We end this section by recalling two results used in the proof of the second-order necessary condition.

Lemma 4.36. *The cone $\mathcal{K}$ is polyhedric, i.e. for all $x \in \mathcal{K}$ and all $\lambda \in N_{\mathcal{K}}(x)$,*

$$T_{\mathcal{K}}(x) \cap \lambda^{\perp} = \mathrm{cl}(\mathcal{R}_{\mathcal{K}}(x) \cap \lambda^{\perp}), \tag{4.211}$$

where $\mathcal{R}_{\mathcal{K}}(x)$ is the radial cone (4.136).

Proof. Let $h \in T_{\mathcal{K}}(x) \cap \lambda^{\perp}$. For $n \in \mathbb{N}^*$, define for a.a. $t \in (0,T)$

$$h_n(t) = \begin{cases} h(t) & \text{a.e. on } [0,T] \setminus \Delta_n(x) \\ h(t)_- & \text{a.e. on } \Delta_n(x) \end{cases}$$

where $h(t)_- = \min(0, h(t))$. For all $0 < \varepsilon < \frac{1}{n\|h\|_\infty}$, it is easily seen that $x + \varepsilon h_n \leq 0$ a.e. on $[0,T]$, and hence $h_n \in \mathcal{R}_{\mathcal{K}}(x)$, for all $n \in \mathbb{N}^*$. Moreover, in view of (4.210), we have that $\langle \lambda, h_n \rangle = \langle \lambda, h_- \rangle$. Since $\langle \lambda, h \rangle = \langle \lambda, h_+ \rangle + \langle \lambda, h_- \rangle = 0$, it follows that

$$|\langle \lambda, h_- \rangle| = |\langle \lambda, h_+ \rangle| = |\langle \lambda, \mathbf{1}_{\Delta_n(x)} h_+ \rangle| \ \leq \ \|\lambda\|_{\infty *} \|\mathbf{1}_{\Delta_n(x)} h_+\|_\infty \ \to \ 0$$

when $n \to +\infty$ by (4.209). Hence $\langle \lambda, h_n \rangle = 0$. Finally, $\|h - h_n\|_\infty = \|\mathbf{1}_{\Delta_n(x)} h_+\|_\infty \to 0$ by (4.209) again. So h_n is a sequence in $\mathcal{R}_K(x) \cap \lambda^{\perp}$ that converges to h in L^∞. □

Lemma 4.37. *Let $x \in \mathcal{K}$. For any $\lambda \in N_{\mathcal{K}}(x) \cap L^2(0,T)$, the set $T_{\mathcal{K}}(x) \cap \lambda^{\perp}$ is dense in the set $\hat{T}(x) \cap \lambda^{\perp}$, with*

$$\hat{T}(x) := \{ w \in L^2(0,T) \ ; \ w \leq 0 \text{ a.e. on } \Delta(x) \}. \tag{4.212}$$

Proof. Let $\hat{w} \in \hat{T}(x) \cap \lambda^{\perp}$. Let w_n be defined a.e. on $[0,T]$ by:

$$w_n(t) = \begin{cases} \max(\min(\hat{w}(t), n), -n) & \text{if } t \in [0,T] \setminus \Delta_n(x) \\ \max(\min(\hat{w}(t), 0), -n) & \text{if } t \in \Delta_n(x). \end{cases}$$

Then $w_n \in L^\infty$, and for all $k \geq n$, $\mathbf{1}_{\Delta_k(x)} w_n \leq 0$ a.e., and hence by (4.209) $w_n \in T_{\mathcal{K}}(x)$. Since $\lambda \in N_{\mathcal{K}}(x) \cap L^2(0,T)$, $\int_0^T \lambda(t) x(t) \mathrm{d}t = 0$ implies that $\lambda(t) = 0$ for a.a. $t \in [0,T] \setminus \Delta(x)$. And then $\int_0^T \lambda(t) \hat{w}(t) \mathrm{d}t = 0$ implies, since $\hat{w}(t) \leq 0$ on $\Delta(x)$, that $\hat{w}(t) = 0$ for a.a. t such that $\lambda(t) \neq 0$. Consequently, we also have that $w_n(t) = 0$ for a.a. t such that $\lambda(t) \neq 0$, and hence, $\langle \lambda, w_n \rangle = \int_0^T \lambda(t) w_n(t) \mathrm{d}t = 0$, i.e. $w_n \in T_{\mathcal{K}}(x) \cap \lambda^{\perp}$. It remains to show that $w_n \to \hat{w}$ for the norm of L^2. If $t \notin \Delta(x)$, for n large enough, $w_n(t) = \max(\min(\hat{w}(t), n), -n) \to \hat{w}(t)$ when $n \to \infty$, and if $t \in \Delta(x)$, since $\hat{w}(t) \leq 0$ a.e. on $\Delta(x)$, for all n we have $w_n(t) = \max(\hat{w}(t), -n) \to \hat{w}(t)$. Hence, $w_n(t) \to \hat{w}(t)$ a.e., and $|w_n(t)| \leq |\hat{w}(t)|$ for all $t \in [0,T]$, with $\hat{w} \in L^2$. It follows then from the Lebesgue's dominated convergence Theorem that $w_n \to \hat{w}$ in L^2, which achieves the proof. □

4.9.2 First-order optimality condition

If u is a local solution of (4.5) satisfying (4.34), then it is well-known that there exist $\eta \in \mathcal{M}([0,T]; \mathbb{R}^{r*})$ and $\lambda \in (L^\infty)^*(0,T; \mathbb{R}^{s*})$ such that

$$DJ(u)v + \langle \eta, DG(u)v \rangle + \langle \lambda, D\mathcal{G}(u)v \rangle = 0, \qquad \forall\, v \in \mathcal{U}, \tag{4.213}$$

$$\eta \in N_K(G(u)), \quad \lambda \in N_{\mathcal{K}}(\mathcal{G}(u)). \tag{4.214}$$

Lemma 4.38. *Assume that u is a local solution of* (4.5) *satisfying* (4.34)*, and that assumption* (4.31) *holds. Then the multiplier λ belongs to $L^\infty(0,T;\mathbb{R}^{s*})$.*

Proof. Let $\tilde{p}$ be the unique solution in $BV(0,T;\mathbb{R}^{n*})$ of:

$$-\mathrm{d}\tilde{p} = H_y(u,y_u,\tilde{p})\mathrm{d}t + \mathrm{d}\eta g_y(y_u); \qquad p(T) = \phi_y(y_u(T)).$$

Then it is not difficult to show that (4.213) writes, with z_v the solution of (4.22):

$$\int_0^T H_u(u,y_u,\tilde{p})v\mathrm{d}t + \langle \lambda, c_y(u,y_u)z_v + c_u(u,y_u)v\rangle = 0, \qquad \forall\, v \in \mathcal{U}. \tag{4.215}$$

Since u, y and $\tilde{p}$ belong to L^∞, so do the functions $H_u(u(\cdot),y_u(\cdot),\tilde{p}(\cdot))$, $c_u(u(\cdot),y_u(\cdot))$ and $c_y(u(\cdot),y_u(\cdot))$. It follows then from (4.215) that for all $v \in \mathcal{U}$,

$$|\langle \lambda, c_u(u,y_u)v\rangle| \;\leq\; \|\lambda\|_{\infty *}\|c_y(u,y_u)\|_\infty\|z_v\|_\infty + \|H_u(u,y_u,\tilde{p})\|_\infty\|v\|_1.$$

By Gronwall's Lemma, there exists a constant $\kappa > 0$ such that $\|z_v\|_\infty \leq \kappa\|v\|_1$, for all $v \in \mathcal{U}$, and hence we obtain that for all $v \in \mathcal{U}$,

$$|\langle \lambda, c_u(u,y_u)v\rangle| \leq (\|\lambda\|_{\infty *}\|c_y(u,y_u)\|_\infty\kappa + \|H_u(u,y_u,\tilde{p})\|_\infty)\|v\|_1 \leq \kappa'\|v\|_1. \tag{4.216}$$

By assumption (4.31), for all $w \in L^\infty(0,T;\mathbb{R}^s)$, there exists $v \in \mathcal{U}$ such that $w_i(t) = c_{i,u}(u(t),y_u(t))v(t)$ for a.a. $t \in \Delta_n(c_i(u,y_u))$, for all $i = r+1,\ldots,r+s$, and $\|v\|_1 \leq M\|w\|_1$ for some constant $M > 0$. Indeed, take e.g. $v(t) = C(t)^\top(C(t)C(t)^\top)^{-1}w(t)$ with $C(t) := c_{I_n^c(t),u}(u(t),y_u(t))$ if $I_n^c(t) \neq \emptyset$, and $v(t) = 0$ otherwise, and $M := \|C^\top(CC^\top)^{-1}\|_\infty$. Since $\lambda \in N_K(\mathcal{G}(u))$, the characterization of the critical cone (4.210) implies that $\langle \lambda, c_u(u,y_u)v\rangle = \langle \lambda, w\rangle$. Then (4.216) yields

$$|\langle \lambda, w\rangle| \;\leq\; \kappa''\|w\|_1, \qquad \forall w \in L^\infty(0,T;\mathbb{R}^s). \tag{4.217}$$

Since L^∞ is dense in L^1 and λ is continuous for the norm of L^1, λ can be extended to a continuous linear form over $L^1(0,T;\mathbb{R}^s)$. Therefore λ belong to the dual space $L^\infty(0,T;\mathbb{R}^{s*})$. □

It is not difficult to derive from this result the first-order optimality condition given in Th. 4.5. See related results in [111, 88].

Chapitre 5

Analyse de stabilité pour les contraintes d'ordre 2*

Abstract This paper gives stability results for nonlinear optimal control problems subject to a regular state constraint of second-order. The strengthened Legendre-Clebsch condition is assumed to hold, and no assumption on the structure of the contact set is made. Under a weak second-order sufficient condition (taking into account the active constraints), we show that the solutions are Lipschitz continuous w.r.t. the perturbation parameter in the L^2 norm, and Hölder continuous in the L^∞ norm. We use a generalized implicit function theorem in metric spaces by Dontchev and Hager [SIAM J. Control Optim., 1998]. The difficulty is that multipliers associated with second-order state constraints have a low regularity (they are only bounded measures). We obtain Lipschitz stability of a "primitive" of the state constraint multiplier.

Résumé Dans cet article on donne un résultat de stabilité pour les problèmes de commande optimale avec une contrainte sur l'état du second ordre régulière. La condition forte de Legendre-Clebsch est supposée satisfaite. Sous une condition suffisante du second ordre faible (prenant en compte les contraintes actives) on montre que les solutions sont lipschitziennes par rapport au paramètre pour la norme L^2, et höldériennes pour la norme L^∞. On utilise un théorème des fonctions implicites généralisé dans des espaces métriques de Dontchev et Hager [SIAM J. Control Optim., 1998]. La difficulté vient du fait que les multiplicateurs associés aux contraintes sur l'état du second ordre sont peu réguliers (ce sont seulement des mesures bornées). On obtient la stabilité lipschitz d'une primitive du multiplicateur associé à la contrainte sur l'état.

5.1 Introduction

This paper deals with stability analysis of nonlinear optimal control problems of an ordinary differential equation with a second-order state constraint. State constraints of second-order occur naturally in applications: For example, in the problem of the atmospheric reentry of a space shuttle, with the back angle as control, the constraints on the thermal flux, normal acceleration and dynamic pressure are second-order state constraints, see [27]. Stability and sensitivity analysis of solutions of optimal control problems is of high interest for the study

*Accepted for publication in SIAM Journal on Optimization, under the title *Stability analysis of optimal control problems with a second-order state constraint.*

of numerical methods, such as e.g. continuation algorithms, see [20], and to analyze the convergence of discretization schemes and obtain errors estimates, see e.g. [54].

For a class of general constrained optimization problems in Banach spaces, when the derivative of the constraint is "onto" and a second-order sufficient condition holds, Lipschitz stability of solutions and multipliers can be obtained by application of Robinson's strong regularity theory [121] to the first-order optimality system. For optimal control problems, this theory does not apply because of the well-known *two-norm discrepancy* (see [99]). Stability results for optimal control problems using variants of Robinson's strong regularity in order to deal with the two-norm approach have been obtained in [52], [87], [55] for control constraints, and [90] for mixed control-state constraints.

Lipschitz stability results for state constraints of first-order have been obtained by Malanowski [88] and Dontchev and Hager [53]. The difficulty of pure state constraints is the low regularity of multipliers, which are bounded Borel measures. These multipliers can be identified with functions of bounded variation, and for first-order state constraints, it is known that under standard hypothesis, they are more regular (they are Lipschitz continuous functions, see Hager [65]). This additional regularity of solutions and multipliers is strongly used in the analysis in [88] and [53]. In those two papers, strong second-order sufficient conditions were used (that do not take into account the active constraints). The sufficient condition was recently weakened by Malanowski [92, 91].

For higher-order state constraints, the multipliers associated with the state constraints are only measures, and are not continuous w.r.t. the perturbation parameter (for the total variation norm). For this reason, the frameworks of [88] or [53] are not directly applicable. The only stability and sensitivity results known for state constraints of higher-order are based on the shooting approach, see Malanowski and Maurer [94] and [19]. Such results require strong assumptions on the structure of the contact set.

The main result of this paper is a stability result for regular second-order state constraints, with no assumption on the structure of the contact set. The control is assumed to be continuous and the strengthened Legendre-Clebsch condition to hold. We use a generalized implicit function theorem in metric spaces by Dontchev and Hager [53], applied to a system equivalent to the first-order optimality condition (the *alternative formulation*). This formulation involves *alternative multipliers* that are "integrals" of the original state constraint multipliers, and therefore are more regular. We obtain Lipschitz continuity of solutions and alternative multipliers in the L^2 norm, and Hölder continuity in the L^∞ norm, under a weak second-order sufficient condition taking into account the active constraints.

The paper is organized as follows. In section 5.2, the problem, optimality conditions, assumptions, and the admissible class of perturbations are introduced. In section 5.3, the second-order sufficient optimality condition is presented. In section 5.4, the main stability results for the nonlinear optimal control problem are given. Section 5.5 is devoted to stability analysis of linear-quadratic problems, that is used to prove the main theorem in section 5.6. Finally, conclusion and comments are given in section 5.7.

5.2 Preliminaries

We consider the following optimal control problem

$$(\mathcal{P}) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell(u(t),y(t))\mathrm{d}t + \phi(y(T)) \tag{5.1}$$

$$\text{subject to} \qquad \dot{y}(t) = f(u(t),y(t)) \quad \text{for a.a. } t\in[0,T], \quad y(0)=y_0 \tag{5.2}$$

$$g(y(t)) \leq 0 \quad \text{for all } t\in[0,T] \tag{5.3}$$

with the control and state spaces $\mathcal{U} := L^\infty(0,T;\mathbb{R}^m)$ and $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$. The following assumptions are assumed to hold throughout the paper and will not be repeated in the various results of the paper.

(A0) The data $\ell : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}$, $\phi : \mathbb{R}^n \to \mathbb{R}$ (resp. $f : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}$) are C^2 (resp. C^3, C^4) mappings, with locally Lipschitz continuous second-order (resp. third-order, fourth order) derivatives, and f is Lipschitz continuous.

(A1) The initial condition $y_0 \in \mathbb{R}^n$ satisfies $g(y_0) < 0$.

We consider in this paper state constraints of *second-order*. This means that the first-order time derivative $g^{(1)} : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}$ of the constraint, defined by

$$g^{(1)}(u,y) := g_y(y)f(u,y)$$

does not depend on the control variable u, i.e. $g_u^{(1)} \equiv 0$ (and hence, we write $g^{(1)}(y) = g^{(1)}(u,y)$), and the second-order time derivative $g^{(2)} : \mathbb{R}^m\times\mathbb{R}^n \to \mathbb{R}$, defined by

$$g^{(2)}(u,y) := g_y^{(1)}(y)f(u,y)$$

depends explicitly on the control, i.e. $g_u^{(2)} \not\equiv 0$.

Remark 5.1. For linear-quadratic control problems of type (5.62)–(5.65) (see section 5.5), with dynamics given by $\dot{z}(t) = A(t)z(t) + B(t)v(t)$ and state constraint by $C(t)z(t) + d(t) \leq 0$, the state constraint is of second-order means that $C(t)B(t) \equiv 0$ on $[0,T]$ and $(\dot{C}(t) + C(t)A(t))B(t) \not\equiv 0$.

Remark 5.2. In this paper the state constraint is assumed to be scalar-valued for simplicity. The results are directly generalizable to several state constraints $g_1,\ldots,g_r$ of second-order (and even of *higher-order* [98, 68] $q_i \geq 2$ for $i = 1,\ldots,r$, see Remark 5.3 further) under the assumption (see [98, 17]) that the gradients of the nearly active constraints $\nabla_u g_i^{(q_i)}(u,y)$ are uniformly linearly independent along the trajectory.

Notation We denote by subscripts Fréchet derivatives w.r.t. the variables u, y, i.e. $f_y(u,y) = D_y f(u,y)$, $f_{yy}(u,y) = D^2_{yy}f(u,y)$, etc. The derivative with respect to the time is denoted by a dot, i.e. $\dot{y} = \frac{\mathrm{d}y}{\mathrm{d}t} = y^{(1)}$. The set of row vectors of dimension n is denoted by $\mathbb{R}^{n*}$. Adjoint or transpose operators are denoted by the symbol $^\top$. The euclidean norm is denoted by $|\cdot|$. By $L^r(0,T)$ we denote the Lebesgue space of measurable functions such that $\|u\|_r := (\int_0^T |u(t)|^r\mathrm{d}t)^{1/r} < \infty$ for $1 \leq r < \infty$, $\|u\|_\infty := \mathrm{supess}_{[0,T]}|u(t)| < \infty$. The space $W^{s,r}(0,T)$ denotes the Sobolev space of functions having their s first weak derivatives in $L^r(0,T)$, with the norm $\|u\|_{s,r} := \sum_{j=0}^s \|u^{(j)}\|_r$. We denote by H^s the space $W^{s,2}$. The space of continuous functions over $[0,T]$ and its dual space, the space of bounded Borel measures,

are denoted respectively by $C[0,T]$ and $\mathcal{M}[0,T]$. The set of nonnegative measures is denoted by $\mathcal{M}_+[0,T]$. The space of functions of bounded variation over $[0,T]$ is denoted by $BV[0,T]$, and the set of normalized BV functions vanishing at T is denoted by $BV_T[0,T]$. Functions of bounded variation are w.l.o.g. assumed to be right-continuous. We identify the elements of $\mathcal{M}[0,T]$ with the distributional derivatives $\mathrm{d}\eta$ of functions η in $BV_T[0,T]$. The support and the total variation of the measure $\mathrm{d}\eta \in \mathcal{M}[0,T]$ are denoted respectively by $\mathrm{supp}(\mathrm{d}\eta)$ and $|\mathrm{d}\eta|_{\mathcal{M}}$. The duality product over $\mathcal{M}[0,T] \times C[0,T]$ is denoted by $\langle \mathrm{d}\eta, x\rangle = \int_0^T x(t)\mathrm{d}\eta(t)$. We denote by $B_X(x,\rho)$ (resp. B_X) the open ball of the space X with center x and radius ρ (resp. the open unit ball of the space X). We write B_r for B_{L^r}, $r = 2, \infty$.

We call a *trajectory* an element $(u,y) \in \mathcal{U} \times \mathcal{Y}$ satisfying the state equation (5.2). A trajectory satisfying the state constraint (5.3) is said to be *feasible*. The *contact set* of a feasible trajectory is defined by

$$I(g(y)) := \{t \in [0,T] : g(y(t)) = 0\}. \tag{5.4}$$

Under assumption (A0), the mapping $\mathcal{U} \to \mathcal{Y}$, $u \mapsto y_u$ where y_u is the unique solution of the state equation (5.2), is well-defined. This leads us to the following abstract formulation of $(\mathcal{P})$:

$$\min_{u \in \mathcal{U}} J(u), \qquad G(u) \in K, \tag{5.5}$$

with the cost function $J(u) := \int_0^T \ell(u, y_u)\mathrm{d}t + \phi(y_u(T))$, the constraint mapping $G(u) := g(y_u)$, and the constraint cone $K := C_-[0,T]$ is the cone of continuous functions taking nonpositive values over $[0,T]$. The polar cone to K, denoted by K^-, is the set of nonnegative measures $\mathcal{M}_+[0,T]$.

Finally, in all the paper the time argument $t \in [0,T]$ is often omitted when there is no ambiguity.

5.2.1 Optimality conditions and Assumptions

Let us first recall the well-known first-order necessary optimality condition of problem $(\mathcal{P})$. The *Hamiltonian* $H : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$ is defined by

$$H(u,y,p) \ := \ \ell(u,y) + pf(u,y). \tag{5.6}$$

We say that a feasible trajectory (u,y) is a *stationary point* of $(\mathcal{P})$, if there exists $(p,\eta) \in BV([0,T];\mathbb{R}^{n*}) \times BV_T[0,T]$ such that

$$\begin{aligned} -\,\mathrm{d}p &= H_y(u,y,p)\mathrm{d}t + g_y(y)\mathrm{d}\eta, \qquad p(T) = \phi_y(y(T)) && (5.7)\\ 0 &= H_u(u(t),y(t),p(t)) \qquad \text{a.e. on } [0,T] && (5.8)\\ \mathrm{d}\eta &\in N_K(g(y)). && (5.9) \end{aligned}$$

Here $N_K(g(y))$ denotes the normal cone to K at point $g(y)$ (in the sense of convex analysis). If $g(y) \in K$, then $N_K(g(y))$ is the set of nonnegative measures in $\mathcal{M}_+[0,T]$ having their support included in the contact set (5.4), otherwise $N_K(g(y))$ is empty.

The *Lagrangian* $L : \mathcal{U} \times \mathcal{M}[0,T] \to \mathbb{R}$ of problem (5.5) is defined by

$$L(u,\eta) := J(u) + \langle \mathrm{d}\eta, G(u)\rangle = J(u) + \int_0^T g(y_u(t))\mathrm{d}\eta(t). \tag{5.10}$$

We may write the first-order optimality condition as follows: $(u, y = y_u)$ is a stationary point of $(\mathcal{P})$ iff there exists $\eta \in BV_T[0,T]$ such that

$$D_u L(u,\eta) = 0, \qquad \mathrm{d}\eta \in N_K(G(u)). \tag{5.11}$$

The costate p is then obtained in function of u, $y = y_u$ and η as the unique solution in $BV([0,T];\mathbb{R}^{n*})$ of the costate equation (5.7).

Robinson's constraint qualification [119, 120] for problem $(\mathcal{P})$ in abstract form (5.5) is as follows:

$$\exists\, \varepsilon > 0, \qquad \varepsilon B_{C[0,T]} \subset G(u) + DG(u)\mathcal{U} - K. \tag{5.12}$$

This condition is equivalent to the existence of some $v \in \mathcal{U}$ such that

$$DG(u)v < 0 \qquad \text{on } I(g(y)).$$

It is well-known that a local solution (weak minimum) of $(\mathcal{P})$ satisfying (5.12) is a stationary point of $(\mathcal{P})$.

Alternative formulation For the stability analysis, it will be convenient to write the optimality condition using alternative multipliers η^2 and p^2, uniquely related to (p,η) in the following way:

$$\eta^1(t) := \int_{(t,T]} \mathrm{d}\eta(s) = -\eta(t), \qquad \eta^2(t) := \int_t^T \eta^1(s)\mathrm{d}s, \tag{5.13}$$

$$p^2(t) := p(t) - \eta^1(t) g_y(y(t)) - \eta^2(t) g_y^{(1)}(y(t)), \qquad t \in [0,T]. \tag{5.14}$$

We see that η^2 belongs to the set $BV_T^2[0,T]$, defined by

$$BV_T^2[0,T] := \{\xi \in W^{1,\infty}(0,T) : \xi(T) = 0,\ \dot\xi \in BV_T[0,T]\}. \tag{5.15}$$

Define the *alternative Hamiltonian* $\tilde H : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^{n*} \times \mathbb{R} \to \mathbb{R}$ by

$$\tilde H(u,y,p^2,\eta^2) := H(u,y,p^2) + \eta^2 g^{(2)}(u,y), \tag{5.16}$$

where H is the classical Hamiltonian (5.6). Using these alternative multipliers, it is not difficult to see by a direct calculation (see [98] or [17, Lemma 3.4][1]) that a feasible trajectory (u,y) is a stationary point of $(\mathcal{P})$ iff there exists $(p^2,\eta^2) \in W^{1,\infty}(0,T;\mathbb{R}^{n*}) \times BV_T^2[0,T]$ such that

$$-\dot p^2 = \tilde H_y(u,y,p^2,\eta^2), \qquad p^2(T) = \phi_y(y(T)) \tag{5.17}$$

$$0 = \tilde H_u(u,y,p^2,\eta^2) \qquad \text{a.e. on } [0,T] \tag{5.18}$$

$$\mathrm{d}\dot\eta^2 \in N_K(g(y)). \tag{5.19}$$

The definition of these multipliers p^2, η^2 is inspired by the ones used in the alternative formulation for the shooting algorithm, see [98, 68, 94, 19], though p^2, η^2 are continuous over $[0,T]$ while the ones in the shooting algorithm have jumps.

Remark 5.3. The results of this paper have a natural generalization to a state constraint of higher-order $q > 2$, considering in the analysis alternative multipliers (η^q, p^q) of order q

[1] Lemma 4.11 of this thesis.

defined below and the resulting alternative formulation of optimality condition of order q. These alternative multipliers of order q, $\eta^q \in BV_T^q[0,T]$ with

$$BV_T^q[0,T] := \{\xi \in W^{q-1,\infty}(0,T) : \xi^{(j)}(T) = 0 \ \forall j = 0,\dots,q-2,\ \xi^{(q-1)} \in BV_T[0,T]\}$$

and $p^q \in W^{1,\infty}(0,T;\mathbb{R}^{n*})$, are defined by

$$\begin{aligned} \eta^1(t) &:= \int_{(t,T]} \mathrm{d}\eta(s), \qquad \eta^j(t) := \int_t^T \eta^{j-1}(s)\mathrm{d}s, \quad j = 2,\dots,q, \\ p^q(t) &:= p(t) - \sum_{j=1}^{q} \eta^j(t) g_y^{(j-1)}(y(t)). \end{aligned}$$

Assumptions Let $(\bar u, \bar y)$ be a local solution of $(\mathcal{P})$. We denote by $\Omega := I(g(\bar y))$ the contact set of the trajectory $(\bar u, \bar y)$, and for a small $\sigma > 0$, let Ω_σ denote a neighborhood of the contact set

$$\Omega_\sigma := \{t \in [0,T] : \mathrm{dist}\{t, \Omega\} < \sigma\}. \tag{5.20}$$

We assume that $(\bar u, \bar y)$ satisfies the assumption below:

(A2) The state constraint is a regular second-order state constraint, i.e. $g_u^{(1)} \equiv 0$ and

$$\exists\, \beta, \sigma > 0, \quad |g_u^{(2)}(\bar u(t), \bar y(t))| \geq \beta, \quad \text{for a.a. } t \in \Omega_\sigma. \tag{5.21}$$

Given $v \in L^r(0,T;\mathbb{R}^m)$, $1 \leq r \leq \infty$, we denote by z_v the unique solution in $W^{1,r}(0,T;\mathbb{R}^n)$ of the linearized state equation

$$\dot z_v(t) = f_y(\bar u(t), \bar y(t)) z_v(t) + f_u(\bar u(t), \bar y(t)) v(t) \quad \text{ a.e. on } [0,T], \quad z_v(0) = 0. \tag{5.22}$$

Note that the derivative of the constraint mapping is given by $DG(\bar u)v = g_y(\bar y) z_v$.

Lemma 5.4. *Let $(\bar u, \bar y)$ be a feasible trajectory of $(\mathcal{P})$ satisfying (A2). Then for all $r \in [1, +\infty]$ and all $\varepsilon \in (0, \sigma)$, with the σ of (5.21), so small that*

$$\Omega_\varepsilon \subset [a,T], \qquad \textit{for some } a > 0, \tag{5.23}$$

the linear mapping

$$L^r(0,T;\mathbb{R}^m) \to W^{2,r}(\Omega_\varepsilon), \qquad v \mapsto (g_y(\bar y(\cdot)) z_v(\cdot))|_{\Omega_\varepsilon}, \tag{5.24}$$

where $|_{\Omega_\varepsilon}$ denotes the restriction to the set Ω_ε, is onto, and therefore has a bounded right inverse by the open mapping theorem.

If u is continuous over $[0,T]$, then Lemma 5.4 is satisfied with $\varepsilon = \sigma$, assuming w.l.o.g. in view of (A1) that σ in (5.21) satisfies (5.23).

Proof. We only recall the main ideas of the proof, given in [17, Lemma 2.2][2]. We have that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\{g_y(\bar y(t)) z_v(t)\} &= g_y^{(1)}(\bar y(t)) z_v(t), \\ \frac{\mathrm{d}^2}{\mathrm{d}t^2}\{g_y(\bar y(t)) z_v(t)\} &= g_y^{(2)}(\bar u(t), \bar y(t)) z_v(t) + g_u^{(2)}(\bar u(t), \bar y(t)) v(t). \end{aligned}$$

Since by hypothesis (5.21) and (A1), $g_u^{(2)}(\bar u(t), \bar y(t))$ is non singular on a left neighborhood of Ω_ε, the result follows from Gronwall's Lemma. □

[2] Lemma 4.3 of this thesis.

By the above lemma, assumption (A2) (together with (A1)) implies that $(\bar{u}, \bar{y})$ satisfies Robinson's constraint qualification (5.12), and hence $(\bar{u}, \bar{y})$ is a stationary point of $(\mathcal{P})$, with multipliers $(\bar{p}, \bar{\eta})$. Moreover, Lemma 5.4 implies that the multipliers $(\bar{p}, \bar{\eta})$ associated with $(\bar{u}, \bar{y})$ are unique. We assume in addition that

(A3) $\bar{u}$ is continuous on $[0,T]$ and the strengthened Legendre-Clebsch condition holds:

$$\exists\, \alpha > 0,\ v^\top H_{uu}(\bar{u}(t), \bar{y}(t), \bar{p}(t))v \geq \alpha |v|^2, \quad \text{for all } t \in [0,T] \text{ and all } v \in \mathbb{R}^m. \tag{5.25}$$

Remark 5.5. A stronger assumption than (5.25), which *implies* the continuity of $\bar{u}$ (see [17, Prop. 3.1][3]), is the uniform strong convexity of the Hamiltonian:

$$\exists\, \alpha > 0,\ v^\top H_{uu}(\hat{u}, \bar{y}(t), \bar{p}(t))v \geq \alpha |v|^2, \quad \text{for all } t \in [0,T] \text{ and all } \hat{u}, v \in \mathbb{R}^m.$$

Denote by $\bar{p}^2$ and $\bar{\eta}^2$ the alternative multipliers related to $\bar{p}$ and $\bar{\eta}$ by (5.13)–(5.14). Assumption (5.25) can be rewritten, using the alternative multipliers $\bar{p}^2$ and $\bar{\eta}^2$ instead of $\bar{p}$ and $\bar{\eta}$ and the alternative Hamiltonian (5.16), by:

$$\exists\, \alpha > 0,\ v^\top \tilde{H}_{uu}(\bar{u}(t), \bar{y}(t), \bar{p}^2(t), \bar{\eta}^2(t))v \geq \alpha |v|^2, \quad \text{for all } t \in [0,T] \text{ and all } v \in \mathbb{R}^m. \tag{5.26}$$

Lemma 5.6. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3). Then $\bar{u}$ belongs to the space $W^{1,\infty}(0,T;\mathbb{R}^m)$.*

Proof. By (A3), implying (5.26), and the implicit function theorem applied to relation (5.18), there exists a C^1 function Υ such that $\bar{u}(t) = \Upsilon(\bar{y}(t), \bar{p}^2(t), \bar{\eta}^2(t))$. Since $\bar{y}, \bar{p}^2, \bar{\eta}^2 \in W^{1,\infty}$, it follows from the chain rule that $\bar{u} \in W^{1,\infty}$. □

Remark 5.7. More precisely, under the assumptions of Lemma 5.6, $\bar{u} \in BV^2([0,T];\mathbb{R}^m)$, where $BV^2[0,T] := \{u \in W^{1,\infty}(0,T) : \dot{u} \in BV[0,T]\}$. Indeed, differentiation of (5.18) w.r.t. time shows that (omitting arguments $(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)$)

$$0 \ = \ \tilde{H}_{uu}\dot{\bar{u}} + \tilde{H}_{uy} f - \tilde{H}_y f_u + \dot{\bar{\eta}}^2 g_u^{(2)}.$$

Since $\dot{\bar{\eta}}^2 = \bar{\eta} \in BV_T[0,T]$ and $\tilde{H}_{uu}$ is uniformly invertible by (5.26), we obtain the result.

5.2.2 Perturbed optimal control problem

We consider perturbed problems in the following form:

$$(\mathcal{P}^\mu) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell^\mu(u(t), y(t))\mathrm{d}t + \phi^\mu(y(T)) \tag{5.27}$$

$$\text{subject to} \qquad \dot{y}(t) \ = \ f^\mu(u(t), y(t)) \quad \text{a.e. on } [0,T], \quad y(0) = y_0^\mu \tag{5.28}$$

$$g^\mu(y(t)) \ \leq \ 0 \quad \text{for all } t \in [0,T]. \tag{5.29}$$

Here μ is the perturbation parameter, belonging to an open subset M_0 of a Banach space M.

Definition 5.8. We say that $(\mathcal{P}^\mu)$ is a *stable extension* of $(\mathcal{P})$, if:
(i) There exists $\bar{\mu} \in M_0$ such that $(\mathcal{P}^{\bar{\mu}}) \equiv (\mathcal{P})$;
(ii) The mappings $\mathbb{R}^m \times \mathbb{R}^n \times M_0 \to \mathbb{R}$, $(u,y,\mu) \mapsto \ell^\mu(u,y)$; $\mathbb{R}^n \times M_0 \to \mathbb{R}$, $(y,\mu) \mapsto \phi^\mu(y)$; $M_0 \to \mathbb{R}^n$, $\mu \mapsto y_0^\mu$ (resp. $\mathbb{R}^m \times \mathbb{R}^n \times M_0 \to \mathbb{R}^n$, $(u,y,\mu) \mapsto f^\mu(u,y)$; $\mathbb{R}^n \times M_0 \to \mathbb{R}$,

[3] Proposition 4.8 of this thesis.

$(y,\mu)\mapsto g^\mu(y)$) are of class C^2 (resp. C^3, C^4), with locally Lipschitz continuous second-order (resp. third-order, fourth order) derivatives, uniformly w.r.t. $\mu\in M_0$;
(iii) The dynamics f^μ is uniformly Lipschitz continuous over $\mathbb{R}^m\times\mathbb{R}^n$ for all $\mu\in M_0$;
(iv) The state constraint is not of first-order, i.e. $(g^\mu)^{(1)}_u(u,y)\equiv 0$ for all $(u,y,\mu)\in\mathbb{R}^m\times\mathbb{R}^n\times M_0$.

Given a stable extension $(\mathcal{P}^\mu)$ and $(u,\mu)\in\mathcal{U}\times M_0$, we denote by y^μ_u the unique solution in $\mathcal{Y}$ of the state equation (5.28), and we have the abstract formulation of $(\mathcal{P}^\mu)$

$$\min_{u\in\mathcal{U}} J^\mu(u),\qquad G^\mu(u)\in K, \tag{5.30}$$

with $J^\mu(u):=\int_0^T \ell^\mu(u,y^\mu_u)\mathrm{d}t+\phi^\mu(y^\mu_u(T))$ and $G^\mu(u):=g^\mu(y^\mu_u)$. When we refer to the data of the reference problem $(\mathcal{P})$, we often omit the superscript $\bar\mu$.

5.3 Second-order sufficient optimality condition

Let $(\bar u,\bar y)$ be a stationary point of $(\mathcal{P})$, with multipliers $(\bar p,\bar\eta)$. Let $\mathcal{V}:=L^2(0,T;\mathbb{R}^m)$. The quadratic form involved in the second-order optimality conditions, defined over $\mathcal{V}$, is as follows:

$$\begin{array}{rcl}\mathcal{Q}(v) & := & \displaystyle\int_0^T D^2_{(u,y)^2}H(\bar u,\bar y,\bar p)(v,z_v)^2\mathrm{d}t \;+\; \phi_{yy}(\bar y(T))(z_v(T),z_v(T))\\ & & \displaystyle+\int_0^T g_{yy}(\bar y)(z_v,z_v)\mathrm{d}\bar\eta.\end{array} \tag{5.31}$$

Recall that z_v is the solution of the linearized state equation (5.22). Here the notation $D^2_{(u,y)^2}H(\bar u,\bar y,\bar p)(v,z_v)^2$ stands for $D^2_{(u,y)(u,y)}H(\bar u,\bar y,\bar p)((v,z_v),(v,z_v))$. The critical cone $\mathcal{C}(\bar u)$ is the set of $v\in\mathcal{V}$ satisfying

$$g_y(\bar y(t))z_v(t) \;=\; 0 \quad \text{on } \mathrm{supp}(\mathrm{d}\bar\eta), \tag{5.32}$$
$$g_y(\bar y(t))z_v(t) \;\le\; 0 \quad \text{on } I(g(\bar y))\setminus\mathrm{supp}(\mathrm{d}\bar\eta). \tag{5.33}$$

A sufficient second-order optimality condition for $(\mathcal{P})$ is, see [21, Th. 18][4] for scalar-valued control and constraint and [17, Th. 6.1][5] for vector-valued ones:

$$\mathcal{Q}(v)>0,\qquad \text{for all } v\in\mathcal{C}(\bar u)\setminus\{0\}. \tag{5.34}$$

When the strengthened Legendre-Clebsch condition (5.25) holds, (5.34) implies that $(\bar u,\bar y)$ is a local solution of $(\mathcal{P})$ satisfying the second-order growth condition:

$$\exists\, c,\rho>0,\quad J(u)\ge J(\bar u)+c\|u-\bar u\|_2^2,\quad \text{for all } u\in\mathcal{U}: G(u)\in K,\ \|u-\bar u\|_\infty<\rho. \tag{5.35}$$

This condition involves two norms, L^2 for the growth condition and L^∞ for the neighborhood.

We will use, in the stability analysis, a natural strengthening of the sufficient condition (5.34), omitting the inequality constraint (5.33) in the critical cone. So let the extended critical cone $\hat{\mathcal{C}}(\bar u)$ be defined as the set of $v\in\mathcal{V}$ satisfying (5.32) (and hence, $\mathcal{C}(\bar u)\subset\hat{\mathcal{C}}(\bar u)$). The strong second-order sufficient condition used in the stability analysis is as follows:

$$\mathcal{Q}(v)>0,\qquad \text{for all } v\in\hat{\mathcal{C}}(\bar u)\setminus\{0\}. \tag{5.36}$$

[4]Theorem 1.18 of this thesis.
[5]Theorem 4.24 of this thesis.

Although we call the above condition the *strong* second-order sufficient condition (in comparison with (5.34)), it takes into account the active constraints so it is weaker than the second-order sufficient condition used in [53] that assumes the strict positivity of $\mathcal{Q}$ over the whole space $\mathcal{V} \setminus \{0\}$.

The strengthened Legendre-Clebsch condition (5.25) implies (see [24, Prop. 3.76(i)]) that the quadratic form $\mathcal{Q}$ is a *Legendre form* (see [74]), i.e. a weakly lower semi-continuous (weakly l.s.c.) quadratic form with the property that if a sequence v_n weakly converges to v in L^2 ($v_n \rightharpoonup v$) and if $Q(v_n) \to Q(v)$, then $v_n \to v$ strongly.

Lemma 5.9. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$. An equivalent expression for the quadratic form $\mathcal{Q}$ defined by (5.31), using the alternative multipliers $(\bar{p}^2, \bar{\eta}^2)$ given by (5.13)–(5.14) instead of $(\bar{p}, \bar{\eta})$ and the alternative Hamiltonian (5.16), is:*

$$\mathcal{Q}(v) \;=\; \int_0^T D^2_{(u,y)^2}\tilde{H}(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)(v, z_v)^2 \mathrm{d}t \;+\; \phi_{yy}(\bar{y}(T))(z_v(T), z_v(T)). \tag{5.37}$$

Proof. Let $v \in \mathcal{V}$. Denote by $\tilde{\mathcal{Q}}(v)$ the right-hand side of (5.37) and set $\Delta := \tilde{\mathcal{Q}}(v) - \mathcal{Q}(v)$. In view of the relations (5.13)–(5.14) between $(\bar{p}^2, \bar{\eta}^2)$ and $(\bar{p}, \bar{\eta})$, we have

$$\begin{aligned}
\Delta \;&=\; \int_0^T (\bar{p}^2 - \bar{p}) D^2 f(\bar{u}, \bar{y})(v, z_v)^2 \mathrm{d}t + \int_0^T D^2 g^{(2)}(\bar{u}, \bar{y})(v, z_v)^2 \bar{\eta}^2 \mathrm{d}t \\
&\quad - \int_0^T g_{yy}(\bar{y})(z_v, z_v) \mathrm{d}\bar{\eta} \\
&=\; -\int_0^T \bar{\eta}^1 g_y(\bar{y}) D^2 f(\bar{u}, \bar{y})(v, z_v)^2 \mathrm{d}t - \int_0^T \bar{\eta}^2 g_y^{(1)}(\bar{y}) D^2 f(\bar{u}, \bar{y})(v, z_v)^2 \mathrm{d}t \\
&\quad + \int_0^T D^2 g^{(2)}(\bar{u}, \bar{y})(v, z_v)^2 \bar{\eta}^2 \mathrm{d}t - \int_0^T g_{yy}(\bar{y})(z_v, z_v) \mathrm{d}\bar{\eta}.
\end{aligned}$$

The integration by parts formula in BV [58, p.154] shows that (the calculus is analogous to Lemma 3.6 in [19][6])

$$\begin{aligned}
\int_0^T g_{yy}(\bar{y})(z_v, z_v) \mathrm{d}\bar{\eta} \;&=\; \int_0^T \frac{\mathrm{d}}{\mathrm{d}t}\{g_{yy}(\bar{y})(z_v, z_v)\} \bar{\eta}^1 \mathrm{d}t + [g_{yy}(\bar{y})(z_v, z_v)\bar{\eta}^1]_0^T \\
&=\; \int_0^T \{g_{yyy}(\bar{y})(f, z_v, z_v) + 2 g_{yy}(\bar{y})(Df(\bar{u}, \bar{y})(v, z_v), z_v)\} \bar{\eta}^1 \mathrm{d}t \\
&=\; \int_0^T g_{yy}^{(1)}(\bar{y})(z_v, z_v) \bar{\eta}^1 \mathrm{d}t - \int_0^T g_y(\bar{y}) D^2 f(\bar{u}, \bar{y})(v, z_v)^2 \bar{\eta}^1 \mathrm{d}t.
\end{aligned}$$

Similarly, we obtain that

$$\int_0^T g_{yy}^{(1)}(\bar{y})(z_v, z_v) \bar{\eta}^1 \mathrm{d}t = \int_0^T D^2 g^{(2)}(\bar{u}, \bar{y})(v, z_v)^2 \bar{\eta}^2 \mathrm{d}t - \int_0^T g_y^{(1)}(\bar{y}) D^2 f(\bar{u}, \bar{y})(v, z_v)^2 \bar{\eta}^2 \mathrm{d}t.$$

Summing the two above equalities, we obtain that $\Delta = 0$, which completes the proof. □

[6] Lemma 2.26 of this thesis.

5.4 Stability analysis for the nonlinear problem

According to Def. 5.16 in [24], adapted to our optimal control framework, we consider the following definition of uniform second-order growth condition.

Definition 5.10. Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$. We say that the *uniform second-order (or quadratic) growth* condition holds, if for all stable extensions $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, there exists $c, \rho > 0$ and a neighborhood $\mathcal{N}$ of $\bar{\mu}$, such that for any stationary point (u^μ, y^μ) of $(\mathcal{P}^\mu)$ with $\mu \in \mathcal{N}$ and $\|u^\mu - \bar{u}\|_\infty < \rho$,

$$J^\mu(u) \geq J^\mu(u^\mu) + c\|u - u^\mu\|_2^2, \quad \text{for all } u \in \mathcal{U} : G^\mu(u) \in K,\ \|u - \bar{u}\|_\infty < \rho. \tag{5.38}$$

The next proposition (proved in subsection 5.4.2) shows that the strong second-order sufficient condition (5.36) implies the uniform second-order growth condition. Therefore, if a stationary point for the perturbed problem $(\mathcal{P}^\mu)$ exists, then the latter is *locally unique* in a L^∞-neighborhood of $\bar{u}$, and is a local solution of $(\mathcal{P}^\mu)$.

Proposition 5.11. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3) and the strong second-order sufficient condition* (5.36)*. Then the uniform second-order growth condition holds.*

The difficult part in the stability analysis here is to prove the *existence* of a stationary point for the perturbed problem. For some general optimization problems, Robinson's constraint qualification (5.12) and the uniform quadratic growth condition imply, for a certain class of perturbations, the existence of a stationary point for the perturbed problem, see Bonnans and Shapiro [24, Th. 5.17]. The proof uses Ekeland's variational principle [59]. However, this result does not apply to our nonlinear optimal control problem, due to the *two-norms discrepancy*, but it does apply to linear-quadratic problems (see the proof of Th. 5.23). For the general nonlinear problem, in order to obtain the existence of a stationary point for the perturbed problem, we need to use a variant of Robinson's strong regularity theory [121].

The main result of the paper is the next theorem (proved in section 5.6).

Theorem 5.12. *Let $(\bar{u}, \bar{y})$ be a local solution of $(\mathcal{P})$, satisfying (A2)–(A3) and the strong second-order sufficient condition* (5.36)*. Then for all stable extensions $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, there exist $c, \rho, \kappa, \tilde{\kappa} > 0$ and a neighborhood $\mathcal{N}$ of $\bar{\mu}$, such that for all $\mu \in \mathcal{N}$, $(\mathcal{P}^\mu)$ has a unique stationary point (u^μ, y^μ) with $\|u^\mu - \bar{u}\|_\infty < \rho$ and unique associated alternative multipliers $(p^{2,\mu}, \eta^{2,\mu})$, and for all $\mu, \mu' \in \mathcal{N}$,*

$$\|u^\mu - u^{\mu'}\|_2,\ \|y^\mu - y^{\mu'}\|_{1,2},\ \|p^{2,\mu} - p^{2,\mu'}\|_{1,2},\ \|\eta^{2,\mu} - \eta^{2,\mu'}\|_2 \ \leq\ \kappa\|\mu - \mu'\|, \tag{5.39}$$

$$\|u^\mu - u^{\mu'}\|_\infty, \|y^\mu - y^{\mu'}\|_{1,\infty}, \|p^{2,\mu} - p^{2,\mu'}\|_{1,\infty}, \|\eta^{2,\mu} - \eta^{2,\mu'}\|_\infty \ \leq\ \tilde{\kappa}\|\mu - \mu'\|^{2/3}. \tag{5.40}$$

Moreover, (u^μ, y^μ) is a local solution of $(\mathcal{P}^\mu)$ satisfying the uniform quadratic growth condition (5.38)*.*

The above theorem is obtained by application of a generalized implicit function theorem by Dontchev and Hager [53] (Th. 5.17 of this paper) to the alternative formulation (5.17)–(5.19) in suitable functional spaces described in subsection 5.4.3. In order to show that the main assumption of this theorem is satisfied (assumption (iv)), we have to show that a perturbed linear-quadratic optimal control problem has a unique solution which is Lipschitz continuous w.r.t. the parameter. For this, we will use Prop. 5.11 (or more precisely, its analogous statement adapted to linear-quadratic problems.) Before giving the proof of Prop. 5.11, we first need to study the stability of multipliers (Prop. 5.13).

5.4.1 Stability of multipliers

The next result shows that under the constraint qualification (A2), the stability of multipliers could be deduced from the stability of solutions. Given $r \in [1, +\infty]$, we denote by $\|\cdot\|_{2,r*}$ the norm of the dual space to $W^{2,r}(0,T)$, i.e., for $\mathrm{d}\eta \in \mathcal{M}[0,T]$ we have

$$\|\mathrm{d}\eta\|_{2,r*} \;:=\; \sup\{\, \frac{|\int_0^T \Phi(t)\mathrm{d}\eta(t)|}{\|\Phi\|_{2,r}},\ \Phi \in W^{2,r}(0,T), \Phi \not\equiv 0\,\}.$$

Proposition 5.13. *Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2). Then for every stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, there exists $\nu > 0$ such that for every stationary point (u, y) of $(\mathcal{P}^\mu)$, with (unique) associated multipliers (p, η) and alternative multipliers (p^2, η^2) given by (5.13)–(5.14), the following hold:*
(i) *If $\|\mu - \bar\mu\|, \|u - \bar u\|_\infty < \nu$, then $\mathrm{d}\eta$ is uniformly bounded in $\mathcal{M}[0,T]$;*
(ii) *There exists $\kappa > 0$ such that, for all $\|\mu - \bar\mu\|, \|u - \bar u\|_\infty < \nu$, we have*

$$\|\mathrm{d}\eta - \mathrm{d}\bar\eta\|_{2,1*},\ \|\eta^2 - \bar\eta^2\|_\infty \;\le\; \kappa(\|u - \bar u\|_\infty + \|\mu - \bar\mu\|).$$

Moreover, when $\|\mu - \bar\mu\|, \|u - \bar u\|_\infty \to 0$:
(iii) *$\mathrm{d}\eta$ weakly-* converges to $\mathrm{d}\bar\eta$ ($\mathrm{d}\eta \overset{*}{\rightharpoonup} \mathrm{d}\bar\eta$) in $\mathcal{M}[0,T]$;*
(iv) *$\eta^1 \to \bar\eta^1$ in L^1;*
(v) *p^2 and η^2 converge uniformly to $\bar p^2$ and $\bar\eta^2$, respectively.*

The proof of the above proposition uses the lemma below.

Lemma 5.14. *For all $1 \le r < \infty$, with $r' := r/(r-1)$ ($1' = \infty$), there exists a positive constant C such that*

$$\|\xi\|_{r'} \;\le\; C\|\mathrm{d}\dot\xi\|_{2,r*} \qquad \text{for all } \xi \in BV^2_T[0,T]. \tag{5.41}$$

Proof. Let $\varphi \in L^r(0,T)$. Set $\Phi^1(t) := \int_0^t \varphi(s)\mathrm{d}s$ and $\Phi(t) := \int_0^t \Phi^1(s)\mathrm{d}s$. Then $\Phi \in W^{2,r}(0,T)$, and $\|\Phi\|_{2,r} \le C\|\varphi\|_r$, with $C = 1 + T/\sqrt[r']{r} + (T/\sqrt[r']{r})^2$. Since $\xi(T) = \dot\xi(T) = 0$, the integration by parts formula in BV [58, p.154] implies that, for all $\xi \in BV^2_T[0,T]$,

$$\int_0^T \varphi(t)\xi(t)\mathrm{d}t \;=\; -\int_0^T \Phi^1(t)\dot\xi(t)\mathrm{d}t \;=\; \int_0^T \Phi(t)\mathrm{d}\dot\xi(t).$$

Therefore,

$$\|\xi\|_{r'} = \sup_{\varphi\in L^r, \varphi\not\equiv 0} \frac{|\int_0^T \varphi(t)\xi(t)\mathrm{d}t|}{\|\varphi\|_r} \le C \sup_{\Phi\in W^{2,r}, \Phi\not\equiv 0} \frac{|\int_0^T \Phi(t)\mathrm{d}\dot\xi(t)|}{\|\Phi\|_{2,r}},$$

which gives the result. □

Proof of Prop. 5.13. Let $(\mathcal{P}^\mu)$ be a stable extension of $(\mathcal{P})$. Note first that for $\|\mu - \bar\mu\|$ and $\|u - \bar u\|_\infty$ small enough, assumptions (A1) and (A2) hold for $(\mathcal{P}^\mu)$. This implies the uniqueness of the multipliers (p, η) associated with a stationary point (u, y) of $(\mathcal{P}^\mu)$. Since $(\bar u, \bar y)$ satisfies Robinson's constraint qualification (5.12), point (i) follows from [24, Prop. 4.43].

Let us show (ii). Since $(u, y = y^\mu_u)$ is a stationary point of $(\mathcal{P}^\mu)$, we have that

$$DJ^\mu(u) + DG^\mu(u)^\top \mathrm{d}\eta = 0, \qquad \mathrm{d}\eta \in N_K(G^\mu(u)).$$

It follows that $DG(\bar{u})^\top(\mathrm{d}\bar{\eta}-\mathrm{d}\eta) = DJ^\mu(u)-DJ(\bar{u})+(DG^\mu(u)-DG(\bar{u}))^\top\mathrm{d}\eta$, and hence, for all $v\in L^1(0,T)$,

$$\langle \mathrm{d}\bar{\eta}-\mathrm{d}\eta, DG(\bar{u})v\rangle = (DJ^\mu(u)-DJ(\bar{u}))v + \langle \mathrm{d}\eta, (DG^\mu(u)-DG(\bar{u}))v\rangle. \tag{5.42}$$

Fix $\varepsilon\in(0,\sigma)$ with the σ of (5.21) satisfying (5.23). By Lemma 5.4, the linear mapping defined in (5.24) for $r=1$ is onto. Since $DG(\bar{u})v = g_y(\bar{y})z_v$, by the open mapping theorem, there exists a constant $C_1>0$ such that for all $\Phi\in W^{2,1}(0,T)$, there exists $v\in L^1(0,T)$ such that $DG(\bar{u})v=\Phi$ on Ω_ε and $\|v\|_1\le C_1\|\Phi\|_{2,1}$. For $\|\mu-\bar{\mu}\|, \|u-\bar{u}\|_\infty$ small enough, the contact set $I(g^\mu(y))$, and hence the support of the measure $\mathrm{d}\eta$, are included in the set Ω_ε. Therefore, $\langle \mathrm{d}\eta-\mathrm{d}\bar{\eta}, DG(\bar{u})v\rangle = \langle \mathrm{d}\eta-\mathrm{d}\bar{\eta},\Phi\rangle$. Consequently, by (5.42),

$$|\langle \mathrm{d}\eta-\mathrm{d}\bar{\eta},\Phi\rangle| \le |(DJ^\mu(u)-DJ(\bar{u}))v| + |\mathrm{d}\eta|_{\mathcal{M}}\|(DG^\mu(u)-DG(\bar{u}))v\|_\infty.$$

By point (i), $|\mathrm{d}\eta|_{\mathcal{M}}$ is uniformly bounded, and it is not difficult to check that

$$|(DJ^\mu(u)-DJ(\bar{u}))v|,\ \|(DG^\mu(u)-DG(\bar{u}))v\|_\infty \ \le\ C(\|u-\bar{u}\|_\infty+\|\mu-\bar{\mu}\|)\|v\|_1,$$

where C denotes (possibly different) positive constants. Therefore, we obtain that

$$\begin{array}{rcl} |\langle \mathrm{d}\eta-\mathrm{d}\bar{\eta},\Phi\rangle| & \le & C(\|u-\bar{u}\|_\infty+\|\mu-\bar{\mu}\|)\|v\|_1 \\ & \le & CC_1(\|u-\bar{u}\|_\infty+\|\mu-\bar{\mu}\|)\|\Phi\|_{2,1}. \end{array}$$

Consequently, $\|\mathrm{d}\eta-\mathrm{d}\bar{\eta}\|_{2,1*}\le CC_1(\|u-\bar{u}\|_\infty+\|\mu-\bar{\mu}\|)$, and since by Lemma 5.14, $\|\eta^2-\bar{\eta}^2\|_\infty\le C\|\mathrm{d}\eta-\mathrm{d}\bar{\eta}\|_{2,1*}$, this proves (ii).

Now consider a sequence $\mu_n\to\bar{\mu}$, and let (u_n,y_n) be a stationary point of $(\mathcal{P}^{\mu_n})$ such that $u_n\to\bar{u}$ in L^∞, with (unique) multipliers (p_n,η_n) and alternative multipliers (p_n^2,η_n^2). Since $W^{2,1}(0,T)$ is dense in $C[0,T]$, we deduce easily from point (ii) that $\mathrm{d}\eta_n \overset{*}{\rightharpoonup} \mathrm{d}\bar{\eta}$ in $\mathcal{M}[0,T]$, which shows (iii). By the compactness Theorem in BV [2, Th. 3.23], it follows that $\eta_n^1\to\bar{\eta}^1$ in L^1, which shows (iv). Finally, since η^2 is given by (5.13), (iv) implies that $\eta_n^2\to\bar{\eta}^2$ uniformly. By (5.17) and by Gronwall's Lemma, we conclude that $p_n^2\to\bar{p}^2$ in $W^{1,\infty}$, which achieves the proof of (v). □

5.4.2 The uniform second-order growth condition (proof of Prop. 5.11)

The proof of Prop. 5.11 uses the auxiliary result below. Given $A,B\subset[0,T]$, denote by $\mathrm{exc}\{A,B\}$ the *Hausdorff excess* of A over B, defined by

$$\mathrm{exc}\{A,B\} \ :=\ \sup_{t\in A}\inf_{s\in B}|t-s|, \tag{5.43}$$

with the convention $\mathrm{exc}\{\emptyset,B\}=0$.

Lemma 5.15. *Let $\mathrm{d}\bar{\eta}\in\mathcal{M}[0,T]$, and a sequence $(\mathrm{d}\eta_n)\subset\mathcal{M}[0,T]$ be such that $\mathrm{d}\eta_n$ weakly-* converges to $\mathrm{d}\bar{\eta}$ in $\mathcal{M}[0,T]$. Then $e_n := \mathrm{exc}\{\mathrm{supp}(\mathrm{d}\bar{\eta}),\mathrm{supp}(\mathrm{d}\eta_n)\}$ converges to zero when $n\to+\infty$.*

Proof. The result follows from classical compactness arguments. By contradiction, assume that the result is false. Then there exist $\varepsilon_0>0$ and a subsequence, still denoted by $\mathrm{d}\eta_n$, such that for all $n\in\mathbb{N}^*$, $e_n>\varepsilon_0$, i.e. there exists $t_n\in\mathrm{supp}(\mathrm{d}\bar{\eta})$ such that for all $s\in\mathrm{supp}(\mathrm{d}\eta_n)$, $|t_n-s|>\varepsilon_0$. The sequence $(t_n)_{n\in\mathbb{N}^*}\subset[0,T]$ being bounded, assume w.l.o.g. that $t_n\to\bar{t}\in[0,T]$. Since $\mathrm{supp}(\mathrm{d}\bar{\eta})$ is closed, $\bar{t}\in\mathrm{supp}(\mathrm{d}\bar{\eta})$. For n large enough, $|t_n-\bar{t}|<\varepsilon_0/2$, and hence,

for all $s \in \mathrm{supp}(\mathrm{d}\eta_n)$, $|\bar{t}-s| \geq |t_n - s| - |t_n - \bar{t}| > \varepsilon_0/2$. Let φ be a continuous function, with support in $[\bar{t}-\varepsilon_0/2, \bar{t}+\varepsilon_0/2]$, and such that $\int_0^T \varphi \mathrm{d}\bar{\eta} \neq 0$. Since $\mathrm{dist}\{\bar{t}, \mathrm{supp}(\mathrm{d}\eta_n)\} > \varepsilon_0/2$ for all large enough n, $\int_0^T \varphi \mathrm{d}\eta_n = 0$. But $\mathrm{d}\eta_n \overset{*}{\rightharpoonup} \mathrm{d}\bar{\eta}$, implying that $\int_0^T \varphi \mathrm{d}\eta_n \to \int_0^T \varphi \mathrm{d}\bar{\eta}$, which gives the desired contradiction. □

Remark 5.16. We may equivalently reformulate Lemma 5.15 as follows: if $\mathrm{d}\eta_n$ weakly-* converges to $\mathrm{d}\bar{\eta}$ in $\mathcal{M}[0,T]$, then

$$\mathrm{supp}(\mathrm{d}\bar{\eta}) \ \subset \ \limsup_{n\to+\infty} \ \mathrm{supp}(\mathrm{d}\eta_n),$$

where the lim sup is in the sense of Painlevé-Kuratowski.

Proof of Prop. 5.11. We argue by contradiction. If the uniform second-order growth condition does not hold, there exist a stable extension $(\mathcal{P}^\mu)$, a sequence $\mu_n \to \bar{\mu}$, a stationary point (u_n, y_n) of $(\mathcal{P}^{\mu_n})$ such that $u_n \to \bar{u}$ in L^∞, with multipliers (p_n, η_n) and alternative multipliers (p_n^2, η_n^2), and a feasible point $(\hat{u}_n, \hat{y}_n)$ of $(\mathcal{P}^{\mu_n})$ such that

$$J^{\mu_n}(\hat{u}_n) \ < \ J^{\mu_n}(u_n) + o(\|\hat{u}_n - u_n\|_2^2). \tag{5.44}$$

Introducing the Lagrangian of $(\mathcal{P}^\mu)$, $L^\mu(u,\eta) = J^\mu(u) + \langle \mathrm{d}\eta, G^\mu(u)\rangle$, and using that $\mathrm{d}\eta_n \in N_K(G^{\mu_n}(u_n))$, (5.44) implies that

$$L^{\mu_n}(\hat{u}_n, \eta_n) - L^{\mu_n}(u_n, \eta_n) \ \leq \ J^{\mu_n}(\hat{u}_n) - J^{\mu_n}(u_n) \ < \ o(\|\hat{u}_n - u_n\|_2^2).$$

Set $\varepsilon_n := \|\hat{u}_n - u_n\|_2 \to 0$ and $v_n := \varepsilon_n^{-1}(\hat{u}_n - u_n)$. A second-order expansion of the Lagrangian shows that $L^{\mu_n}(\hat{u}_n, \eta_n) - L^{\mu_n}(u_n, \eta_n) = \varepsilon_n^2 \mathcal{Q}^{\mu_n}(v_n) + o(\varepsilon_n^2)$, where the quadratic form $\mathcal{Q}^{\mu_n}$ is defined like (5.31) for the stationary point (u_n, y_n) of $(\mathcal{P}^{\mu_n})$. Therefore, dividing the above inequality by ε_n^2, we obtain that

$$\mathcal{Q}^{\mu_n}(v_n) \ \leq \ o(1). \tag{5.45}$$

Since $\|v_n\|_2 = 1$ for all n, taking a subsequence if necessary, we may assume w.l.o.g. that $v_n \rightharpoonup \bar{v}$ weakly in L^2 for some $\bar{v} \in \mathcal{V}$ when $n \to +\infty$. Since by Lemma 5.9, $\mathcal{Q}^{\mu_n}$ can also be expressed by (5.37), and $(u_n, y_n, p_n^2, \eta_n^2) \to (\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)$ uniformly by Prop. 5.13(v), and since v_n is bounded in L^2, it follows that $\mathcal{Q}^{\mu_n}(v_n) - \mathcal{Q}(v_n) \to 0$. Therefore, writing that $\mathcal{Q}^{\mu_n}(v_n) = \mathcal{Q}(v_n) + (\mathcal{Q}^{\mu_n}(v_n) - \mathcal{Q}(v_n))$, and using that $\mathcal{Q}$ is a Legendre form and hence weakly l.s.c., we obtain by (5.45) that

$$\mathcal{Q}(\bar{v}) \ \leq \ 0. \tag{5.46}$$

Moreover, since $v_n \rightharpoonup \bar{v}$ weakly in L^2, and $(u_n, y_n) \to (\bar{u}, \bar{y})$ uniformly, the linearized state z_n, solution of

$$\dot{z}_n = f_y^{\mu_n}(u_n, y_n) z_n + f_u^{\mu_n}(u_n, y_n) v_n \quad \text{a.e. on } [0,T], \quad z_n(0) = 0$$

converges weakly to $\bar{z} := z_{\bar{v}}$ in H^1, and hence uniformly. Since $G^{\mu_n}(\hat{u}_n) \in K$, we have that $0 \geq G^{\mu_n}(\hat{u}_n) - G^{\mu_n}(u_n) = \varepsilon_n DG^{\mu_n}(u_n) v_n + \varepsilon_n r_n$ on $\mathrm{supp}(\mathrm{d}\eta_n)$, with $\|r_n\|_\infty = \mathcal{O}(\varepsilon_n)$. Since $DG^{\mu_n}(u_n)v_n = g_y^{\mu_n}(y_n) z_n$, it follows that

$$g_y^{\mu_n}(y_n) z_n + r_n \leq 0 \quad \text{on } \mathrm{supp}(\mathrm{d}\eta_n). \tag{5.47}$$

Since $\frac{\mathrm{d}}{\mathrm{d}t} g_y^{\mu_n}(y_n(t))z_n(t) = (g^{\mu_n})_y^{(1)}(y_n)z_n$ is uniformly bounded over $[0,T]$, the functions (of time) $g_y^{\mu_n}(y_n)z_n$ are uniformly Lipschitz continuous over $[0,T]$. Therefore,

$$\begin{aligned}\sup_{\mathrm{supp}(\mathrm{d}\bar\eta)} g_y(\bar y)\bar z &\le \|g_y(\bar y)\bar z - g_y^{\mu_n}(y_n)z_n\|_\infty + \|(g^{\mu_n})_y^{(1)}(y_n)z_n\|_\infty e_n + \sup_{\mathrm{supp}(\mathrm{d}\eta_n)} g_y^{\mu_n}(y_n)z_n \\ &\le o(1) + \mathcal{O}(e_n) + \mathcal{O}(\varepsilon_n),\end{aligned}$$

where $e_n := \mathrm{exc}\{\mathrm{supp}(\mathrm{d}\bar\eta), \mathrm{supp}(\mathrm{d}\eta_n)\}$ is defined by (5.43). Since $\mathrm{d}\eta_n \overset{*}{\rightharpoonup} \mathrm{d}\bar\eta$ by Prop. 5.13(iii), it follows from Lemma 5.15 that $e_n \to 0$. Therefore, we obtain that

$$g_y(\bar y)\bar z \le 0 \qquad \text{on } \mathrm{supp}(\mathrm{d}\bar\eta). \tag{5.48}$$

In addition, by (5.44), $DJ^{\mu_n}(u_n)v_n \le \mathcal{O}(\varepsilon_n)$. Since $DJ^{\mu_n}(u_n) + DG^{\mu_n}(u_n)^\top \mathrm{d}\eta_n = 0$, it follows that $\langle \mathrm{d}\eta_n, DG^{\mu_n}(u_n)v_n\rangle = \int_0^T g_y^{\mu_n}(y_n)z_n \mathrm{d}\eta_n \ge \mathcal{O}(\varepsilon_n)$. Since $\mathrm{d}\eta_n \overset{*}{\rightharpoonup} \mathrm{d}\bar\eta$ and $g_y^{\mu_n}(y_n)z_n \to g_y(\bar y)\bar z$ uniformly, we obtain that $\int_0^T g_y(\bar y)\bar z \mathrm{d}\bar\eta \ge 0$. Using that $\mathrm{d}\bar\eta \ge 0$, (5.48) implies that

$$g_y(\bar y)\bar z = 0 \qquad \text{on } \mathrm{supp}(\mathrm{d}\bar\eta),$$

i.e. $\bar v \in \hat{\mathcal{C}}(\bar u)$. The strong second-order sufficient condition (5.36) and (5.46) imply then that $\bar v = 0$. But then $\mathcal{Q}(\bar v) = 0$, and $\mathcal{Q}(v_n) \to \mathcal{Q}(\bar v)$. Since $\mathcal{Q}$ is a Legendre form, we deduce that $v_n \to \bar v = 0$ strongly in L^2, contradicting that $\|v_n\|_2 = 1$ for all n. □

5.4.3 The strong regularity framework

We use the following generalized implicit function theorem in metric spaces by Dontchev and Hager [53], which is a variant of Robinson's strong regularity [121].

Theorem 5.17 ([53], Th. 2.2). *Let X be a complete metric space, $\tilde X$ a closed subset of X, W a linear metric space, Δ a subset of W, P a metric space, $\mathcal{F} : X \times P \to W$, $\mathcal{N} : X \to 2^W$, $\mathcal{L} : X \to W$. Assume that $\mathcal{L}$ is continuous and that there exists $(\bar x, \bar\mu) \in \tilde X \times P$ such that:*
(i) $\mathcal{F}(\bar x, \bar\mu) \in \mathcal{N}(\bar x)$;
(ii) $\mathcal{F}(\bar x, \cdot)$ *is continuous at* $\bar\mu$;
(iii) $\Psi^\mu := \mathcal{F}(\cdot, \mu) - \mathcal{L}(\cdot)$ *is strictly stationary at* $x = \bar x$, *uniformly in* μ *near* $\bar\mu$, *i.e. for all* $\varepsilon > 0$, *there exists* $\nu > 0$ *such that if* $\|x_i - \bar x\|_X$, $\|\mu - \bar\mu\| \le \nu$, $i = 1, 2$,

$$\|\Psi^\mu(x_1) - \Psi^\mu(x_2)\|_W \le \varepsilon \|x_1 - x_2\|_X. \tag{5.49}$$

(iv) *For all* $\delta \in \Delta$, *there exists a unique solution* $x \in \tilde X$ *of*

$$\delta \in \mathcal{L}(x) - \mathcal{N}(x), \tag{5.50}$$

and there exists $\lambda > 0$ *such that, with* x_δ *the unique solution associated with* δ,

$$\|x_\delta - x_{\delta'}\|_X \le \lambda \|\delta - \delta'\|_W, \quad \forall\, \delta, \delta' \in \Delta.$$

(v) $\mathcal{F} - \mathcal{L}$ *maps a neighborhood of* $(\bar x, \bar\mu)$ *into* Δ.
Then for all $\lambda_+ > \lambda$, *there exist neighborhoods* $\mathcal{X}$ *of* $\bar x$ *in* $\tilde X$ *and* $\mathcal{W}$ *of* $\bar\mu$, *such that for each* $\mu \in \mathcal{W}$, *there exists a unique* $x \in \mathcal{X}$ *satisfying* $\mathcal{F}(x, \mu) \in \mathcal{N}(x)$; *moreover, for each* $\mu_i \in \mathcal{W}$, $i = 1, 2$, *if* x_i *denotes the* $x \in \mathcal{X}$ *associated with* μ_i, *then*

$$\|x_2 - x_1\|_X \le \lambda_+ \|\mathcal{F}(x_1, \mu_1) - \mathcal{F}(x_1, \mu_2)\|_W. \tag{5.51}$$

In [53], the theorem is stated with $\tilde{X} = X$, but remains true if we replace the complete metric space X by any closed subset $\tilde{X}$ of X, equipped with the metric of X, since $\tilde{X}$ remains a complete metric space.

This theorem was used for stability analysis of optimal control problems subject to first-order state constraints in [53]. In what follows, we describe a suitable framework to apply Th. 5.17 for second-order state constraints.

Remark 5.18. Our choice of functional spaces to apply Th. 5.17 differs from that of [53] or [88] in the spaces for the state constraint and state constraint multiplier. Whereas in [53, 88] the state constraint is seen in $W^{1,\infty}$, we consider here rather the state constraint in the space of continuous functions $C[0,T]$. Another natural choice for the space of second-order state constraints would be $W^{2,\infty}$ since the constraint is "onto" in this space (Lemma 5.4). The reason for considering here the constraint in $C[0,T]$ is to have multipliers in $\mathcal{M}[0,T]$ instead of in the dual space of $W^{1,\infty}$ or $W^{2,\infty}$. For first-order state constraints it can be shown (see [65]) that the state constraint multiplier η lies in $W^{1,\infty}$ (and therefore a suitable choice for the state constraint multiplier space is the space Lip_k defined below), but this is no more true for higher-order state constraints. Note that since $W^{2,\infty} \subset W^{1,\infty} \subset C[0,T]$ with continuous and dense embeddings, and the constraint is "onto" in $W^{2,\infty}$ by Lemma 5.4, the multipliers in the three possible formulations are one-to-one.

Notation In order to apply Th. 5.17 to prove Th. 5.12 in sections 5.5 and 5.6, we use the following notation. Given $k, l, r, \varrho, k' > 0$, define the spaces

$$\begin{aligned}
\mathrm{Lip}_k(0,T) &:= \{u \in W^{1,\infty}(0,T) : \|\dot{u}\|_\infty \le k\},\\
BV^2_{T,l}[0,T] &:= \{\xi \in BV^2_T[0,T] : |\mathrm{d}\dot{\xi}|_{\mathcal{M}} \le l\},\\
X &:= \mathrm{Lip}_k(0,T;\mathbb{R}^m) \times BV^2_{T,l}[0,T], \qquad (5.52)\\
\tilde{X} &:= \{x = (u,\xi) \in X : \|u - \bar{u}\|_2 \le r\}, \qquad (5.53)\\
W &:= L^2(0,T;\mathbb{R}^{m*}) \times H^2(0,T) \qquad (5.54)
\end{aligned}$$

equipped with its standard norm $\|\delta\|_W := \|\gamma\|_2 + \|\zeta\|_{2,2}$ for $\delta = (\gamma,\zeta) \in W$,

$$\begin{aligned}
\Delta &:= \{\delta \in \mathrm{Lip}_{k'}(0,T;\mathbb{R}^{m*}) \times H^2(0,T),\ \|\delta\|_W \le \varrho\}, \qquad (5.55)\\
P &: \text{ closed neighborhood of } \bar{\mu}, \text{ contained in } M_0,
\end{aligned}$$

and mappings

- $\mathcal{F} : X \times P \to W$,
$$\mathcal{F}(x,\mu) := \begin{pmatrix} \tilde{H}^\mu_u(u, y^\mu_u, p^{2,\mu}_{u,\eta^2}, \eta^2) \\ g^\mu(y^\mu_u) \end{pmatrix},$$
where $\tilde{H}^\mu$ is the alternative Hamiltonian (5.16) of $(\mathcal{P}^\mu)$, y^μ_u is the solution of the state equation (5.28) and $p^{2,\mu}_{u,\eta^2}$ is the solution of the alternative costate equation (5.17) for $(\mathcal{P}^\mu)$, i.e.:
$$-\dot{p}^{2,\mu}_{u,\eta^2} = \tilde{H}^\mu_y(u, y^\mu_u, p^{2,\mu}_{u,\eta^2}, \eta^2) \ \text{ a.e. on } [0,T], \quad p^{2,\mu}_{u,\eta^2}(T) = \phi^\mu_y(y^\mu_u(T)). \qquad (5.56)$$

- $\mathcal{N} : X \to 2^W$, $\mathcal{N}(x) = \{0\} \times (N_{K^-}(\mathrm{d}\dot{\eta}^2) \cap H^2(0,T))$, where
$$N_{K^-}(\mathrm{d}\dot{\eta}^2) = \begin{cases} \{\varphi \in C_-[0,T] : \langle \mathrm{d}\dot{\eta}^2, \varphi\rangle = 0\} & \text{if } \mathrm{d}\dot{\eta}^2 \ge 0,\\ \emptyset & \text{otherwise.} \end{cases}$$

- $\mathcal{L} : X \to W$,

$$\mathcal{L}(x) \;:=\; \mathcal{F}(\bar{x},\bar{\mu}) - D_x\mathcal{F}(\bar{x},\bar{\mu})(x-\bar{x}). \tag{5.57}$$

By Lemma 5.6, we have that $(\bar{u},\bar{\eta}^2) \in X$ for sufficiently large k, l.

Lemma 5.19. *Equipped with the norm*

$$\|(u,\xi)\|_X := \|u\|_2 + \|\xi\|_2, \tag{5.58}$$

X is a complete metric space, and

$$\|u\|_\infty \;\le\; \max\{\sqrt{3/T}\|u\|_2, \sqrt[3]{3k}\|u\|_2^{2/3}\}, \qquad \textit{for all } u \in \mathrm{Lip}_k(0,T). \tag{5.59}$$

Proof. It was shown in [53, Lemma 3.2] that the space $(\mathrm{Lip}_k(0,T), \|\cdot\|_2)$ is a complete metric space, and the estimate (5.59) follows from [53, Lemma 3.1]. We show now that $(BV^2_{T,l}[0,T], \|\cdot\|_2)$ is complete as well. Let (ξ_n) be a Cauchy sequence in $BV^2_{T,l}[0,T]$ (for the norm $\|\cdot\|_2$). Since $L^2(0,T)$ is complete, there exists $\tilde{\xi} \in L^2(0,T)$ such that $\xi_n \to \tilde{\xi}$ in L^2. Let us show that the limit point $\tilde{\xi}$ lies in $BV^2_{T,l}[0,T]$. We have that $|\mathrm{d}\dot{\xi}_n|_{\mathcal{M}} \le l$ for all n, and since $\dot{\xi}_n(T) = 0$, the sequence $(\dot{\xi}_n)$ is bounded in BV for the norm $\|\eta\|_{BV} := \|\eta\|_1 + |\mathrm{d}\eta|_{\mathcal{M}}$. Therefore, by the compactness theorem in BV [2, Th. 3.23], there exists a subsequence $\xi_{\psi(n)}$ and $\zeta \in BV[0,T]$ such that $\mathrm{d}\dot{\xi}_{\psi(n)} \overset{*}{\rightharpoonup} \mathrm{d}\zeta$ weakly-* in $\mathcal{M}[0,T]$ and $\dot{\xi}_{\psi(n)} \to \zeta$ in L^1. Moreover, using the integration by parts formula in BV [58, p.154], we obtain that

$$T\zeta(T) = \int_0^T (\zeta(t) - \dot{\xi}_{\psi(n)}(t))\mathrm{d}t + \int_0^T s(\mathrm{d}\zeta(s) - \mathrm{d}\dot{\xi}_{\psi(n)}(s)) \;\to\; 0,$$

and hence $\zeta(T) = 0$. Setting $\hat{\xi}(t) := -\int_t^T \zeta(s)\mathrm{d}s$, we have that $\hat{\xi} \in BV^2_T[0,T]$, and $\xi_{\psi(n)} \to \hat{\xi}$ in L^∞ and a fortiori in L^2. We deduce that necessarily, $\hat{\xi} = \tilde{\xi} \in BV^2_T[0,T]$, the whole sequence $(\mathrm{d}\dot{\xi}_n)$ weakly-* converges to $\mathrm{d}\dot{\tilde{\xi}}$ in $\mathcal{M}[0,T]$, and then

$$|\mathrm{d}\dot{\tilde{\xi}}|_{\mathcal{M}} \;\le\; \liminf |\mathrm{d}\dot{\xi}_n|_{\mathcal{M}} \;\le\; l.$$

This shows that $\tilde{\xi} \in BV^2_{T,l}[0,T]$, and hence, $(BV^2_{T,l}[0,T], \|\cdot\|_2)$ is a complete metric space. This achieves the proof. □

Note that for all $\xi \in BV^2_{T,l}[0,T]$, we have that $|\mathrm{d}\dot{\xi}|_{\mathcal{M}} \le l$, and since $\dot{\xi}(T) = 0$, it follows that $\|\dot{\xi}\|_\infty \le l$, and hence, $BV^2_{T,l}[0,T] \subset \mathrm{Lip}_l(0,T)$. Therefore, we deduce from (5.59) that

$$\|\xi\|_\infty \;\le\; \max\{\sqrt{3/T}\|\xi\|_2, \sqrt[3]{3l}\|\xi\|_2^{2/3}\}, \qquad \text{for all } \xi \in BV^2_{T,l}[0,T]. \tag{5.60}$$

The space $\tilde{X}$ defined by (5.53) is a closed subset of X, and hence, by Lemma 5.19, $\tilde{X}$ equipped with the norm of X (5.58) is a complete metric space. We need to work with $\tilde{X}$ instead of X in order to obtain the *uniqueness* of a solution of (5.50) in $\tilde{X}$, for small enough $r > 0$. The space of sufficiently smooth variations $\Delta \subset W$, in assumptions (iv) and (v) of Th. 5.17, is defined by (5.55).

Given a stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, our formulation is the following: For μ in the neighborhood of $\bar{\mu}$, find $x = (u,\eta^2) \in \tilde{X}$ solution of

$$\mathcal{F}(x,\mu) \in \mathcal{N}(x), \tag{5.61}$$

where $\mathcal{F}$ and $\mathcal{N}$ are defined as above. Then (u, y_u^μ) is a stationary point of $(\mathcal{P}^\mu)$ with alternative multipliers $(p^{2,\mu}_{u,\eta^2}, \eta^2)$ iff $x = (u,\eta^2)$ is solution of (5.61).

5.5 Stability analysis of linear-quadratic problems

The verification of assumption (iv) of Th. 5.17 is strongly related to stability analysis of linear-quadratic optimal control problems with a second-order state constraint, that we study in this section. Since these results have their own interest, they are stated independently of the rest of the paper. The problem under consideration is of the form:

$$(\mathcal{P}_\delta) \quad \min_{(v,z)\in\mathcal{V}\times\mathcal{Z}} \quad \tfrac{1}{2}\int_0^T (v(t)^\top S(t)v(t) + 2v(t)^\top R(t)z(t) + z(t)^\top Q(t)z(t))\mathrm{d}t \tag{5.62}$$

$$+ \int_0^T (a(t)z(t) + (b(t)-\gamma(t))v(t))\mathrm{d}t + \tfrac{1}{2}z(T)^\top \Phi z(T) \tag{5.63}$$

$$\text{s.t.} \quad \dot z(t) = A(t)z(t) + B(t)v(t) \quad \text{a.e. on } [0,T], \quad z(0) = 0 \tag{5.64}$$

$$C(t)z(t) + d(t) - \zeta(t) \le 0 \quad \text{on } [0,T]. \tag{5.65}$$

The perturbation parameter is here $\delta = (\gamma,\zeta) \in W = L^2(0,T;\mathbb{R}^{m*}) \times H^2(0,T)$, with the norm $\|\delta\|_W = \|\gamma\|_2 + \|\zeta\|_{2,2}$. The control and state spaces for the linearized problem are $\mathcal{V} := L^2(0,T;\mathbb{R}^m)$ and $\mathcal{Z} := H^1(0,T;\mathbb{R}^n)$. The state constraint (5.65) is scalar-valued. The matrix and vectors $S(\cdot), R(\cdot), Q(\cdot), a(\cdot), b(\cdot), A(\cdot), B(\cdot), C(\cdot), d(\cdot)$, of appropriate dimensions, are Lipschitz continuous functions of time. In addition, $C(\cdot)$ and $d(\cdot)$ lie in the space $W^{3,\infty}$ and $A(\cdot)$ in $W^{2,\infty}$. The matrix S and Q are symmetric. We assume in addition in all this section that (recall (A1))

$$d(0) < 0. \tag{5.66}$$

Given $v \in \mathcal{V}$, we denote by z_v the unique solution in $\mathcal{Z}$ of the linearized state equation (5.64). Then we may write $(\mathcal{P}_\delta)$ as follows:

$$(\mathcal{P}_\delta) \qquad \min_{v\in\mathcal{V}} \mathcal{J}^\delta(v), \quad \Gamma^\delta(v) \in K,$$

with $\mathcal{J}^\delta(v) := \int_0^T \{\frac{1}{2}(v^\top Sv + 2v^\top Rz_v + z_v^\top Qz_v) + az + (b-\gamma)v\}\mathrm{d}t + \frac{1}{2}z_v(T)^\top \Phi z_v(T)$, $\Gamma^\delta(v) := Cz_v + d - \zeta$ and $K = C_-[0,T]$.

Assume that $C(t)B(t) \equiv 0$ on $[0,T]$ (state constraint of second-order), and define the matrix:

$$C_1(t) := \dot C(t) + C(t)A(t), \quad C_2(t) := \dot C_1(t) + C_1(t)A(t), \quad N_2(t) := C_1(t)B(t).$$

Then for all $v \in \mathcal{V}$, we have that

$$\frac{\mathrm{d}}{\mathrm{d}t}\{C(t)z_v(t)\} = C_1(t)z_v(t), \qquad \frac{\mathrm{d}^2}{\mathrm{d}t^2}\{C(t)z_v(t)\} = C_2(t)z_v(t) + N_2(t)v(t).$$

The alternative multipliers $(\pi^2,\eta^2) \in W^{1,\infty}(0,T;\mathbb{R}^{n*}) \times BV_T^2[0,T]$ for the linear-quadratic problem are defined by

$$\eta^1(t) := \int_{(t,T]} \mathrm{d}\eta(s), \qquad \eta^2(t) := \int_t^T \eta^1(s)\mathrm{d}s \tag{5.67}$$

$$\pi^2(t) := \pi(t) - \eta^1(t)C(t) - \eta^2(t)C_1(t), \qquad t \in [0,T]. \tag{5.68}$$

Let $(\bar v, \bar z = z_{\bar v})$ be a stationary point of $(\mathcal{P}_0)$, with multipliers $(\bar\pi,\bar\eta)$ and alternative multipliers $(\bar\pi^2,\bar\eta^2)$. Denote the contact set by $\Omega := \{t \in [0,T] : C(t)\bar z(t) + d(t) = 0\}$, and a neighborhood of the contact set by $\Omega_\sigma := \{t \in [0,T] : \mathrm{dist}\{t,\Omega\} < \sigma\}$ for $\sigma > 0$. For linear-quadratic problems, assumptions (A2)–(A3) may be rewritten as follows:

($\tilde{\text{A}}2$) The state constraint is a regular second-order state constraint, i.e. $C(t)B(t) \equiv 0$ on $[0,T]$, and there exists $\beta, \sigma > 0$ (σ satisfying (5.23)) such that

$$|N_2(t)| \geq \beta \quad \text{on } \Omega_\sigma.$$

($\tilde{\text{A}}3$) The matrix $S(t)$ is uniformly positive definite over $[0,T]$, i.e.,

$$\exists\, \alpha > 0, \quad v^\top S(t) v \geq \alpha |v|^2, \quad \text{for all } t \in [0,T] \text{ and all } v \in \mathbb{R}^m.$$

Note that by Rem. 5.5, ($\tilde{\text{A}}3$) is equivalent to (A3). Assumption ($\tilde{\text{A}}2$) (together with (5.66)) imply the following (cf Lemma 5.4):

Lemma 5.20. *Assume that* ($\tilde{\text{A}}2$) *holds. Then there exists a positive constant c such that for all $\varphi \in H^2(0,T)$, there exists $v \in \mathcal{V}$ satisfying*

$$C(t) z_v(t) = \varphi(t) \quad \text{on } \Omega_\sigma \qquad \text{and} \qquad \|v\|_2 \leq c \|\varphi\|_{2,2}. \tag{5.69}$$

Therefore ($\tilde{\text{A}}2$) (and (5.66)) imply that Robinson's constraint qualification holds, and that the multipliers associated with $(\bar{v}, \bar{z})$ are unique.

Propositions 5.21 and 5.22 below hold for a larger set of perturbations, more precisely for $\delta = (\gamma, \zeta) \in \hat{W}$, where

$$\hat{W} := L^2(0,T;\mathbb{R}^m) \times C[0,T],$$

equipped with its standard norm $\|\delta\|_{\hat{W}} := \|\gamma\|_2 + \|\zeta\|_\infty$. We have of course $W \subset \hat{W}$ with continuous embedding. Identical to Prop. 5.13, we obtain the stability of multipliers for linear-quadratic problems (with a slightly modified statement).

Proposition 5.21. *Let $(\bar{v}, \bar{z})$ be a stationary point of $(\mathcal{P}_0)$ satisfying* ($\tilde{\text{A}}2$). *Then there exists $\nu > 0$ such that for every stationary point (v,z) of $(\mathcal{P}_\delta)$, with (unique) multipliers (π, η) and alternative multipliers (π^2, η^2) defined by* (5.68)–(5.67), *the following hold:*
(i) *If $\|\delta\|_{\hat{W}}, \|v - \bar{v}\|_2 < \nu$, then $\mathrm{d}\eta$ is uniformly bounded in $\mathcal{M}[0,T]$;*
(ii) *There exists $\kappa > 0$ such that, for all $\|\delta\|_{\hat{W}}, \|v - \bar{v}\|_2 < \nu$, we have*

$$\|\mathrm{d}\eta - \mathrm{d}\bar{\eta}\|_{2,2*},\ \|\eta^2 - \bar{\eta}^2\|_2 \leq \kappa(\|v - \bar{v}\|_2 + \|\delta\|_{\hat{W}}).$$

Moreover, when $\|\delta\|_{\hat{W}}, \|v - \bar{v}\|_2 \to 0$:
(iii) *$\mathrm{d}\eta$ weakly-* converges to $\mathrm{d}\bar{\eta}$ in $\mathcal{M}[0,T]$;*
(iv) *$\eta^1 \to \bar{\eta}^1$ in L^1;*
(v) *π^2 and η^2 converges uniformly to $\bar{\pi}^2$ and $\bar{\eta}^2$, respectively.*

Second-order optimality conditions

Let $\tilde{\mathcal{Q}}$ denote the quadratic part of the cost $\mathcal{J}^\delta$ (independent of δ):

$$\begin{aligned} \tilde{\mathcal{Q}}(v) \;=\; & \tfrac{1}{2} \int_0^T (v(t)^\top S(t) v(t) + 2 v(t)^\top R(t) z_v(t) + z_v(t)^\top Q(t) z_v(t)) \mathrm{d}t \\ & + \tfrac{1}{2} z_v(T)^\top \Phi z_v(T). \end{aligned} \tag{5.70}$$

The strong second-order sufficient condition is:

$$\tilde{\mathcal{Q}}(v) > 0, \qquad \text{for all } v \in \mathcal{V} \setminus \{0\} \text{ such that } C(t) z_v(t) = 0 \text{ on } \operatorname{supp}(\mathrm{d}\bar{\eta}). \tag{5.71}$$

Identical to Prop. 5.11, we obtain that the second-order sufficient condition (5.71) implies the uniform second-order growth condition for the perturbed problems $(\mathcal{P}_\delta)$ (here again the statement is slightly modified).

Proposition 5.22. *Let $(\bar{v}, \bar{z})$ be a stationary point of $(\mathcal{P}_0)$ satisfying $(\tilde{\text{A}}2)$-$(\tilde{\text{A}}3)$ and the strong second-order sufficient condition (5.71). Then there exist $c, \rho > 0$ and a neighborhood $\mathcal{W}$ of 0 in $\hat{W}$, such that for all $\delta \in \mathcal{W}$ and any stationary point (v_δ, z_δ) of $(\mathcal{P}_\delta)$ with $\|v_\delta - \bar{v}\|_2 < \rho$,*

$$\mathcal{J}^\delta(v) \geq \mathcal{J}^\delta(v_\delta) + c\|v - v_\delta\|_2^2, \quad \forall\, v \in \mathcal{V} : \Gamma^\delta(v) \in K,\ \|v - \bar{v}\|_2 < \rho. \tag{5.72}$$

Stability Analysis

The main result of this section is the theorem below. The key point to show the existence of a stationary point for the perturbed linear-quadratic problem under the weak second-order sufficient condition (5.71), where the active constraints are taken into account. To this end, the uniform growth condition (Prop. 5.22), together with an abstract theorem from Bonnans and Shapiro [24, Th. 5.17 and Rem. 5.19] is used.

Theorem 5.23. *Let $(\bar{v}, \bar{z})$ be a stationary point of $(\mathcal{P}_0)$ satisfying $(\tilde{\text{A}}2)$-$(\tilde{\text{A}}3)$ and the strong second-order sufficient condition (5.71). Then there exist $c, \rho, \lambda > 0$ and a neighborhood $\mathcal{W}$ of 0 in W, such that for all $\delta \in \mathcal{W}$, $(\mathcal{P}_\delta)$ has a unique stationary point (v_δ, z_{v_δ}) with $\|v_\delta - \bar{v}\|_2 < \rho$ and unique associated alternative multipliers $(\pi_\delta^2, \eta_\delta^2)$, and*

$$\|v_\delta - v_{\delta'}\|_2 + \|\eta_\delta^2 - \eta_{\delta'}^2\|_2 \ \leq\ \lambda\|\delta - \delta'\|_W, \quad \forall\, \delta, \delta' \in \mathcal{W}. \tag{5.73}$$

Moreover, (v_δ, z_{v_δ}) is a local solution of $(\mathcal{P}_\delta)$ satisfying the uniform quadratic growth condition (5.72).

Proof. Let us show the existence of a stationary point of problem $(\mathcal{P}_\delta)$. We may write $(\mathcal{P}_\delta)$ as

$$(\mathcal{P}_\delta) \quad \min_{v \in \mathcal{V}}\ \tfrac{1}{2}\langle v, \mathcal{A}v\rangle + \langle b, v\rangle - \langle \gamma, v\rangle \qquad \text{s.t.} \qquad \mathcal{C}v + d - \zeta \in K,$$

where $\mathcal{A}$ is the continuous, self-adjoint bilinear operator over $\mathcal{V}$ associated with the quadratic form (5.70), b is an element in $\mathcal{V}^* \equiv \mathcal{V}$, $\mathcal{C} : v \mapsto Cz_v$ is a linear continuous operator $\mathcal{V} \to C[0,T]$, and $d \in H^2(0,T)$. Here, without ambiguity, we also denote by $\langle \cdot, \cdot \rangle$ the scalar product over $\mathcal{V}$.

Step 1: Reduction to a fixed feasible set. Let us first consider perturbations of the cost function only, i.e. consider the problem $(\mathcal{P}_\gamma)$ defined by

$$(\mathcal{P}_\gamma) \quad \min_{v \in \mathcal{V}}\ \tfrac{1}{2}\langle v, \mathcal{A}v\rangle + \langle b, v\rangle - \langle \gamma, v\rangle \qquad \text{s.t.} \qquad \mathcal{C}v + d \in K.$$

By Prop. 5.22, the uniform second-order growth condition holds for $(\mathcal{P}_\gamma)$, so does Robinson's constraint qualification by $(\tilde{\text{A}}2)$, and the perturbed problem $(\mathcal{P}_\gamma)$ includes the so-called *tilt perturbation* (see [24, p.416]), i.e. additive perturbations of the cost function of type $-\langle \gamma, v\rangle$ with $\gamma \in \mathcal{V}^*$. Therefore, it follows from [24, Th. 5.17 and Rem. 5.19], since the feasible set of $(\mathcal{P}_\gamma)$ is constant, that there exist $\rho_1, \rho_2 > 0$ and a constant $\lambda > 0$, such that for all $\gamma \in B_2(0, \rho_2)$, $(\mathcal{P}_\gamma)$ has a unique stationary point v_γ in $B_2(\bar{v}, \rho_1)$, and

$$\|v_\gamma - v_{\gamma'}\|_2 \ \leq\ \lambda\|\gamma - \gamma'\|_2, \quad \forall\, \gamma, \gamma' \in B_2(0, \rho_2). \tag{5.74}$$

We have of course that $\bar{v} = v_0$.

Step 2: Existence of a stationary point of $(\mathcal{P}_\delta)$. Let now $\delta = (\gamma, \zeta) \in W$. By Lemma 5.20, there exists $v_\zeta \in \mathcal{V}$ such that

$$(\mathcal{C}v_\zeta)(t) = \zeta(t) \quad \text{on } \Omega_\sigma \qquad \text{and} \qquad \|v_\zeta\|_2 \leq c\|\zeta\|_{2,2}.$$

Set $\tilde{\gamma} := \gamma - \mathcal{A}v_\zeta$. We have that $\|\tilde{\gamma}\|_2 \leq \|\gamma\|_2 + c\|\mathcal{A}\|\|\zeta\|_{2,2} < \rho_2$ if $\|\delta\|_W$ is small enough. Therefore, there exists a (unique) stationary point $v_{\tilde{\gamma}} \in B_2(\bar{v}, \rho_1)$ of $(\mathcal{P}_{\tilde{\gamma}})$, with multiplier $\mathrm{d}\eta_{\tilde{\gamma}} \in \mathcal{M}[0,T]$, satisfying the first-order optimality condition

$$\left\{ \begin{array}{l} \mathcal{A}v_{\tilde{\gamma}} + b - \tilde{\gamma} + \mathcal{C}^\top \mathrm{d}\eta_{\tilde{\gamma}} = 0, \\ \mathcal{C}v_{\tilde{\gamma}} + d \leq 0 \text{ on } [0,T], \quad \mathrm{d}\eta_{\tilde{\gamma}} \geq 0, \quad \langle \mathrm{d}\eta_{\tilde{\gamma}}, \mathcal{C}v_{\tilde{\gamma}} + d \rangle = 0. \end{array} \right. \tag{5.75}$$

Since $\|\mathcal{C}v_{\tilde{\gamma}} - \mathcal{C}\bar{v}\|_\infty \leq \|\mathcal{C}\|\|v_{\tilde{\gamma}} - \bar{v}\|_2 \leq \lambda\|\mathcal{C}\|\|\tilde{\gamma}\|_2$ by (5.74), if $\|\delta\|_W$ is small enough then the contact set of $\mathcal{C}v_{\tilde{\gamma}} + d$ is included in Ω_σ, and hence

$$\mathrm{supp}(\mathrm{d}\eta_{\tilde{\gamma}}) \subset \Omega_\sigma. \tag{5.76}$$

Let $v_\delta := v_{\tilde{\gamma}} + v_\zeta$ and $\mathrm{d}\eta_\delta := \mathrm{d}\eta_{\tilde{\gamma}}$. Note that there exists a constant $a > 0$ such that $(\mathcal{C}\bar{v})(t) + d(t) < -a$ on $[0,T] \setminus \Omega_\sigma$. Therefore, on $[0,T] \setminus \Omega_\sigma$, we obtain that (we denote in what follows by C different positive constants)

$$\begin{array}{rcl} \mathcal{C}v_\delta + d - \zeta & = & \mathcal{C}\bar{v} + d - \zeta + \mathcal{C}v_\zeta + \mathcal{C}(v_{\tilde{\gamma}} - \bar{v}) \\ & \leq & -a + \|\zeta\|_\infty + \|\mathcal{C}v_\zeta\|_\infty + \|\mathcal{C}(v_{\tilde{\gamma}} - \bar{v})\|_\infty \\ & \leq & -a + C\|\zeta\|_{2,2} + \|\mathcal{C}\|\|v_\zeta\|_2 + \|\mathcal{C}\|\|v_{\tilde{\gamma}} - \bar{v}\|_2 \\ & \leq & -a + (C + c\|\mathcal{C}\|)\|\zeta\|_{2,2} + \lambda\|\mathcal{C}\|\|\tilde{\gamma}\|_2 \ \leq \ -a + C\|\delta\|_W, \end{array}$$

and hence, if $\|\delta\|_W$ is small enough, then $\mathcal{C}v_\delta + d - \zeta < 0$ on $[0,T] \setminus \Omega_\sigma$. Since on Ω_σ, we have that $\mathcal{C}v_\delta + d - \zeta = \mathcal{C}v_{\tilde{\gamma}} + d \leq 0$, using (5.75) and (5.76), v_δ obviously satisfies

$$\left\{ \begin{array}{l} \mathcal{A}v_\delta + b - \gamma + \mathcal{C}^\top \mathrm{d}\eta_\delta = 0, \\ \mathcal{C}v_\delta + d - \zeta \leq 0 \text{ on } [0,T], \quad \mathrm{d}\eta_\delta \geq 0, \quad \langle \mathrm{d}\eta_\delta, \mathcal{C}v_\delta + d - \zeta \rangle = 0, \end{array} \right.$$

i.e. v_δ is a stationary point of $(\mathcal{P}_\delta)$, with multiplier $\mathrm{d}\eta_\delta$. Consequently, for $\rho_3 > 0$ small enough, reducing ρ_1 if necessary, $(\mathcal{P}_\delta)$ has, for all $\delta \in B_W(0,\rho_3)$, a (*necessarily unique* by Prop. 5.22) stationary point $v_\delta \in B_2(\bar{v},\rho_1)$, with (unique) multiplier $\mathrm{d}\eta_\delta$. That (v_δ, z_{v_δ}) is a local solution of $(\mathcal{P}_\delta)$ satisfying the uniform growth condition (5.72) follows then from Prop. 5.22.

Step 3: Lipschitz continuity of the stationary point. Let $\delta_i = (\gamma_i, \zeta_i) \in B_W(0,\rho_3)$, $i = 1,2$, and v_{ζ_i} be such that

$$\mathcal{C}v_{\zeta_i} = \zeta_i \text{ on } \Omega_\sigma, \ i = 1,2, \quad \text{and} \quad \|v_{\zeta_1}\|_2 \leq c\|\zeta_1\|_{2,2}, \ \|v_{\zeta_1} - v_{\zeta_2}\|_2 \leq c\|\zeta_1 - \zeta_2\|_{2,2}.$$

It follows that $\|v_{\zeta_2}\|_2 \leq c(2\|\zeta_1\|_{2,2} + \|\zeta_2\|_{2,2}) < 3c\rho_3$. Setting $\tilde{\gamma}_i := \gamma_i - \mathcal{A}v_{\zeta_i}$, we obtain as before that if ρ_3 is small enough, then the unique stationary point v_i of $(\mathcal{P}_{\delta_i})$ is given by $v_i = v_{\zeta_i} + v_{\tilde{\gamma}_i}$. Therefore, using (5.74),

$$\begin{array}{rcl} \|v_1 - v_2\|_2 & \leq & \|v_{\zeta_1} - v_{\zeta_2}\|_2 + \lambda\|\tilde{\gamma}_1 - \tilde{\gamma}_2\|_2 \\ & \leq & c(1 + \lambda\|\mathcal{A}\|)\|\zeta_1 - \zeta_2\|_{2,2} + \lambda\|\gamma_1 - \gamma_2\|_2 \\ & \leq & C\|\delta_1 - \delta_2\|_W. \end{array} \tag{5.77}$$

Step 4: Lipschitz continuity of the alternative multiplier η_δ^2 given by (5.67). Using the above notation, denote by $\mathrm{d}\eta_i$ the (unique) multiplier associated with v_i and by η_i^2 the associated alternative multiplier. Since $-\mathcal{C}^\top(\mathrm{d}\eta_2 - \mathrm{d}\eta_1) = \mathcal{A}(v_2 - v_1) + \gamma_2 - \gamma_1$, we have, for all $v \in \mathcal{V}$,

$$|\langle \mathrm{d}\eta_2 - \mathrm{d}\eta_1, \mathcal{C}v \rangle| \ \leq \ (\|\mathcal{A}\|\|v_2 - v_1\|_2 + \|\gamma_2 - \gamma_1\|_2)\|v\|_2. \tag{5.78}$$

By Lemma 5.20, for all $\varphi \in H^2(0,T)$, there exists $v \in \mathcal{V}$ such that $\mathcal{C}v = \varphi$ on Ω_σ and $\|v\|_2 \leq c\|\varphi\|_{2,2}$. It follows from (5.76) that $\int_0^T \varphi(t)(\mathrm{d}\eta_2(t) - \mathrm{d}\eta_1(t)) = \langle \mathrm{d}\eta_2 - \mathrm{d}\eta_1, \mathcal{C}v\rangle$. Therefore, we obtain in view of (5.78) that

$$\|\mathrm{d}\eta_2 - \mathrm{d}\eta_1\|_{2,2*} = \sup_{\varphi\in H^2, \varphi\not\equiv 0} \frac{|\int_0^T \varphi(t)(\mathrm{d}\eta_2(t) - \mathrm{d}\eta_1(t))|}{\|\varphi\|_{2,2}} \leq c(\|\mathcal{A}\|\|v_2 - v_1\|_2 + \|\gamma_2 - \gamma_1\|_2).$$

Since $\|\eta_2^2 - \eta_1^2\|_2 \leq C\|\mathrm{d}\eta_2 - \mathrm{d}\eta_1\|_{2,2*}$ by Lemma 5.14, the above estimate, together with (5.77), shows the existence of a constant $\lambda > 0$ such that (5.73) holds and achieves the proof of the theorem. □

5.6 Proof of Theorem 5.12

In order to prove Th. 5.12, we have to show that assumptions (iii), (iv) and (v) of Th. 5.17 are satisfied, which is done respectively in lemmas 5.24 to 5.26 below. Throughout this section, the assumptions of Th. 5.12 are assumed to hold. We consider a stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, and we use the notations defined in subsection 5.4.3. Moreover, throughout the section, we use the following notations (time dependence is omitted):

$$\begin{array}{rcl}
S & := & \tilde{H}_{uu}(\bar{u},\bar{y},\bar{p}^2,\bar{\eta}^2), \qquad R := \tilde{H}_{uy}(\bar{u},\bar{y},\bar{p}^2,\bar{\eta}^2), \qquad Q := \tilde{H}_{yy}(\bar{u},\bar{y},\bar{p}^2,\bar{\eta}^2),\\
A & := & f_y(\bar{u},\bar{y}), \qquad B := f_u(\bar{u},\bar{y}), \qquad \Phi := \phi_{yy}(\bar{y}(T)),\\
C & := & g_y(\bar{y}), \qquad d := g(\bar{y}), \qquad C_1 = g_y^{(1)}(\bar{y}),\\
C_2 & := & g_y^{(2)}(\bar{u},\bar{y}), \qquad N_2 := g_u^{(2)}(\bar{u},\bar{y}), \qquad a := -C_2\bar{\eta}^2, \qquad b := -N_2\bar{\eta}^2.
\end{array}$$

All the above quantities are bounded and Lipschitz continuous over $[0,T]$.

Let us first make explicit the expression of the derivative $D_x\mathcal{F}(\bar{x},\bar{\mu})(x-\bar{x})$ involved in the definition (5.57) of $\mathcal{L}(x)$, with $x = (u,\eta^2)$ and $\bar{x} = (\bar{u},\bar{\eta}^2)$. Note that the Fréchet derivative of the mapping $(u,\mu) \mapsto y_u^\mu$ w.r.t. u in direction v is the solution $z_{u,v}^\mu$ of

$$\dot{z}_{u,v}^\mu = f_y^\mu(u,y_u^\mu)z_{u,v}^\mu + f_u^\mu(u,y_u^\mu)v, \qquad z_{u,v}^\mu(0) = 0$$

and that of the mapping $(x,\mu) \mapsto p_x^{2,\mu}$ (recall that $p_x^{2,\mu}$ is the solution of (5.56)) w.r.t. $x = (u,\eta^2)$ in direction $h = (v,\xi)$ is the solution $\pi_{x,h}^{2,\mu}$ of (omitting the arguments $(u, y_u^\mu, p_x^{2,\mu}, \eta^2)$):

$$\begin{array}{rcl}
-\dot{\pi}_{x,h}^{2,\mu} & = & \tilde{H}_{yu}^\mu v + \tilde{H}_{yy}^\mu z_{u,v}^\mu + \pi_{x,h}^{2,\mu} f_y^\mu + \xi(g^\mu)_y^{(2)},\\
\pi_{x,h}^{2,\mu}(T) & = & \phi_{yy}^\mu(y_u^\mu(T))z_{u,v}^\mu(T).
\end{array} \tag{5.79}$$

Applications of Gronwall's Lemma shows that, for μ in a neighborhood of $\bar{\mu}$, $x = (u,\eta^2)$ in a L^∞-neighborhood of $\bar{x} = (\bar{u},\bar{\eta}^2)$ and a direction $h = (v,\xi) \in X$,

$$\|z_{u,v}^\mu\|_\infty = \mathcal{O}(\|v\|_2), \qquad \|\pi_{x,h}^{2,\mu}\|_\infty = \mathcal{O}(\|h\|_X), \tag{5.80}$$

$$\|z_{u,v}^\mu - z_{\bar{u},v}^{\bar{\mu}}\|_\infty = \mathcal{O}(\|u-\bar{u}\|_2 + \|\mu-\bar{\mu}\|)\|v\|_2, \tag{5.81}$$

$$\|\pi_{x,h}^{2,\mu} - \pi_{\bar{x},h}^{2,\bar{\mu}}\|_\infty = \mathcal{O}(\|x-\bar{x}\|_X + \|\mu-\bar{\mu}\|)\|h\|_X. \tag{5.82}$$

By the chain rule, we obtain that

$$D_x\mathcal{F}(\bar{x},\bar{\mu})(x-\bar{x}) = \begin{pmatrix} S(u-\bar{u}) + Rz_{u-\bar{u}} + \pi^2_{u-\bar{u},\eta^2-\bar{\eta}^2}B + (\eta^2-\bar{\eta}^2)N_2 \\ Cz_{u-\bar{u}} \end{pmatrix},$$

where $z_{u-\bar u} := z^{\bar\mu}_{\bar u,u-\bar u}$ is the solution of (5.64) for $v = u-\bar u$, and $\pi^2_{u-\bar u,\eta^2-\bar\eta^2} := \pi^{2,\bar\mu}_{\bar x,(x-\bar x)}$ is the solution of (5.79), for $(v,\xi) = (u-\bar u, \eta^2-\bar\eta^2)$:

$$-\dot\pi^2_{v,\xi} \;=\; R^\top v + Qz_v + \pi^2_{v,\xi}A + \xi C_2, \qquad \pi^2_{v,\xi}(T) = \Phi z_v(T).$$

Set $v := u-\bar u$, and let $\delta = (\gamma,\zeta) \in \Delta$. Then (5.50) has a unique solution $x = (u,\eta^2) \in \tilde X$ iff the system of equations below has a unique solution (v,z,π^2,η^2) with $(\bar u+v,\eta^2) \in \tilde X$:

$$\begin{aligned}
\dot z &= Az + Bv, \qquad z(0) = 0,\\
-\dot\pi^2 &= R^\top v + Qz + \pi^2 A + \eta^2 C_2 - \bar\eta^2 C_2, \qquad \pi^2(T) = \Phi z(T)\\
0 &= Sv + Rz + \pi^2 B + \eta^2 N_2 - \bar\eta^2 N_2 - \gamma,\\
0 &\ge d + Cz - \zeta, \qquad \mathrm{d}\dot\eta^2 \ge 0, \qquad \langle \mathrm{d}\dot\eta^2, d + Cz - \zeta\rangle = 0.
\end{aligned}$$

We recognize the first-order necessary optimality condition of linear-quadratic problem $(\mathcal{P}_\delta)$ *in its alternative form*. That is, setting $\mathrm{d}\eta = \mathrm{d}\dot\eta^2$ and $\pi = \pi^2 - C\dot\eta^2 + C_1\eta^2$, we recover the "classical" optimality conditions of $(\mathcal{P}_\delta)$ (note that $C_1 = \dot C + CA$, $C_2 = \dot C_1 + C_1A$, $N_2 = C_1B$ and $CB = g^{(1)}_u(\bar u,\bar y) \equiv 0$):

$$\begin{aligned}
\dot z &= Az + Bv, \qquad z(0) = 0,\\
-\mathrm{d}\dot\pi &= (R^\top v + Qz + \pi A - \bar\eta^2 C_2)\mathrm{d}t + C\mathrm{d}\eta, \qquad \pi(T) = \Phi z(T)\\
0 &= Sv + Rz + \pi B - \bar\eta^2 N_2 - \gamma,\\
0 &\ge d + Cz - \zeta, \qquad \mathrm{d}\eta \ge 0, \qquad \langle \mathrm{d}\eta, d + Cz - \zeta\rangle = 0.
\end{aligned}$$

We see then that $(\bar v,\bar z) := 0$ is a stationary point of $(\mathcal{P}_0)$, with alternative multipliers $\bar\pi^2 := 0$ and $\bar\eta^2$, and classical multipliers $\bar\pi := -C\dot{\bar\eta}^2 + C_1\bar\eta^2$ and $\bar\eta = \dot{\bar\eta}^2$. The second-order optimality condition (5.36), with the quadratic cost expressed by (5.37), is precisely the condition (5.71) and implies that $(\bar v,\bar z) = 0$ is a local solution of $(\mathcal{P}_0)$.

The verifications of assumptions (iii) and (v) in Lemmas 5.24 and 5.26 are only technical, and for assumption (iv) in Lemma 5.25, we use Th. 5.23.

Lemma 5.24. *The mapping $\Psi^\mu = \mathcal{F}(\cdot,\mu) - \mathcal{L}(\cdot)$ is strictly stationary at $x = \bar x$, uniformly in μ near $\bar\mu$.*

Proof. Let $x_1, x_2 \in X$ and $\mu \in P$. We have that

$$\begin{aligned}
\Psi^\mu(x_1) - \Psi^\mu(x_2) &= \mathcal{F}(x_1,\mu) - \mathcal{F}(x_2,\mu) - D_x\mathcal{F}(\bar x,\bar\mu)(x_1-x_2)\\
&= \int_0^1 (D_x\mathcal{F}(\theta x_1 + (1-\theta)x_2,\mu) - D_x\mathcal{F}(\bar x,\bar\mu))\mathrm{d}\theta(x_1-x_2).
\end{aligned}$$

Let $x = (u,\eta^2) \in \tilde X$. Then by (5.59)–(5.60), if x is close to $\bar x = (\bar u,\bar\eta^2)$ for the norm of X, this implies that (u,η^2) belongs to a L^∞-neighborhood of $(\bar u,\bar\eta^2)$. Hence, y^μ_u and $p^{2,\mu}_{u,\eta^2}$ remain also uniformly bounded for μ in a neighborhood of $\bar\mu$. Let $x_i = (u_i,\eta_i^2) \in X$, $i = 1,2$, and given $\theta \in [0,1]$, write $x_\theta := \theta x_1 + (1-\theta)x_2$ and similarly for the other variables. Set

$$\begin{pmatrix} r_1\\ r_2\end{pmatrix} := (D_x\mathcal{F}(x_\theta,\mu) - D_x\mathcal{F}(\bar x,\bar\mu))(x_1-x_2).$$

Let us express the first row r_1. Denoting by $(\cdot)$ the arguments $(u_\theta, y^\mu_{u_\theta}, p^{2,\mu}_{x_\theta}, \eta^2_\theta)$, we obtain that

$$\begin{aligned}
r_1 \;=\;& (\tilde H^\mu_{uu}(\cdot) - S)(u_1-u_2) + (\tilde H^\mu_{uy}(\cdot)z^\mu_{u_\theta,u_1-u_2} - Rz^{\bar\mu}_{\bar u,u_1-u_2})\\
&+ (\pi^{2,\mu}_{x_\theta,x_1-x_2}f^\mu_u(\cdot) - \pi^{2,\bar\mu}_{\bar x,x_1-x_2}B) + (\eta_1^2-\eta_2^2)((g^\mu)^{(2)}_u(\cdot) - N_2).
\end{aligned}$$

For (u_i, η_i^2) in a L^∞-neighborhood of $(\bar{u}, \bar{\eta}^2)$ and μ in the neighborhood of $\bar{\mu}$, we have that $\tilde{H}^\mu_{uu}(\cdot) - S = \tilde{H}^\mu_{uu}(u_\theta, y^\mu_{u_\theta}, p^{2,\mu}_{x_\theta}, \eta^2_\theta) - \tilde{H}^{\bar{\mu}}_{uu}(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)$ is arbitrarily small in the L^∞ norm, and similarly for the terms involving the other derivatives, $\tilde{H}^\mu_{uy}$, f^μ_u, and $(g^\mu)^{(2)}_u$. Therefore, given any $\varepsilon > 0$, for $\|x_i - \bar{x}\|_X, \|\mu - \bar{\mu}\|$ small enough,

$$\begin{aligned}\|r_1\|_2 \;\leq\; & \varepsilon(\|u_1 - u_2\|_2 + \|z^\mu_{u_\theta, u_1 - u_2}\|_2 + \|\pi^{2,\mu}_{x_\theta, x_1 - x_2}\|_2 + \|\eta_1^2 - \eta_2^2\|_2) \\ & + \|R\|_\infty \|z^\mu_{u_\theta, u_1 - u_2} - z^{\bar{\mu}}_{\bar{u}, u_1 - u_2}\|_2 + \|B\|_\infty \|\pi^{2,\mu}_{x_\theta, x_1 - x_2} - \pi^{2,\bar{\mu}}_{\bar{x}, x_1 - x_2}\|_2.\end{aligned}$$

Using (5.80)–(5.82) with $x = x_\theta$ and $h = x_1 - x_2$, we obtain that $\|r_1\|_2 \leq \varepsilon \|x_1 - x_2\|_X$, whenever x_1, x_2 are close enough to $\bar{x}$ in X and μ is close enough to $\bar{\mu}$. For the second row r_2, we have that

$$\begin{aligned} r_2 &= g^\mu_y(y^\mu_{u_\theta}) z^\mu_{u_\theta, u_1 - u_2} - g^{\bar{\mu}}_y(\bar{y}) z^{\bar{\mu}}_{\bar{u}, u_1 - u_2}, \\ \dot{r}_2 &= (g^\mu)^{(1)}_y(y^\mu_{u_\theta}) z^\mu_{u_\theta, u_1 - u_2} - (g^{\bar{\mu}})^{(1)}_y(\bar{y}) z^{\bar{\mu}}_{\bar{u}, u_1 - u_2}, \\ \ddot{r}_2 &= ((g^\mu)^{(2)}_u(u_\theta, y^\mu_{u_\theta}) - (g^{\bar{\mu}})^{(2)}_u(\bar{u}, \bar{y}))(u_1 - u_2) \\ &\quad + (g^\mu)^{(2)}_y(u_\theta, y^\mu_{u_\theta}) z^\mu_{u_\theta, u_1 - u_2} - (g^{\bar{\mu}})^{(2)}_y(\bar{u}, \bar{y}) z^{\bar{\mu}}_{\bar{u}, u_1 - u_2}. \end{aligned}$$

Therefore, we conclude with the same arguments that $\|r_2\|_{2,2} \leq \varepsilon \|u_1 - u_2\|_2$, whenever $\|x_i - \bar{x}\|_X$, $i = 1, 2$ and $\|\mu - \bar{\mu}\|$ are small enough. This shows the desired property. □

Lemma 5.25. *For k sufficiently large w.r.t. l in definition* (5.52) *of the space X, r small enough in definition* (5.53) *of the space $\tilde{X}$, and small enough positive constants ϱ and k' in definition* (5.55) *of the set Δ,* (5.50) *has a unique solution $x_\delta = (u_\delta, \eta^2_\delta)$ in $\tilde{X}$, for all $\delta \in \Delta$, and this solution is Lipschitz continuous w.r.t. δ.*

Proof. We have that $x = (u, \eta^2)$ is solution of (5.50) iff $(v := u - \bar{u}, z_v)$ is solution of the first-order optimality condition of $(\mathcal{P}_\delta)$ with alternative multipliers $\pi^2_{v, \eta^2 - \bar{\eta}^2}$ and η^2. By the hypotheses of Th. 5.12, $(\bar{v}, \bar{z}) = 0$ is a stationary point of $(\mathcal{P}_0)$ satisfying the assumptions of Th. 5.23. Choose ϱ small enough, so that $B_W(0, \varrho)$ is included in the neighborhood $\mathcal{W}$ of Th. 5.23. By this theorem, for all $\delta \in B_W(0, \varrho)$, $(\mathcal{P}_\delta)$ has a unique stationary point (v_δ, z_{v_δ}) with $\|v_\delta\|_2 < \rho$ and unique associated alternative multipliers $(\pi^2_{v_\delta, \eta^2_\delta - \bar{\eta}^2}, \eta^2_\delta)$. Therefore, (5.50) has a unique solution $(u_\delta := \bar{u} + v_\delta, \eta^2_\delta)$ with $\|u_\delta - \bar{u}\|_2 < \rho$. We have to show that $(u_\delta, \eta^2_\delta)$ belongs to the space $\tilde{X}$. Throughout the proof, we denote by C different positive constants.

By Prop. 5.21(i), shrinking ϱ if necessary, we immediately obtain that η^2_δ belongs to the space $BV^2_{T,l}[0, T]$, for large enough l. Therefore, by (5.60) and (5.73), for all $\delta \in B_W(0, \varrho)$,

$$\|\eta^2_\delta - \bar{\eta}^2\|_\infty \;\leq\; \sqrt[3]{6l} \|\eta^2_\delta - \bar{\eta}^2\|_2^{2/3} \;\leq\; \sqrt[3]{6l} \lambda^{2/3} \|\delta\|_W^{2/3}.$$

For $\delta = (\gamma, \zeta) \in \Delta$ (then $\gamma \in \mathrm{Lip}_{k'}$), let us show now that $u_\delta = \bar{u} + v_\delta \in \mathrm{Lip}_k$. From the first-order alternative optimality condition of $(\mathcal{P}_\delta)$, we have that

$$S v_\delta + R z_{v_\delta} + \pi^2_{v_\delta, \eta^2_\delta - \bar{\eta}^2} B + N_2(\eta^2_\delta - \bar{\eta}^2) - \gamma = 0. \tag{5.83}$$

Since S is uniformly invertible by (A3), using (5.80), (5.73), and (5.59), we deduce that

$$\begin{aligned}\|v_\delta\|_\infty &\leq C(\|z_{v_\delta}\|_\infty + \|\pi^2_{v_\delta, \eta^2_\delta - \bar{\eta}^2}\|_\infty + \|\eta^2_\delta - \bar{\eta}^2\|_\infty) + \|\gamma\|_\infty \\ &\leq C(2\lambda \|\delta\|_W + \sqrt[3]{6l} \lambda^{2/3} \|\delta\|_W^{2/3}) + \sqrt[3]{3k'} \|\gamma\|_2^{2/3} \\ &\leq (C(l) + \sqrt[3]{3k'}) \|\delta\|_W^{2/3}.\end{aligned}$$

We denote here and in what follows by $C(l)$ different positive constants that depend on l (but not on k). Since $\gamma \in \mathrm{Lip}_{k'}$, $\eta_\delta^2, \bar\eta^2 \in BV_{T,l}^2 \subset \mathrm{Lip}_l$, $z_{v_\delta}, \pi^2_{v_\delta,\eta_\delta^2-\bar\eta^2} \in W^{1,\infty}$, S, R, B, N_2 are Lipschitz continuous, and S is uniformly invertible, we can differentiate (5.83) in time and we get

$$S\dot v_\delta + \dot S v_\delta + R\dot z_{v_\delta} + \dot R z_{v_\delta} + \dot\pi^2_{v_\delta,\eta_\delta^2-\bar\eta^2} B + \pi^2_{v_\delta,\eta_\delta^2-\bar\eta^2}\dot B + N_2(\dot\eta^2 - \dot{\bar\eta}^2) + \dot N_2(\eta^2-\bar\eta^2) - \dot\gamma = 0.$$

Since $\|z_{v_\delta}\|_\infty$, $\|\pi^2_{v_\delta,\eta_\delta^2-\bar\eta^2}\|_\infty$, $\|\dot z_{v_\delta}\|_\infty$, $\|\dot\pi^2_{v_\delta,\eta_\delta^2-\bar\eta^2}\|_\infty \le C(\|v_\delta\|_\infty + \|\eta_\delta^2-\bar\eta^2\|_\infty)$, and S has the inverse uniformly bounded over $[0,T]$, whereas $\|\dot\eta_\delta^2\|_\infty, \|\dot{\bar\eta}^2\|_\infty \le l$, we obtain that

$$\begin{aligned}\|\dot v_\delta\|_\infty &\le C(\|v_\delta\|_\infty + \|\eta_\delta^2-\bar\eta^2\|_\infty + \|\dot\eta_\delta^2 - \dot{\bar\eta}^2\|_\infty) + \|\dot\gamma\|_\infty \\ &\le (C(l) + C\sqrt[3]{3k'})\|\delta\|_W^{2/3} + 2Cl + k'.\end{aligned}$$

Therefore, we have that $\|\dot v_\delta\|_\infty \le k/2$ if, fixing a suitable l, we take k so large that $k > \max\{4Cl; 2\|\dot{\bar u}\|_\infty\}$, and choose ϱ and k' in (5.55) small enough. It follows that the solution $x_\delta = (u_\delta = \bar u + v_\delta, \eta_\delta^2)$ of (5.50) belongs to the space X. In addition, if we choose $r = \rho$, with the ρ of Th. 5.23, then $x_\delta \in \tilde X$ for $\|\delta\|_W$ small enough, and is the unique solution of (5.50) in $\tilde X$. Moreover, by Th. 5.23,

$$\|u_\delta - u_{\delta'}\|_2 + \|\eta_\delta^2 - \eta_{\delta'}^2\|_2 \ \le\ \lambda\|\delta-\delta'\|_W, \quad \forall\, \delta,\delta' \in \Delta.$$

This achieves the proof of assumption (iv) of Th. 5.17. □

Lemma 5.26. *There exists a neighborhood of $(\bar x, \bar\mu)$, such that $\mathcal{F}(x,\mu) - \mathcal{L}(x)$ belongs to Δ, for all (x,μ) in this neighborhood.*

Proof. We have to show that for $\|x-\bar x\|_X, \|\mu-\bar\mu\|$ small enough, $\mathcal{F}(x,\mu) - \mathcal{L}(x) \in \Delta$, where Δ is our set of smooth variations defined by (5.55). Throughout the proof, we denote by C different positive constants. For $\theta \in [0,1]$, set $x_\theta := \theta x + (1-\theta)\bar x$ and similarly define μ_θ. We have that

$$\begin{aligned}\mathcal{F}(x,\mu) - \mathcal{L}(x) &= \mathcal{F}(x,\mu) - \mathcal{F}(\bar x,\bar\mu) - D_x\mathcal{F}(\bar x,\bar\mu)(x-\bar x) \\ &= \int_0^1 (D_x\mathcal{F}(x_\theta,\mu_\theta) - D_x\mathcal{F}(\bar x,\bar\mu))\mathrm{d}\theta(x-\bar x) \\ &\quad + \int_0^1 D_\mu\mathcal{F}(x_\theta,\mu_\theta)\mathrm{d}\theta(\mu-\bar\mu) \ =:\ \begin{pmatrix} r_1 \\ r_2\end{pmatrix}.\end{aligned}$$

Let us show that $\|r_1\|_2 + \|r_2\|_{2,2} \le \varrho$ and $\|\dot r_1\|_\infty \le k'$, for $\|x-\bar x\|_X$ and $\|\mu-\bar\mu\|$ small enough. By the arguments of Lemma 5.24, given any $\varepsilon > 0$, for $\|x-\bar x\|_X$ and $\|\mu-\bar\mu\|$ small enough, we have that $\|\int_0^1 (D_x\mathcal{F}(x_\theta,\mu_\theta) - D_x\mathcal{F}(\bar x,\bar\mu))\mathrm{d}\theta(x-\bar x)\|_W \le \varepsilon\|x-\bar x\|_X$. Moreover, since $D_\mu\mathcal{F}(x,\mu)$ is uniformly bounded for (x,μ) in a neighborhood of $(\bar x,\bar\mu)$ by definition of a stable extension, we deduce that

$$\|r_1\|_2 + \|r_2\|_{2,2} \ \le\ \varepsilon\|x-\bar x\|_X + C\|\mu-\bar\mu\| \ \le\ \varrho, \tag{5.84}$$

for $\|x-\bar x\|_X$ and $\|\mu-\bar\mu\|$ small enough. Making now explicit the expression of r_1, we obtain that (recall the notations $S = \tilde H^{\bar\mu}_{uu}$, $R = \tilde H^{\bar\mu}_{uy}$, $B = f^{\bar\mu}_u$, $N_2 = (g^{\bar\mu})^{(2)}_u$):

$$\begin{aligned}r_1 &= \tilde H^\mu_u(u, y^\mu_u, p^{2,\mu}_{u,\eta^2}, \eta^2) - \tilde H^{\bar\mu}_u(\bar u, \bar y, \bar p^2, \bar\eta^2) - S(u-\bar u) - Rz_{u-\bar u} \\ &\quad - \pi^2_{u-\bar u,\eta^2-\bar\eta^2} B - N_2(\eta^2-\bar\eta^2).\end{aligned}$$

Time derivation yields (omitting arguments and reorganizing the terms)

$$\begin{aligned}\dot r_1 &= (\tilde H^\mu_{uu}-\tilde H^{\bar\mu}_{uu})\dot u+(\tilde H^\mu_{uy}f^\mu-\tilde H^{\bar\mu}_{uy}f^{\bar\mu})-(\tilde H^\mu_y f^\mu_u-\tilde H^{\bar\mu}_y f^{\bar\mu}_u)+((g^\mu)^{(2)}_u-(g^{\bar\mu})^{(2)}_u)\dot\eta^2\\ &\quad - R\dot z_{u-\bar u}-\dot\pi^2_{u-\bar u,\eta^2-\bar\eta^2}B-\dot S(u-\bar u)-\dot R z_{u-\bar u}-\pi^2_{u-\bar u,\eta^2-\bar\eta^2}\dot B-\dot N_2(\eta^2-\bar\eta^2).\end{aligned}$$

For (u,η^2) close to $(\bar u,\bar\eta^2)$ in X, and μ in a neighborhood of $\bar\mu$, we have by (5.59)–(5.60) that $\|(u,y^\mu_u,p^{2,\mu}_{u,\eta^2},\eta^2)-(\bar u,\bar y,\bar p^2,\bar\eta^2)\|_\infty$ is arbitrarily small, and hence, by continuity of $\tilde H^\mu_{uu}$, etc, given any $\varepsilon>0$, we obtain that

$$\begin{aligned}\|\dot r_1\|_\infty &\le \varepsilon(\|\dot u\|_\infty+\|\dot\eta^2\|_\infty+1)+C(\|\dot z_{u-\bar u}\|_\infty+\|\dot\pi^2_{u-\bar u,\eta^2-\bar\eta^2}\|_\infty)\\ &\quad + C(\|u-\bar u\|_\infty+\|z_{u-\bar u}\|_\infty+\|\pi^2_{u-\bar u,\eta^2-\bar\eta^2}\|_\infty+\|\eta^2-\bar\eta^2\|_\infty)\\ &\le \varepsilon(k+l+1)+C(\|u-\bar u\|_\infty+\|\eta^2-\bar\eta^2\|_\infty)\\ &\le \varepsilon(k+l+1)+C(\sqrt[3]{6k}+\sqrt[3]{6l})\|x-\bar x\|_X^{2/3}\ \le\ k',\end{aligned}$$

if $\|x-\bar x\|_X$ and $\|\mu-\bar\mu\|$ are small enough. It follows that $r_1\in \mathrm{Lip}_{k'}(0,T;\mathbb{R}^m)$, and with (5.84), this achieves the proof. □

Proof of Th. 5.12. We apply Th. 5.17 with the spaces X, $\tilde X$, W, Δ, P and mappings $\mathcal{F}$, $\mathcal{N}$, $\mathcal{L}$ defined in subsection 5.4.3. We set $\bar x:=(\bar u,\bar\eta^2)$. The assumptions (i) and (ii) of Th. 5.17 are obviously fulfilled by our hypotheses and the definition of a stable extension. For an appropriate choice of the constants k,l,r,k',ϱ involved in the definition of the spaces X, $\tilde X$ and Δ, assumptions (iii), (iv) and (v) hold by Lemmas 5.24, 5.25 and 5.26, respectively. It follows that for all μ in a neighborhood of $\bar\mu$, there exists a unique stationary point (u^μ,y^μ) of $(\mathcal{P}^\mu)$ and unique associated alternative multipliers $(p^{2,\mu},\eta^{2,\mu})$ with $(u^\mu,\eta^{2,\mu})$ in a X-neighborhood of $\bar x$, and (5.51) is satisfied. Since by definition of a stable extension, $\mathcal{F}$ is Lipschitz continuous w.r.t. μ, uniformly w.r.t. x, this implies that (5.39) holds, while (5.40) follows from (5.59)–(5.60). Finally, by (5.40), taking if necessary a smaller neighborhood of $\bar\mu$, u^μ belongs to the L^∞-neighborhood of $\bar u$ on which the uniform quadratic growth condition holds (Prop. 5.11). Therefore, (u^μ,y^μ) is the unique stationary point of $(\mathcal{P}^\mu)$ with u^μ in a L^∞-neighborhood of $\bar u$ and is a local solution of $(\mathcal{P}^\mu)$ satisfying (5.38). □

5.7 Conclusion and Remarks

In this paper, we obtain for the first time stability results for optimal control problems with a state constraint of order greater than one without any assumption on the structure of the contact set. For this we use a generalized implicit function theorem in metric spaces [53] applied to a system equivalent to the first-order optimality condition, involving *alternative multipliers* obtained by integrating the original state constraint multiplier. In the stability analysis of linear-quadratic problems, we use [24, Th. 5.17] to obtain the existence of a stationary point for the perturbed problem under a weak second-order sufficient condition taking into account the active constraints. In this way the method for weakening the second-order sufficient condition is different from the method used in [92, 91].

Due to the low regularity of state constraint multipliers, we use a framework that differs from the ones used for first-order state constraints in [88] or in [53] in the choice of the spaces for the state constraint and state constraint multiplier. We keep the idea of [53] to use as control space the space of Lipschitz continuous functions with a bound on the Lipschitz constant.

Though the analysis is restricted to a scalar state constraint of second-order, the framework and results presented in this paper have a natural extension to several state constraints of orders ≥ 2 (see Remarks 5.2 and 5.3). Taking into account both components of first-order and higher-order is more delicate since then the arguments used in [88, 53, 91] and in the present paper would have to be combined.

Making additional assumptions on the structure of the contact set, L^∞ Lipschitz stability of solutions can be obtained, see [94, 19], improving (5.40), as it is the case for first-order state constraints (see [53, Section 4]). In [94, 19] it was also shown using a shooting approach that the solutions are directionally differentiable w.r.t. the parameter. It would be interesting as well to obtain sensitivity results without assumption on the structure of the contact set, extending to higher-order state constraints the sensivity results obtained by Malanowski [88] for state constraints of first-order.

Finally, let us note that the second-order sufficient condition (5.36) used in the stability analysis might be weakened by taking into account the curvature term of the constraint (see [21, Th. 27], [17, Th. 6.1] and [19, Th. 4.3][7]).

Acknowledgments The author thanks J.F. Bonnans for his comments on the manuscript and the anonymous referees for their useful remarks.

[7] Theorems 1.27, 4.24, and 2.34 of this thesis.

Chapitre 6

Méthode d'homotopie pour les contraintes d'ordre 2*

Abstract This chapter is devoted to optimal control problems with a regular second-order state constraint and a scalar control, when the strengthened Legendre-Clebsch condition holds. It is shown that under a uniform strict complementarity assumption, boundary arcs are stable under sufficiently smooth perturbations of the data. On the contrary, nonreducible touch points are not stable under perturbations. We show that under some reasonable conditions, either a boundary arc or a second touch point may appear. Those results, combined with the stability analysis of Chapter 5, allow us to design an homotopy algorithm that automatically detects the structure of the trajectory and initializes the shooting parameters associated with boundary arcs and touch points, extending the continuation method of Chapter 3 to second-order state constraints.

Résumé Ce chapitre est consacré aux problèmes de commande optimale avec une contrainte sur l'état scalaire du second ordre régulière et une commande scalaire, lorsque la condition forte de Legendre-Clebsch est satisfaite. On montre que sous une hypothèse de complémentarité stricte uniforme, les arcs frontières sont stables sous des perturbations suffisamment régulières des données. Au contraire, les points de contact isolés non réductibles ne sont pas stables. Sous des conditions raisonables, on montre que soit un arc frontière soit un second point de contact isolé peut apparaître. Ces résultats, combinés avec l'analyse de stabilité du chapitre 5, nous permettent de concevoir un algorithme d'homotopie qui détecte automatiquement la structure de la trajectoire et initialise les paramètres de tir associés aux arcs frontière et points de contact isolés, étendant la méthode de continuation du chapitre 3 aux contraintes du second ordre.

6.1 Introduction

This paper deals with optimal control problems with a state constraint of second-order (see [29, 98]). Many papers devoted to optimal control problems with state constraints deal with state constraints of first-order (see e.g. [65, 88, 92, 53, 93, 54, 20]), i.e. when the control appears explicitly after one time derivation of the state constraint along the dynamics. This assumption may not be satisfied in applications. For example, in the problem of the atmospheric reentry of a space shuttle, if the control is the bank angle (the angle of attack being fixed), the

*Rapport de Recherche INRIA RR-6626 (2008). Submitted for publication under the title *Homotopy algorithm for optimal control problems with a second-order state constraint.*

constraints on the thermal flux, normal acceleration and dynamic pressure are second-order state constraints, see [27].

When the strengthened Legendre-Clebsch condition holds, the shooting algorithm enables to solve optimal control problems with a very high accuracy at low cost. This algorithm (see [125]) is based on the parametrization of the trajectory by a finite-dimensional vector of *shooting parameters* and the resolution of the resulting multi-point boundary value problem by a Newton's method. Shooting methods are very sensitive to the initial conditions, and require a careful initialization of all parameters. Moreover, in presence of constraints, the structure of constraints (the number and order of boundary arcs and touch points) has to be known a priori. This makes the shooting algorithm generally hard to apply. However, when the precision is a strong requirement, such as e.g. to compute aerospace trajectories, shooting algorithms may be preferred to others methods, less accurate.

In order to determine the structure of the trajectory, which is generally unknown, and facilitate the initialization of parameters, homotopy (or continuation) methods can be used. Their well-known principle (see [1]) is to solve a sequence of problems depending continuously on a parameter, such that the first problem is "easy" to solve (e.g. the problem without the state constraint) and the last problem is the original problem. Doing so the structure of solutions may vary in the course of iterations. Homotopy methods have been applied to control problems with control constraints in e.g. [63, 97] and with state constraints in e.g. [11, 31]. The difficulty to apply classical continuation methods is connected with the changes of structure of the trajectory. Moreover, when the structure of the trajectory changes, the dimension of the vector of shooting parameters changes as well. In [20], an homotopy algorithm has been proposed for first-order state constraints, whose novelty is to automatically detect the changes in the structure of the trajectory and initialize the associated shooting parameters. It is well-known that the structure of a trajectory highly depends on the order of the constraint (see [29]). In this paper, we aim to extend the homotopy algorithm of [20] to second-order state constraints.

They are two main tools in the analysis of the homotopy method. Firstly, stability results which guarantee the existence and local uniqueness of a solution for the perturbed problem, and insure that the homotopy path is locally well-defined. Secondly, an analysis of the structure of solutions of the perturbed problem. New results concerning the first point (stability analysis) have been obtained recently in [71]. Contrary to previous stability results known for second- (and higher-)order state constraints ([94, 19]), no assumptions on the structure of the trajectory are made. This allows us precisely to deal with situations encountered in the homotopy method, when the structure of solution is not stable and hence, where the stability and sensitivity results of [94, 19] do not apply.

In this paper, results are obtained on the second point, i.e. we study the evolution of structure of solutions under small perturbations of the data. We show that when a strict complementarity hypothesis is satisfied on boundary arcs, then the latter are stable for a class of sufficiently smooth perturbations. Then we study the case of nonreducible touch points, which are excluded from the analysis based on shooting methods in [94] and [19]. In that case the structure of the trajectory is not stable. We show that under some rather general conditions, either a boundary arc or a second touch point may appear. Finally, we follows [20] in order to describe an homotopy method for second-order state constraints. The analysis is more involved than for first-order state constraints, since the structure of second-order state constraints is more complex (both essential touch points and boundary arcs are possible, while first-order state constraints typically do not have essential touch points).

The paper is organized as follows. Preliminaries (optimality conditions, assumptions) are

recalled in section 6.2. In section 6.3, the stability of boundary arcs is studied. In section 6.4, the case of nonreducible touch points is dealt with. In section 6.5, the stability result of [71] is recalled. In section 6.6, lemmas used in the analysis of the homotopy method are given. In section 6.7, the homotopy algorithm is presented and analyzed. Finally, in section 6.8, some comments are given. The contributions of the paper are the structural analysis of stationary points in sections 6.3 and 6.4 and the analysis of the homotopy algorithm. The application of this homotopy algorithm to the atmospheric reentry of a space shuttle is presented in [70].

6.2 Preliminaries

We consider the following optimal control problem with a scalar control and scalar state constraint:

$$(\mathcal{P}) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell(u(t),y(t))\mathrm{d}t + \phi(y(T)) \tag{6.1}$$

$$\text{subject to} \qquad \dot{y}(t) \;=\; f(u(t),y(t)) \quad \text{for a.a. } t\in[0,T], \quad y(0)=y_0 \tag{6.2}$$

$$g(y(t)) \;\leq\; 0 \quad \text{for all } t\in[0,T] \tag{6.3}$$

with the control and state spaces $\mathcal{U} := L^\infty(0,T;\mathbb{R})$ and $\mathcal{Y} := W^{1,\infty}(0,T;\mathbb{R}^n)$. Throughout the paper, it is assumed that assumptions (A0) and (A1) below hold:

(A0) The data $\ell : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$, $\phi : \mathbb{R}^n \to \mathbb{R}$ (resp. $f : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \to \mathbb{R}$) are C^3 (resp. C^4) mappings, with locally Lipschitz continuous third-order (resp. fourth-order) derivatives, and f is Lipschitz continuous.

(A1) The initial condition $y_0 \in \mathbb{R}^n$ satisfies $g(y_0) < 0$.

The state constraint is assumed to be of *second-order*. This means that the first-order time derivative $g^{(1)} : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$ of the constraint, defined by

$$g^{(1)}(u,y) := g_y(y)f(u,y)$$

does not depend on the control variable u, i.e. $g^{(1)}_u \equiv 0$ (and hence, we may write $g^{(1)}(y) = g^{(1)}(u,y)$), and the second-order time derivative $g^{(2)} : \mathbb{R}\times\mathbb{R}^n \to \mathbb{R}$, defined by

$$g^{(2)}(u,y) := g^{(1)}_y(y)f(u,y)$$

depends explicitly on the control, i.e. $g^{(2)}_u \not\equiv 0$.

Notation We denote by subscripts Fréchet derivatives w.r.t. the variables u, y, i.e. $f_y(u,y) = D_y f(u,y)$, $f_{yy}(u,y) = D^2_{yy} f(u,y)$, etc. The derivative with respect to the time is denoted by a dot, i.e. $\dot{y} = \frac{\mathrm{d}y}{\mathrm{d}t} = y^{(1)}$. The set of row vectors of dimension n is denoted by $\mathbb{R}^{n*}$. Adjoint or transpose operators are denoted by the symbol $^\top$. The euclidean norm is denoted by $|\cdot|$. By $L^r(0,T)$ we denote the Lebesgue space of measurable functions such that $\|u\|_r := (\int_0^T |u(t)|^r \mathrm{d}t)^{1/r} < \infty$ for $1 \leq r < \infty$, $\|u\|_\infty := \operatorname{supess}_{[0,T]} |u(t)| < \infty$. The space $W^{s,r}(0,T)$ denotes the Sobolev space of functions in $L^r(0,T)$ having their s first weak derivatives in $L^r(0,T)$, with the norm $\|u\|_{s,r} := \sum_{j=0}^s \|u^{(j)}\|_r$. We denote by H^s the space $W^{s,2}$. The space of continuous functions over $[0,T]$ and its dual space, the space of bounded Borel measures, are denoted respectively by $C[0,T]$ and $\mathcal{M}[0,T]$. The cone of continuous functions with nonpositive values over $[0,T]$ is denoted by $K := C_-[0,T]$ and its dual space, the

set of nonnegative measures, is denoted by $\mathcal{M}_+[0,T]$. The space of functions of bounded variation over $[0,T]$ is denoted by $BV[0,T]$, and the set of normalized BV functions vanishing at T is denoted by $BV_T[0,T]$. Functions of bounded variation are w.l.o.g. assumed to be right-continuous. We identify the elements of $\mathcal{M}[0,T]$ with the distributional derivatives $\mathrm{d}\eta$ of functions η in $BV_T[0,T]$. The support and the total variation of the measure $\mathrm{d}\eta \in \mathcal{M}[0,T]$ are denoted respectively by $\mathrm{supp}(\mathrm{d}\eta)$ and $|\mathrm{d}\eta|_{\mathcal{M}}$. Left- and right limits of a function of bounded variation φ will be denoted by $\varphi(\tau^{\pm}) := \lim_{t\to\tau^{\pm}} \varphi(t)$, and jumps by $[\varphi(\tau)] := \varphi(\tau^+) - \varphi(\tau^-)$. The cardinal of a finite set $\mathcal{T}$ is denoted by $|\mathcal{T}|$.

We call a *trajectory* an element $(u,y) \in \mathcal{U} \times \mathcal{Y}$ satisfying the state equation (6.2). A trajectory satisfying the state constraint (6.3) is said to be *feasible*. The contact set of a feasible trajectory is defined by

$$I(g(y)) := \{t \in [0,T] : g(y(t)) = 0\} \tag{6.4}$$

and for a small $\varepsilon > 0$, a neighborhood of the contact set is denoted by

$$I_\varepsilon(g(y)) := \{t \in [0,T] : \mathrm{dist}\{t, I(g(y))\} < \varepsilon\}. \tag{6.5}$$

A *boundary arc* (resp. *interior arc*) of a feasible trajectory (u,y) is a maximal (open) interval of positive measure (τ_1,τ_2) such that $g(y(t)) = 0$ (resp. $g(y(t)) < 0$) for all $t \in (\tau_1,\tau_2)$. The left- and right endpoints of a boundary arc (τ_{en},τ_{ex}) are called respectively *entry* and *exit* point. A *touch point* τ_{to} is an isolated contact point, i.e. such that $g(y(\tau_{to})) = 0$ and $g(y(t)) < 0$ for $t \neq \tau_{to}$ in a neighborhood of τ_{to}. An entry (resp. exit) point is said to be *regular*, if it belongs to $(0,T)$ and if there exists $\delta > 0$ such that $g(y(t)) < 0$ on $(\tau_{en} - \delta, \tau_{en})$ (resp. on $(\tau_{ex}, \tau_{ex} + \delta)$). A boundary arc is regular, if its entry and exit points are regular. The *structure* of a trajectory is the number and order of its boundary arcs and touch points.

Optimality conditions Let us first recall the well-known first-order necessary optimality condition of problem $(\mathcal{P})$. The *Hamiltonian* $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n*} \to \mathbb{R}$ is defined by

$$H(u,y,p) \;:=\; \ell(u,y) + pf(u,y). \tag{6.6}$$

We say that a feasible trajectory (u,y) is a *stationary point* of $(\mathcal{P})$, if there exists $(p,\eta) \in BV([0,T];\mathbb{R}^{n*}) \times BV_T[0,T]$ such that

$$\begin{aligned}
\dot{y} &= f(u,y), \qquad y(0) = y_0, && (6.7)\\
-\mathrm{d}p &= H_y(u,y,p)\mathrm{d}t + g_y(y)\mathrm{d}\eta, \qquad p(T) = \phi_y(y(T)) && (6.8)\\
0 &= H_u(u(t),y(t),p(t)) \qquad \text{a.e. on } [0,T] && (6.9)\\
0 &\geq g(y(t)) \;\text{ for all } t \in [0,T], \quad \mathrm{d}\eta \in \mathcal{M}_+[0,T], \quad \mathrm{supp}(\mathrm{d}\eta) \subset I(g(y)). && (6.10)
\end{aligned}$$

Alternative formulation For the stability analysis, it is convenient to write the optimality condition using alternative multipliers η^2 and p^2, uniquely related to (p,η) in the following way:

$$\begin{aligned}
\eta^1(t) &:= \int_{(t,T]} \mathrm{d}\eta(s) = -\eta(t), \qquad \eta^2(t) := \int_t^T \eta^1(s)\mathrm{d}s, && (6.11)\\
p^2(t) &:= p(t) - \eta^1(t)g_y(y(t)) - \eta^2(t)g_y^{(1)}(y(t)), \qquad t \in [0,T]. && (6.12)
\end{aligned}$$

We see that η^2 belongs to the set $BV_T^2[0,T]$, defined by

$$BV_T^2[0,T] := \{\xi \in W^{1,\infty}(0,T) : \xi(T) = 0,\ \dot{\xi} \in BV_T[0,T]\}. \tag{6.13}$$

Define the *alternative Hamiltonian* $\tilde{H} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{n*} \times \mathbb{R} \to \mathbb{R}$ by

$$\tilde{H}(u, y, p^2, \eta^2) := H(u, y, p^2) + \eta^2 g^{(2)}(u, y), \tag{6.14}$$

where H is the classical Hamiltonian (6.6). Using these alternative multipliers, we obtain easily by a direct calculation (see e.g. [98] or [17, Lemma 3.4][1]) that a feasible trajectory $(u, y) \in \mathcal{U} \times \mathcal{Y}$ is a stationary point of $(\mathcal{P})$ iff there exists $(p^2, \eta^2) \in W^{1,\infty}(0, T; \mathbb{R}^{n*}) \times BV_T^2[0, T]$ such that

$$\begin{aligned}
\dot{y}(t) &= f(u(t), y(t)) \quad \text{a.e. on } [0, T], \quad y(0) = y_0, && (6.15)\\
-\dot{p}^2(t) &= \tilde{H}_y(u(t), y(t), p^2(t), \eta^2(t)) \quad \text{a.e. on } [0, T], \quad p^2(T) = \phi_y(y(T)) && (6.16)\\
0 &= \tilde{H}_u(u(t), y(t), p^2(t), \eta^2(t)) \quad \text{a.e. on } [0, T] && (6.17)\\
0 &\geq g(y(t)) \ \text{ for all } t \in [0, T], \quad \mathrm{d}\dot{\eta}^2 \in \mathcal{M}_+[0, T], \quad \mathrm{supp}(\mathrm{d}\dot{\eta}^2) \subset I(g(y)). && (6.18)
\end{aligned}$$

Assumptions Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$, with alternative multipliers $(\bar{p}^2, \bar{\eta}^2)$. We make the following assumptions:

(A2) The state constraint is a regular second-order state constraint, i.e. $g_u^{(1)} \equiv 0$ and

$$\exists\, \beta, \sigma > 0, \quad |g_u^{(2)}(\bar{u}(t), \bar{y}(t))| \geq \beta \quad \text{for a.a. } t \in I_\sigma(g(\bar{y})). \tag{6.19}$$

(A3) $\bar{u}$ is continuous on $[0, T]$ and the strengthened Legendre-Clebsch condition holds:

$$\exists\, \alpha > 0, \quad \tilde{H}_{uu}(\bar{u}(t), \bar{y}(t), \bar{p}^2(t), \bar{\eta}^2(t)) \geq \alpha \quad \text{for all } t \in [0, T]. \tag{6.20}$$

Lemma 6.1. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ with alternative multipliers $(\bar{p}^2, \bar{\eta}^2)$ satisfying (A2)–(A3). Then $\bar{u}$ and $\bar{\eta}^2$ are of class C^2 on the interior of the (interior and boundary) arcs of the trajectory, with Lipschitz continuous second-order time derivatives.*

Proof. By the implicit function theorem applied to (6.17) on interior arcs, using that $\dot{\eta}^2$ is constant, and to $g^{(2)}(u(t), y(t)) = 0$ and (6.17) on boundary arcs, the control and alternative state constraint multipliers can be expressed, on the interior of arcs, as C^2 functions of the state and alternative costate (y, p^2). The result follows. □

Assume now that $(\bar{u}, \bar{y})$ has a (regular) boundary arc $(\bar{\tau}_{en}, \bar{\tau}_{ex})$. We consider the uniform strict complementarity assumption on boundary arcs below:

$$\exists\, \beta > 0, \qquad \ddot{\bar{\eta}}^2(t) \geq \beta \qquad \text{on } (\bar{\tau}_{en}, \bar{\tau}_{ex}). \tag{6.21}$$

Remark 6.2. Using the classical multipliers $(\bar{p}, \bar{\eta})$ associated with $(\bar{u}, \bar{y})$ in (6.7)–(6.10), assumption (6.21) can equivalently be rewritten as (recall that $\bar{\eta} = \dot{\bar{\eta}}^2$):

$$\exists\, \beta > 0, \qquad \frac{\mathrm{d}\bar{\eta}}{\mathrm{d}t}(t) \geq \beta \qquad \text{on } (\bar{\tau}_{en}, \bar{\tau}_{ex}). \tag{6.22}$$

Lemma 6.3. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3) and having a regular boundary arc $(\bar{\tau}_{en}, \bar{\tau}_{ex})$. Then the uniform strict complementarity assumption* (6.21) *implies that*

$$\frac{\mathrm{d}^3}{\mathrm{d}t^3} g(\bar{y}(t))|_{t=\bar{\tau}_{en}^-} > 0, \qquad \frac{\mathrm{d}^3}{\mathrm{d}t^3} g(\bar{y}(t))|_{t=\bar{\tau}_{ex}^+} < 0. \tag{6.23}$$

For convenience, Lemma 6.3 will be proved in section 6.3, after the suitable notation has been introduced.

[1] Lemma 4.11 of this thesis.

Perturbed optimal control problem We consider perturbed problems in the following form:

$$(\mathcal{P}^\mu) \qquad \min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^T \ell^\mu(u(t),y(t))\mathrm{d}t + \phi^\mu(y(T)) \tag{6.24}$$

$$\text{subject to} \qquad \dot{y}(t) \;=\; f^\mu(u(t),y(t)) \quad \text{a.e. on } [0,T], \quad y(0) = y_0^\mu \tag{6.25}$$

$$g^\mu(y(t)) \;\leq\; 0 \quad \text{for all } t\in[0,T]. \tag{6.26}$$

Here μ is the perturbation parameter, living in an open subset M_0 of a Banach space M. In what follows, we consider *stable extensions* $(\mathcal{P}^\mu)$ of problem $(\mathcal{P})$ in the following sense.

Definition 6.4. We say that $(\mathcal{P}^\mu)$ is a *stable extension* of $(\mathcal{P})$ if:
(i) There exists $\bar{\mu}\in M_0$ such that $(\mathcal{P}^{\bar{\mu}}) \equiv (\mathcal{P})$;
(ii) The mappings $\mathbb{R}\times\mathbb{R}^n\times M_0\to\mathbb{R}$, $(u,y,\mu)\mapsto\ell^\mu(u,y)$; $\mathbb{R}^n\times M_0\to\mathbb{R}$, $(y,\mu)\mapsto\phi^\mu(y)$; $M_0\to\mathbb{R}^n$, $\mu\mapsto y_0^\mu$ (resp. $\mathbb{R}\times\mathbb{R}^n\times M_0\to\mathbb{R}^n$, $(u,y,\mu)\mapsto f^\mu(u,y)$; $\mathbb{R}^n\times M_0\to\mathbb{R}$, $(y,\mu)\mapsto g^\mu(y)$) are of class C^3 (resp. C^4), with locally Lipschitz continuous third-order (resp. fourth-order) derivatives, uniformly w.r.t. $\mu\in M_0$;
(iii) The dynamics f^μ is uniformly Lipschitz continuous over $\mathbb{R}\times\mathbb{R}^n$ for all $\mu\in M_0$;
(iv) The state constraint is not of first-order, i.e. $(g^\mu)_u^{(1)}(u,y)\equiv 0$ for all $(u,y,\mu)\in\mathbb{R}\times\mathbb{R}^n\times M_0$.

Abstract formulation Given a stable extension $(\mathcal{P}^\mu)$, the mapping $\mathcal{U}\times M_0\to\mathcal{Y}$, $(u,\mu)\mapsto y_u^\mu$, where y_u^μ is the unique solution in $\mathcal{Y}$ of the state equation (6.25), is well-defined, and we may write the following abstract formulation of $(\mathcal{P}^\mu)$

$$\min_{u\in\mathcal{U}} J^\mu(u), \qquad G^\mu(u)\in K, \tag{6.27}$$

with the cost function $J^\mu(u) := \int_0^T \ell^\mu(u,y_u^\mu)\mathrm{d}t + \phi^\mu(y_u^\mu(T))$, the constraint mapping $G^\mu(u) := g^\mu(y_u^\mu)$, and the constraint cone $K = C_-[0,T]$.

Given a stationary point $(\bar{u},\bar{y})$ of $(\mathcal{P})$, we say that the *uniform quadratic growth* condition holds, if for all stable extensions $(\mathcal{P}^\mu)$ of $(\mathcal{P})$, there exists $c,\rho>0$ and a neighborhood $\mathcal{N}$ of $\bar{\mu}$, such that for any stationary point (u^μ,y^μ) of $(\mathcal{P}^\mu)$ with $\mu\in\mathcal{N}$ and $\|u^\mu-\bar{u}\|_\infty<\rho$,

$$J^\mu(u)\geq J^\mu(u^\mu)+c\|u-u^\mu\|_2^2, \quad \text{for all } u\in\mathcal{U}: G^\mu(u)\in K,\ \|u-\bar{u}\|_\infty<\rho. \tag{6.28}$$

Qualification condition and stability of multipliers Robinson's constraint qualification for problem $(\mathcal{P})$ in abstract form (6.27) is as follows (omitting the perturbation parameter at the reference point $\mu=\bar{\mu}$):

$$\exists\,\varepsilon>0, \qquad \varepsilon B_{C[0,T]}\subset G(\bar{u})+DG(\bar{u})\mathcal{U}-K, \tag{6.29}$$

where $B_{C[0,T]}$ denotes the open unit ball of the space $C[0,T]$. It is well-known that a local solution (weak minimum) of $(\mathcal{P})$ satisfying (6.29) is a stationary point of $(\mathcal{P})$. Given $v\in L^r(0,T)$, $1\leq r\leq\infty$, denote by z_v the unique solution in $W^{1,r}(0,T;\mathbb{R}^n)$ of the linearized state equation

$$\dot{z}_v(t) = f_y(\bar{u}(t),\bar{y}(t))z_v(t) + f_u(\bar{u}(t),\bar{y}(t))v(t) \quad \text{a.e. on } [0,T], \quad z_v(0)=0. \tag{6.30}$$

Assumption (A2) implies that Robinson's constraint qualification (6.29) holds, and that the multipliers associated with a stationary point are unique. This is a consequence of the lemma below.

Lemma 6.5 ([21, Prop. 10]). *Let $(\bar u,\bar y)$ be a feasible trajectory of $(\mathcal{P})$ satisfying (A2). Then for all $r\in[1,+\infty]$ and all $\varepsilon\in(0,\sigma)$, with the σ of (6.19), so small that $\Omega_\varepsilon\subset[a,T]$ for some $a>0$, the linear mapping*

$$L^r(0,T)\to W^{2,r}(\Omega_\varepsilon),\qquad v\mapsto (g_y(\bar y(\cdot))z_v(\cdot))|_{\Omega_\varepsilon}, \tag{6.31}$$

where $|_{\Omega_\varepsilon}$ denotes the restriction to the set Ω_ε, is onto, and therefore has a bounded right inverse by the open mapping theorem.

Let us end this section by recalling two results that will be used in the paper.

Proposition 6.6 ([71, Prop. 4.4]). *Let $(\bar u,\bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2). Then for every stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$ and for every stationary point (u,y) of $(\mathcal{P}^\mu)$, with (unique) associated multipliers (p,η) and alternative multipliers (p^2,η^2) given by (6.11)–(6.12), we have:*
(i) *If $\|\mu-\bar\mu\|,\|u-\bar u\|_\infty$,are small enough, then $\mathrm{d}\eta$ is uniformly bounded in $\mathcal{M}[0,T]$;*
Moreover, when $\|\mu-\bar\mu\|,\|u-\bar u\|_\infty\to 0$:
(ii) *$\mathrm{d}\eta$ weakly-* converges to $\mathrm{d}\bar\eta$ in $\mathcal{M}[0,T]$;*
(iii) *p^2 and η^2 converge uniformly to $\bar p^2$ and $\bar\eta^2$, respectively.*

Given $A,B\subset[0,T]$, we denote by $\mathrm{exc}\{A,B\}$ the *Hausdorff excess* of A over B, defined by

$$\mathrm{exc}\{A,B\}\ :=\ \sup_{t\in A}\inf_{s\in B}|t-s|, \tag{6.32}$$

with the convention $\mathrm{exc}\{\emptyset,B\}=0$.

Lemma 6.7 ([71, Lemma 4.6]). *Let $\mathrm{d}\bar\eta\in\mathcal{M}[0,T]$, and a sequence $(\mathrm{d}\eta_n)\subset\mathcal{M}[0,T]$ be such that $\mathrm{d}\eta_n$ weakly-* converges to $\mathrm{d}\bar\eta$ in $\mathcal{M}[0,T]$. Then $e_n:=\mathrm{exc}\{\mathrm{supp}(\mathrm{d}\bar\eta),\mathrm{supp}(\mathrm{d}\eta_n)\}$ converges to zero when $n\to+\infty$.*

6.3 Stability of boundary arcs

The aim of this section is to show that boundary arcs are "stable" under perturbations, for sufficiently smooth perturbations (the *stable extensions* satisfying Def. 6.4). Here is the main result of this section.

Theorem 6.8. *Let $(\bar u,\bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3). Assume that $(\bar u,\bar y)$ has a regular boundary arc $(\bar\tau_{en},\bar\tau_{ex})$ and that (6.21) holds. Then, for every stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$ and for all small enough $\delta>0$, there exist $\rho,\varrho>0$ such that if (u,y) is a stationary point of $(\mathcal{P}^\mu)$ with $\|\mu-\bar\mu\|<\varrho$ and $\|u-\bar u\|_\infty<\rho$, then (u,y) has on $(\bar\tau_{en}-\delta,\bar\tau_{ex}+\delta)$ a unique boundary arc (τ_{en},τ_{ex}) (and no touch point). Moreover, we have that $|\tau_{en}-\bar\tau_{en}|,|\tau_{ex}-\bar\tau_{ex}|<\delta$ and (u,y) satisfies the uniform strict complementarity assumption (6.21) on (τ_{en},τ_{ex}).*

We derive next some useful relations for the proof of Th. 6.8 and Lemma 6.3, and for other results of the paper. Let $(\bar u,\bar y)$ be a stationary point of $(\mathcal{P})\equiv(\mathcal{P}^{\bar\mu})$ satisfying (A2)–(A3) with alternative multipliers $(\bar p^2,\bar\eta^2)$, and let (u,y) be a stationary point of $(\mathcal{P}^\mu)$ with alternative multipliers (p^2,η^2). If $\|\mu-\bar\mu\|$ and $\|u-\bar u\|_\infty$ are small enough, then by (6.19), (6.20), and Prop. 6.6(iii), we have that

$$|(g^\mu)^{(2)}_u(u,y)|\ \ge\ \beta/2\ >\ 0,\qquad \text{a.e. on } I_\sigma(g(\bar y))\supset I(g^\mu(y)), \tag{6.33}$$

$$\tilde H^\mu_{uu}(u,y,p^2,\eta^2)\ \ge\ \alpha/2\ >\ 0\qquad \text{on } [0,T], \tag{6.34}$$

with $\tilde{H}^\mu$ the alternative Hamiltonian (6.14) for $(\mathcal{P}^\mu)$. Moreover, by the implicit function theorem applied locally to (6.17) under hypothesis (A3), we may write that $u(t) = \Upsilon(y(t), p^2(t), \eta^2(t))$ for some C^2 function Υ, and hence u is continuous over $[0,T]$. It follows from Lemma 6.1 that u and η^2 are C^2 on the interior of arcs of the trajectory (u,y). So we may consider the time derivatives of the state constraint of order 3 and 4, defined on the interior of (interior and boundary) arcs by:

$$(g^\mu)^{(3)}(\dot{u},u,y) := (g^\mu)^{(2)}_u(u,y)\dot{u} + (g^\mu)^{(2)}_y(u,y)f^\mu(u,y) \tag{6.35}$$

$$(g^\mu)^{(4)}(\ddot{u},\dot{u},u,y) := (g^\mu)^{(2)}_u(u,y)\ddot{u} + (g^\mu)^{(3)}_u(\dot{u},u,y)\dot{u} + (g^\mu)^{(3)}_y(\dot{u},u,y)f^\mu(u,y). \tag{6.36}$$

Time derivations of (6.17) shows that, on the interior of arcs, where u and η^2 are C^2 (arguments (u,y,p^2,η^2) and time are omitted as well as the superscript μ to simplify the notation)

$$0 = \tilde{H}_{uu}\dot{u} + \tilde{H}_{uy}f - \tilde{H}_y f_u + \dot{\eta}^2 g^{(2)}_u \tag{6.37}$$

$$0 = \tilde{H}_{uu}\ddot{u} + \ddot{\eta}^2 g^{(2)}_u + \Phi_1(\dot{u},\dot{\eta}^2,u,y,p^2,\eta^2,\mu), \tag{6.38}$$

where Φ_1 is a locally Lipschitz continuous function w.r.t. its arguments. By (6.34), multiplying (6.38) by $g^{(2)}_u/\tilde{H}_{uu}$ and using (6.36) we may write that for all $t \in (0,T)$ in the interior of arcs,

$$0 = g^{(4)} + \frac{(g^{(2)}_u)^2}{\tilde{H}_{uu}}\ddot{\eta}^2 + \Phi_2(\dot{u},\dot{\eta}^2,u,y,p^2,\eta^2,\mu), \tag{6.39}$$

where Φ_2 is a locally Lipschitz continuous function w.r.t. its arguments. Moreover, by (6.33), it follows from (6.35) and (6.37) that we may express $\dot{u}$ and $\dot{\eta}^2$ as locally Lipschitz continuous functions of $(g^{(3)},u,y,p^2,\eta^2,\mu)$, i.e. more precisely

$$\begin{aligned}\dot{u} &= (g^{(2)}_u)^{-1}(g^{(3)} - g^{(2)}_y f),\\ \dot{\eta}^2 &= -(g^{(2)}_u)^{-1}(\tilde{H}_{uu}(g^{(2)}_u)^{-1}(g^{(3)} - g^{(2)}_y f) + \tilde{H}_{uy}f - \tilde{H}_y f_u).\end{aligned}$$

Therefore, (6.39) yields, on the interior of arcs,

$$g^{(4)} + \frac{(g^{(2)}_u)^2}{\tilde{H}_{uu}}\ddot{\eta}^2 + \Lambda(g^{(3)},u,y,p^2,\eta^2,\mu) = 0 \tag{6.40}$$

where Λ is a locally Lipschitz continuous function w.r.t. its arguments.

In the sequel, we abbreviate the notation as follows:

$$g^{(3)}(t) := (g^\mu)^{(3)}(\dot{u}(t),u(t),y(t)), \quad g^{(4)}(t) := (g^\mu)^{(4)}(\ddot{u}(t),\dot{u}(t),u(t),y(t)) \tag{6.41}$$

$$\bar{g}^{(3)}(t) := (g^{\bar{\mu}})^{(3)}(\dot{\bar{u}}(t),\bar{u}(t),\bar{y}(t)), \quad \bar{g}^{(4)}(t) := (g^{\bar{\mu}})^{(4)}(\ddot{\bar{u}}(t),\dot{\bar{u}}(t),\bar{u}(t),\bar{y}(t)), \tag{6.42}$$

$$\tilde{H}_{uu}(t) := \tilde{H}^\mu_{uu}(u(t),y(t),p^2(t),\eta^2(t)), \quad \bar{H}_{uu}(t) := \tilde{H}^{\bar{\mu}}_{uu}(\bar{u}(t),\bar{y}(t),\bar{p}^2(t),\bar{\eta}^2(t)), \tag{6.43}$$

$$\Lambda(t) := \Lambda(g^{(3)}(t),u(t),y(t),p^2(t),\eta^2(t),\mu),$$

$$\bar{\Lambda}(t) := \Lambda(\bar{g}^{(3)}(t),\bar{u}(t),\bar{y}(t),\bar{p}^2(t),\bar{\eta}^2(t),\bar{\mu}).$$

We start by the proof of Lemma 6.3 and then give that of Th. 6.8.

Proof of Lemma 6.3. Assume that (6.21) holds. Assume by contradiction that (6.23) does not hold, i.e. $\bar{g}^{(3)}$ is continuous at entry or exit point τ. Then by continuity of $(\bar{u},\bar{y},\bar{p}^2,\bar{\eta}^2)$, (6.40) implies that

$$[\bar{g}^{(4)}(\tau)] + \frac{(\bar{g}^{(2)}_u)^2}{\bar{H}_{uu}}[\dddot{\bar{\eta}}^2(\tau)] = 0. \tag{6.44}$$

In the neighborhood of τ, on the side of the interior arc, we have

$$g(\bar{y}(t)) = \bar{g}^{(4)}(\tau^{\pm})\frac{(t-\tau)^4}{24} + o((t-\tau)^4) \;\leq\; 0,$$

where $\tau^{\pm}$ denotes τ^- if $\tau = \bar{\tau}_{en}$ and τ^+ if $\tau = \bar{\tau}_{ex}$. Since $\bar{g}^{(4)} = 0$ on the interior of the boundary arc, it follows that

$$[\bar{g}^{(4)}(\bar{\tau}_{en})] \geq 0 \qquad \text{and} \qquad [\bar{g}^{(4)}(\bar{\tau}_{ex})] \leq 0. \tag{6.45}$$

Moreover, (6.21) implies that

$$[\ddot{\bar{\eta}}^2(\bar{\tau}_{en})] \geq \beta > 0 \qquad \text{and} \qquad [\ddot{\bar{\eta}}^2(\bar{\tau}_{ex})] \leq -\beta < 0. \tag{6.46}$$

Since $\frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}} > 0$ by (A2)–(A3), the above display and (6.44) yield

$$[\bar{g}^{(4)}(\bar{\tau}_{en})] < 0 \qquad \text{and} \qquad [\bar{g}^{(4)}(\bar{\tau}_{ex})] > 0,$$

contradicting (6.45). Therefore (6.23) holds, which completes the proof. □

Proof of Th. 6.8. Let (u, y) be a stationary point of $(\mathcal{P}^{\mu})$ with u in a L^{∞}-neighborhood of $\bar{u}$ and μ in a neighborhood of $\bar{\mu}$. Assume by contradiction that (u, y) has an interior arc $(\tau_1, \tau_2) \subset (\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta)$. On the interior arc (τ_1, τ_2), u and η^2 are C^2, and $g(t) := g^{\mu}(y(t))$ attains its minimum on (τ_1, τ_2) at a point where the second-order derivative $g^{(2)}$ is nonnegative. Since $g^{(2)}(\tau_i) \leq 0$, $i = 1, 2$, the continuous function $g^{(2)}$ attains its maximum over $[\tau_1, \tau_2]$ at some point $t_m \in (\tau_1, \tau_2)$, and we have at this point of maximum of $g^{(2)}$

$$g^{(3)}(t_m) \;=\; 0 \qquad \text{and} \qquad g^{(4)}(t_m) \;\leq\; 0. \tag{6.47}$$

Assume first that $t_m \in (\bar{\tau}_{en}, \bar{\tau}_{ex})$. By Prop. 6.6(iii), $(y, p^2, \eta^2) \to (\bar{y}, \bar{p}^2, \bar{\eta}^2)$ uniformly over $[0, T]$ when $\|\mu - \bar{\mu}\| \to 0$ and $\|u - \bar{u}\|_{\infty} \to 0$, and $g^{(3)}(t_m) = 0 = \bar{g}^{(3)}(t_m)$ since $t_m \in (\bar{\tau}_{en}, \bar{\tau}_{ex})$. Therefore, $\Lambda(t_m) - \bar{\Lambda}(t_m) \to 0$, and hence (6.40) implies that when $\|\mu - \bar{\mu}\| \to 0$ and $\|u - \bar{u}\|_{\infty} \to 0$,

$$g^{(4)}(t_m) + \frac{(g_u^{(2)})^2}{\tilde{H}_{uu}}\ddot{\eta}^2(t_m) \;-\; (\bar{g}^{(4)}(t_m) + \frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}}\ddot{\bar{\eta}}^2(t_m)) \;\to\; 0.$$

But $\ddot{\eta}^2(t_m) = 0$ since we are on an interior arc for (u, y), and $\bar{g}^{(4)}(t_m) = 0$ since we are on a boundary arc for $(\bar{u}, \bar{y})$. It follows that when $\|\mu - \bar{\mu}\| \to 0$ and $\|u - \bar{u}\|_{\infty} \to 0$,

$$g^{(4)}(t_m) \;-\; \frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}}\ddot{\bar{\eta}}^2(t_m) \;\to\; 0.$$

Since $\frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}} \geq C > 0$ by (6.19) and (6.20), we obtain by (6.21) that $\frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}}\ddot{\bar{\eta}}^2(t_m) \geq C\beta > 0$. Therefore, for $\|\mu - \bar{\mu}\|$ and $\|u - \bar{u}\|_{\infty}$ small enough, $g^{(4)}(t_m) \geq C\beta/2 > 0$, contradicting (6.47).

Assume now that $t_m \in (\bar{\tau}_{en} - \delta, \bar{\tau}_{en}]$ (the case when $t_m \in [\bar{\tau}_{ex}, \bar{\tau}_{ex} + \delta)$ is analogous). For all $0 < \varepsilon < \delta$, if $\|\mu - \bar{\mu}\|$ and $\|u - \bar{u}\|_{\infty}$ are small enough, then $g^{\mu}(y(t)) < 0$ on the interval $[\bar{\tau}_{en} - \delta, \bar{\tau}_{en} - \varepsilon]$. This implies that $t_m \uparrow \bar{\tau}_{en}$ when $\|\mu - \bar{\mu}\| \to 0$ and $\|u - \bar{u}\|_{\infty} \to 0$. Therefore, since $g^{(3)}(t_m) = 0 = \bar{g}^{(3)}(\bar{\tau}_{en}^+)$ and $(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)$ is continuous over $[0, T]$, we obtain by Prop. 6.6(iii) that $\Lambda(t_m) \to \bar{\Lambda}(\bar{\tau}_{en}^+)$. It follows then from (6.40) that

$$g^{(4)}(t_m) + \frac{(g_u^{(2)})^2}{\tilde{H}_{uu}}\ddot{\eta}^2(t_m) \;\to\; \bar{g}^{(4)}(\bar{\tau}_{en}^+) + \frac{(\bar{g}_u^{(2)})^2}{\bar{H}_{uu}}\ddot{\bar{\eta}}^2(\bar{\tau}_{en}^+) \;\geq\; 0 + C\beta > 0,$$

contradicting (6.47) again since $g^{(4)}(t_m) \leq 0$ and $\ddot{\eta}^2(t_m) = 0$. This shows that for all small $\delta > 0$, if $\|u - \bar{u}\|_\infty$ and $\|\mu - \bar{\mu}\|$ are small enough, then (u, y) has *no interior arc* contained in $(\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta)$.

It follows that $I(g^\mu(y)) \cap (\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta)$ is either empty, or a touch point, or a boundary arc. Let us refute the two first possibilities. For all small $\varepsilon > 0$, if $\|u - \bar{u}\|_\infty$ and $\|\mu - \bar{\mu}\|$ are small enough, then $I(g^\mu(y)) \subset I_\varepsilon(g(\bar{y}))$, and by hypothesis (6.21), Prop. 6.6(ii) and Lemma 6.7 (recall that $\mathrm{d}\dot{\eta}^2 = \mathrm{d}\eta$), for all $t \in [\bar{\tau}_{en}, \bar{\tau}_{ex}]$, there exists $s \in \mathrm{supp}(\mathrm{d}\dot{\eta}^2) \subset I(g^\mu(y))$ such that $|t - s| < \varepsilon$. Therefore, we deduce that $I(g^\mu(y)) \cap (\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta)$ is necessarily a boundary arc (τ_{en}, τ_{ex}), and that $|\tau_{en} - \bar{\tau}_{en}|, |\tau_{ex} - \bar{\tau}_{ex}| < \varepsilon$.

It remains to show that uniform strict complementarity holds on that boundary arc. By (6.40), it holds for all t in boundary arc (τ_{en}, τ_{ex}) that

$$\ddot{\eta}^2(t) = -\frac{\tilde{H}_{uu}(t)}{(g_u^{(2)}(t))^2}\Lambda(0, u(t), y(t), p^2(t), \eta^2(t), \mu). \tag{6.48}$$

The same relation applied to $(\bar{u}, \bar{y})$, the uniform strict complementarity assumption (6.21) and (A2)–(A3) imply that $\Lambda(0, \bar{u}(t), \bar{y}(t), \bar{p}^2(t), \bar{\eta}^2(t), \bar{\mu}) \leq -C$ for some positive constant C, for all $t \in [\bar{\tau}_{en}, \bar{\tau}_{ex}]$. Therefore, by continuity $\Lambda(0, \bar{u}(t), \bar{y}(t), \bar{p}^2(t), \bar{\eta}^2(t), \bar{\mu}) \leq -C/2$ for all $t \in (\bar{\tau}_{en} - \delta, \bar{\tau}_{ex} + \delta) \supset (\tau_{en}, \tau_{ex})$ for $\delta > 0$, $\|u - \bar{u}\|_\infty$ and $\|\mu - \bar{\mu}\|$ small enough. By Prop. 6.6(iii), for small enough $\|u - \bar{u}\|_\infty$ and $\|\mu - \bar{\mu}\|$, (u, y, p^2, η^2) is arbitrarily close to $(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)$ in L^∞ and hence $\Lambda(0, u(t), y(t), p^2(t), \eta^2(t), \mu) \leq -C/4$ on (τ_{en}, τ_{ex}). It follows then from (6.33)–(6.34) and (6.48) that $\ddot{\eta}^2$ is uniformly positive over (τ_{en}, τ_{ex}). This achieves the proof of the theorem. □

Remark 6.9. The regularity of the class of perturbations considered (here, satisfying Def. 6.4) is crucial to show the stability of boundary arcs, as it is the case for first-order state constraints (see [20, Th. 2.1][2]). If the perturbation is not sufficiently smooth, then boundary arcs are not stable, even if the uniform strict complementarity assumption (6.21) holds, as it is shown in [92, section 2] for a first-order state constraint and a perturbation that goes to zero in the L^2 norm but not in the $W^{1,\infty}$ norm.

6.4 Instability of nonreducible touch points

Definition 6.10. Let $\bar{\tau}_{to} \in (0, T)$ be a touch point of a stationary point $(\bar{u}, \bar{y})$ of $(\mathcal{P})$, with alternative multipliers $(\bar{p}^2, \bar{\eta}^2)$.
(a) We say that $\bar{\tau}_{to}$ is *reducible*, if (i) $t \mapsto g^{(2)}(\bar{u}(t), \bar{y}(t))$ is continuous at point $\bar{\tau}_{to}$ (which always holds under assumption (A3)) and (ii)

$$g^{(2)}(\bar{u}(\bar{\tau}_{to}), \bar{y}(\bar{\tau}_{to})) < 0. \tag{6.49}$$

(b) We say that $\bar{\tau}_{to}$ is *essential*, if

$$[\dot{\bar{\eta}}^2(\bar{\tau}_{to})] > 0. \tag{6.50}$$

Remark 6.11. Using the classical multipliers $(\bar{p}, \bar{\eta})$ associated with $(\bar{u}, \bar{y})$ in (6.7)–(6.10) (recall that $\bar{\eta} = \dot{\bar{\eta}}^2$), (6.50) is equivalent to

$$[\bar{\eta}(\bar{\tau}_{to})] > 0, \tag{6.51}$$

which is in accordance with the classical definition of essential touch points.

[2]Theorem 3.4 of this thesis.

Let $(\bar{u},\bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3). Assume that $(\bar{u},\bar{y})$ has a reducible touch point $\bar{\tau}_{to}$. Then given a stationary point (u,y) of $(\mathcal{P}^\mu)$ such that $\|\mu-\bar{\mu}\|$ and $\|u-\bar{u}\|_\infty$ are small enough, it is easy to see (see e.g. [21, section 5.1][3]) that the mapping $t\mapsto g^\mu(y(t))$ attains its maximum over a neighborhood $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ of $\bar{\tau}_{to}$, $\delta>0$, at a unique point τ_{to}. Therefore, if $g^\mu(y(\tau_{to}))=0$, (u,y) has a unique touch point in $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$, and if $g^\mu(y(\tau_{to}))<0$, the state constraint is locally not active in a neighborhood of $\bar{\tau}_{to}$. Moreover, by Prop. 6.6(ii) and relation (6.11), $\mathrm{d}\dot{\eta}^2$ weakly-* converges in $\mathcal{M}[0,T]$ to $\mathrm{d}\dot{\bar{\eta}}^2$ when $\|\mu-\bar{\mu}\|,\|u-\bar{u}\|_\infty\to 0$. Therefore, if strict complementarity holds at $\bar{\tau}_{to}$, i.e. if $\bar{\tau}_{to}$ is an essential touch point, this implies that for $\delta>0$, $\|\mu-\bar{\mu}\|$ and $\|u-\bar{u}\|_\infty$ small enough, $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)\cap\mathrm{supp}(\mathrm{d}\dot{\eta}^2)\neq\emptyset$. Hence by (6.18) we necessarily have $g^\mu(y(\tau_{to}))=0$, i.e. τ_{to} is a (essential) touch point of (u,y).

The above discussion shows that touch points that are both reducible and essential are stable. When strict complementarity does not hold, there are two possibilities for nonessential reducible touch points: either the state constraint of the perturbed problem is not active on a neighborhood of $\bar{\tau}_{to}$, or it is active in a neighborhood of $\bar{\tau}_{to}$ at a unique touch point, the latter being essential or not.

We see that the reducibility hypothesis (6.49) excludes other structural changes. In what follows, we release this reducibility hypothesis and and show that two possible changes in the structure of perturbed stationary points may happen in the neighborhood of a nonreducible touch point: The apparition of a boundary arc or the apparition of a second touch point.

Let now $\bar{\tau}_{to}$ be a nonreducible touch point of $(\bar{u},\bar{y})$, i.e. such that

$$g^{(2)}(\bar{u}(\bar{\tau}_{to}),\bar{y}(\bar{\tau}_{to}))\;=\;0. \tag{6.52}$$

We consider the following assumption (compare to (6.23))

$$\begin{aligned}\frac{\mathrm{d}^3}{\mathrm{d}t^3}g(\bar{y}(t))|_{t=\bar{\tau}_{to}^-}&=g^{(3)}(\dot{\bar{u}}(\bar{\tau}_{to}^-),\bar{u}(\bar{\tau}_{to}),\bar{y}(\bar{\tau}_{to}))\;>\;0,\\ \frac{\mathrm{d}^3}{\mathrm{d}t^3}g(\bar{y}(t))|_{t=\bar{\tau}_{to}^+}&=g^{(3)}(\dot{\bar{u}}(\bar{\tau}_{to}^+),\bar{u}(\bar{\tau}_{to}),\bar{y}(\bar{\tau}_{to}))\;<\;0.\end{aligned} \tag{6.53}$$

By (6.35) and (6.37), the jumps of $g^{(3)}$ and $\dot{\eta}^2$ at a touch point τ_{to} are related by

$$[g^{(3)}(\dot{u},u,y)(\tau_{to})]\;=\;g_u^{(2)}(u,y)[\dot{u}(\tau_{to})]\;=\;-\frac{(g_u^{(2)}(u,y))^2}{\tilde{H}_{uu}(u,y,p^2,\eta^2)}[\dot{\eta}^2(\tau_{to})]\;\leq\;0, \tag{6.54}$$

where we have $[\dot{\eta}^2(\tau_{to})]=[\eta(\tau_{to})]$ by (6.11). Therefore, if (6.53) holds, this implies by (A2)–(A3) that $[\dot{\bar{\eta}}^2(\bar{\tau}_{to})]>0$. We obtain then the following result.

Lemma 6.12. *Let $(\bar{u},\bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3) and having a nonreducible touch point $\bar{\tau}_{to}\in(0,T)$ satisfying* (6.53)*. Then $\bar{\tau}_{to}$ is an essential touch point, i.e. satisfies* (6.50)*.*

Let (u,y) be a stationary point of $(\mathcal{P}^\mu)$, with $\|\mu-\bar{\mu}\|$ and $\|u-\bar{u}\|_\infty$ arbitrarily small. We use the notations (6.41)–(6.43). At a nonreducible touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, we cannot ensure that the state constraint of the perturbed problem $g(t):=g^\mu(y(t))$ will have a unique maximum point in a neighborhood $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ of $\bar{\tau}_{to}$, for small $\delta>0$.

[3]Section 1.5.2 of this thesis.

So let us assume that $g(t)$ has either a boundary arc or an interior arc included in $(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$. We deduce in both cases the existence of a time $t_m\in(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$ where $g^{(2)}$ is maximum (similar to the proof of Th. 6.8) such that

$$g^{(3)}(t_m)=0.$$

For all $\delta>0$, if $\|\mu-\bar\mu\|$ and $\|u-\bar u\|_\infty$ are small enough, then $I(g^\mu(y))\subset I_\delta(g(\bar y))$. Letting $\delta\downarrow 0$, we obtain that $t_m\to\bar\tau_{to}$ when $\|\mu-\bar\mu\|\to 0$ and $\|u-\bar u\|_\infty\to 0$. Hence, (6.40) implies that when $\|\mu-\bar\mu\|\to 0$ and $\|u-\bar u\|_\infty\to 0$, using Prop. 6.6(iii),

$$\begin{aligned} g^{(4)}(t_m)+\frac{(g_u^{(2)}(t_m))^2}{\tilde H_{uu}(t_m)}\ddot\eta^2(t_m) &= -\Lambda(0,u(t_m),y(t_m),p^2(t_m),\eta^2(t_m),\mu)\\ &\to -\Lambda(0,\bar u(\bar\tau_{to}),\bar y(\bar\tau_{to}),\bar p^2(\bar\tau_{to}),\bar\eta^2(\bar\tau_{to}),\bar\mu). \end{aligned}\tag{6.55}$$

Therefore, if (u,y) has a boundary arc in $(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$, we have that $g^{(4)}(t_m)=0$ and $\ddot\eta^2(t_m)\ge 0$, which implies that

$$\Lambda(0,\bar u(\bar\tau_{to}),\bar y(\bar\tau_{to}),\bar p^2(\bar\tau_{to}),\bar\eta^2(\bar\tau_{to}),\bar\mu)\ \le\ 0.\tag{6.56}$$

If (u,y) has an interior arc in $(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$, then $g^{(4)}(t_m)\le 0$ (this was shown in the proof of Th. 6.8, recall (6.47)) and $\ddot\eta^2(t_m)=0$. This implies that

$$\Lambda(0,\bar u(\bar\tau_{to}),\bar y(\bar\tau_{to}),\bar p^2(\bar\tau_{to}),\bar\eta^2(\bar\tau_{to}),\bar\mu)\ \ge\ 0.\tag{6.57}$$

Conversely, if (6.56) holds with a strict inequality, then for $\|\mu-\bar\mu\|$ and $\|u-\bar u\|_\infty$ small enough, $g^{(4)}(t_m)+\frac{(g_u^{(2)}(t_m))^2}{\tilde H_{uu}(t_m)}\ddot\eta^2(t_m)>0$, excluding the possibility of an interior arc. Similarly, if (6.57) holds with a strict inequality, this excludes the possibility of a boundary arc. Using the above arguments, we are able to obtain the following result.

Theorem 6.13. *Let $(\bar u,\bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3). Assume that $(\bar u,\bar y)$ has a nonreducible and essential touch point $\bar\tau_{to}\in(0,T)$. Set*

$$\bar\lambda(\bar\tau_{to})\ :=\ \Lambda(0,\bar u(\bar\tau_{to}),\bar y(\bar\tau_{to}),\bar p^2(\bar\tau_{to}),\bar\eta^2(\bar\tau_{to}),\bar\mu).\tag{6.58}$$

Then, for every stable extension $(\mathcal{P}^\mu)$ and for all $\delta>0$ small enough, there exist $\rho,\varrho>0$ such that:
(i) *If $\bar\lambda(\bar\tau_{to})<0$ holds, then all stationary points (u,y) of the perturbed problem $(\mathcal{P}^\mu)$ with $\|\mu-\bar\mu\|<\varrho$ and $\|u-\bar u\|_\infty<\rho$ have either a single touch point τ_{to} or a single boundary arc (τ_{en},τ_{ex}) in $(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$. Moreover, in case of a boundary arc (τ_{en},τ_{ex}), (u,y) satisfies the uniform strict complementarity assumption* (6.21) *on (τ_{en},τ_{ex}).*
(ii) *If $\bar\lambda(\bar\tau_{to})>0$ holds, then all stationary points (u,y) of the perturbed problem $(\mathcal{P}^\mu)$ with $\|\mu-\bar\mu\|<\varrho$ and $\|u-\bar u\|_\infty<\rho$ have either one or two touch points in $(\bar\tau_{to}-\delta,\bar\tau_{to}+\delta)$ and no boundary arc.*

Remark 6.14. Under the assumptions of the above theorem, if $\bar\lambda(\bar\tau_{to})=0$ holds, then we cannot conclude and any structure in the neighborhood of $\bar\tau_{to}$ is a priori possible for a stationary point (u,y) of the perturbed problem $(\mathcal{P}^\mu)$, however small $\|u-\bar u\|_\infty$ and $\|\mu-\bar\mu\|$ are (see Example 6.15 below).

Proof of Th. 6.13. Note first that since $\bar\tau_{to}$ is essential, it follows from Prop. 6.6(ii) and Lemma 6.7 that for $\delta > 0$ and $\|u-\bar u\|_\infty$ and $\|\mu-\bar\mu\|$ small enough, $I(g^\mu(y)) \cap (\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$ is not empty. In view of what precedes, it remains to show in the case (ii) when $\bar\lambda(\bar\tau_{to}) > 0$ that (u,y) cannot have more than one interior arc included in $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$. Since boundary arcs are not possible either, this will show that the only two possibilities for (u,y) is to have one or two touch points in $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$.

If $\bar\lambda(\bar\tau_{to}) > 0$, then we see by (6.55) that on an interior arc included in $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$, for $\|u-\bar u\|_\infty$ and $\|\mu-\bar\mu\|$ small enough, for t in the interior arc, the functions (of time) $(g^\mu)^{(4)}(\ddot u, \dot u, u, y)$ being Lipschitz continuous on interior arcs by Lemma 6.1, uniformly w.r.t. μ by Definition 6.4 of a stable extension,

$$g^{(4)}(t) \le -\tfrac{1}{2}\bar\lambda(\bar\tau_{to}) < 0,$$

and hence $g^{(3)}$ is strictly decreasing along an interior arc. In addition, $g^{(3)}$ vanishes at some point t_m on the interior of an interior arc where $g^{(2)}$ is maximum and satisfying (6.47). Now assume that (u,y) has two interior arcs in $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$, say (τ_1,τ_2) and (τ_2,τ_3). Since $g^{(3)}$ is strictly decreasing on the interior arcs and vanishes at an interior point of these arcs, this implies that $g^{(3)}(\tau_2^-) < 0$ and $g^{(3)}(\tau_2^+) > 0$, and hence, $[g^{(3)}(\tau_2)] > 0$. But at the touch point τ_2, $[g^{(3)}(\tau_2)] \le 0$ by (6.54), which gives the desired contradiction and shows that (u,y) can only have a single interior arc in $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$, for small enough $\|u-\bar u\|_\infty$ and $\|\mu-\bar\mu\|$ and $\delta > 0$.

We end the proof by checking that in the case (i), uniform strict complementarity holds on the boundary arc (τ_{en}, τ_{ex}). By (6.40) and (6.33), for all t in boundary arc (τ_{en}, τ_{ex}) we have that

$$\ddot\eta^2(t) = -\frac{\tilde H_{uu}(t)}{(g_u^{(2)}(t))^2}\Lambda(0, u(t), y(t), p^2(t), \eta^2(t), \mu). \tag{6.59}$$

Since $c := \bar\lambda(\bar\tau_{to}) < 0$, it follows that for $\delta > 0$ small enough, $\Lambda(0, \bar u(t), \bar y(t), \bar p^2(t), \bar\eta^2(t)) < c/2 < 0$ on $(\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$. For $\|u-\bar u\|_\infty$ and $\|\mu-\bar\mu\|$ small enough, (u, y, p^2, η^2) is arbitrarily close to $(\bar u, \bar y, \bar p^2, \bar\eta^2)$ in L^∞ by Prop. 6.6(iii), so if (u,y) has a boundary arc $(\tau_{en}, \tau_{ex}) \subset (\bar\tau_{to}-\delta, \bar\tau_{to}+\delta)$, we deduce that $\Lambda(0, u(t), y(t), p^2(t), \eta^2(t), \mu) \le c/4 < 0$ on (τ_{en}, τ_{ex}). With (6.34)–(6.33) and (6.59) this shows that $\ddot\eta^2$ is uniformly positive on (τ_{en}, τ_{ex}). This achieves the proof of the theorem. □

Example 6.15. Consider the problem below:

$$\min_{(u,y)\in\mathcal{U}\times\mathcal{Y}} \int_0^1 \left(\frac{u(t)^2}{2} + \mu_1 y_1(t)\right) \mathrm{d}t$$

subject to the dynamics and boundary conditions[4]

$$\dot y_1(t) = y_2(t), \quad \dot y_2(t) = u(t), \tag{6.60}$$

$$y_1(0) = y_1(1) = 0, \quad \dot y_1(0) = 1 = -\dot y_2(1) \tag{6.61}$$

and second-order state constraint

$$y_1(t) \le \mu_2.$$

[4]Extension of the results of this paper when there are constraints on the final and/or the initial state is possible if a strong controllability condition is assumed, see [17, Section 8][5].

The perturbation parameter is $(\mu_1, \mu_2) \in \mathbb{R} \times \mathbb{R}_+^*$. The above problem was studied in [29] for $\mu_1 = 0$ and in [4] for $\mu_1 \neq 0$. By convexity, the first-order optimality condition is necessary and sufficient and the problem has a unique optimal solution.

For the unconstrained problem, the optimality condition reduces to $y_1^{(4)} \equiv -\mu_1$, together with the boundary conditions (6.61). Therefore the unconstrained optimal trajectory is given by

$$y_1^{uncons}(t) = -\frac{\mu_1}{24}t^4 + \frac{\mu_1}{12}t^3 - \left(1 + \frac{\mu_1}{24}\right)t^2 + t.$$

Its derivatives being given by $\dot{y}_1^{uncons}(t) = \left(t - \frac{1}{2}\right)\left(-\frac{\mu_1}{6}t^2 + \frac{\mu_1}{6}t - 2\right)$ and $\ddot{y}_1^{uncons}(t) = \frac{\mu_1}{2}t(1-t) - 2 - \frac{\mu_1}{12}$, this fourth-order polynomial has on $[0,1]$ a maximum at $t = \frac{1}{2}$ for $\mu_1 \leq 48$, and one local minimum at $t = \frac{1}{2}$ and two maxima, one in $(0, \frac{1}{2})$ and the other in $(\frac{1}{2}, 1)$, for $\mu > 48$. For $\mu_1 \leq 48$ and $\mu_2 = y_1^{uncons}(\frac{1}{2}) = \frac{1}{4} - \frac{\mu_1}{384}$, we have therefore a nonessential touch point at $\tau_{to} = \frac{1}{2}$, which is reducible for $\mu_1 < 48$.

In the sequel we shall consider the case when $\mu_1 < 48$. When μ_2 decreases beyond the value $\frac{1}{4} - \frac{\mu_1}{384}$, the optimal trajectory has one touch point at $\tau_{to} = \frac{1}{2}$ and is given by

$$y_1^{onetouch}(t) = \begin{cases} -\frac{\mu_1}{24}t^4 + \frac{a}{6}t^3 + \frac{b}{2}t^2 + t & \text{on } [0, \frac{1}{2}] \\ -\frac{\mu_1}{24}(t-1)^4 - \frac{a}{6}(t-1)^3 + \frac{b}{2}(t-1)^2 - (t-1) & \text{on } [\frac{1}{2}, 1] \end{cases}$$

with $a = 24 + \frac{\mu_1}{4} - 96\mu_2$ and $b = -8 - \frac{\mu_1}{48} + 24\mu_2$. This touch point becomes nonreducible when $\ddot{y}_1^{onetouch}(\tau_{to}) = 0$ i.e. when $\mu_2 = \frac{1}{6} - \frac{\mu_1}{1152}$, and satisfies (6.53).

So let us compute the term (6.58) at the optimal trajectory for a given value of $\bar{\mu}_1 \in (-\infty, 48)$ and $\bar{\mu}_2 := \frac{1}{6} - \frac{\bar{\mu}_1}{1152}$. We have that

$$g(y) = y_1 - \mu_2, \quad g^{(1)}(y) = y_2, \quad g^{(2)}(u, y) = u, \quad g^{(3)}(\dot{u}, u, y) = \dot{u}, \quad g^{(4)}(\ddot{u}, \dot{u}, u, y) = \ddot{u}.$$

The alternative Hamiltonian (6.14) is given by

$$\tilde{H}^{\mu}(u, y, p^2, \eta^2) \;=\; \frac{u^2}{2} + \mu_1 y_1 + p_1^2 y_2 + p_2^2 u + \eta^2 u$$

and the costate and control equations (6.16) and (6.17) are given by

$$\begin{aligned} -\dot{p}_1^2 &= \mu_1, \qquad -\dot{p}_2^2 \;=\; p_1^2, \\ 0 &= u + p_2^2 + \eta^2. \end{aligned}$$

Differentiating twice the last above relation, we obtain

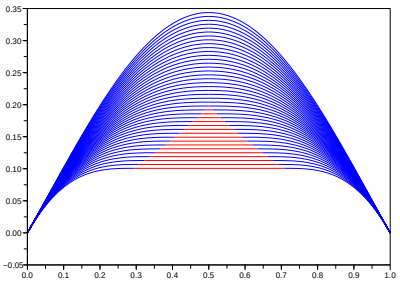
$$0 \;=\; \ddot{u} + \mu_1 + \ddot{\eta}^2 \;=\; g^{(4)} + \mu_1 + \ddot{\eta}^2.$$

Identifying with (6.40), we simply have that $\Lambda(g^{(3)}, u, y, p^2, \eta^2, \mu) = \mu_1$, and hence, at the nonreducible touch point $\bar{\tau}_{to} = \frac{1}{2}$,

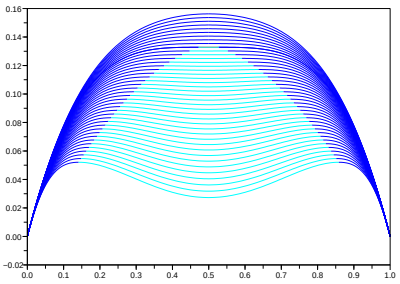
$$\bar{\lambda}(\bar{\tau}_{to}) = \bar{\mu}_1.$$

Conditions (i) and (ii) of Th. 6.13 are satisfied respectively for $\bar{\mu}_1 < 0$ and for $\bar{\mu}_1 > 0$ (see figure 6.1 below). Therefore, for $\mu_2 < \frac{1}{6} - \frac{\mu_1}{1152}$, the touch point turns into two touch points if $\mu_1 > 0$ and turns into a boundary arc if $\mu_1 < 0$, and strict complementarity holds on that boundary arc since $\ddot{\eta}^2 \equiv -\mu_1 > 0$.

If $\bar{\mu}_1 = 0$, then $\bar{\lambda}(\bar{\tau}_{to}) = 0$ and we cannot conclude for the structure of the solutions of the perturbed problem. For $\mu_2 < \frac{1}{6}$, a boundary arc appears but strict complementarity does not hold on that boundary arc since $\ddot{\eta}^2 \equiv -\mu_1 = 0$. If we take e.g. $\mu_2 = \frac{1}{6} - \frac{\mu_1}{1152} - \varepsilon\mu_1^2$, with $\varepsilon > 0$ a fixed parameter, we have in the neighborhood of the nonreducible touch point $\bar{\tau}_{to}$ a boundary arc for $\mu_1 < 0$ and two touch points for $\mu_1 > 0$.



(a) State constraint for $\bar{\mu}_1 = -36$ and varying μ_2.



(b) State constraint for $\bar{\mu}_1 = 36$ and varying μ_2.

Figure 6.1: Transformation of a nonreducible touch point into a boundary arc or into two touch points for $\bar{\mu}_1 \neq 0$ when μ_2 decreases.

6.5 Stability analysis

Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ with alternative multipliers $(\bar{p}^2, \bar{\eta}^2)$. Let $\mathcal{V} := L^2(0,T)$. For $v \in \mathcal{V}$, recall that we denote by z_v the unique solution in $H^1(0,T;\mathbb{R}^n)$ of the linearized state equation (6.30). The quadratic form involved in the second-order optimality conditions in [71] is as follows: For $v \in \mathcal{V} = L^2(0,T)$,

$$\mathcal{Q}(v) := \int_0^T D^2_{(u,y)(u,y)} \tilde{H}(\bar{u}, \bar{y}, \bar{p}^2, \bar{\eta}^2)((v, z_v), (v, z_v))\mathrm{d}t \; + \; \phi_{yy}(\bar{y}(T))(z_v(T), z_v(T)). \tag{6.62}$$

The *extended critical cone* used in the stability analysis is defined as the set of $v \in \mathcal{V}$ such that

$$g_y(\bar{y}(t))z_v(t) = 0 \qquad \text{for all } t \in \operatorname{supp}(\mathrm{d}\dot{\bar{\eta}}^2). \tag{6.63}$$

This set is obtained from the classical critical cone, defined as the set of $v \in \mathcal{V}$ satisfying (6.63) and

$$g_y(\bar{y}(t))z_v(t) \leq 0 \qquad \text{for all } t \in I(g(\bar{y})) \setminus \operatorname{supp}(\mathrm{d}\dot{\bar{\eta}}^2), \tag{6.64}$$

by omission of the inequality constraint (6.64). The strong second-order sufficient condition used in the stability analysis is:

$$\mathcal{Q}(v) > 0, \quad \text{for all } v \in \mathcal{V},\ v \neq 0,\ \text{satisfying (6.63)}. \tag{6.65}$$

This condition is a natural strengthening of the second-order sufficient condition of [21, Th. 18][6]

$$\mathcal{Q}(v) > 0, \quad \text{for all } v \in \mathcal{V},\ v \neq 0,\ \text{satisfying (6.63)-(6.64)}. \tag{6.66}$$

The strengthened Legendre-Clebsch condition (6.20) implies that the quadratic form $\mathcal{Q}$ is a *Legendre form*, i.e. a weakly lower semi-continuous quadratic form with the property that if a sequence v_n weakly converges to v in L^2 and if $\mathcal{Q}(v_n) \to \mathcal{Q}(v)$, then v_n converges to v strongly in L^2. Consequently, (6.65) (resp. (6.66)) is equivalent to the existence of some $c > 0$ such that $\mathcal{Q}(v) \geq c\|v\|_2^2$ for all $v \in \mathcal{V}$ satisfying (6.63) (resp. satisfying (6.63)-(6.64)).

[6] Theorem 1.18 of this thesis.

For first-order state constraints, the stability analysis for the homotopy algorithm in [20] was conducted using a shooting approach. For second-order state constraints, a shooting approach can be used for the stability analysis if all the touch points are reducible, see [94, 19], but not in presence of nonreducible touch points, since in that case the structure is not stable by Th. 6.13. For this reason, a stability result has been obtained in [71] (Th. 6.16 below) that makes no assumptions on the structure of the trajectory, and hence applies when the structure of the trajectory is not stable. This result is based on a variant of Robinson's strong regularity theory [121] and extends the stability results known for first-order state constraints, see [53, 88].

Theorem 6.16 ([71, Th. 4.3]). *Let $(\bar{u}, \bar{y})$ be a local solution of $(\mathcal{P})$, satisfying (A2)–(A3) and the strong second-order sufficient condition* (6.65)*, and let $(\mathcal{P}^\mu)$ be a stable extension of $(\mathcal{P})$. Then there exist $c, \rho, \kappa, \tilde{\kappa} > 0$ and a neighborhood $\mathcal{N}$ of $\bar{\mu}$, such that for all $\mu \in \mathcal{N}$, $(\mathcal{P}^\mu)$ has a unique stationary point (u^μ, y^μ) with $\|u^\mu - \bar{u}\|_\infty < \rho$ and unique associated alternative multipliers $(p^{2,\mu}, \eta^{2,\mu})$, and for all $\mu, \mu' \in \mathcal{N}$,*

$$\|u^\mu - u^{\mu'}\|_2,\ \|y^\mu - y^{\mu'}\|_{1,2},\ \|p^{2,\mu} - p^{2,\mu'}\|_{1,2},\ \|\eta^{2,\mu} - \eta^{2,\mu'}\|_2 \ \leq\ \kappa\|\mu - \mu'\|, \tag{6.67}$$

$$\|u^\mu - u^{\mu'}\|_\infty, \|y^\mu - y^{\mu'}\|_{1,\infty}, \|p^{2,\mu} - p^{2,\mu'}\|_{1,\infty}, \|\eta^{2,\mu} - \eta^{2,\mu'}\|_\infty \ \leq\ \tilde{\kappa}\|\mu - \mu'\|^{2/3}. \tag{6.68}$$

Moreover, (u^μ, y^μ) is a local solution of $(\mathcal{P}^\mu)$ satisfying the uniform quadratic growth condition (6.28) *and the strong second-order sufficient condition* (6.65)*.*

Proof. The theorem follows from [71, Th. 4.3][7], excepted for the fact that (u^μ, y^μ) satisfies the strong second-order sufficient condition (6.65). The latter can be proved by contradiction, by a slight modification of the proof of [71, Prop. 4.2][8], using Prop. 6.6, Lemma 6.7, and the fact that $\mathcal{Q}$ is a Legendre form. □

6.6 The shooting algorithm

By Th. 6.16, the perturbed problem $(\mathcal{P}^\mu)$ has a locally unique local solution. The objective of this section is to see, under additional assumptions, how we could use the shooting algorithm and the results of Theorems 6.8 and 6.13 to obtain in practice in the homotopy algorithm the solution of the perturbed problem.

Let us first recall the shooting algorithm for a second-order scalar state constraint (see [29, 115, 94, 19]). The alternative multipliers used in the shooting algorithm are denoted by (p_2, η_2), with the '2' as subscript, not to be confused with the multipliers (p^2, η^2) (with the '2' as superscript) used in the stability analysis. Let us recall that the multipliers used in the shooting algorithm (p_2, η_2) are defined, on each boundary arc (τ_{en}, τ_{ex}) of the trajectory, by

$$\eta_1(t) \ :=\ \int_{(t,\tau_{ex}]} \mathrm{d}\eta(s) = \eta(\tau_{ex}^+) - \eta(t^+), \qquad \eta_2(t) \ :=\ \int_t^{\tau_{ex}} \eta_1(s)\mathrm{d}s, \tag{6.69}$$

$$p_2(t) \ :=\ p(t) - \eta_1(t) g_y(y(t)) - \eta_2 g_y^{(1)}(y(t)) \tag{6.70}$$

and $\eta_1(t), \eta_2(t), p_2(t) = 0$ outside boundary arcs. Here p and η denote the multipliers associated with a stationary point (u, y) in the classical optimality condition (6.7)–(6.10).

Why do we use so many different multipliers? The multipliers η^2, p^2 are very useful in the stability analysis because they are continuous and converge uniformly. The multipliers (p_2, η_2)

[7] Theorem 5.12 of this thesis.
[8] Proposition 5.11 of this thesis.

used in the shooting algorithm have jumps, and these jumps are used as additional degrees of freedom in the shooting algorithm, in order to have as many free parameters as conditions to satisfy. An explicit relation between these multipliers (p_2, η_2) and (p^2, η^2) is made precise later, see (6.113)–(6.115).

Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A3) and the assumption below:

(A4) $(\bar{u}, \bar{y})$ has finitely many boundary arcs and finitely many touch points and the state constraint is not active at final time, i.e. $g(\bar{y}(T)) < 0$.

Denote by $\bar{\mathcal{T}}_{en}$, $\bar{\mathcal{T}}_{ex}$ and $\bar{\mathcal{T}}_{to}$ the (finite and possibly empty) sets of respectively entry, exit and touch times of the trajectory $(\bar{u}, \bar{y})$, and its set of junction points by $\bar{\mathcal{T}} := \bar{\mathcal{T}}_{en} \cup \bar{\mathcal{T}}_{ex} \cup \mathcal{T}_{to}$. Let $N_{ba} := |\bar{\mathcal{T}}_{en}| = |\bar{\mathcal{T}}_{ex}|$ and $N_{to} := |\bar{\mathcal{T}}_{to}|$. Moreover let us introduce the following notation. Given a real-valued function φ over $[0, T]$ and a finite subset $\mathcal{S}$ of $(0, T)$, assuming w.l.o.g. the elements of $\mathcal{S}$ in increasing order, we may define $\varphi(\mathcal{S}) := (\varphi(\tau))_{\tau \in \mathcal{S}} \in \mathbb{R}^{\mathrm{Card}\,\mathcal{S}}$. We adopt a similar convention for vectors and define $\nu_{\mathcal{S}} := (\nu_\tau)_{\tau \in \mathcal{S}} \in \mathbb{R}^{\mathrm{Card}\,\mathcal{S}}$.

The shooting algorithm is as follows. The unknown are the initial value of the costate p_0, the (finite) sets of entry, exit and touch points of the trajectory, respectively $\mathcal{T}_{en}$, $\mathcal{T}_{ex}$ and $\mathcal{T}_{to}$, and the jump parameters of the costate. More precisely, there are two jump parameters $\nu^1_{\tau_{en}}$ and $\nu^2_{\tau_{en}}$ for each entry point $\tau_{en} \in \mathcal{T}_{en}$ and one jump parameter $\nu_{\tau_{to}}$ for each touch point $\tau_{to} \in \mathcal{T}_{to}$. The shooting mapping $\mathcal{F}$ in a neighborhood of $(\bar{u}, \bar{y})$ is defined by

$$\begin{array}{rcl} \mathcal{F} : \mathbb{R}^n \times (\mathbb{R}^{N_{ba}})^4 \times (\mathbb{R}^{N_{to}})^2 & \to & \mathbb{R}^n \times (\mathbb{R}^{N_{ba}})^4 \times (\mathbb{R}^{N_{to}})^2, \\ \begin{pmatrix} p_0 \\ \nu^1_{\mathcal{T}_{en}} \\ \nu^2_{\mathcal{T}_{en}} \\ \mathcal{T}_{en} \\ \mathcal{T}_{ex} \\ \nu_{\mathcal{T}_{to}} \\ \mathcal{T}_{to} \end{pmatrix} & \mapsto & \begin{pmatrix} p_2(T) - \phi_y(y(T)) \\ g(y(\mathcal{T}_{en})) \\ g^{(1)}(y(\mathcal{T}_{en})) \\ g^{(2)}(u(\mathcal{T}^-_{en}), y(\mathcal{T}_{en})) \\ g^{(2)}(u(\mathcal{T}^+_{ex}), y(\mathcal{T}_{ex})) \\ g(y(\mathcal{T}_{to})) \\ g^{(1)}(y(\mathcal{T}_{to})) \end{pmatrix} \end{array}$$

where (u, y, p_2, η_2) are the solution of:

$$\begin{aligned} \dot{y} &= f(u, y) \quad \text{on } [0, T], \quad y(0) = y_0 && (6.71) \\ -\dot{p}_2 &= \tilde{H}_y(u, y, p_2, \eta_2) \quad \text{on } [0, T] \setminus \mathcal{T}, \quad p_2(0) = p_0, && (6.72) \\ 0 &= \tilde{H}_u(u, y, p_2, \eta_2) \quad \text{on } [0, T] \setminus \mathcal{T}, && (6.73) \\ 0 &= g^{(2)}(u, y) \quad \text{on boundary arcs} && (6.74) \\ 0 &= \eta_2 \quad \text{on interior arcs} && (6.75) \\ [p_2(\tau_{en})] &= -\nu^1_{\tau_{en}} g_y(y(\tau_{en})) - \nu^2_{\tau_{en}} g^{(1)}_y(y(\tau_{en})) \quad \text{at entry times } \tau_{en} \in \mathcal{T}_{en} && (6.76) \\ [p_2(\tau_{to})] &= -\nu_{\tau_{to}} g_y(y(\tau_{to})) \quad \text{at touch points } \tau_{to} \in \mathcal{T}_{to}. && (6.77) \end{aligned}$$

A vector of shooting parameters will be denoted by θ. With a stationary point of $(\mathcal{P})$ satisfying (A2)–(A4) is associated a unique set of shooting parameters, which is a zero of the shooting mapping. The vector of shooting parameters of $(\bar{u}, \bar{y})$ will be denoted by $\bar{\theta}$. More generally the 'bar' will refer in what follows to shooting parameters associated with the reference trajectory $(\bar{u}, \bar{y})$. Let us recall (see [19, Rem. 2.11(ii)][9]) that using the multipliers $(\bar{p}_2, \bar{\eta}_2)$ uniquely

[9] Remark 2.11 of this thesis.

associated with $(\bar u, \bar y)$ in the shooting algorithm, assumption (A3) is equivalent to

$$\begin{array}{l} \bar u \text{ is continuous over } [0,T] \text{ and} \\ \exists\, \alpha > 0, \quad \tilde H_{uu}(\bar u(t), \bar y(t), \bar p_2(t^\pm), \bar\eta_2(t^\pm)) \geq \alpha \quad \text{for all } t \in [0,T]. \end{array} \tag{6.78}$$

If (u,y) is a trajectory associated with a zero of the shooting mapping, with alternative shooting multipliers (p_2,η_2), then u, p_2 and η_2 are piecewise continuous on $[0,T]$ and have their set of discontinuity times included in the set of junction times $\mathcal{T} := \mathcal{T}_{en} \cup \mathcal{T}_{ex} \cup \mathcal{T}_{to}$. Let us recall the additional conditions that are automatically satisfied by a zero of the shooting mapping and the additional conditions, under which a zero of the shooting mapping is associated with a stationary point of the optimal control problem. Given $a, b \in \mathbb{R}$, set $[a,b] := \{(1-\lambda)a + \lambda b \ ;\ \lambda \in [0,1]\}$.

Lemma 6.17 ([19, Prop. 2.15 and Rem. 2.16]). *Let (u,y) be the trajectory associated with a zero of the shooting mapping, with alternative shooting multipliers (p_2,η_2). Assume that there exists $\beta, \alpha > 0$ such that*

$$\beta \ \leq\ |g_u^{(2)}(\hat u, y(t))| \quad \textit{for all } \hat u \in [u(t^-), u(t^+)] \textit{ and all } t \in I(g(y)); \tag{6.79}$$

$$\alpha \ \leq\ \tilde H_{uu}(\hat u, y(t), p_2(t^\pm), \eta_2(t^\pm)) \quad \textit{for all } \hat u \in [u(t^-), u(t^+)] \textit{ and all } t \in [0,T]. \tag{6.80}$$

Then: (i) *u is continuous over $[0,T]$.*
(ii) *For each boundary arc (τ_{en}, τ_{ex}) of (u,y), the following holds:*

$$\eta_2(\tau_{en}^+) = \nu^2_{\tau_{en}} \qquad \textit{and} \qquad \eta_2(\tau_{ex}^-) = 0. \tag{6.81}$$

Proposition 6.18 ([19, Corollary 2.17]). *A zero of the shooting mapping is associated with a stationary point (u,y) of $(\mathcal{P})$ satisfying (A2), (6.78), and (A4), with alternative shooting multipliers (p_2,η_2), iff:*

$$g(y(t)) \ \leq\ 0 \quad \textit{on interior arcs}, \tag{6.82}$$

$$0 \ \leq\ \ddot\eta_2(t) \quad \textit{on boundary arcs}, \tag{6.83}$$

$$0 \ \leq\ \nu^1_{\tau_{en}} + \dot\eta_2(\tau_{en}^+) \quad \textit{for each entry point } \tau_{en}, \tag{6.84}$$

$$\dot\eta_2(\tau_{ex}^-) \ \leq\ 0 \quad \textit{for each exit point } \tau_{ex} \tag{6.85}$$

$$0 \ \leq\ \nu_{\tau_{to}} \quad \textit{for each touch point } \tau_{to}. \tag{6.86}$$

Lemma 6.19. *Let (u,y) be the trajectory associated with a zero of the shooting mapping satisfying (A2), (6.78), and (A4). Then the additional conditions* (6.84) *and* (6.85) *are equivalent, respectively, to*

$$g^{(3)}(\dot u(\tau_{en}^-), u(\tau_{en}), y(\tau_{en})) \geq 0 \qquad \textit{and} \qquad g^{(3)}(\dot u(\tau_{ex}^+), u(\tau_{ex}), y(\tau_{ex})) \leq 0 \tag{6.87}$$

where the function $g^{(3)}$ is defined by (6.35).

Proof. By time differentiation of (6.73) on the interior of arcs, we have (omitting the arguments (u,y,p_2,η_2))

$$0 \ =\ \tilde H_{uu}\dot u + \tilde H_{uy} f - \tilde H_y f_u + \dot\eta_2 g_u^{(2)}. \tag{6.88}$$

Taking the jumps at entry time τ_{en}, we have by (6.76) and (6.81) (omitting arguments)

$$\begin{aligned} [\tilde H_{uu}] &= [p_2] f_{uu} + [\eta_2] g^{(2)}_{uu} \ =\ -\nu^1_{\tau_{en}} g_y f_{uu} - \nu^2_{\tau_{en}} g_y^{(1)} f_{uu} + \nu^2_{\tau_{en}} g^{(2)}_{uu} \\ &= -\nu^1_{\tau_{en}} g^{(1)}_{uu} - \nu^2_{\tau_{en}} g^{(2)}_{uu} + \nu^2_{\tau_{en}} g^{(2)}_{uu} \\ &= 0, \\ [\tilde H_{uy}] f - [\tilde H_y] f_u &= [p_2] f_{uy} f + [\eta_2] g^{(2)}_{uy} f - [p_2] f_y f_u - [\eta_2] g^{(2)}_y f_u \\ &= -\nu^1_{\tau_{en}} (g_y f_{uy} f - g_y f_y f_u) - \nu^2_{\tau_{en}} (g_y^{(1)} f_{uy} f - g^{(2)}_{uy} f - g_y^{(1)} f_y f_u + g_y^{(2)} f_u). \end{aligned}$$

Using that $g_{uy}^{(j)} = g_{yy}^{(j-1)} f_u + g_y^{(j-1)} f_{uy}$, $j = 1, 2$, that $g_u^{(2)} = g_y^{(1)} f_u = g_{yy} f f_u + g_y f_y f_u$ and that $g_{uy}^{(1)} \equiv 0$, we obtain

$$\begin{aligned}[\tilde{H}_{uy}]f - [\tilde{H}_y]f_u &= -\nu^1_{\tau_{en}}(g_{uy}^{(1)} f - g_{yy} f_u f - g_u^{(2)} + g_{yy} f f_u) - \nu^2_{\tau_{en}}(-g_{yy}^{(1)} f_u f + g_{yy}^{(1)} f f_u) \\ &= \nu^1_{\tau_{en}} g_u^{(2)}.\end{aligned}$$

Therefore, taking the jump of (6.88) at τ_{en}, we obtain

$$0 = \tilde{H}_{uu}[\dot{u}(\tau_{en})] + (\nu^1_{\tau_{en}} + [\dot{\eta}_2(\tau_{en})]) g_u^{(2)}.$$

By (6.35), we have that $[g^{(3)}(\dot{u}(\tau_{en}), u(\tau_{en}), y(\tau_{en}))] = g_u^{(2)}[\dot{u}(\tau_{en})]$ and hence, since $g^{(3)}$ vanishes on the interior of the boundary arc,

$$\nu^1_{\tau_{en}} + \dot{\eta}_2(\tau_{en}^+) = \frac{\tilde{H}_{uu}}{(g_u^{(2)})^2} g^{(3)}(\dot{u}(\tau_{en}^-), u(\tau_{en}), y(\tau_{en})). \tag{6.89}$$

Since $\tilde{H}_{uu}/(g_u^{(2)})^2$ is positive by (6.78) and (A2), the additional condition (6.84) is equivalent to the first condition of (6.87). Using similar arguments at exit points, the result follows. □

Remark 6.20. It follows from the above lemma that (6.82) together with the continuity of u imply that (6.84)–(6.85) are satisfied, since a Taylor expansion of the state constraint near entry/exit of boundary arcs yields

$$0 \geq g(y(t)) = g^{(3)}(\dot{u}(\tau^{\pm}), u(\tau), y(\tau)) \frac{(t-\tau)^3}{6} + o(|t-\tau|^3),$$

where $\tau^{\pm}$ stands for τ_{en}^- or τ_{ex}^+, implying (6.87), and in turn (6.84)–(6.85).

In what follows, $(\mathcal{P}^\mu)$ denotes a stable extension of $(\mathcal{P})$, and to indicate the dependence on μ of the data g, f, ℓ, ϕ and $\tilde{H}$, we will denote in what follows the shooting mapping by $\mathcal{F}(\cdot, \mu)$.

6.6.1 Well-posedness with nonreducible touch points

We assume in addition to (A2)–(A4) that

(A5) The strict complementarity assumption (6.21) holds on each (regular) boundary arc $(\bar{\tau}_{en}, \bar{\tau}_{ex})$ of $(\bar{u}, \bar{y})$;

(A6) **(i)** Each nonreducible touch point $\bar{\tau}_{to}$ of $(\bar{u}, \bar{y})$ satisfies (6.53);

(ii) Each nonreducible touch point $\bar{\tau}_{to}$ of $(\bar{u}, \bar{y})$ satisfies $\bar{\lambda}(\bar{\tau}_{to}) < 0$, where $\bar{\lambda}(\bar{\tau}_{to})$ is defined by (6.58).

Assumption (A6)(i) implies by Lemma 6.12 that all *nonreducible* touch points of $(\bar{u}, \bar{y})$ are *essential*. Therefore, by (A6)(i) all *nonessential* touch points of $(\bar{u}, \bar{y})$ are *reducible*, i.e. satisfy (6.49).

We exclude in (A6)(ii) the case when $\bar{\lambda}(\bar{\tau}_{to}) = 0$, since in that case, by Remark 6.14, we have no information on the structure of solutions of the perturbed problem, which is not very useful for the homotopy algorithm. We also exclude the case when $\bar{\lambda}(\bar{\tau}_{to}) > 0$, though we know by Th. 6.13 that in that case the solutions of the perturbed problem have either one or two touch points in the neighborhood of $\bar{\tau}_{to}$. The reason to leave aside this case in the following analysis is that singularities happen in the shooting algorithm when a touch point turns into two touch points (this is discussed more precisely in Remark 6.33 at the end of the paper).

Definition 6.21. Let $(\bar{u},\bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A6) and let $(\mathcal{P}^\mu)$ be a stable extension of $(\mathcal{P})$. We say that a stationary point (u,y) of $(\mathcal{P}^\mu)$ has a *neighboring structure* to that of $(\bar{u},\bar{y})$ if there exists a small $\delta>0$, $\delta<\min_{\tau,\tau'\in\bar{\mathcal{T}},\tau\neq\tau'}|\tau-\tau'|$, such that (a)–(e) below hold:

(a) The contact set $I(g^\mu(y))$ is included in $I_\delta(g(\bar{y}))=\{t\in[0,T]:\mathrm{dist}\{t,I(g(\bar{y}))\}<\delta\}$;

(b) For each *boundary arc* $(\bar{\tau}_{en},\bar{\tau}_{ex})$ of $(\bar{u},\bar{y})$, (u,y) has on $(\bar{\tau}_{en}-\delta,\bar{\tau}_{ex}+\delta)$ a unique boundary arc (τ_{en},τ_{ex});

(c) For each *essential* and *reducible* touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, (u,y) has on $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ a unique touch point τ_{to};

(d) For each *nonessential* touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, either the state constraint $g^\mu(y)$ is not active on $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ or (u,y) has on $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ a unique touch point τ_{to};

(e) For each *nonreducible* touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, (u,y) has on $(\bar{\tau}_{to}-\delta,\bar{\tau}_{to}+\delta)$ either a unique touch point τ_{to} or a unique boundary arc (τ_{en},τ_{ex}).

We denote by $\bar{\mathcal{T}}^{ess}_{red}$, $\bar{\mathcal{T}}^{nes}$, and $\bar{\mathcal{T}}_{nrd}$ the sets of respectively essential and reducible, nonessential, and nonreducible touch points of the trajectory $(\bar{u},\bar{y})$. Set $N_{nes}:=|\bar{\mathcal{T}}^{nes}|$ and $N_{nrd}:=|\bar{\mathcal{T}}_{nrd}|$. By the above definition, there are $N_s:=2^{N_{nes}+N_{nrd}}$ different neighboring structures to that of $(\bar{u},\bar{y})$. For $j=1,\dots,N_s$, denote by $\mathcal{F}_j$ the shooting mappings corresponding to each of those different neighboring structures. For each nonessential touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, the latter is introduced or not in the shooting mapping $\mathcal{F}_j$ (with a zero jump parameter $\bar{\nu}_{\tau_{to}}$), and for each nonreducible touch point $\bar{\tau}_{to}$ of $(\bar{u},\bar{y})$, the latter is introduced as a touch point or as a boundary arc (of zero length) in the shooting mapping $\mathcal{F}_j$. More precisely, similarly to first-order state constraints (see [20, section 4.2][10]) since $g^{(2)}(\bar{u}(\bar{\tau}^\pm_{to}),\bar{y}(\bar{\tau}_{to}))=0$ a nonreducible touch point $\bar{\tau}_{to}$ can be seen as a boundary arc of zero length, by taking

$$\bar{\tau}_{en}:=\bar{\tau}_{to}=:\bar{\tau}_{ex} \tag{6.90}$$

and, in view of the jump conditions (6.76)–(6.77),

$$\bar{\nu}^1_{\tau_{en}}:=\bar{\nu}_{\tau_{to}}\qquad\text{and}\qquad\bar{\nu}^2_{\tau_{en}}:=0. \tag{6.91}$$

For $j=1,\dots,N_s$, denote by $\bar{\theta}_j$ the vector of shooting parameters, of appropriate dimension, associated with $(\bar{u},\bar{y})$ in the shooting mapping $\mathcal{F}_j$.

For $v\in\mathcal{V}$ in the extended critical cone (i.e. satisfying (6.63)), we consider the additional constraint below:

$$g^{(1)}_y(y(\bar{\tau}_{to}))z_v(\bar{\tau}_{to})=0\qquad\text{for all }\bar{\tau}_{to}\in\bar{\mathcal{T}}_{nrd}. \tag{6.92}$$

Recall that z_v is the solution of (6.30). A sufficient condition ensuring the well-posedness of the shooting algorithm, as we will see, is

$$\mathcal{Q}(v)-\sum_{\tau\in\bar{\mathcal{T}}^{ess}_{red}}\bar{\nu}_\tau\frac{(g^{(1)}_y(\bar{y}(\tau))z_v(\tau))^2}{g^{(2)}(\bar{u}(\tau),\bar{y}(\tau))}>0,\quad\text{for all }v\in\mathcal{V},\ v\neq0,\ \text{satisfying (6.63) and (6.92)}, \tag{6.93}$$

where $\mathcal{Q}$ is given by (6.62). Note that the sum in (6.93) is nonpositive. Therefore, the strong second-order sufficient condition (6.65) used in the stability analysis implies that the weaker condition (6.93) is satisfied.

[10]Section 3.5.2 of this thesis.

Lemma 6.22. *Let $(\bar{u},\bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A6) and (6.93). Then there exists a neighborhood W of $\bar{\mu}$ and, for each $j=1,\dots,N_s$, a neighborhood V_j of $\bar{\theta}_j$ such that for each $\mu\in W$, the equation*

$$\mathcal{F}_j(\theta,\mu)=0 \tag{6.94}$$

has a unique solution θ_j^μ in V_j, which is C^1 w.r.t. μ.

Of course, if nonreducible touch points are converted into boundary arcs in the shooting mapping $\mathcal{F}_j$, it may happen that for μ in the neighborhood of $\bar{\mu}$, the solution θ_j^μ of (6.94) is such that some entry times are greater than the corresponding exit times. In that case the trajectory associated with θ_j^μ by (6.71)–(6.77) has no physical meaning since not single-valued. In Lemma 6.25 we will give necessary and sufficient conditions so that a solution θ_j^μ of (6.94) is associated with a stationary point of $(\mathcal{P}^\mu)$.

Proof of Lemma 6.22. We follows the ideas of the proof of [19, Th. 3.3][11] and include the presence of nonreducible touch points. Let us show that the Jacobian $D_\theta\mathcal{F}_j(\bar{\theta}_j,\bar{\mu})$ is invertible, for all $j=1,\dots,N_s$. It will then follow from the implicit function theorem that (6.94) has a locally unique solution for μ in a neighborhood of $\bar{\mu}$ which is C^1 w.r.t. μ.

Let $\mathcal{F}_j$ be one of these shooting mappings. Let $\omega:=\big(\pi_0,\gamma^1_{\mathcal{T}_{en}},\gamma^2_{\mathcal{T}_{ex}},\sigma_{\mathcal{T}_{en}},\sigma_{\mathcal{T}_{ex}},\gamma_{\mathcal{T}_{to}},\sigma_{\mathcal{T}_{to}}\big)^\top$ be such that $D_\theta\mathcal{F}_j(\bar{\theta}_j,\bar{\mu})\omega=0$. Then, by differentiation of the shooting mapping, we have

$$\begin{aligned}
0 &= \pi_2(T)-\phi_{yy}(y(T))z(T), && (6.95)\\
0 &= g_y(\bar{y}(\bar{\tau}_{en}))z(\bar{\tau}_{en}) \quad \text{for all entry points } \bar{\tau}_{en}, && (6.96)\\
0 &= g_y^{(1)}(\bar{y}(\bar{\tau}_{en}))z(\bar{\tau}_{en}) \quad \text{for all entry points } \bar{\tau}_{en}, && (6.97)\\
0 &= Dg^{(2)}(\bar{u}(\bar{\tau}_{en}),\bar{y}(\bar{\tau}_{en}))(v(\bar{\tau}_{en}^-),z(\bar{\tau}_{en}))+\sigma_{\bar{\tau}_{en}}\frac{\mathrm{d}}{\mathrm{d}t}g^{(2)}(\bar{u}(t),\bar{y}(t))|_{t=\bar{\tau}_{en}^-}\\
&\qquad \text{for all entry points } \bar{\tau}_{en}, && (6.98)\\
0 &= Dg^{(2)}(\bar{u}(\bar{\tau}_{ex}),\bar{y}(\bar{\tau}_{ex}))(v(\bar{\tau}_{ex}^+),z(\bar{\tau}_{ex}))+\sigma_{\bar{\tau}_{ex}}\frac{\mathrm{d}}{\mathrm{d}t}g^{(2)}(\bar{u}(t),\bar{y}(t))|_{t=\bar{\tau}_{ex}^+}\\
&\qquad \text{for all exit points } \bar{\tau}_{ex}, && (6.99)\\
0 &= g_y(\bar{y}(\bar{\tau}_{to}))z(\bar{\tau}_{to}) \quad \text{for all touch points } \bar{\tau}_{to}, && (6.100)\\
0 &= g_y^{(1)}(\bar{y}(\bar{\tau}_{to}))z(\bar{\tau}_{to})+\sigma_{\bar{\tau}_{to}}g^{(2)}(\bar{u}(\bar{\tau}_{to}),\bar{y}(\bar{\tau}_{to})) \quad \text{for all touch points } \bar{\tau}_{to}, && (6.101)
\end{aligned}$$

where (v,z,π_2,ζ_2) are the solutions of the variational system below (the arguments $(\bar{u},\bar{y},\bar{p}_2,\bar{\eta}_2)$ are omitted)

$$\begin{aligned}
\dot{z} &= f_u v+f_y z \quad \text{on } [0,T], \quad z(0)=0, && (6.102)\\
-\dot{\pi}_2 &= \tilde{H}_{yu}v+\tilde{H}_{yy}z+\pi_2 f_y+\zeta_2 g_y^{(2)} \quad \text{on } [0,T]\setminus\mathcal{T}, \quad \pi_2(0)=\pi_0, && (6.103)\\
0 &= \tilde{H}_{uu}v+\tilde{H}_{uy}z+\pi_2 f_u+\zeta_2 g_u^{(2)} \quad \text{on } [0,T]\setminus\mathcal{T}, && (6.104)\\
0 &= g_u^{(2)}v+g_y^{(2)}z \quad \text{on boundary arcs}, && (6.105)\\
0 &= \zeta_2 \quad \text{on interior arcs}, && (6.106)\\
[\pi_2(\bar{\tau}_{en})] &= -\bar{\nu}^1_{\bar{\tau}_{en}}g_{yy}(\bar{y}(\bar{\tau}_{en}))z(\bar{\tau}_{en})-\bar{\nu}^2_{\bar{\tau}_{en}}g^{(1)}_{yy}(\bar{y}(\bar{\tau}_{en}))z(\bar{\tau}_{en})-\gamma^1_{\bar{\tau}_{en}}g_y(\bar{y}(\bar{\tau}_{en}))\\
&\quad -(\gamma^2_{\bar{\tau}_{en}}+\sigma_{\bar{\tau}_{en}}\bar{\nu}^1_{\bar{\tau}_{en}})g_y^{(1)}(\bar{y}(\bar{\tau}_{en})) \quad \text{for all entry points } \bar{\tau}_{en}, && (6.107)\\
[\pi_2(\bar{\tau}_{to})] &= -\bar{\nu}_{\bar{\tau}_{to}}g_{yy}(\bar{y}(\bar{\tau}_{to}))z(\bar{\tau}_{to})-\gamma_{\bar{\tau}_{to}}g_y(\bar{y}(\bar{\tau}_{to}))-\sigma_{\bar{\tau}_{to}}\bar{\nu}_{\bar{\tau}_{to}}g_y^{(1)}(\bar{y}(\bar{\tau}_{to}))\\
&\quad \text{for all touch points } \bar{\tau}_{to}. && (6.108)
\end{aligned}$$

[11] Theorem 2.23 of this thesis.

The jump condition of the costate (6.107) follows from [19, Lemma 3.7][12]. Recall that for nonreducible touch points $\bar\tau_{to} = \bar\tau_{en}$ converted into a boundary arc in $\mathcal{F}_j$, we have $\bar\nu^2_{\bar\tau_{en}} = 0$ in (6.107) by (6.91). For a nonreducible touch point $\bar\tau_{to}$ introduced as a touch point in $\mathcal{F}_j$, (6.101) becomes

$$0 = g^{(1)}_y(\bar y(\bar\tau_{to}))z(\bar\tau_{to}). \tag{6.109}$$

The above constraint holds as well for nonreducible touch points converted into boundary arcs by (6.97). Moreover, we substitute $\sigma_{\bar\tau_{to}}$ using (6.101) into the jump condition (6.108) for reducible touch points, and we consider for nonreducible touch points introduced as touch points the constraint (6.109) with associated multiplier $\sigma_{\bar\tau_{to}}\bar\nu_{\bar\tau_{to}}$ in (6.108). In this way we obtain that (6.95)–(6.97) and (6.100)–(6.108) constitute the first-order optimality condition of the linear-quadratic problem (PQ) of minimizing

$$\begin{aligned}\mathcal{Q}^2(v) \;:=\; & \int_0^T D^2_{(u,y)(u,y)}\tilde H(\bar u,\bar y,\bar p_2,\bar\eta_2)((v,z_v),(v,z_v))\mathrm{d}t + \phi_{yy}(\bar y(T))(z_v(T),z_v(T)) \\ & + \sum_{\bar\tau_{en}\in\bar{\mathcal{T}}_{en}} \Big(\nu^1_{\bar\tau_{en}} g_{yy}(\bar y(\bar\tau_{en}))(z_v(\bar\tau_{en}),z_v(\bar\tau_{en})) + \nu^2_{\bar\tau_{en}} g^{(1)}_{yy}(\bar y(\bar\tau_{en}))(z_v(\bar\tau_{en}),z_v(\bar\tau_{en}))\Big) \\ & + \sum_{\bar\tau_{to}\in\bar{\mathcal{T}}_{to}} \bar\nu_{\bar\tau_{to}} g_{yy}(\bar y(\bar\tau_{to}))(z_v(\bar\tau_{to}),z_v(\bar\tau_{to})) \;-\; \sum_{\bar\tau_{to}\in\bar{\mathcal{T}}^{ess}_{red}} \bar\nu_{\bar\tau_{to}} \frac{(g^{(1)}_y(\bar y(\bar\tau_{to}))z_v(\bar\tau_{to}))^2}{g^{(2)}(\bar u(\bar\tau_{to}),\bar y(\bar\tau_{to}))},\end{aligned}$$

subject to the constraints (6.96), (6.97), (6.100), (6.105), and (6.109) at nonreducible touch points. Since $\frac{\mathrm{d}}{\mathrm{d}t} g_y(\bar y(t))z_v(t) = g^{(1)}_y(\bar y)z_v$ and $\frac{\mathrm{d}^2}{\mathrm{d}t^2} g_y(\bar y(t))z_v(t) = g^{(2)}_y(\bar u,\bar y)z_v + g^{(2)}_u(\bar u,\bar y)v$, the constraints (6.96), (6.97), and (6.105) are equivalent to $g_y(\bar y(t))z(t) = 0$ on boundary arcs (of positive length) $[\bar\tau_{en},\bar\tau_{ex}]$. Consequently, the constraints (6.96), (6.97), (6.100), (6.105), and (6.109) of (PQ) are equivalent to (6.63), (6.92), and $g_y(\bar y(\bar\tau_{to}))z(\bar\tau_{to}) = 0$ for all nonessential touch point $\bar\tau_{to}$ introduced in the shooting mapping $\mathcal{F}_j$.

By straightforward calculation (see [19, Lemma 3.6] and [71, Lemma 3.1][13]), we can show that the quadratic form $\mathcal{Q}^2(v)$ is equal to the left-hand side of (6.93). Since the latter is a Legendre form by assumption (6.20), (6.93) implies that (PQ) has a weakly lower semicontinuous and strongly convex cost function on its closed and convex feasible set. Moreover, the constraints of (PQ) are onto by assumption (A2) (see Lemma 6.5) and hence the unique solution and associated multipliers of the first-order optimality condition of (PQ) are zero. This implies that $(v,z,\pi_2,\zeta_2) \equiv 0$. Therefore, $\pi_0 = 0$ and the multipliers associated with the constraints (6.96)–(6.97), (6.100), and (6.109) for nonreducible touch points introduced as touch points are equal to zero, implying that

$$\gamma^1_{\bar\tau_{en}} = 0, \qquad \gamma^2_{\bar\tau_{en}} + \sigma_{\bar\tau_{en}}\bar\nu^1_{\bar\tau_{en}} = 0, \qquad \gamma_{\bar\tau_{to}} = 0, \tag{6.110}$$

and, for nonreducible touch points $\bar\tau_{to}$ introduced as touch points,

$$\sigma_{\bar\tau_{to}}\bar\nu_{\bar\tau_{to}} = 0. \tag{6.111}$$

By (6.98)–(6.99), since $\frac{\mathrm{d}}{\mathrm{d}t} g^{(2)}(\bar u(t),\bar y(t))|_{t=\bar\tau^-_{en},\bar\tau^+_{ex}} \neq 0$ both for entry/exit points of boundary arcs by Lemma 6.3, and for nonreducible touch points converted into boundary arcs by hypothesis (A6)(i), we have that $\sigma_{\bar\tau_{en}} = 0 = \sigma_{\bar\tau_{ex}}$, and by (6.101), $\sigma_{\bar\tau_{to}} = 0$ for reducible touch points $\bar\tau_{to}$. Finally, with (6.110)–(6.111), since $\bar\nu_{\bar\tau_{to}} \neq 0$ at nonreducible touch points $\bar\tau_{to}$ by (A6)(i) and Lemma 6.12, it follows that $\omega = 0$, i.e. the Jacobian of the shooting mapping $\mathcal{F}_j$ is one-to-one, and hence invertible. □

[12] Lemma 2.27 of this thesis.

[13] Lemmas 2.26 and 5.9 of this thesis.

6.6.2 Stability of shooting parameters

Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A6) and the strong second-order sufficient condition (6.65) and let $(\mathcal{P}^\mu)$ be a stable extension of $(\mathcal{P})$. For μ in a neighborhood of $\bar{\mu}$, the perturbed problem $(\mathcal{P}^\mu)$ has by Th. 6.16 a locally unique stationary point (u^μ, y^μ), which has by Theorems 6.8 and 6.13 a neighboring structure to that of $(\bar{u}, \bar{y})$, in the sense of Def. 6.21. Therefore it makes sense to speak about the shooting parameters associated with (u^μ, y^μ). Note that its set of shooting parameters may not necessarily be unique if (u^μ, y^μ) has nonessential or nonreducible touch points, since a nonessential touch point may or not be introduced in the set of shooting parameters, with an associated zero jump parameter, and a nonreducible touch point may be considered either as a boundary arc (of zero length) or as a touch point. The next lemma shows that the stationary point (u^μ, y^μ) of $(\mathcal{P}^\mu)$ has its shooting parameters in the neighborhood of the shooting parameters of $(\bar{u}, \bar{y})$.

For this it will be useful to make explicit the relation between the multipliers η_2 and η^2 used respectively in the shooting algorithm and in the stability analysis. Recall that the multipliers used in the shooting algorithm are defined by (6.69)–(6.70) while those used in the stability analysis are defined by (6.11)–(6.12). Moreover, by [19, Prop. 2.10][14], for all boundary arcs (τ_{en}, τ_{ex}) (including the case $\tau_{en} = \tau_{ex}$), we have that

$$\nu^1_{\tau_{en}} = \int_{[\tau_{en}, \tau_{ex}]} \mathrm{d}\eta = [\eta(\tau_{en})] + \eta_1(\tau^+_{en}), \tag{6.112}$$

and the condition (6.81) holds *a fortiori* for a stationary point. Combining the above relations, we obtain that

$$\eta^1(t) = \eta_1(t) + \sum_{\tau_{en} \in \mathcal{T}_{en}} \nu^1_{\tau_{en}} \mathbf{1}_{[0,\tau_{en}]}(t) + \sum_{\tau_{to} \in \mathcal{T}_{to}} \nu_{\tau_{to}} \mathbf{1}_{[0,\tau_{to}]}(t), \tag{6.113}$$

$$\begin{aligned} \eta^2(t) &= \int_t^T \eta^1(s)\mathrm{d}s \\ &= \eta_2(t) + \sum_{\tau_{en} \in \mathcal{T}_{en}} \mathbf{1}_{[0,\tau_{en}]}(t)(\nu^2_{\tau_{en}} + \nu^1_{\tau_{en}}(\tau_{en} - t)) + \sum_{\tau_{to} \in \mathcal{T}_{to}} \nu_{\tau_{to}} \mathbf{1}_{[0,\tau_{to}]}(t)(\tau_{to} - t). \end{aligned} \tag{6.114}$$

Here $\mathbf{1}_{[a,b]}(\cdot)$ denotes the indicator function of the interval $[a, b] \subset [0, T]$ equal to 1 on $[a, b]$ and zero outside. Then p_2 and p^2 defined respectively by (6.12) and (6.70) are related by

$$p^2 = p_2 - (\eta^1 - \eta_1) g_y(y) - (\eta^2 - \eta_2) g^{(1)}_y(y). \tag{6.115}$$

Lemma 6.23. *Let $(\bar{u}, \bar{y})$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A6) and the strong second-order sufficient condition (6.65) and let $(\mathcal{P}^\mu)$ be a stable extension of $(\mathcal{P})$. Then for each $\varepsilon > 0$, there exist neighborhoods W of $\bar{\mu}$ and V_∞ of $\bar{u}$ (in L^∞) such that for each $\mu \in W$, the locally unique stationary point (u^μ, y^μ) of $(\mathcal{P}^\mu)$ with $u^\mu \in V_\infty$ has a neighboring structure to that of $(\bar{u}, \bar{y})$. Moreover, any vector of shooting parameters θ^μ associated with (u^μ, y^μ), of appropriate dimension, satisfy*

$$|\theta^\mu - \bar{\theta}_j| < \varepsilon$$

where $\bar{\theta}_j$ is the vector of shooting parameters associated with $(\bar{u}, \bar{y})$ matching the structure of θ^μ.

[14] Proposition 2.10 of this thesis.

It follows from the above lemma and Lemma 6.22 that for a given neighboring structure $\mathcal{F}_j$ of $(\bar{u},\bar{y})$, the vector of shooting parameters θ^μ associated with the stationary point (u^μ,y^μ) is locally unique.

Proof. It follows from Theorems 6.8 and 6.13 that the locally unique (by Th. 6.16) stationary point (u^μ,y^μ) of $(\mathcal{P}^\mu)$ has a neighboring structure to that of $(\bar{u},\bar{y})$. The convergence of junction times was proved in Theorems 6.8 and 6.13. So let us show the convergence of jump parameters. For this we use the formula (6.114) that links the multiplier η^2 used in the stability analysis to the shootings parameters and the uniform convergence of $\eta^{2,\mu}$ towards $\bar{\eta}^2$ by Prop. 6.6(iii). The proof is by finite induction.

Let N denote the total number of boundary arcs and touch points of the trajectory $(\bar{u},\bar{y})$. We may write that $\{1,\ldots,N\}=\mathcal{N}_{ba}\cup\mathcal{N}_{to}$, where $\mathcal{N}_{ba}\cap\mathcal{N}_{to}=\emptyset$ and $\mathcal{N}_{ba}$ and $\mathcal{N}_{to}$ denote the sets of index corresponding respectively to boundary arcs (possibly of zero length) and to touch points (possibly nonreducible or nonessential). This partition is not unique since a nonreducible touch point can be considered either as a boundary arc of zero length or as a touch point. We have then that $I(g(\bar{y}))=\cup_{i=1}^N\bar{I}_i$, where $\bar{I}_i:=[\bar{\tau}_{en,i},\bar{\tau}_{ex,i}]$ for $i\in\mathcal{N}_{ba}$ (with possibly $\bar{\tau}_{en,i}=\bar{\tau}_{ex,i}$), $\bar{I}_i:=\{\bar{\tau}_{to,i}\}$ for $i\in\mathcal{N}_{to}$, $\bar{I}_i\cap\bar{I}_j=\emptyset$ for $i\neq j$, and $\bar{I}_i<\bar{I}_{i+1}$ for all $i<N$ (in the sense that $t<t'$ for all $(t,t')\in\bar{I}_i\times\bar{I}_{i+1}$). The jump parameters associated with a boundary arc $[\bar{\tau}_{en,i},\bar{\tau}_{ex,i}]$ are denoted by $\bar{\nu}_i^1$ and $\bar{\nu}_i^2$ and that associated with a touch point $\bar{\tau}_{to,i}$ by $\bar{\nu}_i$. Since (u^μ,y^μ) has a neighboring structure to that of $(\bar{u},\bar{y})$, we can choose the partition $(\mathcal{N}_{ba},\mathcal{N}_{to})$ such that $I(g^\mu(y^\mu))=\cup_{i=1}^N I_i^\mu$ for a sequence $\mu_n\to_{n\to\infty}\bar{\mu}$, where $I_i^\mu=[\tau_{en,i}^\mu,\tau_{ex,i}^\mu]$ for $i\in\mathcal{N}_{ba}$, with associated jump parameters $\nu_i^{1,\mu}$ and $\nu_i^{2,\mu}$, and $I_i^\mu=\{\tau_{to,i}^\mu\}$ with jump parameter ν_i^μ or possibly $I_i^\mu=\emptyset$ (if $\bar{\tau}_{to,i}$ is a nonessential touch point) for $i\in\mathcal{N}_{to}$.

Given $k\in\{1,\ldots,N\}$, assume by induction that the jump parameters associated with $I_i^{\mu_n}$ converge to those associated with $\bar{I}_i$ for all $i\in\{k+1,\ldots,N\}$. (For $k=N$ we assume nothing.) Let us show that the jump parameters associated with $I_k^{\mu_n}$ converges to those associated with $\bar{I}_k$. There are two cases to consider.

Case 1: $k\in\mathcal{N}_{to}$. If $I_k^{\mu_n}=\emptyset$, there is nothing to prove, so assume that $I_k^{\mu_n}=\{\tau_{to,k}^{\mu_n}\}$. Recall that by definition, η_1 and η_2 vanish on interior arcs. Then for a fixed $\varepsilon>0$ small enough ($\varepsilon<\min_{\tau,\tau'\in\bar{\mathcal{T}}\cup\{0\},\tau\neq\tau'}\frac{1}{2}|\tau-\tau'|$), for all $t\in[\bar{\tau}_{to,k}-2\varepsilon,\bar{\tau}_{to,k}-\varepsilon]$, we have by (6.114) and Th. 6.16 for n large enough that

$$\begin{aligned}\eta^{2,\mu_n}(t) \quad &= \sum_{i\in\mathcal{N}_{ba},i>k}(\nu_i^{2,\mu_n}+\nu_i^{1,\mu_n}(\tau_{en,i}^{\mu_n}-t))+\sum_{i\in\mathcal{N}_{to},i>k}\nu_i^{\mu_n}(\tau_{to,i}^{\mu_n}-t)\ +\ \nu_k^{\mu_n}(\tau_{to,k}^{\mu_n}-t)\\ \xrightarrow[n\to\infty]{}\ \bar{\eta}^2(t) &= \sum_{i\in\mathcal{N}_{ba},i>k}(\bar{\nu}_i^2+\bar{\nu}_i^1(\bar{\tau}_{en,i}-t))+\sum_{i\in\mathcal{N}_{to},i>k}\bar{\nu}_i(\bar{\tau}_{to,i}-t)\ +\ \bar{\nu}_k(\bar{\tau}_{to,k}-t).\end{aligned}$$

Since the junction times of (u^{μ_n},y^{μ_n}) converge to those of $(\bar{u},\bar{y})$, as well as the jump parameters associated with $I_i^{\mu_n}$ for $i>k$ by the induction hypothesis, we deduce immediately that $\nu_k^{\mu_n}$ converges to $\bar{\nu}_k$.

Case 2: $k\in\mathcal{N}_{ba}$. Then $I_k^{\mu_n}=[\tau_{en,k}^{\mu_n},\tau_{ex,k}^{\mu_n}]$ and reasoning similarly, for a fixed $\varepsilon>0$ small

enough, for all $t \in [\bar\tau_{en,k} - 2\varepsilon, \min\{\bar\tau_{en,k}, \tau^{\mu_n}_{en,k}\})$ and n large enough, we have that

$$\begin{aligned}\eta^{2,\mu_n}(t) \;&=\; \sum_{i\in\mathcal{N}_{ba}, i>k} (\nu_i^{2,\mu_n} + \nu_i^{1,\mu_n}(\tau^{\mu_n}_{en,i} - t)) + \sum_{i\in\mathcal{N}_{to}, i>k} \nu_i^{\mu_n}(\tau^{\mu_n}_{to,i} - t) \\ &\qquad + \nu_k^{2,\mu_n} + \nu_k^{1,\mu_n}(\tau^{\mu_n}_{en,k} - t) \\ &\xrightarrow[n\to\infty]{} \bar\eta^2(t) = \sum_{i\in\mathcal{N}_{ba}, i>k} (\bar\nu_i^{2} + \bar\nu_i^{1}(\bar\tau_{en,i} - t)) + \sum_{i\in\mathcal{N}_{to}, i>k} \bar\nu_i(\bar\tau_{to,i} - t) \\ &\qquad + \bar\nu_k^{2} + \bar\nu_k^{1}(\bar\tau_{en,k} - t).\end{aligned}$$

By letting $t \uparrow \bar\tau_{en,k}$, $t < \min\{\bar\tau_{en,k}, \tau^{\mu_n}_{en,k}\}$, and $n \to +\infty$, using that the convergence of η^{2,μ_n} towards $\bar\eta^2$ is uniform and that ν_k^{1,μ_n} is bounded (since $\nu_k^{1,\mu_n} = \int_{[\tau^{\mu_n}_{en,k}, \tau^{\mu_n}_{ex,k}]} \mathrm{d}\eta^{\mu_n}$ by (6.112) and $\mathrm{d}\eta^{\mu_n}$ is uniformly bounded by Prop. 6.6(i)), we deduce as previously that $\nu_k^{2,\mu_n} \to \bar\nu_k^2$. Taking then $t \in [\bar\tau_{en,k} - 2\varepsilon, \bar\tau_{en,k} - \varepsilon]$, it follows that $\nu_k^{1,\mu_n} \to \bar\nu_k^1$. This completes the induction step and achieves to show the converge of jump parameters of (u^μ, y^μ) towards those of $(\bar u, \bar y)$.

It remains to show the convergence of the initial costate. For a small $\varepsilon > 0$ and $\|\mu - \bar\mu\|$ small enough, the state constraint $g^\mu(y^\mu)$ is not active on $[0, \varepsilon]$. Therefore, by (6.113), $\eta^{1,\mu}$ converges uniformly to $\bar\eta^1$ on $[0, \varepsilon]$ and $\eta_1^\mu = \eta_2^\mu = 0$ since we are on an interior arc. It follows that for all $t \in [0, \varepsilon]$, using (6.115),

$$\begin{aligned} p_2^\mu(t) \;&=\; p^2(t) + \eta^{1,\mu}(t) g_y^\mu(y^\mu(t)) + \eta^{2,\mu}(t)(g^\mu)_y^{(1)}(y^\mu(t)) \\ &\xrightarrow[\mu\to\bar\mu]{} \bar p^2(t) + \bar\eta^1(t) g_y(\bar y(t)) + \bar\eta^2(t) g_y^{(1)}(\bar y(t)) \;=\; \bar p_2(t)\end{aligned}$$

since $p^{2,\mu}$, $\eta^{2,\mu}$ and y^μ converges uniformly to $\bar p^2$, $\bar\eta^2$ and $\bar y$, respectively. For $t = 0$ this gives the convergence of the initial costate $p_2^\mu(0) \to \bar p_2(0)$. This achieves the proof of the lemma. □

6.6.3 Additional conditions for a stationary point

By Lemma 6.23, we know that the locally unique stationary point (u^μ, y^μ) of the perturbed problem $(\mathcal{P}^\mu)$ has its shooting parameters in the neighborhood of those of the reference trajectory $(\bar u, \bar y)$. By Lemma 6.22, the shooting algorithm is then well-posed to find a vector of shooting parameters associated with (u^μ, y^μ). Lemma 6.23 ensures that *at least one* of the solutions θ_j^μ obtained in Lemma 6.22 for the neighboring structures to that of $(\bar u, \bar y)$ is associated to this (locally unique) stationary point of $(\mathcal{P}^\mu)$. Of course we do not know a priori what the structure of (u^μ, y^μ) is. We only know that it is a neighboring structure to that of $(\bar u, \bar y)$. In Lemma 6.25 below we give necessary and sufficient conditions in order to recognize a vector of shooting parameters associated with a stationary point of the perturbed problem, among all the solutions of (6.94). Let us first note the following.

Remark 6.24. The statement of Lemma 6.17 extends without difficulty to the case when there are nonreducible touch points converted into boundary arcs of zero length (with $\tau_{en} = \tau_{ex}$). In that case $\eta_2(\tau_{en}^+) = 0$ and (6.81) yields that $\nu^2_{\tau_{en}} = 0$ automatically holds at nonreducible touch points converted into boundary arcs. The statement of Prop. 6.18 extend as well. For nonreducible touch points τ_{to} converted into boundary arcs of zero length, since $\dot\eta_2(\tau_{en}^+) = 0$, (6.84) amounts to the classical condition $\nu_{\tau_{to}} = \nu^1_{\tau_{en}} \geq 0$, while (6.85) is automatically satisfied (with equality).

Lemma 6.25. *Let $(\bar u, \bar y)$ be a stationary point of $(\mathcal{P})$ satisfying (A2)–(A6) and* (6.93)*. For $j \in \{1, \ldots, N_s\}$, let $\mathcal{F}_j$ denote one of the shooting mappings associated with a neighboring*

structure to $(\bar{u}, \bar{y})$. *Then there exist a neighborhood* W *of* $\bar{\mu}$ *and a neighborhood* V_j *of* $\bar{\theta}_j$, *such that a solution* θ *in* V_j *of* (6.94) *for* $\mu \in W$ *is associated with a stationary point of* $(\mathcal{P}^\mu)$ *iff, denoting by* (u, y, p_2, η_2) *the trajectory and multipliers associated with* θ, *the following conditions are satisfied:*

$$\begin{aligned} 0 &\geq g^\mu(y(t)) \text{ on } [0,T], && (6.116)\\ 0 &\geq (g^\mu)^{(2)}(u(\tau_{to}), y(\tau_{to})) \quad \text{for each touch point } \tau_{to} \text{ of } \theta, && (6.117)\\ 0 &\leq \nu_{\tau_{to}} \quad \text{for each touch point } \tau_{to} \text{ of } \theta, && (6.118)\\ \tau_{en} &\leq \tau_{ex} \quad \text{for each boundary arc of } \theta. && (6.119) \end{aligned}$$

Proof. By Prop. 6.18, it is obvious that the conditions (6.116), (6.118), and (6.119) are necessary for a stationary point. The condition (6.117) is necessary as well, since in the neighborhood of a touch point τ, we have that

$$g^\mu(y(t)) = (g^\mu)^{(2)}(u(\tau), y(\tau))\frac{(t-\tau)^2}{2} + o(|t-\tau|^2) \leq 0.$$

Now we show that the conditions (6.116)–(6.119) are sufficient to have a stationary point of $(\mathcal{P}^\mu)$. In order to show that the trajectory and multipliers (u, y, p_2, η_2) associated with θ are a stationary point of $(\mathcal{P}^\mu)$ and its associated multipliers in the shooting algorithm, we have to show by Prop. 6.18 and Remark 6.24 that the additional conditions (6.82)–(6.86) are satisfied.

The conditions (6.82) and (6.86) at touch points follow immediately from (6.116) and (6.118). Let us show now (6.84)–(6.85). By (A2) and (A3), implying (6.78), for μ in the vicinity of $\bar{\mu}$, (u, y, p_2, η_2) satisfies by continuity (6.79)–(6.80) and hence, it follows from Lemma 6.17 and Remark 6.24 that u is continuous over $[0, T]$. Therefore the conditions (6.84)–(6.85) at entry and exit points of boundary arcs of nonzero length (τ_{en}, τ_{ex}) are satisfied by Rem. 6.20 as a consequence of (6.116). For possible boundary arcs of zero length $\tau_{en} = \tau_{ex}$, (6.84)–(6.85) amounts to check that $\nu^1_{\tau_{en}} \geq 0$. By the same arguments than in the proof of Lemma 6.19, this last condition is equivalent to $[g^{(3)}(\dot{u}(\tau_{en}), u(\tau_{en}), y(\tau_{en}))] < 0$, which holds by continuity for $\|\mu - \bar{\mu}\|$ and $|\theta - \bar{\theta}_j|$ small enough by (A6)(i).

Let us end the proof by showing that (6.83) is satisfied on boundary arcs (τ_{en}, τ_{ex}) with $\tau_{en} < \tau_{ex}$. *Define* the multipliers η^2 and p^2 by respectively (6.114) and (6.115). By (6.81), we have that η^2 is continuous over $[0, T]$. By (6.113)–(6.115) and (6.76)–(6.77), we see directly that p^2 is continuous over $[0, T]$ as well. Moreover, (6.72)–(6.73) imply by straightforward calculations that the following hold over $[0, T]$

$$\begin{aligned} -\dot{p}^2 &= \tilde{H}^\mu_y(u, y, p^2, \eta^2), && (6.120)\\ 0 &= \tilde{H}^\mu_u(u, y, p^2, \eta^2). && (6.121) \end{aligned}$$

On the interior of each arc, (u, η_2) can be expressed as a C^1 function of (y, p_2) and μ. Therefore, for $\|\mu - \bar{\mu}\|$ and $|\theta - \bar{\theta}_j|$ small enough, we have that $|u(t) - \bar{u}(t)|$, $|y(t) - \bar{y}(t)|$, $|\eta_2(t) - \bar{\eta}_2(t)|$, $|p_2(t) - \bar{p}_2(t)|$ are arbitrarily small, uniformly on an interior of each arc. Since u, y, η^2, and p^2 are continuous, uniformly over $[0, T]$, we deduce that $\|u - \bar{u}\|_\infty, \|y - \bar{y}\|_\infty, \|\eta^2 - \bar{\eta}^2\|_\infty, \|p^2 - \bar{p}^2\|_\infty$ are arbitrarily small for μ and θ in the neighborhood of $\bar{\mu}$ and $\bar{\theta}_j$, respectively. Using the relations (6.120)–(6.121), we obtain like in section 6.3 that the relation (6.40) holds. From now, the end of the proof is similar to the end of the proof of Th. 6.8 or 6.13 to show that the uniform strict complementarity assumption holds on boundary arc, depending on whether (τ_{en}, τ_{ex}) is in the neighborhood of a boundary arc $(\bar{\tau}_{en}, \bar{\tau}_{ex})$ or in the neighborhood of a nonreducible touch point $\bar{\tau}_{to}$ of $(\bar{u}, \bar{y})$. □

6.7 Application to homotopy methods

In this section, we extend to second-order state constraints the homotopy algorithm of [20] that detects automatically the structure of the trajectory for first-order state constraints, in the case when assumptions (A2)–(A6) and the strong second-order sufficient condition (6.65) are satisfied along the homotopy path.

6.7.1 Description of the algorithm

We consider the natural homotopy on the state constraint

$$g^\mu(y) \;:=\; g(y) - (1-\mu)M \qquad \text{and} \qquad (\ell^\mu, \phi^\mu, f^\mu, y_0^\mu) \equiv (\ell, \phi, f, y_0), \tag{6.122}$$

where $M > 0$ is large enough, so that the state constraint of problem $(\mathcal{P}^0)$ is not active, and we have that $(\mathcal{P}^1) \equiv (\mathcal{P})$. More generally, the algorithm below can be extended to any stable extension $(\mathcal{P}^\mu)$ of $(\mathcal{P})$ satisfying the assumption (H0) below, if a solution of $(\mathcal{P}^0)$ can be easily obtained:

(H0) $(\mathcal{P}^\mu)$ is a stable extension of $(\mathcal{P})$, defined for $\mu \in [0,1]$, such that $(\mathcal{P}^1) \equiv (\mathcal{P})$ and satisfying $g^\mu(y_0^\mu) < 0$ for all $\mu \in [0,1]$.

The homotopy algorithm is as follows. We denote the current structure of the trajectory by $\mathcal{S}$, i.e. the variable $\mathcal{S}$ indicates the number and order of boundary arcs and touch points. The shooting mapping associated with the structure $\mathcal{S}$ is denoted by $\mathcal{F}_\mathcal{S}$. Given a vector of shooting parameters θ, of dimension appropriate with $\mathcal{S}$, and a value $\mu \in [0,1]$ of the homotopy parameter, we will denote by $(u^\mu_{\mathcal{S},\theta}, y^\mu_{\mathcal{S},\theta})$ the trajectory associated with θ in the shooting algorithm for the structure $\mathcal{S}$ and the homotopy parameter μ.

Algorithm 6.26 (Homotopy Algorithm).

Input *p_0 initial costate candidate for the unconstrained problem $(\mathcal{P}^0)$ and $\delta \in (0,1)$.*

Initialization *Let $\mathcal{S}$ be the empty structure (with no boundary arc and no touch point). Solve by the Newton algorithm (initialized by the value p_0) $\mathcal{F}_\mathcal{S}(\theta, 0) = 0$ and obtain a vector of shooting parameters θ associated with a solution of the unconstrained problem $(\mathcal{P}^0)$. Set $M := \max_{t\in[0,T]} g(y^1_{\mathcal{S},\theta}(t))$. If $M \le 0$ then $\mu := 1$ else $\mu := 0$. Set $\Delta\mu := \delta$.*

While $\mu < 1$ **do**

Prediction Step *Set $\bar\mu := \min\{\mu + \Delta\mu; 1\}$ and compute*

$$\bar\theta := \theta - D_\theta \mathcal{F}_\mathcal{S}(\theta,\mu)^{-1} D_\mu \mathcal{F}_\mathcal{S}(\theta,\mu)\Delta\mu. \tag{6.123}$$

Correction Step *Solve, with the Newton algorithm initialized by the value $\bar\theta$,*

$$\mathcal{F}_\mathcal{S}(\hat\theta, \bar\mu) = 0. \tag{6.124}$$

If *the Newton algorithm fails, set $\Delta\mu := \Delta\mu/2$ and* **go to** *the* prediction step*;*
Else *obtain a vector of shooting parameters $\hat\theta$ solution of* (6.124)*.*

Update the structure

[TO→BA] **If** *there exists a touch point* τ_{to} *of* $\hat{\theta}$ *such that*

$$(g^{\bar{\mu}})^{(2)}(u^{\bar{\mu}}_{\mathcal{S},\hat{\theta}}(\tau_{to}), y^{\bar{\mu}}_{\mathcal{S},\hat{\theta}}(\tau_{to})) \geq 0, \tag{6.125}$$

let $\hat{\mathcal{S}}$ *be the structure obtained by replacing in* $\mathcal{S}$ *the touch point* τ_{to} *by a boundary arc, set* $\mathcal{S} := \hat{\mathcal{S}}$*, and let* $\bar{\theta}$ *be the vector of shooting parameters obtained from* θ *by replacing the touch point* τ_{to} *and its jump parameter* $\nu_{\tau_{to}}$ *by a boundary arc, with shooting parameters*

$$\tau_{en} := \tau_{to}, \qquad \tau_{ex} := \tau_{to}, \qquad \nu^1_{\tau_{en}} := \nu_{\tau_{to}}, \qquad \nu^2_{\tau_{en}} := 0. \tag{6.126}$$

Go to *the* CORRECTION STEP*;*

[ADD TO] **Else if** $m := \max_{t\in[0,T]} g^{\bar{\mu}}(y^{\bar{\mu}}_{\mathcal{S},\hat{\theta}}(t)) > 0$*, set* $\tau_{to} := \operatorname{argmax}_{t\in[0,T]} g^{\bar{\mu}}(y^{\bar{\mu}}_{\mathcal{S},\hat{\theta}}(t))$*, let* $\hat{\mathcal{S}}$ *be the structure obtained from* $\mathcal{S}$ *by adding the touch point* τ_{to}*, set* $\mathcal{S} := \hat{\mathcal{S}}$*, and let* $\bar{\theta}$ *be the vector of shooting parameters obtained from* θ *by adding the touch point* τ_{to} *with a zero jump parameter* $\nu_{\tau_{to}}$*.* **Go to** *the* CORRECTION STEP*;*

[REM TO] **Else if** *there exists a touch point* τ_{to} *of* $\hat{\theta}$ *such that its jump parameter* $\nu_{\tau_{to}}$ *is negative, then let* $\hat{\mathcal{S}}$ *be the structure obtained from* $\mathcal{S}$ *by deleting the touch point* τ_{to}*, set* $\mathcal{S} := \hat{\mathcal{S}}$*, and let* $\bar{\theta}$ *be the vector of shooting parameters obtained from* θ *by deleting the touch point* τ_{to} *and its jump parameter* $\nu_{\tau_{to}}$*.* **Go to** *the* CORRECTION STEP*;*

[BA→TO] **Else if** *there exists a boundary arc* (τ_{en}, τ_{ex}) *of* $\hat{\theta}$ *such that* $\tau_{en} > \tau_{ex}$*, then let* $\hat{\mathcal{S}}$ *be the structure obtained from* $\mathcal{S}$ *by replacing the boundary arc* (τ_{en}, τ_{ex}) *by a touch point, set* $\mathcal{S} := \hat{\mathcal{S}}$*, and let* $\bar{\theta}$ *be the vector of shooting parameters obtained from* θ *by replacing the shooting parameters associated with the boundary arc* (τ_{en}, τ_{ex}) *by a touch point and its jump parameter,*

$$\tau_{to} := \tau_{en}, \qquad \nu_{\tau_{to}} := \nu^1_{\tau_{en}}. \tag{6.127}$$

Go to *the* CORRECTION STEP*;*

[OK] **Else** *set* $\theta := \hat{\theta}$*,* $\mu := \bar{\mu}$*.*

End While

6.7.2 Construction of the homotopy path

The analysis of the existence of the homotopy path is analogous to that of [20] for first-order state constraints. Let $(\mathcal{P}^\mu)$ satisfy (H0) and assume that

(H1) The problem $(\mathcal{P}^0)$ has a local solution (u^0, y^0) satisfying (A2)–(A6) and the strong second-order sufficient condition (6.65).

By Th. 6.16, there exists $\delta > 0$ such that for all $\mu \in [0, \delta)$, $(\mathcal{P}^\mu)$ has a stationary point (u^μ, y^μ), locally unique in a L^∞-neighborhood of (u^0, y^0), which is Hölder continuous w.r.t. μ in the L^∞ norm and is a local solution of $(\mathcal{P}^\mu)$. By assumptions (A4)–(A6) and Theorems 6.8 and 6.13, (u^μ, y^μ) has a neighboring structure to that of (u^0, y^0) (in the sense of Def. 6.21), i.e. satisfies (A4). Moreover, reducing δ if necessary, assumptions (A2)–(A3) are satisfied, as

well as (A5) by Theorems 6.8 and 6.13 and (A6) by continuity. Finally, by Th. 6.16, (u^μ, y^μ) satisfies the strong second-order sufficient condition (6.65). So let

$$\mu_{max} := \sup\{\mu \in [0,1] : \text{for all } \mu' \in [0,\mu], \text{ the locally unique solution } (u^{\mu'}, y^{\mu'}) \text{ of } (\mathcal{P}^{\mu'}) \text{ satisfy (A2)–(A6) and the strong second-order sufficient condition (6.65).}\}$$

Under assumption (H1), we have that $\mu_{max} > 0$. We obtain the following result.

Lemma 6.27. *Assume that (H0)–(H1) are satisfied, and that there exist $L, \beta, \sigma > 0$ such that for all $\mu \in [0, \mu_{max})$,*

$$\|u^\mu\|_{1,1} \leq L, \tag{6.128}$$

$$|(g^\mu)^{(2)}_u(u^\mu(t), y^\mu(t))| \geq \beta \quad \text{for all } t \in I_\sigma(g^\mu(y^\mu)). \tag{6.129}$$

Then for all sequences $\mu_n \uparrow \mu_{max}$, there exists a subsequence, still denoted by (μ_n), such that $(u^{\mu_n}, y^{\mu_n}, p^{2,\mu_n}, \eta^{2,\mu_n})$ converges uniformly to some $(\tilde{u}, \tilde{y}, \tilde{p}^2, \tilde{\eta}^2)$, and $(\tilde{u}, \tilde{y}, \tilde{p}^2, \tilde{\eta}^2)$ is a stationary point and its alternative multipliers of $(\mathcal{P}^{\mu_{max}})$. Moreover, if $(\tilde{u}, \tilde{y}, \tilde{p}^2, \tilde{\eta}^2)$ satisfies assumptions (A2)–(A6) and the strong second-order sufficient condition (6.65), then $(u^\mu, y^\mu, p^{2,\mu}, \eta^{2,\mu})$ converges uniformly when $\mu \uparrow \mu_{max}$ to a locally unique local solution of $(\mathcal{P}^{\mu_{max}})$ and its alternative multipliers $(\tilde{u}, \tilde{y}, \tilde{p}^2, \tilde{\eta}^2) =: (u^{\mu_{max}}, y^{\mu_{max}}, p^{2,\mu_{max}}, \eta^{2,\mu_{max}})$, and $\mu_{max} = 1$, i.e. the homotopy path is well-defined over $\mu \in [0,1]$.

Proof. The proof follows from that of [20, Lemma 8.4][15]. By the compactness Theorem in BV [2, Th. 3.23], the weak-* convergence in $\mathcal{M}[0,T]$ of the multiplier $\mathrm{d}\eta^{\mu_n}$ associated with (u^{μ_n}, y^{μ_n}) in the optimality conditions (6.7)–(6.10) implies the uniform convergence of the alternative multiplier η^{2,μ_n} defined by (6.11). The uniform convergence of p^{2,μ_n} follows then from (6.16). □

Given a stable extension $(\mathcal{P}^\mu)$ satisfying (H0) and (H1), we make the following assumptions that guarantee the existence (and local uniqueness) of the homotopy path over $\mu \in [0,1]$:

(H2) There exists $L, \beta, \sigma > 0$ such that for all $\mu \in [0,1]$, (u^μ, y^μ) satisfies (6.128)–(6.129);

(H3) For all $\mu \in [0,1]$, (u^μ, y^μ) satisfies the assumptions (A3)–(A6);

(H4) For all $\mu \in [0,1]$, (u^μ, y^μ) satisfies the strong second-order sufficient condition (6.65).

6.7.3 Proof of convergence

In addition to hypotheses (H0)–(H4), we make the assumptions below in the proof of correctness of Algorithm 6.26. Note that a change in the structure of the trajectories (u^μ, y^μ), $\mu \in [0,1]$, may occur only at some values $\tilde{\mu} \in [0,1)$ having either a nonessential or a nonreducible touch point.

(H5) There exist finitely values of μ, $0 \leq \tilde{\mu}_1 < \ldots < \tilde{\mu}_N < 1$ for which the structure of the trajectory changes.

(H6) For each $\tilde{\mu}_k$, $k = 1, \ldots, N$, $(u^{\tilde{\mu}_k}, y^{\tilde{\mu}_k})$ has either one (single) nonessential touch point or one (single) nonreducible touch point.

[15] Lemma 3.37 of this thesis.

When (H6) holds, there are only two different neighboring structures to that of $(u^{\tilde{\mu}_k}, y^{\tilde{\mu}_k})$, for each $\tilde{\mu}_k$. The algorithm 6.26 could be generalized to the case when (H6) does not hold, but in that case the UPDATE THE STRUCTURE step is more delicate. A possibility is to enumerate all the possible neighboring structures until the conditions (6.116)–(6.119) of Lemma 6.25 are satisfied.

Proposition 6.28. *Let $(\mathcal{P}^\mu)$ be given by* (6.122) *and assume that assumptions (H1)–(H6) are satisfied. Then there exist a neighborhood V_0 of $p_2^0(0)$, the initial costate of the unconstrained problem $(\mathcal{P}^0)$, and $\bar{\delta} > 0$ such that for all $p_0 \in V_0$ and all $\delta \in (0, \bar{\delta})$, the homotopy algorithm 6.26 follows the homotopy path and ends with a vector of shooting parameters θ, of appropriate dimension, associated with a local solution (u^1, y^1) of $(\mathcal{P}^1) \equiv (\mathcal{P})$. In addition, if δ is small enough, then the steps $\Delta\mu$ are not reduced by the algorithm in the* CORRECTION STEP*, i.e. Newton's algorithm does not fail.*

Remark 6.29. In practice, at the end of the instruction labelled [OK], when the homotopy step has succeeded, it is possible to increase $\Delta\mu$, so that the algorithm adapts itself to the largest possible value of the homotopy step $\Delta\mu$ allowing the convergence in the CORRECTION step.

Proof. The proof follows the ideas of [20, Prop. 8.11][16]. Note that the value of μ is increased only in the instruction labelled [OK] in the UPDATE THE STRUCTURE step. Therefore, if the algorithm ends with $\mu = 1$, this means that all the conditions (6.116)–(6.119) are satisfied, and hence, by Lemma 6.25, θ is a vector of shooting parameters associated with a stationary point (u^1, y^1) of $(\mathcal{P}^1)$. Note that when there is no change in the structure of solutions, then Algorithm 6.26 is a classical predictor-corrector algorithm. We therefore have to show that the algorithm ends with $\mu = 1$, i.e.

- There is no failure in the Newton's algorithm in the CORRECTION STEP if δ is small enough;
- The algorithm finishes off at the (finitely many by (H5)) changes in the structure of the trajectories along the homotopy path, i.e. after finitely many iterations in the UPDATE THE STRUCTURE step, succeeds in finding the new structure $\mathcal{S}$ and a vector of shooting parameter $\hat{\theta}$ associated with $(u^{\bar{\mu}}, y^{\bar{\mu}})$ that satisfies the conditions (6.116)–(6.119) of Lemma 6.25.

For the current value of $\mu \in (0, 1)$, assume by induction that the current value θ is a vector of shooting parameters associated with the stationary point (u^μ, y^μ) of $(\mathcal{P}^\mu)$, and that $\mathcal{S}$ denotes the corresponding structure of (u^μ, y^μ). Assume that

$$\theta \text{ and } \mathcal{S} \text{ are such that } \textit{nonreducible touch points are introduced as boundary arcs.} \quad (6.130)$$

(We still do not have the uniqueness of θ and $\mathcal{S}$ whenever nonessential touch points are present, that can or not be introduced in the shooting mapping.) This holds for $\mu = 0$ if p_0 is chosen sufficiently close to $p_2^0(0)$ by (H1).

Let $\bar{\theta}$ and $\bar{\mu}$ be defined as in the PREDICTION STEP. Let $\hat{\theta}$ be the solution of (6.124). By (6.123), $|\bar{\theta} - \hat{\theta}| \leq C|\bar{\mu} - \mu|^2$ for some positive constant C. Since $|\bar{\mu} - \mu| \leq \Delta\mu \leq \delta$, for δ small enough, $\bar{\theta}$ belongs to the domain of convergence of the Newton algorithm, which converges to $\hat{\theta}$. Note that the constant C and the size of the domain of convergence of the Newton

[16] Proposition 3.44 of this thesis.

algorithm are uniform along the homotopy path for $\mu \in [0,1]$, see e.g. [20, Prop. 8.11][17], so that we do not have $\delta \to 0$.

Let us show that if δ is small enough, there is at most one passage in one of the instructions [TO→BA], [ADD TO], [REM TO], [BA→TO] before the value of μ is increased. Assume by (H5) that

$$0 \ < \ \delta \ < \ \min_{1\le k\le N-1} \tilde\mu_{k+1} - \tilde\mu_k. \tag{6.131}$$

If one of the tests [TO→BA], [ADD TO], [REM TO], [BA→TO] is satisfied, this means by Lemma 6.25 and (6.130) that the current structure $\mathcal{S}$ is not correct, and hence by (H5) and (6.131) there exists $k \in \{1,\dots,N\}$ such that

$$\mu \ \le \ \tilde\mu_k \ \le \bar\mu,$$

with at least one of the two above inequalities being strict, and we have $\bar\mu < \tilde\mu_{k+1}$ if $k < N$ and $\bar\mu \le 1$ if $k = N$ and $\bar\mu > \tilde\mu_N$.

Let us start by the case [TO→BA] when (6.125) is satisfied. This can occur only in the neighborhood of a nonreducible touch point $\bar\tau_{to}$ of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$. If $(g^{\bar\mu})^{(2)}(u^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to}), y^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to})) > 0$, a second-order Taylor expansion of $g^{\bar\mu}(y^{\bar\mu}_{\mathcal{S},\hat\theta})$ at the touch point τ_{to} shows that $g^{\bar\mu}(y^{\bar\mu}_{\mathcal{S},\hat\theta}(t)) > 0$ for t in the neighborhood of τ_{to}, $t \ne \tau_{to}$. If $(g^{\bar\mu})^{(2)}(u^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to}), y^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to})) = 0$, then τ_{to} is a nonreducible touch point. In view of (6.130), in both cases the structure $\mathcal{S}$ where τ_{to} is considered as a touch point is not correct. By (H6), there exist only two different neighboring structures to that of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$, so having eliminated $\mathcal{S}$, it remains only the other possible structure $\hat{\mathcal{S}}$ where $\bar\tau_{to}$ is introduced as a boundary arc. The associated new vector of shooting parameters $\bar\theta$ is obtained from θ by (6.126). Since we know that $\theta^{\bar\mu}$, the vector of shooting parameters associated with $(u^{\bar\mu}, y^{\bar\mu})$, is solution of

$$\mathcal{F}_{\hat{\mathcal{S}}}(\theta^{\bar\mu}, \bar\mu) = 0, \tag{6.132}$$

it remains to show that the Newton algorithm initialized with the value $\bar\theta$ converges to $\theta^{\bar\mu}$. Denote by $\theta^{\tilde\mu_k}_{\mathcal{S}}$ and $\theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}$ the vector of shooting parameters associated with $\tilde\mu_k$ for the structures $\mathcal{S}$ and $\hat{\mathcal{S}}$, respectively, and $\bar\tau_{en}$, $\bar\tau_{ex}$, $\bar\nu^1_{\tau_{en}}$, $\bar\nu^2_{\tau_{en}}$ the shooting parameters associated with the nonreducible touch point $\bar\tau_{to}$ introduced as a boundary arc in $\theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}$. Recall that the latter are given by (6.90)–(6.91). Therefore, in view of (6.126),

$$\begin{aligned} |\bar\theta - \theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}| \ &\le \ |\theta - \theta^{\tilde\mu_k}_{\mathcal{S}}| + |\tau_{en} - \bar\tau_{en}| + |\tau_{ex} - \bar\tau_{ex}| + |\nu^1_{\tau_{en}} - \bar\nu^1_{\tau_{en}}| + |\nu^2_{\tau_{en}} - \bar\nu^2_{\tau_{en}}| \\ &\le \ |\theta - \theta^{\tilde\mu_k}_{\mathcal{S}}| + 2|\tau_{to} - \bar\tau_{to}| + |\nu_{\tau_{to}} - \bar\nu_{\tau_{to}}| \ \le \ 4|\theta - \theta^{\tilde\mu_k}_{\mathcal{S}}|. \end{aligned}$$

Since θ is the solution of $\mathcal{F}_{\mathcal{S}}(\theta,\mu) = 0$, it follows from Lemma 6.22 applied with $(\bar u, \bar y) = (u^{\tilde\mu_k}, y^{\tilde\mu_k})$ that there exists $\kappa > 0$ such that $|\theta - \theta^{\tilde\mu_k}_{\mathcal{S}}| \le \kappa|\mu - \tilde\mu_k| \le \kappa\delta$. By Lemma 6.22 again, there exists a constant κ' such that $|\theta^{\bar\mu} - \theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}| \le \kappa'|\bar\mu - \tilde\mu_k| \le \kappa'\delta$. It follows that $|\bar\theta - \theta^{\bar\mu}| \le |\bar\theta - \theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}| + |\theta^{\bar\mu} - \theta^{\tilde\mu_k}_{\hat{\mathcal{S}}}| \le (4\kappa + \kappa')\delta$. Therefore, for δ small enough, $\bar\theta$ belong to the domain of convergence of the Newton algorithm which converges to $\hat\theta := \theta^{\bar\mu}$, and all the conditions (6.116)–(6.119) are satisfied, as well as (6.130), so we may set $\theta := \hat\theta$, $\mu := \bar\mu$, and $\hat{\mathcal{S}} = S$ and the induction step is completed. (Here again, the constants κ, κ' can be chosen uniform w.r.t. μ along the homotopy path so that $\delta \not\to 0$.)

[17] Proposition 3.44 of this thesis.

For the other cases, the discussion is similar so will be less detailed. In the case [ADD TO], the state constraint is violated. But then $(g^{\bar\mu})^{(2)}(u^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to}), y^{\bar\mu}_{\mathcal{S},\hat\theta}(\tau_{to})) < 0$ for all touch points τ_{to}, since otherwise we would have been in the previous case [TO→BA]. Therefore, by a second-order Taylor expansion of $g^{\bar\mu}(y^{\bar\mu}_{\mathcal{S},\hat\theta})$, the state constraint is not violated in the neighborhood of a touch point. Consequently, it may only be violated in the neighborhood of a nonessential touch point $\bar\tau_{to}$ of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$ which is not introduced in the shooting mapping. By (H6), the only other possible structure $\hat{\mathcal{S}}$ is when $\bar\tau_{to}$ is introduced as a touch point in the shooting mapping.

In the case [REM TO], a jump parameter associated with a touch point is negative. This cannot happen in the neighborhood of a nonreducible touch point of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$, since nonreducible touch points are assumed to be essential by (A6)(i) and Lemma 6.12. Therefore, this can only happen in the neighborhood of a nonessential touch point $\bar\tau_{to}$ of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$. By (H6), the only other possible structure $\hat{\mathcal{S}}$ is to remove this touch point from the shooting mapping. Finally, in the last case [BA→TO], we have a boundary arc whose entry point τ_{en} is greater than the corresponding exit point τ_{ex}. This can only happen in the neighborhood of a nonreducible touch point $\bar\tau_{to}$ of $(u^{\tilde\mu_k}, y^{\tilde\mu_k})$ that was converted in a boundary arc, and therefore by (H6) the only other possible structure $\hat{\mathcal{S}}$ is to introduce this nonreducible touch point as a touch point instead. We conclude with similar arguments as before that for δ small enough, the Newton algorithm initialized by $\bar\theta$ converges to the solution of (6.132), which is a vector of shooting parameters associated with the stationary point $(u^{\bar\mu}, y^{\bar\mu})$ of $(\mathcal{P}^{\bar\mu})$. This completes the induction step.

This shows that if δ is small enough, the algorithm follows the homotopy path, the Newton algorithm does not fail, and the algorithm ends with $\mu = 1$. By (H4), the second-order sufficient condition (6.65) holds and therefore (u^1, y^1) is a local solution of $(\mathcal{P})$. □

6.8 Remarks

Remark 6.30. It would of course be interesting to test the homotopy algorithm on numerical applications. This is the subject of the report [70]. The homotopy algorithm is based on the strong assumptions (A5) and (A6)(ii), that would have to be checked in practice in order to guarantee the validity of the algorithm, as well as the second-order sufficient condition (6.65). Moreover, the same restrictions as for first-order state constraints hold, see [20, Remarks 8.12 and 8.13][18]. In particular, a value of δ that guarantee the convergence by Prop. 6.28 is not known in practice, and may be small if the problem is ill-conditioned.

Remark 6.31. It is expected that the homotopy algorithm can be extended to vector-valued control and several state constraints of first- and second order if the constraints are linearly independent (see [98, 17]). The difficulty in the theoretical justification of the algorithm is the extension of Theorems 6.8, 6.13, and [20, Th. 2.1][19]. For control constraints, the extension of this homotopy algorithm is not immediate (see [20, Remark 6.3][20]) and is an interesting open question. In contrast, it seems not to be possible to extend this algorithm to state constraints of order greater than or equal to three, since in that case optimal trajectories typically exhibit infinitely many touch points near entry/exit of boundary arcs, see [118].

Remark 6.32. The sufficient second-order condition (6.65) used in Th. 6.16 is not the weakest possible since it does not take into account the curvature of the constraint. The curvature

[18] Remarks 3.45 and 3.46 of this thesis.
[19] Theorem 3.4 of this thesis.
[20] Remark 3.32 of this thesis.

term of the constraint (see [21]) is the term with the sum in (6.93). It would therefore be interesting to see if Th. 6.16 is still true under the weaker second-order sufficient condition

$$\mathcal{Q}(v) - \sum_{\tau \in \bar{\mathcal{T}}_{red}^{ess}} \bar{\nu}_\tau \frac{(g_y^{(1)}(\bar{y}(\tau))z_v(\tau))^2}{g^{(2)}(\bar{u}(\tau), \bar{y}(\tau))} > 0, \quad \text{for all } v \in \mathcal{V},\ v \neq 0, \text{ satisfying (6.63).} \tag{6.133}$$

With additional assumptions (A4)–(A5) on the structure of the trajectory and in the absence of nonreducible touch points, it was shown in [19, Th. 4.3][21] (see also [20]) that (6.133) characterizes the uniform quadratic growth condition (6.28), and implies L^∞-Lipchitz continuity and directional differentiability of solutions in L^r, $r < \infty$, (see [94, 19]), improving the Hölder continuity in L^∞ only obtained in Th. 6.16. Directional differentiability of all shooting parameters is also obtained. It would be interesting to extend those results in presence of nonreducible touch points as well.

Remark 6.33. Let us discuss the case when the term $\bar{\lambda}(\bar{\tau}_{to})$ defined by (6.58) at a nonreducible touch point $\bar{\tau}_{to}$ is positive. In that case, by Th. 6.13, a second touch point may appear for stationary points of the perturbed problem. The first idea is therefore to introduce a second touch point in the shooting mapping, and at the reference trajectory $(\bar{u}, \bar{y})$, the values of both touch points would be equal to the value of the nonreducible touch point $\bar{\tau}_{to}$. The problem is that doing so, it is easy to see that the Jacobian of the shooting mapping becomes singular (two rows are equal). Moreover, the jump parameters associated with each touch point at the reference trajectory are not well-defined, only the sum of the two jumps parameters must be equal to $\bar{\nu}_{\bar{\tau}_{to}}$. There exist indeed several zeros of the perturbed shooting function in the neighborhood of a nonreducible touch point splitting into two touch points, and one of them is such that the values of both touch points remain equal to each other (as if we had a single touch point). In that case of course the state constraint may be violated.

For this reason, it would be necessary to initialize the two touch points with distinct values and it is an open question how to do so in order to insure to be into the domain of convergence of the Newton algorithm for the new structure. To solve the academic problem in Fig. 6.1(b), the nonreducible touch point was first converted into a boundary arc. We thus obtained a zero of the resulting shooting function with a boundary arc satisfying $\tau_{en} < \tau_{ex}$, but the condition $\ddot{\eta}^2 \geq 0$ was of course violated. We used the obtained values of τ_{en} and τ_{ex} to initialize the two touch points, and the heuristic formula below (recall (6.54))

$$\nu_{\tau_{en}} := \frac{\tilde{H}_{uu}}{(g_u^{(2)})^2} g^{(3)}(\dot{u}, u, y)(\tau_{en}^-), \qquad \nu_{\tau_{ex}} := -\frac{\tilde{H}_{uu}}{(g_u^{(2)})^2} g^{(3)}(\dot{u}, u, y)(\tau_{ex}^+)$$

to initialize the associated jump parameters.

[21] Theorem 2.34 of this thesis.

Chapitre 7

Conclusion

Cette thèse a apporté des résultats théoriques nouveaux pour les problèmes de commande optimale avec contraintes sur l'état, dans le cas où la condition forte de Legendre-Clebsch est satisfaite. Ces résultats portent sur les conditions du second ordre (des conditions sans saut entre conditions nécessaire et suffisante ont été obtenues), l'analyse de l'algorithme de tir et l'analyse de stabilité et sensibilité des solutions, en particulier pour des contraintes d'ordre élevé (supérieur ou égal à deux), cas relativement peu traité dans la littérature. Une méthode d'homotopie a également été proposée, dont la nouveauté est de déterminer automatiquement, sous certaines hypothèses, la structure de la trajectoire. Cette méthode reste encore à être validée sur des applications numériques, et éventuellement généralisée à des cas plus généraux (plusieurs contraintes sur l'état et sur la commande). Une première application au problème de la rentrée atmosphérique d'une navette spatiale avec contrainte sur le flux thermique (contrainte sur l'état du second ordre) a été réalisée dans [70].

Dans la suite de cette conclusion on présente quelques questions ouvertes qui se sont posées au cours de la thèse et pourraient faire l'objet de futures recherches.

7.1 Questions non résolues

7.1.1 Vérification de la condition suffisante du second ordre

Un problème important pour les applications est d'être capable de vérifier numériquement la condition suffisante du second ordre no-gap (4.133) du théorème 4.24. Il s'agit d'un problème ouvert. On sait cependant vérifier une condition suffisante plus forte que (4.133), ce qui peut parfois s'avérer suffisant pour les applications dans le cas où cette condition plus forte est satisfaite.

On utilise dans cette section les notations du chapitre 4. Utilisant les techniques basées sur les équations de Riccati (voir Maurer [99]), il est possible de vérifier numériquement la stricte positivité du membre de gauche de (4.133) (noté $Q(v)$ dans la suite et donné par (4.135) avec $\mathcal{T}_{to}^{i,ess}$ remplacé par $\mathcal{T}_{red}^{i}$) sur l'ensemble — plus grand que le cône critique $\hat{C}_{L^2}(u)$ — des $v \in L^2(0,T;\mathbb{R}^m)$ satisfaisant

$$g_{i,u}^{(q_i)}(u,y)v + g_{i,y}^{(q_i)}(u,y)z_v = 0 \quad \text{p.p. sur } \operatorname{int}\Delta_i, \quad \forall\, i = 1,\dots,r+s. \tag{7.1}$$

On considère alors la condition du second ordre suivante, plus forte que (4.133),

$$Q(v) > 0, \quad \forall\, v \in L^2(0,T;\mathbb{R}^m) \setminus \{0\} \text{ satisfaisant (7.1)}. \tag{7.2}$$

Lorsque l'hypothèse de complémentarité stricte (A6) est satisfaite, le cône critique $\hat{C}_{L^2}(u)$ étant donné par (4.203)–(4.205), on omet donc dans (7.2) les contraintes "pures sur l'état" (4.204)–(4.205). En effet les techniques connues basées sur les équations de Riccati permettent seulement de prendre en compte les contraintes mixtes sur la commande et l'état. Ainsi la difficulté dans la vérification de la condition suffisante du second ordre par cette méthode ne provient pas du terme de courbure, mais des contraintes (4.204)–(4.205) liées aux contraintes pures sur l'état, intrinsèquement présentes dans le cône critique.

Soit $\overset{o}{I}(t) := \{i \in \{1, \dots, r+s\} : t \in \operatorname{int} \Delta_i\}$. Considérons l'équation de Riccati suivante (f_y, H_{yy}, etc. étant évalués le long de la trajectoire considérée)

$$\begin{aligned} -\mathrm{d}X(t) &= (Xf_y + f_y^\top X + H^0_{yy})\mathrm{d}t + \sum_{i=1}^{r} g_{i,yy}\mathrm{d}\eta_i(t) - \sum_{i=1}^{r}\sum_{\tau\in\mathcal{T}^i_{red}} \nu^i_\tau \frac{(g^{(1)}_{i,y})^\top g^{(1)}_{i,y}}{g^{(2)}_i}\delta_\tau(t) \\ &\quad - \left(\begin{pmatrix} H^0_{uy} \\ G^{(q)}_{\overset{o}{I}(t),y} \end{pmatrix}^\top + X\begin{pmatrix} f_u^\top \\ 0 \end{pmatrix}^\top\right)\begin{pmatrix} H^0_{uu} & (G^{(q)}_{\overset{o}{I}(t),u})^\top \\ G^{(q)}_{\overset{o}{I}(t),u} & 0 \end{pmatrix}^{-1} \\ &\qquad * \left(\begin{pmatrix} f_u^\top \\ 0 \end{pmatrix} X + \begin{pmatrix} H^0_{uy} \\ G^{(q)}_{\overset{o}{I}(t),y} \end{pmatrix}\right)\mathrm{d}t, \\ X(T) &= \phi_{yy}(y(T)), \end{aligned} \tag{7.3}$$

où δ_τ désigne la mesure de Dirac en τ et $\nu^i_\tau = [\eta_i(\tau)]$. On a le résultat suivant, dont la preuve est donnée dans l'annexe de ce chapitre (section 7.2.1). Il serait intéressant de disposer d'une caractérisation analogue de la condition (4.133), mais la prise en compte des contraintes sur l'état (4.204)–(4.205) par les techniques utilisées dans ce résultat ne semble pas immédiate.

Proposition 7.1. *Soit (u, y) un point stationnaire de $(\mathcal{P})$ satisfaisant les hypothèses (A1)–(A4) du chapitre 4, et pour chaque contrainte sur l'état g_i d'ordre $q_i \geq 2$, soit $\mathcal{T}^i_{red}$ un ensemble fini de points de contact isolés essentiels réductibles de la contrainte g_i. Si l'équation de Riccati* (7.3) *a une solution bornée X sur $[0, T]$, alors* (7.2) *(et donc* a fortiori (4.133)*) est satisfaite.*

De plus, la réciproque est vraie, i.e. si (7.2) *est satisfaite, alors l'équation de Riccati* (7.3) *admet une solution bornée X sur $[0, T]$.*

7.1.2 Extensions du résultat sur les conditions du second ordre aux équations aux dérivées partielles

Lorsque la dynamique du problème n'est plus décrite par une équation différentielle ordinaire (0.2), comme considéré dans cette thèse, mais par une équation aux dérivées partielles, la question d'obtenir des conditions du second ordre no-gap est largement ouverte. Des conditions suffisantes du second ordre sont obtenues dans par exemple [64, 38, 117], mais des conditions nécessaires ou suffisantes no-gap sont connues seulement dans le cas d'un nombre fini de contraintes d'égalité ou d'inégalité sur l'état (théorie de la polyédricité) [25, 36, 37]. Lorsqu'il y a des contraintes distribuées sur l'état, le problème est ouvert. On peut alors se demander si l'on peut étendre la méthode développée dans le chapitre 1 pour des EDP avec contraintes distribuées sur l'état. Le calcul du terme de courbure de Kawasaki s'avère dans ce cas plus difficile, car l'état y_u et l'état linéarisé z_v sont moins réguliers que pour une EDO. De plus, les hypothèses (A4)-(A6) intervenant dans la preuve deviennent plus délicates à formuler pour une EDP.

Des hypothèses de type (A4) sur la structure de l'ensemble de contact pour les problèmes de commande optimale des EDP elliptiques ont été faites dans [10]. Sous ces hypothèses les auteurs montrent que le multiplicateur associé à la contrainte sur l'état μ est composé d'une partie distribuée dans L^2 sur l'intérieur de l'ensemble de contact, et d'une partie mesure concentrée sur la frontière de l'ensemble de contact. Ce résultat est l'analogue de la proposition 0.2 pour les EDO, qui dit que les sauts du multiplicateur η ne peuvent se produire qu'aux points de jonction. Les hypothèses et résultats de [10] pourraient donc être intéressants pour étendre l'analyse du chapitre 1 aux EDP. Par ailleurs, dans [16], un problème de programmation semi-infinie (avec contraintes dans un espace de fonctions continues sur un espace métrique compact) est considéré. Les auteurs font des hypothèses sur l'ensemble de contact et sur la croissance des solutions au voisinage de l'ensemble de contact qui leur permettent de calculer le terme de courbure et d'obtenir des conditions du second ordre sans saut. Il est peut-être possible d'étendre ces résultats à certains problèmes de commande optimale des EDP.

Regardons ce que donne le calcul du terme de courbure et les difficultés qui se posent sur un exemple simple. Pour cela on considère le problème elliptique suivant

$$\min_{u\in L^2(\Omega)} \quad \frac{1}{2}\int_\Omega (y_u(x)-y_d(x))^2\mathrm{d}x + \frac{\gamma}{2}\int_\Omega u^2(x)\mathrm{d}x \tag{7.4}$$

$$\text{s.c.} \quad -\Delta y_u = u \text{ dans } \Omega, \quad y_u = 0 \text{ sur } \partial\Omega, \tag{7.5}$$

$$y_u(x) \le 1 \text{ dans } \Omega \tag{7.6}$$

où Ω désigne un ouvert borné de $\mathbb{R}^d$ avec $d \le 3$, à bord $\partial\Omega$ suffisamment régulier, γ est une constante strictement positive et $y_d \in L^2(\Omega)$. Pour $d = 1$, (7.5) est une EDO donc le calcul du terme de courbure se ramène à celui effectué pour une contrainte sur l'état d'ordre 2. On s'intéresse donc au cas où $d = 2$ ou 3. On note dans la suite $C^{r,\nu}(\bar\Omega)$ l'espace des fonctions de classe C^r sur $\bar\Omega$ et dont la dérivée d'ordre r est höldérienne d'exposant ν.

La première étape est d'étudier la régularité des solutions. L'opérateur $L^s(\Omega) \to W^{2,s}(\Omega) \cap W_0^{1,s}(\Omega)$ qui à u associe l'unique solution y_u de (7.5) est bien défini et c'est un isomorphisme, pour tout $s \in [2, +\infty[$. De plus on a les inclusions de Sobolev (voir par exemple [85, chap. 2, Th. 6.2])

$$\begin{array}{lllll} \text{si } \frac{d}{2} < s < d & W^{2,s}(\Omega) & \subset & C^{0,\nu}(\bar\Omega) & \text{avec } \nu = 2 - \frac{d}{s} \\ \text{si } s = d & W^{2,s}(\Omega) & \subset & C^{0,\nu}(\bar\Omega) & \text{pour tout } 0 < \nu < 1 \\ \text{si } s > d & W^{2,s}(\Omega) & \subset & C^{1,\nu}(\bar\Omega) & \text{avec } \nu = 1 - \frac{d}{s}. \end{array} \tag{7.7}$$

En particulier, pour $u \in L^2(\Omega)$ et $d = 2, 3$, l'état y_u est continu sur $\bar\Omega$ et on peut considérer la contrainte sur l'état (7.6) dans l'espace des fonctions continues.

La condition d'optimalité du premier ordre est la suivante : en plus de (7.5)-(7.6), il existe p, μ tels que

$$-\Delta p = (y_u - y_d) + \mu \text{ dans } \Omega, \qquad \partial p = 0 \text{ sur } \partial\Omega, \tag{7.8}$$

$$0 = p + \gamma u \tag{7.9}$$

$$\mu \in \mathcal{M}_+(\bar\Omega), \qquad \mathrm{supp}(\mu) \subset I(y_u) := \{x \in \Omega : y_u(x) = 1\}. \tag{7.10}$$

D'après [35], l'équation adjointe (7.8)[1] a, pour tout $\mu \in \mathcal{M}(\bar\Omega)$, une unique solution $p \in W_0^{1,r^*}(\Omega)$ avec $1 \le r^* < d/(d-1)$. Avec (7.9), on en déduit que $u \in W_0^{1,r^*}(\Omega) \subset L^q(\Omega)$ avec

[1] L'équation (7.8) est à comprendre au sens des distributions, i.e. $-\int_\Omega p\Delta\varphi = \int_\Omega (y_u - y_d)\varphi + \int_\Omega \varphi \mathrm{d}\mu$ pour toute fonction φ C^∞ à support compact dans Ω.

$1/q = 1/r^* - 1/d$, d'où $q < +\infty$ si $d = 2$ et $q < 3$ si $d = 3$. Par cette analyse on améliore la régularité de u, et donc aussi celle de y_u. On en déduit que $y_u \in W^{2,q}(\Omega)$, d'où avec (7.7), $y_u \in C^{1,\nu}(\bar\Omega)$ avec $\nu < 1$ si $d = 2$ et $y_u \in C^{0,\nu}(\bar\Omega)$ avec $\nu < 1$ si $d = 3$.

Passons maintenant au calcul du terme de courbure. Pour $v \in L^2(\Omega)$, soit z_v l'unique solution dans $H^2(\Omega) \cap H_0^1(\Omega)$ de

$$-\Delta z_v = v \text{ dans } \Omega, \qquad z_v = 0 \text{ sur } \partial\Omega$$

et soit $I^2(y_u, z_v)$ l'ensemble de contact du second ordre

$$I^2(y_u, z_v) := \{x \in I(y_u) : z_v(x) = 0\}.$$

Supposons pour simplifier l'hypothèse de complémentarité stricte $\mathrm{supp}(\mu) = I(y_u)$ satisfaite. Alors le cône critique est donné par

$$C(u) = \{v \in L^2(\Omega) : z_v(x) = 0 \text{ sur } I(y_u)\}.$$

Le terme de courbure est donné par (voir la section 1.2.1)

$$\sigma(\mu, T_K^{2,i}(y_u - 1, z_v)) \tag{7.11}$$

avec $\sigma(\mu, S) = \sup_{w\in S}\langle\mu, w\rangle$ la fonction support de l'ensemble S et $T_K^{2,i}(y_u - 1, z_v)$ l'ensemble tangent intérieur du second ordre à $K = C_-(\bar\Omega)$, caractérisé par (voir [42])

$$T_K^{2,i}(y_u - 1, z_v) = \{w \in C(\bar\Omega) : w(x) \le \iota_{y_u,z_v}(x) \ \forall\, x \in \bar\Omega\}$$

où

$$\iota_{y_u,z_v}(x) = \begin{cases} 0 & \text{si } x \in \operatorname{int} I(y_u) \cap I^2(y_u, z_v) \\ \liminf\limits_{x'\to x,\ y_u(x')<1} \dfrac{(z_v(x')_+)^2}{2(y_u(x') - 1)} & \text{si } x \in \partial I(y_u) \cap I^2(y_u, z_v) \\ +\infty & \text{à l'extérieur de } I^2(y_u, z_v). \end{cases} \tag{7.12}$$

Le terme de courbure (7.11) est $> -\infty$ ssi $T_K^{2,i}(y_u - 1, z_v) \neq \emptyset$, et donc ssi $\iota_{y_u,z_v}(x) > -\infty$ pour tout x. Comme pour les EDO, la difficulté dans le calcul de ι_{y_u,z_v} provient du terme avec la lim inf pour $x \in \partial I(y_u) \cap I^2(y_u, z_v)$. Analysons par exemple la contribution d'un point de contact isolé $\xi \in \Omega$. Pour $v \in L^2(\Omega)$, on a $z_v \in C^{0,\sigma}(\bar\Omega)$ avec $\sigma \in (0, 1)$ si $d = 2$ et $\sigma = 1/2$ si $d = 3$. Il vient qu'au voisinage de ξ, on a, pour $v \in C(u)$ et x au voisinage de ξ,

$$z_v(x) = \mathcal{O}(|x - \xi|^\sigma).$$

Le terme au numérateur de (7.12) est donc un $\mathcal{O}(|x - \xi|^{2\sigma})$.

Pour assurer que la contribution de ξ dans le terme de courbure est finie, il faudrait une hypothèse (analogue de (A5) du chapitre 1) du type

$$|y_u(x) - 1| \ge a|x - \xi|^\alpha \tag{7.13}$$

au voisinage de ξ, avec $a, \alpha > 0$ et, pour que la lim inf soit finie,

$$\alpha \le 2\sigma < 2 \text{ si } d = 2 \quad \text{et} \quad \alpha \le 2\sigma \le 1 \text{ si } d = 3. \tag{7.14}$$

Par ailleurs, l'étude de la régularité de y_u implique que $y_u \in C^{1,\nu}(\bar\Omega)$ pour tout $\nu < 1$ si $d = 2$ et $y_u \in C^{0,\nu}(\bar\Omega)$ pour tout $\nu < 1$ si $d = 3$, et donc $|y_u(x) - 1| = \mathcal{O}(|x - \xi|^{1+\nu})$ pour tout $\nu < 1$

si $d = 2$ (puisque qu'alors $\nabla y_u(\xi) = 0$) et $|y_u(x) - 1| = \mathcal{O}(|x - \xi|^\nu)$ pour tout $\nu < 1$ si $d = 3$. Cela implique qu'on a nécessairement, pour $d = 2$, $\alpha \geq 2$, ce qui n'est pas compatible avec la condition (7.14). Pour $d = 3$, on a nécessairement $\alpha \geq 1$, et donc avec (7.14), $\alpha = 1$. Ainsi, sur cet exemple dans le cas où $d = 2$ on ne peut pas conclure que la contribution d'un point de contact isolé dans le terme de courbure est finie. Et si le terme de courbure est égal à $-\infty$, la condition nécessaire du second ordre de Kawasaki n'apporte pas d'information. En revanche, pour $d = 3$, la contribution du point de contact isolé ξ dans le terme de courbure est finie sous l'hypothèse (7.13) avec $\alpha = 1$ (qui est bien satisfaite, voir la remarque 7.2). Il resterait encore à évaluer précisément cette contribution en calculant la lim inf dans (7.12).

Remarque 7.2. Plus précisément, au voisinage d'un point de contact isolé ξ, μ est une mesure de Dirac concentrée en ξ, $\mu = \nu\delta_\xi$ avec $\nu > 0$. Formellement, on a donc d'après (7.5) et (7.8)-(7.9) que y_u se comporte, au voisinage de ξ, comme la solution ϕ de

$$\Delta(\Delta\phi) = -\frac{\nu}{\gamma}\delta_\xi.$$

Les solutions fondamentales du bilaplacien pour $d = 2, 3$ donnent donc qu'au voisinage de ξ,

$$y_u(x) - 1 \sim \frac{1}{8\pi}\frac{\nu}{\gamma}|x - \xi|^2 \ln|x - \xi| \quad \text{si } d = 2, \qquad y_u(x) - 1 \sim -\frac{1}{8\pi}\frac{\nu}{\gamma}|x - \xi| \quad \text{si } d = 3.$$

L'hypothèse (7.13) est donc satisfaite avec $\alpha = 2$ si $d = 2$ et $\alpha = 1$ si $d = 3$.

Remarque 7.3. Modulons la conclusion apparemment négative à laquelle on arrive à la suite de cet exemple pour $d = 2$. Pour des problèmes non linéaires ou avec contraintes sur la commande, on peut prendre pour espace de commande $u \in L^s(\Omega)$ avec $s \in (2, +\infty]$. Ceci permet d'augmenter la régularité de z_v au numérateur du terme de courbure (7.12), ce qui pourrait permettre de faire des hypothèses au voisinage de l'ensemble de contact du type (7.13) réalistes et assurant la finitude du terme de courbure. Pour obtenir des conditions du second ordre sans saut, il faudrait ensuite jouer avec les deux normes (L^s et L^2) comme on l'a fait pour les EDO (avec L^∞ et L^2). En particulier on aurait besoin d'un résultat de densité du cône critique dans L^s dans le cône critique dans L^2.

7.1.3 Cas d'un nombre infini de points de contact isolés

La condition nécessaire du second ordre du chapitre 1 suppose un nombre fini d'arcs frontière et points de contact isolés, hypothèse restrictive pour des contraintes d'ordre élevé ($q \geq 3$) qui peuvent typiquement exhiber un nombre infini de points de contact isolés. Une question qui se pose alors est la suivante : peut-on généraliser le calcul du terme de courbure effectué dans la section 1.3 au cas où l'on a un nombre infini de points de contact isolés ? En particulier, est-ce que la somme qui intervient dans le terme de courbure dans (1.34) est finie lorsque l'on somme sur un nombre infini (mais dénombrable, puisque η a un nombre au plus dénombrable de points de discontinuité) de points de contact isolés essentiels réductibles ? Remarquons que si l'on a un nombre infini de points de contact isolés $(\tau_n)_{n\in\mathbb{N}^*}$, alors notant $\nu_n := [\eta(\tau_n)]$ le saut associé, comme $\mathrm{d}\eta$ est une mesure finie sur $[0, T]$, on en déduit que nécessairement, $\nu_n \to 0$.

Regardons ce que donne le calcul du terme de courbure dans l'exemple de Robbins [118], où la solution est calculée explicitement pour une contrainte sur l'état d'ordre trois ayant une infinité de points de contact isolés réductibles suivie d'un arc frontière. Les calculs sont

rappelés dans l'annexe (section 7.2.2). Dans cet exemple, l'auteur montre que la distance entre deux points de contact isolés successifs est géométrique,

$$\tau_{n+1} - \tau_n = ar^n,$$

avec $a > 0$ une constante dépendant des conditions initiales et $0 < r < 1$. On peut alors montrer que ν_n est de l'ordre de r^n, alors que le terme au dénominateur du terme de courbure $g^{(2)}(y(\tau_n))$ est de l'ordre de r^{4n} (voir (7.60) et la remarque 7.6). Ainsi, le terme de courbure est de l'ordre de (à une constante multiplicative près)

$$\sum_{n\in\mathbb{N}^*} \left(\frac{1}{r^3}\right)^n (g_y^{(1)}(y_u(\tau_n))z_v(\tau_n))^2, \tag{7.15}$$

où z_v est solution de l'équation d'état linéarisée (1.8), avec pour tout v dans le cône critique $C(u)$ dans L^∞

$$g_y(y_u(\tau_n))z_v(\tau_n) = 0 \quad \forall\, n \in \mathbb{N}^*. \tag{7.16}$$

Comme $\frac{\mathrm{d}}{\mathrm{d}t} g_y(y_u(t))z_v(t) = g_y^{(1)}(y_u(t))z_v(t)$, la fonction $g_y^{(1)}(y_u)z_v$ admet un zero s_n dans l'intervalle (τ_{n-1}, τ_n) pour tout $n \geq 2$. Ainsi

$$\begin{aligned} g_y^{(1)}(y_u(\tau_n))z_v(\tau_n) &= \int_{s_n}^{\tau_n} g_y^{(2)}(y_u(\theta))z_v(\theta)\mathrm{d}\theta \;\leq\; |\tau_n - \tau_{n-1}| \sup_{[\tau_{n-1},\tau_n]} |g_y^{(2)}(y_u)z_v| \\ &\leq ar^{n-1} \sup_{[\tau_{n-1},\tau_n]} |g_y^{(2)}(y_u)z_v|. \end{aligned}$$

Par le même raisonnement, la fonction continue $g_y^{(2)}(y_u)z_v = \frac{\mathrm{d}}{\mathrm{d}t} g_y^{(1)}(y_u(t))z_v(t)$ admet un zéro σ_n dans l'intervalle (s_{n-1}, s_n) pour tout $n \geq 2$. Ainsi, comme $\sigma_n \in (\tau_{n-1}, \tau_{n+1})$,

$$\begin{aligned} \sup_{[\tau_{n-1},\tau_n]} |g_y^{(2)}(y_u)z_v| &\leq \sup_{t\in[\tau_{n-1},\tau_{n+1}]} |\int_{\sigma_n}^{t} (g_y^{(3)}(u(\theta), y_u(\theta))z_v(\theta) + g_u^{(3)}(u(\theta), y_u(\theta))v(\theta))\mathrm{d}\theta| \\ &\leq |\tau_{n+1} - \tau_{n-1}| \|g_y^{(3)}(u, y_u)z_v + g_u^{(3)}(u, y_u)v\|_\infty \\ &\leq Car^{n-1}\|v\|_\infty, \end{aligned}$$

où C est une constante strictement positive. Ainsi, pour tout $n \geq 2$ on obtient que

$$(g_y^{(1)}(y_u(\tau_n))z_v(\tau_n))^2 \;\leq\; C^2(a/r)^4 r^{4n}\|v\|_\infty^2.$$

La série (7.15) a un terme générique en r^n, et est donc convergente, pour tout $v \in C(u)$.

Pour conclure le calcul du terme de courbure il resterait à calculer la contribution du point d'accumulation $T_c = \lim_{n\to+\infty} \tau_n$ qui marque l'entrée sur un arc frontière. Ceci s'avère plus délicat car numérateur et dénominateur de (1.36) s'annulent une infinité de fois dans tout voisinage de T_c. On ne peut donc pas utiliser de développement de Taylor. Sur l'exemple de Robbins, on peut montrer que pour $t \in [\tau_1, T_c)$,

$$g(y(t)) = -(1 - \theta(t))\frac{(t - T_c)^6}{6!} \tag{7.17}$$

où θ est une fonction de classe C^4 sur $[\tau_1, T_c)$, C^∞ sur chaque arc intérieur (τ_n, τ_{n+1}), telle que $\theta(\tau_n) = 1$ pour tout n et $0 < \theta(t) < 1$ pour tout $t \in (\tau_n, \tau_{n+1})$. Cette relation est démontrée dans la section 7.2.3 et une expression explicite de θ est fournie.

Étant donné $v \in C(u)$ — donc vérifiant (7.16) — on en déduit que la contribution de T_c dans le terme de courbure est finie si et seulement si il existe une constante $C > 0$ et un voisinage ϑ de T_c tels que pour tout $t \in \vartheta$,

$$(g_y(y(t))z_v(t))_+^2 \ \leq \ C(1-\theta(t))|t-T_c|^6. \tag{7.18}$$

Si ceci est vérifié pour tout $v \in C(u)$, alors comme $[\eta(T_c)] = 0$, on obtiendrait que le terme de courbure est donné par

$$\sigma(\eta, T_K^{2,i}(G(u), DG(u)v)) = \sum_{n\in\mathbb{N}^*} \nu_n \frac{(g_y^{(1)}(y(\tau_n))z_v(\tau_n))^2}{g^{(2)}(y(\tau_n))}.$$

Il resterait encore à étendre le lemme 1.17 (dont les arguments ne s'appliquent plus) de densité du cône critique dans L^∞ dans le cône critique dans L^2 pour étendre la condition nécessaire (1.34) aux v dans L^2. Ainsi on obtiendrait, avec la condition suffisante (1.81) qui fait intervenir un nombre fini, mais arbitrairement grand, de points de contact isolés essentiels réductibles, des conditions nécessaires ou suffisantes du second ordre "arbitrairement" proches sur cet exemple particulier ayant un nombre infini de points de contact isolés. Dans le cas contraire (i.e. (7.18) n'est pas satisfait), le terme de courbure est égal à $-\infty$ et donc la condition nécessaire du second ordre de Kawasaki n'apporte pas d'information.

7.1.4 Cas de contraintes linéairement dépendantes

On se place dans le cadre du chapitre 4. Considérons l'exemple suivant, où la dynamique est donnée par

$$\left\{ \begin{array}{lcll} \dot{y}_1 & = & u_1, & y_1(0) = y_1^0 \\ \dot{y}_2 & = & y_1 - u_2, & y_2(0) = y_0^2 \\ \dot{y}_3 & = & u_2, & y_3(0) = y_3^0 \end{array} \right. \tag{7.19}$$

et les contraintes sur l'état par

$$\left\{ \begin{array}{l} g_1(y(t)) \ = \ y_2(t) - 1 \ \leq \ 0, \\ g_2(y(t)) \ = \ y_3(t) - 1 \ \leq \ 0. \end{array} \right. \tag{7.20}$$

Chaque contrainte est, séparément, d'ordre 1 et régulière, mais l'hypothèse d'indépendance linéaire (4.30) n'est pas satisfaite si ces deux contraintes sont actives en même temps car $G_u^{(q)}(u(t), y(t)) = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}$ n'est pas de rang plein (mais il est de *rang constant* égal à 1). Or sur un arc où les deux contraintes sont actives, on a (par exemple) $u_2(t) = 0$ fixé par $g_2(y(t)) = 0$, et alors $g_1^{(1)}(u(t), y(t)) = y_1(t) - u_2(t) = y_1(t)$ et g_1 se comporte comme une contrainte du second ordre, régulière car $g_{1,u}^{(2)}(u(t), y(t)) = \begin{pmatrix} 1 & 0 \end{pmatrix}^\top$, et le système est régulier, dans le sens où $\begin{pmatrix} g_{1,u}^{(2)}(u(t), y(t)) \\ g_{2,u}^{(1)}(u(t), y(t)) \end{pmatrix} = I_2$ est de rang plein. La commande permettant de réaliser $g_1(y(t)) = 0$ et $g_2(y(t)) = 0$ sur un arc est alors parfaitement déterminée, ici donnée par $u_1(t) = 0$ et $u_2(t) = 0$.

Les cas de contraintes mixtes linéairement dépendantes ou de contraintes sur l'état et contraintes mixtes linéairement dépendantes peuvent aussi se rencontrer. Il semble alors restrictif de définir l'ordre des différentes contraintes indépendamment les unes des autres, puisqu'en "augmentant l'ordre", on pourrait se ramener à un système régulier. On pourrait alors

affaiblir l'hypothèse d'indépendance linéaire, et la remplacer par une hypothèse de type "rang constant", généralisant l'exemple ci-dessus.

Plus précisément, supposons que l'on ait deux contraintes sur l'état g_1 et g_2, chacune d'entre elles étant d'ordre $q_1 = q_2 = 1$ et (séparément) régulière, i.e.

$$\exists\, \gamma, \varepsilon > 0, \qquad |g^{(1)}_{i,u}(u(t), y(t))| \geq \gamma \quad \text{pour tout } t : \operatorname{dist}(t, \Delta_i) < \varepsilon, \qquad \forall\, i = 1, 2, \tag{7.21}$$

avec $\Delta_i := \{t \in [0, T] : g_i(y(t)) = 0\}$, et satisfaisant l'hypothèse de *rang constant*

$$\exists\, \varepsilon > 0, \qquad \operatorname{rang} \begin{pmatrix} g^{(1)}_{1,u}(u(t), y(t)) \\ g^{(1)}_{2,u}(u(t), y(t)) \end{pmatrix} = 1, \qquad \text{pour tout } t : \operatorname{dist}(t, \Delta_1 \cap \Delta_2) < \varepsilon. \tag{7.22}$$

Supposons de plus que u est continue. Alors, d'après le théorème des fonctions implicites, pour tout $t \in \Delta_1 \cap \Delta_2$, il existe $\delta > 0$ tel que l'on peut localement sur $(t - \delta, t + \delta)$ partitionner u et g en $u = (u_1, \bar{u}) \in \mathbb{R} \times \mathbb{R}^{m-1}$ et $g = (g_1, g_2)$ de sorte que $g^{(1)}_{1,u_1}$ soit (localement) inversible, et il existe une fonction régulière $\Phi : \mathbb{R}^{m-1} \times \mathbb{R}^n \supset \mathcal{V} \to \mathbb{R}$ telle que, localement sur $(t - \delta, t + \delta)$, la relation $g^{(1)}_1(u(t), y(t)) = 0$ est équivalente à $u_1(t) = \Phi(\bar{u}(t), y(t))$. Les décompositions $g = (g_1, g_2)$ et $u = (u_1, \bar{u})$ ne sont pas uniques. D'après (7.22), la fonction $g^{(1)}_2((\Phi(\bar{u}, y), \bar{u}), y)$ ne peut dépendre de $\bar{u}$, ainsi on peut poser

$$\tilde{g}^{(1)}_2(y) \;:=\; g^{(1)}_2((\Phi(\bar{u}, y), \bar{u}), y). \tag{7.23}$$

On définit de même sa dérivée temporelle $\tilde{g}^{(2)}_2(u, y) := \tilde{g}^{(1)}_{2,y}(y) f(u, y)$. Si l'on suppose qu'il existe $\gamma, \varepsilon > 0$ tel que pour tout $t \in \Delta_1 \cap \Delta_2$,

$$\left| \begin{pmatrix} g^{(1)}_{1,u}(u(t), y(t)) \\ \tilde{g}^{(2)}_{2,u}(u(t), y(t)) \end{pmatrix}^{\top} \xi \right| \geq \gamma |\xi| \quad \text{pour tout } \xi \in \mathbb{R}^2 \tag{7.24}$$

alors on peut étendre la formulation alternative et le résultat de la proposition 4.13, en considérant dans l'analyse g_2 comme une contrainte du second ordre. Ces hypothèses peuvent être étendues au cas général, voir par exemple [23, p.132–134].

Remarque 7.4. Noter que si les contraintes mixtes sont linéairement indépendantes et (4.44) satisfaite, la proposition 4.8 est toujours valable, i.e. u est continue sur $[0, T]$. En particulier, dans le cas de deux contraintes du premier ordre g_1 et g_2 (et pas de contraintes mixtes) vérifiant (7.21)-(7.22), les multiplicateurs η_1 et η_2 associés aux contraintes du premier ordre g_1 et g_2 ne sont plus continus sur $[0, T]$, mais la relation (4.47) montre que

$$g^{(1)}_{1,u}(u, y)\eta_1 + g^{(1)}_{2,u}(u, y)\eta_2 \tag{7.25}$$

est continu sur $[0, T]$, et donc les sauts de η_1 et η_2 sont liés.

Pour étendre l'algorithme de tir au cas de contraintes linéairement dépendantes, une difficulté est d'écrire soigneusement les conditions de jonction entre arcs et les conditions supplémentaires, en particulier parce que les points de jonction de deux contraintes linéairement dépendantes peuvent génériquement coïncider, comme le montre l'exemple 7.5 ci-après.

Exemple 7.5. Pour un paramètre μ réel au voisinage de zéro, on considère le problème $(\mathcal{P}^\mu)$

$$\min_{u,y} \int_0^T \left(\frac{u_1(t)^2 + u_2(t)^2}{2} + \mu y_3(t) \right) \mathrm{d}t + y_1(T) \quad \text{s.t. (7.19)-(7.20)} \tag{7.26}$$

avec $T = 6$, $y_1^0 = 5$, $y_2^0 = y_3^0 = 0$.

Les solutions et multiplicateurs ont été tracés pour $\mu = 0$, $\mu = 0.1$ et $\mu = -0.1$ sur les figures 7.1, 7.2 et 7.3, respectivement. Ces solutions ont été obtenus par un algorithme de tir, étendu à cet exemple particulier, basé sur la résolution de la condition nécessaire du premier ordre. La structure des trajectoires a été déterminée par tâtonnement de façon à satisfaire les conditions supplémentaires non prises en compte dans l'algorithme de tir. Par convexité du problème (7.26), la condition du premier ordre est aussi suffisante, et comme de plus le coût est fortement convexe sur l'ensemble admissible, la solution est unique. La quantité (7.25), égale sur cet exemple à $\eta_2 - \eta_1$, a été également tracée sur la figure 7.3(b) et on vérifie que cette quantité est bien continue.

Pour $\mu = 0$, la contrainte g_1 a un arc frontière $[\tau_1, \tau_2]$ et la contrainte g_2 a également un arc frontière $[\tau_2, \tau_3]$, dont le point d'entrée coïncide avec le point de sortie de g_1. Le multiplicateur $\mathrm{d}\eta_2$ est une mesure de Dirac concentrée en τ_2.

Pour $\mu > 0$, g_1 a toujours un arc frontière $[\tau_1, \tau_2]$, mais g_2 a maintenant un point de contact isolé essentiel $\tau_0 \in (\tau_1, \tau_2)$. On constate que quand g_1 est active, la contrainte sur l'état du premier ordre g_2 se comporte bien comme une contrainte d'ordre 2.

Pour $\mu < 0$, g_1 et g_2 ont toutes les deux un arc frontière, et comme dans le cas $\mu = 0$, l'instant de sortie de g_1 coïncide avec l'instant d'entrée de g_2, ce qui nous amène à penser que pour des contraintes linéairement dépendantes, *les instants de jonctions peuvent génériquement coïncider*, dans le sens où le fait que les instants de jonctions coïncident est stable par petites perturbations des données.

Précisons les conditions de jonctions que nous avons utilisées dans l'algorithme de tir pour obtenir ces solutions. Les multiplicateurs alternatifs p^q, η^q sont définis comme dans la section 4.7.1. Les équations (4.146)-(4.150) sont résolues sur chaque arc, avec la condition de saut de l'adjoint alternatif p^q donnée par (4.154). Pour déterminer l'instant d'entrée τ_1 de g_1 et le paramètre de saut associé de p^q, on utilise les conditions classiques (4.152) (point d'entrée d'une contrainte du premier ordre). Pour déterminer le point de contact isolé τ_0 de g_2 sur l'arc frontière de g_1 (dans le cas $\mu \geq 0$) et le paramètre de saut associé de p^q, par analogie avec un point de contact isolé pour une contrainte d'ordre 2, on utilise les conditions (avec $\tilde{g}_2^{(1)}$ définie comme dans (7.23))

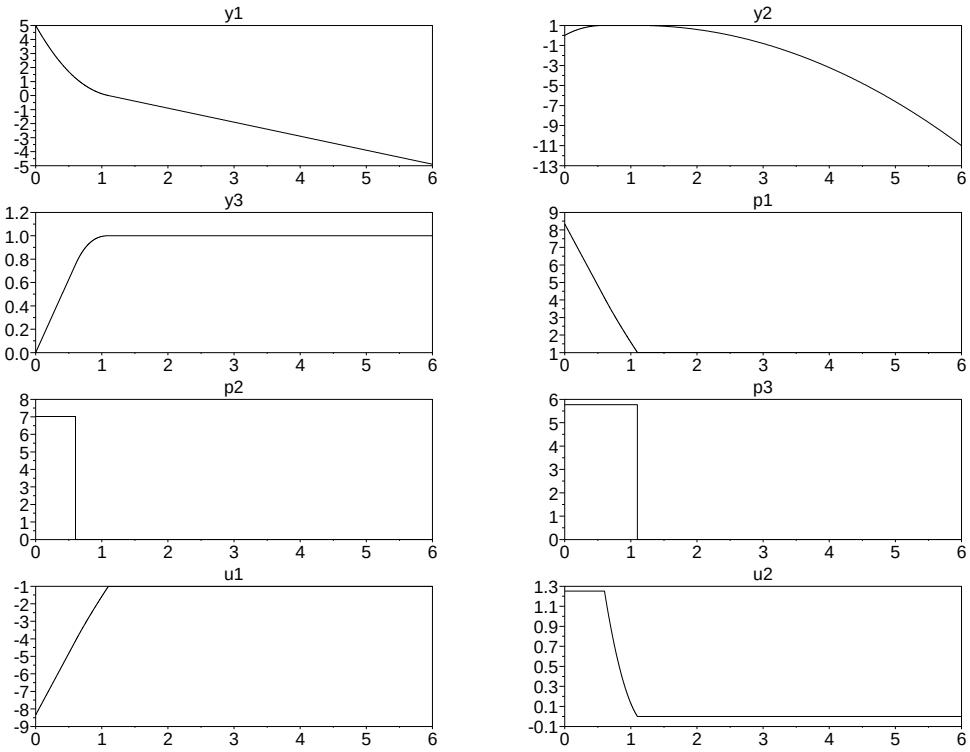
$$g_2(y(\tau_0)) = y_3(\tau_0) - 1 = 0, \qquad \tilde{g}_2^{(1)}(y(\tau_0)) = y_1(\tau_0) = 0.$$

Enfin, pour déterminer le point d'entrée τ_2 de g_2 qui est aussi le point de sortie de g_1 (dans le cas $\mu \leq 0$) et le paramètre de saut associé de p^q, on utilise les conditions

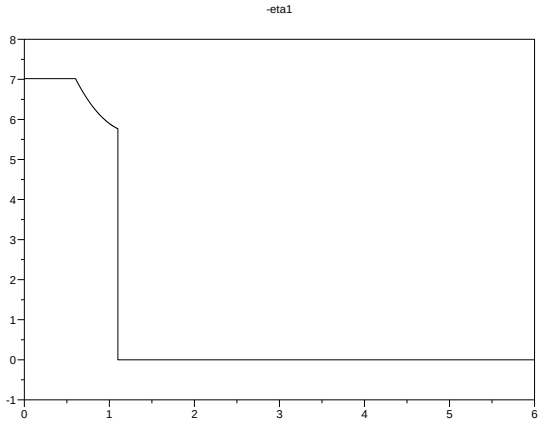
$$g_2(y(\tau_2)) = y_3(\tau_2) - 1 = 0, \qquad \tilde{g}_2^{(1)}(y(\tau_2)) = y_1(\tau_2) = 0. \tag{7.27}$$

Noter que cette dernière condition est bien équivalente à la continuité de u en τ_2, puisque l'on a $[u_1(\tau_2)] = -[p_1^q(\tau_2)] = 0$, $u_2(\tau_2^+) = 0$ car on est sur un arc frontière de g_2 et $u_2(\tau_2^-) = y_1(\tau_2) = 0$ puisque qu'on est sur un arc frontière de g_1 et $y_1(\tau_2) = 0$ par (7.27).

Un raisonnement analogue à celui de la proposition 4.29 montre que, sur cet exemple, si les conditions supplémentaires (4.166), (4.168), (4.169), (4.171), (4.173) et (4.175) sont satisfaites, alors on a un point stationnaire, qui correspond nécessairement à l'unique solution du problème. Les multiplicateurs η_1 et η_2 du principe du minimum ont été reconstitués sur les figures 7.1, 7.2 et 7.3 (on a tracé $-\eta_1$ et $-\eta_2$). À la différence de contraintes du premier ordre linéairement indépendantes, les conditions supplémentaires (4.169), (4.171), (4.173) peuvent être satisfaites avec une inégalité stricte pour des contraintes du premier ordre linéairement dépendantes. C'est ce que l'on observe sur les figures 7.1, 7.2 et 7.3 car η_1 et η_2 présentent

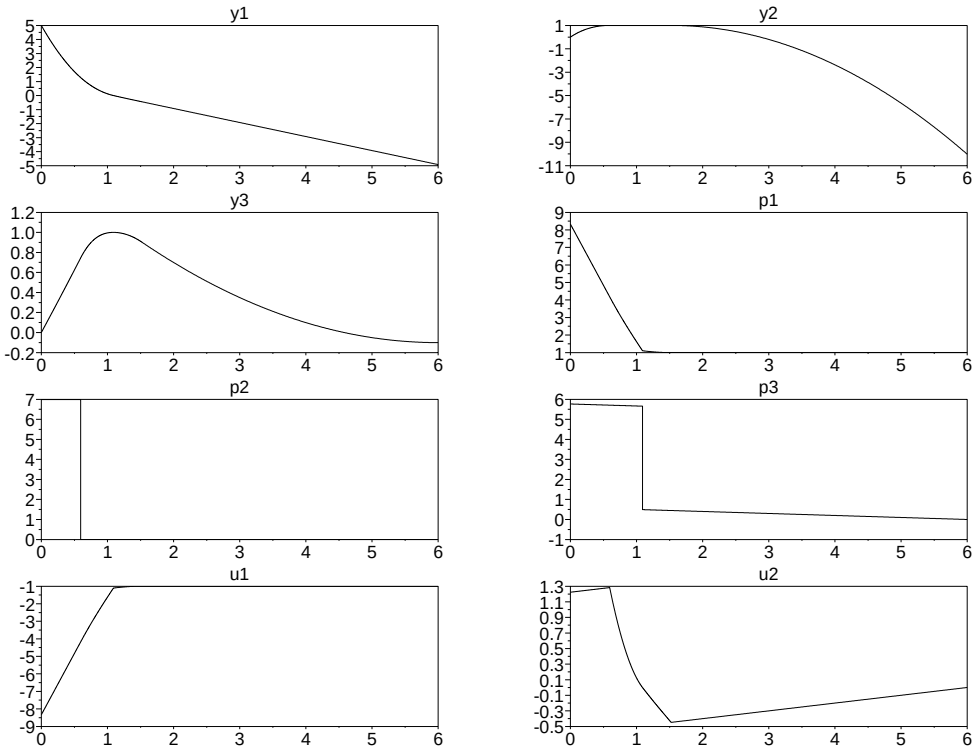


(a) État, état adjoint alternatif et commande

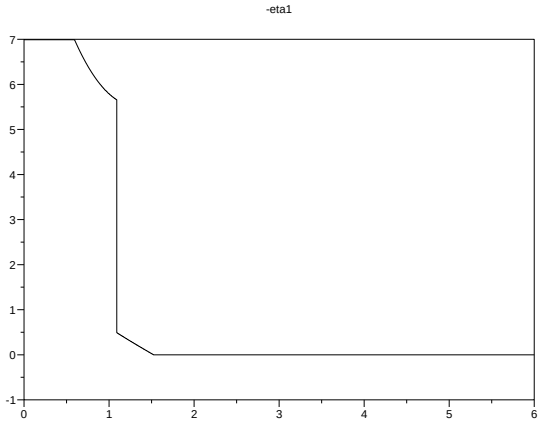


(b) Multiplicateur $-\eta_1$ associé à la contrainte g_1

FIG. 7.1 – Solution et multiplicateurs pour $\mu = 0$.

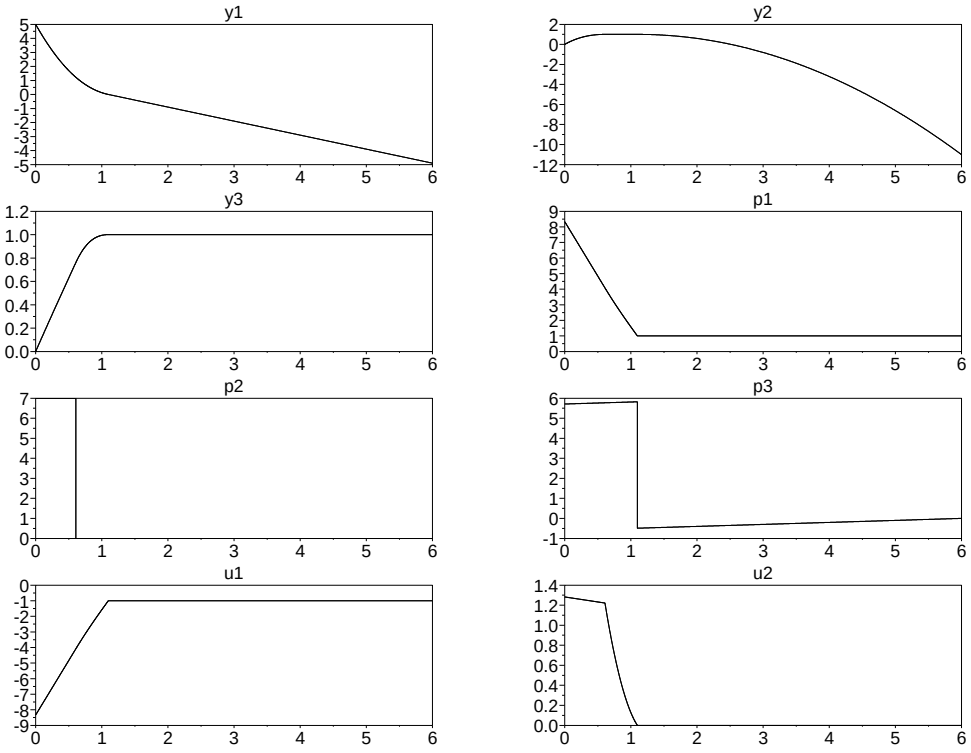


(a) État, état adjoint alternatif et commande

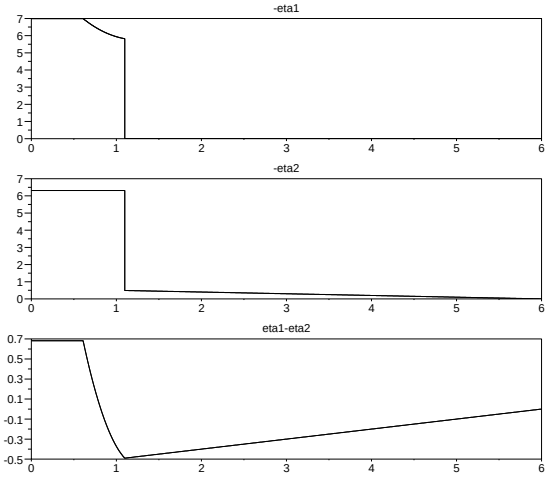


(b) Multiplicateur $-\eta_1$ associé à la contrainte g_1

FIG. 7.2 – Solution et multiplicateurs pour $\mu = 0.1$.



(a) État, état adjoint alternatif et commande



(b) Multiplicateurs $-\eta_1$ et $-\eta_2$ associés respectivement aux contraintes g_1 et g_2 et différence $\eta_1 - \eta_2$.

FIG. 7.3 – Solution et multiplicateurs pour $\mu = -0.1$.

des sauts aux instants où les deux contraintes sont actives en même temps. Enfin, on constate que les contraintes g_1 et g_2 ne peuvent pas être actives en même temps (sur un intervalle de temps de longueur non nulle) sur cet exemple pour μ au voisinage de zéro car cela requerrait $\tilde{g}_2^{(2)}(u,y) = u_1 = 0$ or u_1 ne s'annule pas.

7.2 Annexe

7.2.1 Preuve de la proposition 7.1

Démonstration. On commence par montrer que si (7.3) a une solution bornée X sur $[0,T]$, alors (7.2) est satisfaite. La preuve (calculatoire) de ce point est inspirée par [99, 89]. Comme l'équation de Riccati (7.3) est symétrique et la condition finale $\phi_{yy}(y(T))$ aussi, on en déduit que X est symétrique. En effet, si X est solution de (7.3), alors $X^\top$ aussi et donc par unicité de la solution (théorème de Cauchy-Lipschitz) on en déduit que $X = X^\top$. Par la formule d'intégration par parties dans BV, utilisant que $z_v(0) = 0$ et $X(T) = \phi_{yy}(y(T))$, on a

$$2\int_0^T z_v^\top X \dot{z}_v \mathrm{d}t + \int_0^T z_v^\top \mathrm{d}X(t) z_v \;=\; z_v^\top(T)\phi_{yy}(y(T))z_v(T).$$

Ajoutant le terme nul $2\int_0^T z_v^\top X(f_y z_v + f_u v - \dot{z}_v)$ à la fonction quadratique $Q(v)$ (on rappelle que Q est donnée par (4.135) où $\mathcal{T}_{to}^{i,ess}$ est remplacé par $\mathcal{T}_{red}^i$), on en déduit que

$$\begin{aligned} Q(v) \;=\; & \int_0^T (z_v^\top((H_{uy}^0)^\top + Xf_u)v + v^\top(H_{uy}^0 + f_u^\top X)z_v + v^\top H_{uu}^0 v)\mathrm{d}t \\ & + \int_0^T z_v^\top((H_{yy}^0 + Xf_y + f_y^\top X)\mathrm{d}t + \mathrm{d}X(t))z_v \\ & + \int_0^T z_v^\top(\sum_{i=1}^r g_{i,yy}\mathrm{d}\eta_i(t) - \sum_{i=1}^r \sum_{\tau\in\mathcal{T}_{red}^i} \nu_\tau^i \frac{(g_{i,y}^{(1)})^\top g_{i,y}^{(1)}}{g_i^{(2)}}\delta_\tau(t))z_v. \end{aligned}$$

Introduisons la variable artificielle $\varpi(t) \in \mathbb{R}^{|\overset{o}{I}(t)|}$, et posons, pour simplifier l'écriture,

$$\begin{aligned} \mathrm{d}\mathcal{A}(t) \;&:=\; (H_{yy}^0 + Xf_y + f_y^\top X)\mathrm{d}t + \mathrm{d}X(t) + \sum_{i=1}^r g_{i,yy}\mathrm{d}\eta_i(t) - \sum_{i=1}^r \sum_{\tau\in\mathcal{T}_{red}^i} \nu_\tau^i \frac{(g_{i,y}^{(1)})^\top g_{i,y}^{(1)}}{g_i^{(2)}}\delta_\tau(t), \\ \mathcal{C}(t) \;&:=\; \begin{pmatrix} H_{uy}^0 + f_u^\top X \\ G^{(q)}_{\overset{o}{I}(t),y} \end{pmatrix}, \qquad \mathcal{D}(t) \;:=\; \begin{pmatrix} H_{uu}^0 & (G^{(q)}_{\overset{o}{I}(t),u})^\top \\ G^{(q)}_{\overset{o}{I}(t),u} & 0 \end{pmatrix}. \end{aligned}$$

Alors on a

$$\begin{aligned} Q(v) \;=\; & \int_0^T \begin{pmatrix} z_v \\ \begin{pmatrix} v \\ \varpi \end{pmatrix} \end{pmatrix}^\top \begin{pmatrix} \mathrm{d}\mathcal{A}(t) & \mathcal{C}^\top \mathrm{d}t \\ \mathcal{C}\mathrm{d}t & \mathcal{D}\mathrm{d}t \end{pmatrix} \begin{pmatrix} z_v \\ \begin{pmatrix} v \\ \varpi \end{pmatrix} \end{pmatrix} \\ & - 2\int_0^T \varpi(t)^\top (G^{(q)}_{\overset{o}{I}(t),y} z_v + G^{(q)}_{\overset{o}{I}(t),u} v)\mathrm{d}t. \end{aligned}$$

Pour tout $v \in L^2$ satisfaisant (7.1), on a que $G^{(q)}_{\substack{o\\ I(t),y}} z_v + G^{(q)}_{\substack{o\\ I(t),u}} v = 0$ p.p., et donc, comme d'après l'équation de Riccati (7.3) $\mathrm{d}\mathcal{A} = \mathcal{C}^\top \mathcal{D}^{-1}\mathcal{C}\mathrm{d}t$, on obtient finalement

$$Q(v) \;=\; \int_0^T \left(\mathcal{D}^{-1}\mathcal{C}z_v + \begin{pmatrix} v \\ \varpi \end{pmatrix} \right)^\top \mathcal{D} \left(\mathcal{D}^{-1}\mathcal{C}z_v + \begin{pmatrix} v \\ \varpi \end{pmatrix} \right) \mathrm{d}t,$$

et ce, quelque soit $v \in L^2$ satisfaisant (7.1) et quelque soit la variable $\varpi(t) \in \mathbb{R}^{|\overset{o}{I}(t)|}$. Noter que par les hypothèses (4.30) et (4.43), la matrice $\mathcal{D}(t)$ est bien inversible. Choisissons $\varpi(t)$ égal à l'opposé des $|\overset{o}{I}(t)|$ dernières composantes du vecteur $\mathcal{D}^{-1}(t)\mathcal{C}(t)z_v(t) \in \mathbb{R}^{m+|\overset{o}{I}(t)|}$, et notons $(\mathcal{D}^{-1}(t)\mathcal{C}(t)z_v(t))|_{1:m}$ les composantes 1 à m de ce vecteur. On a alors

$$Q(v) = \int_0^T \left(\; (\mathcal{D}^{-1}\mathcal{C}z_v)|_{1:m} + v \; \right)^\top H^0_{uu} \left(\; (\mathcal{D}^{-1}\mathcal{C}z_v)|_{1:m} + v \; \right) \;\geq\; 0,$$

pour tous les $v \in L^2$ satisfaisant (7.1) puisque H^0_{uu} est uniformément définie positive par (4.43). Pour obtenir l'uniforme positivité de Q (et donc la stricte positivité) sur l'espace défini par (7.1), il suffit de remplacer dans les calculs précédents H^0_{uu} par $H^0_{uu} - \varepsilon I_m$, avec I_m la matrice identité de taille m, qui reste définie positive uniformément sur $[0,T]$ pour $\varepsilon > 0$ suffisamment petit. Pour ε suffisamment petit, l'équation de Riccati obtenue en remplaçant H^0_{uu} par $H^0_{uu} - \varepsilon I_m$ dans (7.3) admet aussi une solution bornée sur $[0,T]$ (résultat standard de perturbation des équations différentielles, voir par exemple [99, p.176]). On obtient ainsi de même que $Q(v) - \varepsilon\|v\|_2^2 \geq 0$ pour tout $v \in L^2$ satisfaisant (7.1), impliquant (7.2).

Montrons maintenant la réciproque. Pour $s \in [0,T)$, soit Q_s la forme quadratique définie sur $L^2(s,T;\mathbb{R}^m)$ comme Q mais en intégrant sur $[s,T]$, i.e.

$$\begin{aligned} Q_s(v) \;&:=\; \int_s^T (v^\top H^0_{uu} v + 2v^\top H^0_{uy} z_{v,s})\mathrm{d}t + z_{v,s}(T)^\top \phi_{yy}(y(T)) z_{v,s}(T) \\ &+\; \int_s^T z_{v,s}^\top (H^0_{yy}\mathrm{d}t + \sum_{i=1}^r g_{i,yy}\mathrm{d}\eta_i(t) - \sum_{i=1}^r \sum_{\tau \in \mathcal{T}^i_{red}} \nu^i_\tau \frac{(g^{(1)}_{i,y})^\top g^{(1)}_{i,y}}{g^{(2)}_i}\delta_\tau(t)) z_{v,s} \end{aligned}$$

où $z_{v,s}$ est la solution de $\dot{z}_{v,s} = f_y z_{v,s} + f_u v$ p.p. sur $[s,T]$ avec condition initiale $z_{v,s}(s) = 0$, et soit le problème défini par

$$(\mathcal{P}^s) \qquad \min_{v \in L^2(s,T;\mathbb{R}^m)} Q_s(v) \quad \text{s.t.} \quad G^{(q)}_{\substack{o\\ I(t),u}} v + G^{(q)}_{\substack{o\\ I(t),y}} z_{v,s} = 0 \text{ pour p.p. } t \in [s,T].$$

Alors (7.2) implique que

$$Q_s(v) > 0, \quad \forall\, v \in L^2(s,T;\mathbb{R}^m) \setminus \{0\} \text{ vérifiant } G^{(q)}_{\substack{o\\ I(t),u}} v + G^{(q)}_{\substack{o\\ I(t),y}} z_{v,s} = 0 \text{ p.p. sur } [s,T]. \quad (7.28)$$

En effet, soit $v \in L^2(s,T;\mathbb{R}^m) \setminus \{0\}$ vérifiant la contrainte ci-dessus. Comme $z_{v,s}(s) = 0$, on peut prolonger v par zéro sur $[0,s)$ et on obtient alors $\bar{v} \in L^2(0,T;\mathbb{R}^m)$ tel que $(\bar{v}, z_{\bar{v}}) = 0$ sur $[0,s)$ et $(\bar{v}, z_{\bar{v}}) = (v, z_{v,s})$ sur $[s,T]$. Ainsi $\bar{v}$ vérifie (7.1), $\bar{v} \neq 0$ et $Q_s(v) = Q(\bar{v}) > 0$ d'après (7.2). Ceci prouve (7.28). De plus, comme Q_s est, comme Q, une forme de Legendre, $(\mathcal{P}^s)$ a un coût fortement convexe semi-continu inférieurement sur son ensemble admissible convexe fermé, et donc (7.28) implique que $(v, z_{v,s}) = 0$ est l'unique solution et l'unique point stationnaire de $(\mathcal{P}^s)$.

Pour simplifier les notations, posons

$$\mathrm{d}\mathcal{M}(t) \;:=\; H^0_{yy}\mathrm{d}t + \sum_{i=1}^{r} g_{i,yy}\mathrm{d}\eta_i(t) - \sum_{i=1}^{r}\sum_{\tau\in\mathcal{T}^i_{red}} \nu^i_\tau \frac{(g^{(1)}_{i,y})^\top g^{(1)}_{i,y}}{g^{(2)}_i}\delta_\tau(t).$$

Soit $(v, z = z_{v,s})$ un point stationnaire de $(\mathcal{P}^s)$. Ce dernier vérifie, notant π et λ les multiplicateurs associés respectivement à l'équation d'état et à la contrainte mixte de $(\mathcal{P}^s)$,

$$\begin{aligned}
-\mathrm{d}\pi &= \mathrm{d}\mathcal{M}(t)z + ((H^0_{uy})^\top v + f_y^\top \pi + (G^{(q)}_{o_{I(t),y}})^\top\lambda)\mathrm{d}t, \qquad \pi(T) \;=\; \phi_{yy}(y(T))z(T),\\
0 &= H^0_{uu}v + H^0_{uy}z + f_u^\top\pi + (G^{(q)}_{o_{I(t),u}})^\top\lambda,\\
0 &= G^{(q)}_{o_{I(t),u}} v + G^{(q)}_{o_{I(t),y}} z.
\end{aligned}$$

Les deux dernières équations permettent d'exprimer v et λ en fonction de z et π

$$\begin{pmatrix} v \\ \lambda \end{pmatrix} \;=\; -\mathcal{D}(t)^{-1}\begin{pmatrix} H^0_{uy} & f_u^\top \\ G^{(q)}_{o_{I(t),y}} & 0 \end{pmatrix}\begin{pmatrix} z \\ \pi \end{pmatrix},$$

avec la matrice $\mathcal{D}(t)$ définie plus haut, et on en déduit que la dynamique état-état adjoint est donnée par

$$\begin{aligned}
\begin{pmatrix} \dot z\mathrm{d}t \\ -\mathrm{d}\pi \end{pmatrix} &= \Bigg[\begin{pmatrix} f_y\mathrm{d}t & 0 \\ \mathrm{d}\mathcal{M}(t) & f_y^\top\mathrm{d}t \end{pmatrix}\\
&\qquad - \begin{pmatrix} f_u & 0 \\ (H^0_{uy})^\top & (G^{(q)}_{o_{I(t),y}})^\top \end{pmatrix}\mathcal{D}(t)^{-1}\begin{pmatrix} H^0_{uy} & f_u^\top \\ G^{(q)}_{o_{I(t),y}} & 0 \end{pmatrix}\mathrm{d}t\Bigg]\begin{pmatrix} z \\ \pi \end{pmatrix}.
\end{aligned} \tag{7.29}$$

Une solution (z,π) du système ci-dessus sur $[s,T]$ avec condition initiale $(z(s),\pi(s)) = (x,q)$ est donc un point stationnaire de $(\mathcal{P}^s)$ ssi $(\pi(T) - \phi_{yy}(y(T))z(T), x) = 0$.

Soit F l'application $\mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}^n\times\mathbb{R}^n$, $(x,q)\mapsto(\pi(T)-\phi_{yy}(y(T))z(T),x)$ où (z,π) est la solution de (7.29) sur $[s,T]$ avec la condition initiale $(z(s),\pi(s)) = (x,q)$. Cette application est linéaire, et, puisque $(\mathcal{P}^s)$ admet un unique point stationnaire, elle est inversible. Ceci implique que pour tout $x\in\mathbb{R}^n$, il existe un unique $q\in\mathbb{R}^n$ tel que $\pi(T) - \phi_{yy}(y(T))z(T) = 0$, où (z,π) est la solution de (7.29) sur $[s,T]$ avec la condition initiale $(z(s),\pi(s)) = (x,q)$, et q est une fonction linéaire de x. Il existe donc une matrice $X(s)$ telle que $q = X(s)x$. Ce raisonnement étant valide pour tout s, on a que pour tout $x\in\mathbb{R}^n$ et tout $s\in[0,T]$, l'unique solution (z,π) de (7.29) sur $[s,T]$ vérifiant $z(s) = x$ et $\pi(T) = \phi_{yy}(y(T))z(T)$ est telle que

$$\pi(t) = X(t)z(t), \qquad \forall\, t\in[s,T]. \tag{7.30}$$

En particulier, en $t = T$ on obtient que $X(T) = \phi_{yy}(y(T))$. Reportant (7.30) dans la dynamique de π, on trouve

$$\begin{aligned}
-\mathrm{d}X(t)z &= X\dot z\mathrm{d}t + \mathrm{d}\mathcal{M}(t)z + ((H^0_{uy})^\top v + f_y^\top Xz + (G^{(q)}_{o_{I(t),y}})^\top\lambda)\mathrm{d}t\\
&= (Xf_y\mathrm{d}t + f_y^\top X\mathrm{d}t + \mathrm{d}\mathcal{M}(t))z + \begin{pmatrix} Xf_u + (H^0_{uy})^\top & (G^{(q)}_{o_{I(t),y}})^\top \end{pmatrix}\begin{pmatrix} v \\ \lambda \end{pmatrix}\\
&= \left[(Xf_y + f_y^\top X)\mathrm{d}t + \mathrm{d}\mathcal{M}(t) - \begin{pmatrix} H^0_{uy} + f_u^\top X^\top \\ G^{(q)}_{o_{I(t),y}} \end{pmatrix}^\top \mathcal{D}(t)^{-1}\begin{pmatrix} H^0_{uy} + f_u^\top X \\ G^{(q)}_{o_{I(t),y}} \end{pmatrix}\right] z.
\end{aligned}$$

Comme cette relation est vraie quelque soit $z(t) = x\in\mathbb{R}^n$ et quelque soit $t\in[0,T]$, on en déduit que X vérifie (7.3), ce qui achève la preuve. □

7.2.2 Rappel de l'exemple de Robbins [118]

L'exemple étudié par Robbins dans [118] est le suivant :

$$
\begin{aligned}
(\mathcal{P}_{rob}) \qquad \min_{(u,y)} \quad & \int_0^T \left(\frac{u(t)^2}{2} + y_1(t) \right) \mathrm{d}t, && (7.31)\\
\text{s.t.} \quad & \dot{y}_1(t) = y_2(t), \quad \dot{y}_2(t) = y_3(t), \quad \dot{y}_3(t) = u(t) \quad \text{p.p. } t \in [0,T], && (7.32)\\
& y_j(0) = y_j^0, \quad j = 1,2,3, && (7.33)\\
& -y_1(t) \leq 0 \qquad t \in [0,T]. && (7.34)
\end{aligned}
$$

On suppose que la condition initiale $y^0 = (y_1^0, y_2^0, y_3^0) \in \mathbb{R}^3$ est telle que

$$y_1^0 > 0. \tag{7.35}$$

Le coût de $(\mathcal{P}_{rob})$ est continu et fortement convexe sur son ensemble admissible qui est convexe fermé, ce qui garantit l'existence et l'unicité d'une solution (u, y) dans l'espace $L^2(0,T) \times H^1(0,T;\mathbb{R}^3)$, caractérisée par la condition d'optimalité du premier ordre donnée par le principe du minimum. De plus, la contrainte sur l'état est régulière d'ordre 3 et donc les multiplicateurs associés (p, η) sont uniques.

Supposons que la trajectoire ait un arc frontière avec point d'entrée régulier $\tau \in (0,T)$. Alors il est facile de voir qu'il n'est pas optimal de quitter l'arc, et on a donc un arc frontière du type $(\tau, T]$. Le Hamiltonien est uniformément fortement convexe par rapport à la commande et la contrainte sur l'état est régulière d'ordre $q = 3$. Les conditions de jonction au point d'entrée τ impliquent alors que les dérivées de y_1 sont continues jusqu'à l'ordre $2q - 1 = 5$, i.e.

$$y_1^{(j)}(\tau) = 0, \qquad j = 0, \ldots, 5. \tag{7.36}$$

Par ailleurs, sur un arc intérieur le principe du minimum implique que $y_1^{(6)} \equiv 1$. Ainsi, sur l'arc intérieur précédant τ, y_1 est donné par

$$y_1(t) = \frac{(t-\tau)^6}{6!}. \tag{7.37}$$

On voit que y_1 ne peut pas s'annuler sur $[0,\tau)$ et donc la trajectoire optimale est donnée par (7.37) sur $[0,\tau)$ et par $y_1(t) = 0$ sur $[\tau, T]$. La condition initiale (7.33) implique alors que

$$y_1^0 = \frac{\tau^6}{6!}, \qquad y_2^0 = -\frac{\tau^5}{5!}, \qquad y_3^0 = \frac{\tau^4}{4!}. \tag{7.38}$$

On ne peut clairement pas trouver un instant τ vérifiant (7.38) quelque soit la condition initiale $y^0 \in \mathbb{R}^3$ vérifiant (7.35), et donc sauf pour le cas particulier où la condition initiale se met sous la forme (7.38), la trajectoire optimale n'a pas d'arc frontière avec point d'entrée régulier.

Robbins a étudié la forme générale des solutions pour une condition initiale $y^0 \in \mathbb{R}^3$ quelconque vérifiant (7.35). Pour ceci, on laisse le temps final tendre vers l'infini et on s'intéresse au problème de commande optimale en horizon infini

$$
\begin{aligned}
(\mathcal{P}^{\infty}_{rob}) \qquad \min_{(u,y)} \quad & \int_0^{\infty} \left(\frac{u(t)^2}{2} + y_1(t) \right) \mathrm{d}t, && (7.39)\\
\text{s.t.} \quad & \dot{y}_1(t) = y_2(t), \quad \dot{y}_2(t) = y_3(t), \quad \dot{y}_3(t) = u(t) \quad \text{p.p. } t \in [0,+\infty), && (7.40)\\
& y_j(0) = y_j^0, \quad j = 1,2,3, && (7.41)\\
& -y_1(t) \leq 0 \qquad t \in [0,+\infty). && (7.42)
\end{aligned}
$$

Ce problème a une valeur finie car le système (7.40) étant commandable, on peut construire une trajectoire admissible s'annulant sur $[A, +\infty)$ pour $A > 0$. De plus, le coût (7.39) est fortement convexe semi-continu inférieurement sur l'ensemble admissible (7.40)-(7.42) et donc $(\mathcal{P}^{\infty}_{rob})$ admet une unique solution (u, y) avec $u \in L^2(0, \infty)$.

La condition d'optimalité du premier ordre pour les problèmes de commande optimale en horizon infini est analogue à celle connue en horizon fini, à l'exception de la condition finale de l'état adjoint qui est plus délicate à énoncer (voir [66] et [68, Rem. 4.4]). En revanche, si le Hamiltonien $H(\cdot, \cdot, p(t))$ est convexe et si la contrainte sur l'état g est convexe (ce qui est satisfait pour notre problème $(\mathcal{P}^{\infty}_{rob})$), alors le principe du minimum sur $(0, \infty)$ avec la condition limite suivante pour l'état adjoint

$$\liminf_{T \to +\infty} \; p(T)(\tilde{y}(T) - y(T)) \leq 0 \quad \text{pour toute trajectoire admissible } (\tilde{u}, \tilde{y}) \tag{7.43}$$

est une condition *suffisante* d'optimalité (voir [96, 122] et [68, Th. 8.4]) [2].

Dans le cas où les conditions initiales ne se mettent pas sous la forme (7.38), la trajectoire optimale ne présente pas d'arc frontière avec point d'entrée régulier mais un nombre *infini* de points de contact isolés. En effet, on ne peut pas avoir un nombre fini (éventuellement nul) de points de contact isolés sans arc frontière puisque sur un arc intérieur, y_1 est un polynôme de degré 6 et de coefficient dominant égal à 1, donc si on avait un arc intérieur de longueur infinie, on aurait $y_1(t) \to +\infty$ quand $t \to +\infty$, ce qui rendrait le critère infini et ne serait pas optimal. L'étude précédente montre qu'un arc frontière avec point d'entrée régulier, précédé d'un nombre fini (éventuellement nul) de points de contact isolés, est exclu si les conditions initiales ne se mettent pas sous la forme (7.38). Enfin, il n'est pas optimal de quitter un arc frontière. On a donc au plus un arc frontière sur la trajectoire optimale, et pas de point de sortie. La trajectoire optimale de $(\mathcal{P}^{\infty}_{rob})$ est donc composée d'un nombre *infini* de points de contacts isolés, suivis on non d'un arc frontière.

Un point de contact isolé essentiel correspondant à une discontinuité du multiplicateur η qui est une fonction croissante, il en existe un nombre dénombrable. On les note $(\tau_1, ..., \tau_n, ...)$. En point de contact isolé τ_n on vérifie nécessairement :

$$y_1(\tau_n) = 0 = \dot{y}_1(\tau_n), \qquad \ddot{y}_1(\tau_n) \geq 0. \tag{7.44}$$

On a en fait $\ddot{y}_1(\tau_n) > 0$ pour tout n (car si on avait $\ddot{y}_1(\tau_n) = 0$, alors il serait optimal de prolonger pour $t \geq \tau_n$ l'état et la commande par zéro).

L'étude qui suit est basée sur l'article de Robbins [118]. Elle s'appuie sur deux éléments : d'une part le *Principe de la Programmation Dynamique* qui dit que si $(\bar{u}, \bar{y})$ est une solution optimale d'un problème de commande optimale sur $[0, T]$ avec la condition initiale $y(0) = y^0$, alors quelque soit $0 < \tau < T$, $(\bar{u}(\cdot - \tau), \bar{y}(\cdot - \tau))$ restreinte à l'intervalle $[\tau, T]$ est une solution optimale du même problème sur $[0, T - \tau]$ avec la condition initiale $y(0) = \bar{y}(\tau)$; d'autre part, le paramétrage par la valeur de la dérivée seconde du problème démarrant à l'instant τ_n avec conditions initiales (7.44).

[2] Ce résultat est basé sur l'inégalité suivante, facile à démontrer. Notant $\ell(u, y)$ le coût dans l'intégrale de (7.39), on a pour toute trajectoire admissible $(\tilde{u}, \tilde{y})$ et tout $T \in (0, +\infty)$

$$\int_0^T \ell(u(t), y(t))\mathrm{d}t - \int_0^T \ell(\tilde{u}(t), \tilde{y}(t))\mathrm{d}t \;\leq\; p(T)(\tilde{y}(T) - y(T)).$$

.

Étape 1 : Etude du problème paramétré (P_α). On s'intéresse, pour $\alpha > 0$, au problème

$$(P_\alpha) \quad \begin{cases} \min \displaystyle\int_0^{+\infty} \left(y(t) + \frac{u(t)^2}{2} \right) \mathrm{d}t \\ y^{(3)}(t) = u(t) \quad \text{p.p. } t \in [0, \infty), \quad y(0) = \dot{y}(0) = 0, \quad \ddot{y}(0) = \alpha \\ -y(t) \leq 0 \quad t \in [0, \infty). \end{cases} \tag{7.45}$$

Par les mêmes arguments que précédemment pour $(\mathcal{P}^\infty_{rob})$, ce problème admet, pour tout $\alpha > 0$, une unique solution y_α. Pour $\alpha_0 > 0$, supposons connue la solution $\bar{y}$ du problème (P_{α_0}). On cherche à relier la solution y_α de (P_α) pour α quelconque à $\bar{y}$. Soit $\lambda > 0$. On effectue le changement en temps $s = \frac{t}{\lambda}$ et on pose $z(s) := \lambda^{-6} y(t)$ et $v(s) := \lambda^{-3} u(t)$. Le problème (P_α) se réécrit alors

$$(P_\alpha) \quad \begin{cases} \min \lambda^5 \displaystyle\int_0^{+\infty} \left(z(s) + \frac{v(s)^2}{2} \right) \mathrm{d}s \\ z^{(3)}(s) = v(s) \quad \text{p.p. } s \in [0, \infty), \quad z(0) = \dot{z}(0) = 0, \quad \ddot{z}(0) = \lambda^{-4}\alpha \\ -z(s) \leq 0 \quad s \in [0, \infty). \end{cases}$$

Si l'on choisit λ tel que $\lambda^{-4}\alpha = \alpha_0$ on reconnait le problème (P_{α_0}). On en déduit donc que

$$y_\alpha(t) = \lambda^6 \bar{y}\left(\frac{t}{\lambda}\right) \quad \text{avec} \quad \lambda = \left(\frac{\alpha}{\alpha_0}\right)^{\frac{1}{4}}. \tag{7.46}$$

Étape 2 : Relation avec le problème $(\mathcal{P}^\infty_{rob})$. Soit y_1 la solution du problème $(\mathcal{P}^\infty_{rob})$. Soit τ_1 le premier point de contact isolé de y_1 et $\beta_1 := \ddot{y}_1(\tau_1)$. On a alors pour $t \geq \tau_1$, d'après le principe de la programmation dynamique, $y_1(t) = y_{\beta_1}(t - \tau_1)$ où y_{β_1} est la solution de (P_{β_1}). D'après (7.46), on a $y_1(t) = \lambda^6 \bar{y}\left(\frac{t-\tau_1}{\lambda}\right)$ pour $t \geq \tau_1$ avec $\lambda = (\beta_1/\alpha_0)^{1/4}$. Soit τ_2 le deuxième point de contact isolé de y_1. On choisit de poser

$$\alpha_0 := (\tau_2 - \tau_1)^{-4} \beta_1. \tag{7.47}$$

La raison de ce choix est que l'on a alors avec (7.46) $\lambda = \tau_2 - \tau_1$ et

$$y_1(t) = (\tau_2 - \tau_1)^6 \bar{y}\left(\frac{t - \tau_1}{\tau_2 - \tau_1}\right) \qquad \text{pour } t \geq \tau_1, \tag{7.48}$$

où on rappelle que $\bar{y}$ désigne la solution de (P_{α_0}) pour α_0 donné par (7.47). D'après (7.48), les points de contact isolés de y_1 notés $(\tau_1, ..., \tau_n, ...)$ sont reliés à ceux de $\bar{y}$ que l'on note $(\sigma_0 = 0, \sigma_1, ..., \sigma_n, ...)$ par la relation

$$\tau_{n+1} = \tau_1 + (\tau_2 - \tau_1)\sigma_n. \tag{7.49}$$

En particulier on a que $\sigma_1 = 1$. Il suffit donc d'étudier la suite des points de contact isolés de $\bar{y}$ pour en déduire celle de y_1.

Étape 3 : Etude de la suite des instants de contact (σ_n) de $\bar{y}$. Posons, pour $n \in \mathbb{N}^*$, $\alpha_n := \ddot{\bar{y}}(\sigma_n)$. Pour $s \geq \sigma_n$, on a $\bar{y}(s) = y_{\alpha_n}(s - \sigma_n)$ où y_{α_n} est la solution de (P_{α_n}) (principe de la programmation dynamique). D'après (7.46), on en déduit, pour $s \geq \sigma_n$,

$$\bar{y}(s) = \lambda_n^6 \bar{y}\left(\frac{s - \sigma_n}{\lambda_n}\right) \quad \text{avec} \quad \lambda_n^4 = \frac{\alpha_n}{\alpha_0}. \tag{7.50}$$

Cette expression évaluée au premier instant de contact implique que

$$\frac{\sigma_{n+1}-\sigma_n}{\lambda_n}=\sigma_1=1. \tag{7.51}$$

Avec (7.50), cela donne

$$\bar{y}(s)=(\sigma_{n+1}-\sigma_n)^6\bar{y}\left(\frac{s-\sigma_n}{\sigma_{n+1}-\sigma_n}\right)\qquad\forall\, s\geq\sigma_n. \tag{7.52}$$

De plus, par (7.52), on a pour tout $s\geq\sigma_n$

$$\bar{y}(s)=(\sigma_{n+1}-\sigma_n)^6\bar{y}\left(\frac{s-\sigma_n}{\sigma_{n+1}-\sigma_n}\right)=(\sigma_n-\sigma_{n-1})^6\bar{y}\left(\frac{s-\sigma_{n-1}}{\sigma_n-\sigma_{n-1}}\right).$$

On dérive deux fois cette expression par rapport à s et on prend la valeur en σ_n. On obtient alors que $(\sigma_{n+1}-\sigma_n)^4\alpha_0=(\sigma_n-\sigma_{n-1})^4\alpha_1$ et donc

$$\frac{\sigma_{n+1}-\sigma_n}{\sigma_n-\sigma_{n-1}}=\left(\frac{\alpha_1}{\alpha_0}\right)^{\frac{1}{4}}=C^{ste}=:r>0\qquad\forall\, n\in\mathbb{N}^*. \tag{7.53}$$

Montrons que $r<1$ et donc que la longueur des arcs non contraints décroît géométriquement. On a d'après (7.52)

$$\int_{\sigma_n}^{\sigma_{n+1}}\bar{y}(s)\mathrm{d}s=(\sigma_{n+1}-\sigma_n)^7\int_0^1\bar{y}(\theta)\mathrm{d}\theta$$

d'où, comme $\sigma_{n+1}-\sigma_n=r^n$ et $\bar{y}$ est la solution de (P_{α_0}),

$$\sum_{n\in\mathbb{N}}r^{7n}\left(\int_0^1\bar{y}(\theta)\mathrm{d}\theta\right)\;\leq\;\int_0^\infty\bar{y}(s)\mathrm{d}s\;\leq\;\int_0^\infty\left(\bar{y}(s)+\frac{\bar{u}(s)^2}{2}\right)\mathrm{d}s\;<\;+\infty,$$

avec $\bar{u}:=\bar{y}^{(3)}$. Ceci implique que $r<1$.

Étape 4 : Calcul de de r et $\bar{y}$. Commençons par déterminer $\bar{y}$ sur $[0,1]$. Comme d'après (7.48), $\bar{y}(s)=(\tau_2-\tau_1)^{-6}y_1(\tau_1+s(\tau_2-\tau_1))$ pour tout $s\geq 0$, d'après les conditions d'optimalité vérifiées par y_1 on a $\bar{y}^{(6)}\equiv 1$ sur $(0,1)$, et $\bar{y}$ s'annule ainsi que sa dérivée première en 0 et 1. Ainsi $\bar{y}$ est de la forme

$$\bar{y}(s)=\frac{1}{6!}s^2(1-s)^2\left(s^2+as+b\right)\qquad s\in(0,1) \tag{7.54}$$

avec $a,b\in\mathbb{R}$. De plus, d'après (7.52), utilisant que $\sigma_{n+1}-\sigma_n=r^n$ et faisant le changement de temps $\theta=(s-\sigma_n)/(\sigma_{n+1}-\sigma_n)$, on obtient que

$$\bar{y}(\sigma_n+r^n\theta)=r^{6n}\bar{y}(\theta)\quad\forall\,\theta\in[0,1],\quad\forall\, n\geq 1. \tag{7.55}$$

En dérivant (7.55) pour $n=1$ j fois par rapport à θ $(j=0,1,..,4)$ et en prenant les valeurs en $\theta=0$ on obtient

$$\bar{y}^{(j)}(1)=r^{6-j}\bar{y}^{(j)}(0) \tag{7.56}$$

et comme $[\bar{y}^{(5)}(1)]=(\tau_2-\tau_1)^{-1}[y_1(\tau_2)]\leq 0$, d'après le principe du minimum on doit aussi avoir

$$\bar{y}^{(5)}(1^-)\;\geq\;\bar{y}^{(5)}(1^+)=r\bar{y}^{(5)}(0^+). \tag{7.57}$$

On développe (7.54) et on dérive 5 fois. On prend ensuite les valeurs de ces dérivées en 0 et 1 et (7.56) pour $j = 2, 3, 4$ implique que r vérifie

$$\left\{ \begin{array}{lcl} 1+a+b & = & r^4 b \\ 4+3a+2b & = & r^3(a-2b) \\ 6+3a+b & = & r^2(1-2a+b) \end{array} \right. \tag{7.58}$$

donc

$$\left(\begin{array}{ccc} 1 & 1 & 1-r^4 \\ 4 & 3-r^3 & 2(1+r^3) \\ 6-r^2 & 3+2r^2 & 1-r^2 \end{array} \right) \left(\begin{array}{c} 1 \\ a \\ b \end{array} \right) = 0. \tag{7.59}$$

Cela implique que le déterminant de la matrice est nulle car $\left(\begin{array}{ccc} 1 & a & b \end{array} \right)^\top \neq 0$. On trouve (numériquement) que les valeurs réelles de r annulant ce déterminant sont :

$$\mathcal{R} = \{ r^-;\ r^+;\ 1;\ -1 \} \quad \text{avec} \quad r^- = 0.3194887 \ \text{ et } \ r^+ = 3.130001.$$

La valeur $r = r^-$, la seule dans l'intervalle $(0, 1)$, convient, avec les valeurs associées de a et b égales respectivement à $a = -2.1728586$ et $b = 1.1852072$. On vérifie de plus que $s^2 + as + b$ est bien strictement positif sur $[0, 1]$ et que la condition (7.57) donnant la condition supplémentaire $4 + a \geq r(a - 2)$, est également satisfaite (avec inégalité stricte).

Étape 5 : Retour à la solution de $(\mathcal{P}^\infty_{rob})$. D'après (7.48), (7.52) et (7.49), la solution y_1 de $(\mathcal{P}^\infty_{rob})$ satisfait

$$y_1(t) = (\tau_{n+1} - \tau_n)^6 \bar{y} \left(\frac{t - \tau_n}{\tau_{n+1} - \tau_n} \right) \qquad \forall\, t \geq \tau_n \tag{7.60}$$

où avec (7.53),

$$\tau_{n+1} - \tau_n = (\tau_2 - \tau_1) r^{n-1}. \tag{7.61}$$

Ceci implique que

$$\lim_{n \to +\infty} \tau_n \ = \ (\tau_2 - \tau_1) \frac{1}{1-r} + \tau_1 \ =: \ T_c < +\infty. \tag{7.62}$$

Connaissant la fonction $\bar{y}$ sur $[0, 1]$ d'après l'étape 4, par (7.60) on a y_1 sur l'intervalle (τ_1, T_c). De plus, on a par (7.60) et (7.61), pour $t \in [\tau_n, \tau_{n+1}]$,

$$|y_1^{(j)}(t)| \leq (\tau_2 - \tau_1)^{6-j} r^{(6-j)(n-1)} \left(\sup_{[0,1]} \bar{y}^{(j)} \right) \leq M r^{(6-j)n} \qquad j = 0, .., 5. \tag{7.63}$$

Ainsi y_1 et ses dérivées jusqu'à l'ordre 5 tendent vers 0 quand $t \to T_c$. On peut donc prolonger y_1 par zéro sur $(T_c, +\infty)$.

Il reste à déterminer y_1 sur $[0, \tau_1)$. Pour cela il faut vérifier les conditions de recollement en τ_1 avec (7.48), c'est-à-dire la continuité de y_1 et de ses dérivées jusqu'à l'ordre 4 en τ_1 ainsi que la condition de saut $\nu = -[y_1^{(5)}(\tau_1)] \geq 0$, i.e., notant $\Delta := \tau_2 - \tau_1$,

$$y_1(\tau_1) = \dot{y}_1(\tau_1) = 0, \quad \ddot{y}_1(\tau_1) = \frac{b}{360} \Delta^4, \quad y_1^{(3)}(\tau_1) = \frac{a - 2b}{120} \Delta^3,$$
$$y_1^{(4)}(\tau_1) = \frac{1 - 2a + b}{30} \Delta^2, \quad y_1^{(5)}(\tau_1^-) = \frac{a-2}{6} \Delta + \nu.$$

On a trois inconnues que sont $\tau_1 > 0$, $\Delta = \tau_2 - \tau_1 > 0$ et $\nu \geq 0$ à déterminer pour satisfaire les trois conditions initiales (7.41), i.e. dans la base de Taylor en τ_1 :

$$\begin{aligned}
y_1^0 &= \frac{\tau_1^6}{720} - \left(\frac{a-2}{6}\Delta + \nu\right)\frac{\tau_1^5}{120} + \frac{1-2a+b}{30}\Delta^2\frac{\tau_1^4}{24} - \frac{a-2b}{120}\Delta^3\frac{\tau_1^6}{3} + \frac{b}{360}\Delta^4\frac{\tau_1^2}{2},\\
y_2^0 &= -\frac{\tau_1^5}{120} + \left(\frac{a-2}{6}\Delta + \nu\right)\frac{\tau_1^4}{24} - \frac{1-2a+b}{30}\Delta^2\frac{\tau_1^3}{6} + \frac{a-2b}{120}\Delta^3\frac{\tau_1^2}{2} - \frac{b}{360}\Delta^4\tau_1,\\
y_3^0 &= \frac{\tau_1^4}{24} - \left(\frac{a-2}{6}\Delta + \nu\right)\frac{\tau_1^3}{6} + \frac{1-2a+b}{30}\Delta^2\frac{\tau_1^2}{2} - \frac{a-2b}{120}\Delta^3\tau_1 + \frac{b}{360}\Delta^4.
\end{aligned}$$

(Lorsque $\Delta = \tau_2 - \tau_1 = 0$ et $\nu = 0$, on retrouve les conditions initiales particulières (7.38) correspondant à un arc frontière.)

Enfin, comme $y_1(t) = 0$ sur $[T_c, \infty)$, on a aussi $p(t) = 0$ sur $[T_c, \infty)$, et donc (7.43) est aussi vérifié. Ainsi construite, y_1 vérifie le principe du minimum sur $(0, +\infty)$ ainsi que la condition aux limites (7.43), de plus le hamiltonien est convexe par rapport à (u, y) et la contrainte sur l'état est convexe, d'où y_1 vérifie la condition d'optimalité suffisante de $(\mathcal{P}_{rob}^{\infty})$. C'est donc bien l'unique solution de $(\mathcal{P}_{rob}^{\infty})$.

Remarque 7.6. Comme $\nu_n = -[y_1^{(5)}(\tau_n)]$, avec (7.60)–(7.61) on a que

$$\nu_{n+1} \;=\; -[y_1^{(5)}(\tau_{n+1})] \;=\; \gamma(\tau_2 - \tau_1)r^{n-1}$$

avec

$$\gamma \;:=\; -[\bar{y}^{(5)}(1)] \;=\; \frac{1}{6}\left(4 + a - r(a-2)\right) > 0.$$

Étape 6 : Retour à la solution de $(\mathcal{P}_{rob})$. Repassant en temps T fini, pour un temps final suffisamment grand $T > T_c$, la solution y_1 ainsi construite, restreinte à l'intervalle $[0, T]$, vérifie le principe du minimum de Pontryaguine sur $[0, T]$. En particulier, comme on termine sur un arc frontière, la condition finale sur l'état adjoint $p(T) = 0$ est satisfaite. La condition d'optimalité du premier ordre étant nécessaire et suffisante, on a bien trouvé l'unique solution du problème $(\mathcal{P}_{rob})$ pour des conditions initiales ne se mettant pas sous la forme (7.38), et cette solution a une infinité de points de contact isolés dont le point limite est un point d'entrée sur un arc frontière.

On a tracé sur la figure 7.4 la solution optimale de $(\mathcal{P}_{rob})$ correspondant aux conditions initiales

$$y_1^0 = 1, \quad y_2^0 = -1, \quad y_3^0 = -2.$$

On trouve alors (numériquement) $\tau_1 = 1.4110209$, $\Delta = \tau_2 - \tau_1 = 3.5497156$ et $\nu = 19.144858$. On a tracé les cinq premiers arcs intérieurs de y_1. Par (7.63) pour $j = 0$, on voit que

$$\max_{t\in[\tau_n,\tau_{n+1}]} y_1(t) = (\tau_2 - \tau_1)^6 r^{6(n-1)} \max_{s\in[0,1]} \bar{y}(s)$$

avec $\max_{s\in[0,1]} \bar{y}(s)$ de l'ordre de 4.10^{-5} et $r^6 = 0.0010635$. Ainsi $y_1(t)$ décroît très rapidement en pratique, ce que l'on observe sur la figure 7.4 (seul les deux premiers arcs intérieurs sont visibles). Ainsi, résolvant le problème en utilisant par exemple une méthode directe (pour lesquelles la présence d'un nombre infini de points de contact isolés ne pose pas de difficulté), on obtiendra numériquement une solution semblant présenter un arc frontière avec point d'entrée régulier, précédé ou non d'un nombre fini de points de contact isolés, bien que cela semble contredire la théorie. En cela on rejoint la conclusion de Betts et al. [13] sur la résolution numérique des problèmes de commande optimale avec contrainte sur l'état d'ordre élevé.

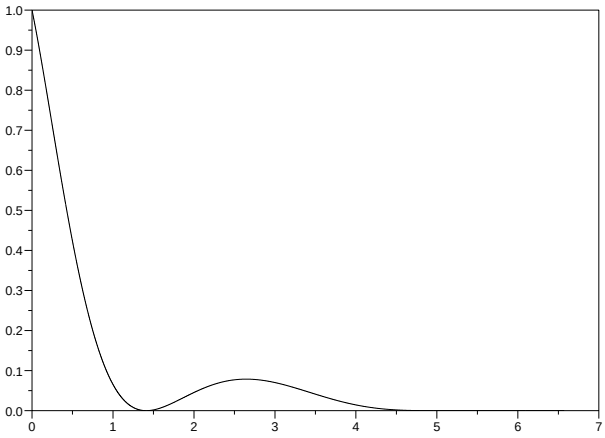


FIG. 7.4 – Trajectoire optimale y_1 dans l'exemple de Robbins pour la condition initiale $y^0 = (1, -1, -2)$.

7.2.3 Preuve de (7.17)

On utilise dans cette section les notations de l'exemple de Robbins de la section précédente. On note par y_1 la trajectoire solution de $(\mathcal{P}_{rob})$. Par le principe du minimum, on a que $y_1^{(5)}$ est une fonction à variation bornée et satisfait

$$\mathrm{d}y_1^{(5)}(t) = \mathrm{d}t - \mathrm{d}\eta(t) = \mathbf{1}_{[0,T_c)}\mathrm{d}t - \sum_{n\in\mathbb{N}^*} \nu_n \delta_{\tau_n}(t).$$

Comme toutes les dérivées de y_1 jusqu'à l'ordre 5 s'annulent en T_c par (7.63), en intégrant sur (t, T_c) cinq fois on obtient que

$$y_1(t) = \frac{(t-T_c)^6}{720} - \sum_{n\in\mathbb{N}^*} \nu_n \mathbf{1}_{t<\tau_n} \frac{(\tau_n - t)^5}{120}. \tag{7.64}$$

Nous allons chercher à expliciter le terme $\sum_n \nu_n \mathbf{1}_{t<\tau_n} \frac{(\tau_n-t)^5}{120}$. D'après (7.61), on a que

$$\tau_n = (\tau_2 - \tau_1)\left(\frac{1-r^{n-1}}{1-r}\right) + \tau_1. \tag{7.65}$$

D'où $t < \tau_n$ si et seulement si

$$n \;>\; 1 + \frac{\ln\left(1 - (1-r)\frac{t-\tau_1}{\tau_2-\tau_1}\right)}{\ln(r)} \;=:\; \psi(t).$$

Utilisant l'expression de T_c donnée par (7.62), on a que $\Psi(t) = 1 + \ln(r)^{-1}\ln\left(\frac{1-r}{\tau_2-\tau_1}(T_c - t)\right)$ d'où

$$r^{\psi(t)-1} \;=\; \frac{1-r}{\tau_2-\tau_1}(T_c - t). \tag{7.66}$$

On a donc

$$\sum_{n\in\mathbb{N}^*} \nu_n \mathbf{1}_{t<\tau_n} \frac{(\tau_n - t)^5}{120} \;=\; \sum_{n>\Psi(t)} \nu_n \frac{(t-\tau_n)^5}{120}. \tag{7.67}$$

De plus, d'après (7.65), (7.62) et (7.66), on a pour tout n

$$\begin{aligned} t-\tau_n &= t-T_c+(\tau_2-\tau_1)\frac{r^{n-1}}{1-r} \\ &= t-T_c+\frac{(\tau_2-\tau_1)}{1-r}r^{n-\psi(t)}\frac{1-r}{\tau_2-\tau_1}(T_c-t) \\ &= (t-T_c)(1-r^{n-\psi(t)}). \end{aligned} \tag{7.68}$$

Enfin, d'après la remarque 7.6 on a que

$$\nu_n = \gamma(\tau_2-\tau_1)r^{n-2} \tag{7.69}$$

où γ est une constante strictement positive. D'après (7.67), (7.68) et (7.69), on obtient donc que

$$\sum_{n\in\mathbb{N}^*}\nu_n\mathbf{1}_{t<\tau_n}\frac{(\tau_n-t)^5}{120} = \gamma(\tau_2-\tau_1)\frac{(t-T_c)^5}{120}\sum_{n>\Psi(t)}r^{n-2}(1-r^{n-\psi(t)})^5. \tag{7.70}$$

Cette expression évaluée en $t=\tau_N$ (i.e. $\psi(t)=N$) donne

$$\begin{aligned} \sum_{n\in\mathbb{N}^*}\nu_n\mathbf{1}_{\tau_N<\tau_n}\frac{(\tau_n-\tau_N)^5}{120} &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-\tau_N)^5}{120}\sum_{n>N}r^{n-1}(1-r^{n-N})^5 \\ &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-\tau_N)^5}{120}r^{N-1}\sum_{k>0}r^k(1-r^k)^5. \end{aligned}$$

Posons

$$\Lambda := \sum_{k>0}r^k(1-r^k)^5 = \frac{r}{1-r}-\frac{5r^2}{1-r^2}+\frac{10r^3}{1-r^3}-\frac{10r^4}{1-r^4}+\frac{5r^5}{1-r^5}-\frac{r^6}{1-r^6}.$$

Utilisant (7.66), on obtient que

$$\begin{aligned} \sum_{n\in\mathbb{N}^*}\nu_n\mathbf{1}_{\tau_N<\tau_n}\frac{(\tau_n-\tau_N)^5}{120} &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-\tau_N)^5}{120}\frac{1-r}{\tau_2-\tau_1}(T_c-\tau_N)\Lambda \\ &= \gamma\frac{1-r}{r}\frac{(T_c-\tau_N)^6}{120}\Lambda. \end{aligned}$$

Or on a aussi $y_1(\tau_n)=0$ pour tout n, donc par (7.64),

$$\sum_{n\in\mathbb{N}^*}\nu_n\mathbf{1}_{\tau_N<\tau_n}\frac{(\tau_n-\tau_N)^5}{120} = \frac{(T_c-\tau_N)^6}{720}.$$

On déduit des deux expressions ci-dessus que

$$\Lambda = \frac{r}{6\gamma(1-r)}. \tag{7.71}$$

Repassons maintenant au calcul de (7.70) pour $t\notin\{\tau_n\}_{n\in\mathbb{N}^*}$. Soit N le plus grand entier supérieur à $\psi(t)$, $N:=\lceil\psi(t)\rceil$. On a

$$\begin{aligned} \sum_{n\in\mathbb{N}^*}\nu_n\mathbf{1}_{t<\tau_n}\frac{(\tau_n-t)^5}{120} &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-t)^5}{120}\sum_{n>\psi(t)}r^{n-1}(1-r^{n-\psi(t)})^5 \\ &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-t)^5}{120}r^{\psi(t)-1}\sum_{k\geq 0}r^{k+N-\psi(t)}(1-r^{k+N-\psi(t)})^5. \end{aligned}$$

Notons $\rho(t) := N - \psi(t) = \lceil \psi(t) \rceil - \psi(t) \in (0,1)$. Soit

$$\Xi(t) \;:=\; \sum_{k\geq 0} r^{k+\rho(t)}(1-r^{k+\rho(t)})^5.$$

Développant $(1-r^{k+\rho(t)})^5$ en utilisant la formule du binôme de Newton, on obtient que

$$\begin{aligned}\Xi(t) &= \sum_{k\geq 0}\Big(r^k r^{\rho(t)} - 5r^{2k}r^{2\rho(t)} + 10r^{3k}r^{3\rho(t)} - 10r^{4k}r^{4\rho(t)} + 5r^{5k}r^{5\rho(t)} - r^{6k}r^{6\rho(t)}\Big)\\ &= \frac{r^{\rho(t)}}{1-r} - \frac{5r^{2\rho(t)}}{1-r^2} + \frac{10r^{3\rho(t)}}{1-r^3} - \frac{10r^{4\rho(t)}}{1-r^4} + \frac{5r^{5\rho(t)}}{1-r^5} - \frac{r^{6\rho(t)}}{1-r^6}.\end{aligned}$$

Alors, utilisant (7.66) et (7.71),

$$\begin{aligned}\sum_{n\in\mathbb{N}^*}\nu_n \mathbf{1}_{t<\tau_n}\frac{(\tau_n - t)^5}{120} &= \gamma\frac{\tau_2-\tau_1}{r}\frac{(T_c-t)^5}{120}\frac{1-r}{\tau_2-\tau_1}(T_c-t)\Xi(t)\\ &= \frac{(T_c-t)^6}{720}\frac{\Xi(t)}{\Lambda}. \qquad (7.72)\end{aligned}$$

La fonction $\Phi : [0,1] \to \mathbb{R}$,

$$\rho \;\mapsto\; \frac{r^\rho}{1-r} - \frac{5r^{2\rho}}{1-r^2} + \frac{10r^{3\rho}}{1-r^3} - \frac{10r^{4\rho}}{1-r^4} + \frac{5r^{5\rho}}{1-r^5} - \frac{r^{6\rho}}{1-r^6} \;=\; \sum_{k\geq 0} r^{k+\rho}(1-r^{k+\rho})^5$$

est de classe C^∞ sur $[0,1]$, telle que $\Phi^{(j)}(0) = \Phi^{(j)}(1)$ pour tout $j = 0,\ldots,4$ (s'obtient facilement à partir de la formule sommatoire) et $0 < \Phi(\rho) < \Phi(1)$ pour tout $\rho \in (0,1)$ (voir la figure 7.5). Comme $\Lambda = \phi(1)$, posant

$$\theta(t) \;:=\; \frac{\Xi(t)}{\Lambda} \;=\; \frac{\Phi(\rho(t))}{\Lambda},$$

on a que θ est C^4 sur $[\tau_1, T_c)$, C^∞ sur chaque arc intérieur (τ_n, τ_{n+1}), telle que $\theta(\tau_n) = 1$ pour tout n et $0 < \theta(t) < 1$ pour tout $t \in (\tau_n, \tau_{n+1})$. D'où avec (7.64) et (7.72), cela achève la preuve de (7.17).

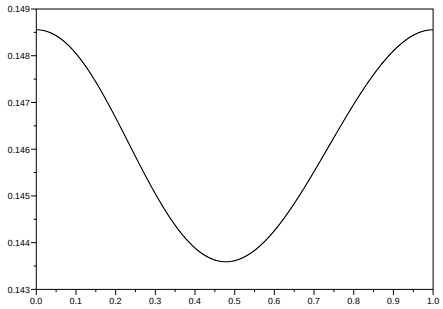


FIG. 7.5 – Fonction Φ.

Remarque 7.7. Noter que par (7.66), avec $N = \lceil \psi(t) \rceil$,

$$r^{\rho(t)} = r^{N-\psi(t)} = \frac{r^{N-1}}{r^{\psi(t)-1}} = \frac{T_c - \tau_N}{T_c - t}$$

où τ_N est le plus petit instant de jonction supérieur strictement à t. Ainsi on peut réécrire θ sur chaque arc (τ_{N-1}, τ_N) par

$$\begin{aligned}\theta(t) &= \Lambda^{-1}\left(\frac{1}{1-r}\frac{T_c-\tau_N}{T_c-t} - \frac{5}{1-r^2}\left(\frac{T_c-\tau_N}{T_c-t}\right)^2 + \frac{10}{1-r^3}\left(\frac{T_c-\tau_N}{T_c-t}\right)^3\right.\\ &\quad \left. - \frac{10}{1-r^4}\left(\frac{T_c-\tau_N}{T_c-t}\right)^4 + \frac{5}{1-r^5}\left(\frac{T_c-\tau_N}{T_c-t}\right)^5 - \frac{1}{1-r^6}\left(\frac{T_c-\tau_N}{T_c-t}\right)^6\right)\end{aligned}$$

avec

$$T_c - \tau_N = \frac{\tau_2 - \tau_1}{1-r} r^{N-1}.$$

Bibliographie

[1] E.L. Allgower and K. Georg. *Numerical continuation methods*, volume 13 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1990. An introduction.

[2] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[3] J.P. Aubin. Comportement lipschitzien des solutions de problèmes de minimisation convexes. *Comptes Rendus de l'Académie des Sciences de Paris Série I*, 295 :235–238, 1982.

[4] D. Augustin and H. Maurer. Computational sensitivity analysis for state constrained optimal control problems. *Ann. Oper. Res.*, 101 :75–99, 2001.

[5] D. Augustin and H. Maurer. Second order sufficient conditions and sensitivity analysis for the optimal control of a container crane under state constraints. *Optimization. A Journal of Mathematical Programming and Operations Research*, 49(4) :351–368, 2001.

[6] M. Bardi and I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Systems and Control : Foundations and Applications. Birkhäuser, Boston, 1997.

[7] G. Barles. *Solutions de viscosité des équations de Hamilton-Jacobi*, volume 17 of *Mathématiques et Applications*. Springer, Paris, 1994.

[8] R. Bellman. *Dynamic programming*. Princeton University Press, Princeton, 1961.

[9] N. Bérend, J.F. Bonnans, M. Haddou, J. Laurent-Varin, and Ch. Talbot. An interior-point approach to trajectory optimization. to appear in J. Guidance, Control and Dynamics, 2006.

[10] M. Bergounioux and K. Kunisch. On the structure of Lagrange multipliers for state-constrained optimal control problems. *Systems Control Lett.*, 48(3-4) :169–176, 2003. Optimization and control of distributed systems.

[11] P. Berkmann and H.J. Pesch. Abort landing in windshear : optimal control problem with third-order state constraint and varied switching structure. *J. of Optimization Theory and Applications*, 85, 1995.

[12] J.T. Betts. *Practical methods for optimal control using nonlinear programming*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001.

[13] J.T. Betts, S.L. Campbell, and A. Engelsone. Direct transcription solution of optimal control problems with higher order state constraints : theory vs practice. *Optim. Eng.*, 8(1) :1–19, 2007.

[14] J.F. Bonnans. Second order analysis for control constrained optimal control problems of semilinear elliptic systems. *Journal of Applied Mathematics & Optimization*, 38 :303–325, 1998.

[15] J.F. Bonnans, R. Cominetti, and A. Shapiro. Sensitivity analysis of optimization problems under abstract constraints. *Mathematics of Operations Research*, 23 :806–831, 1998.

[16] J.F. Bonnans, R. Cominetti, and A. Shapiro. Second order optimality conditions based on parabolic second order tangent sets. *SIAM Journal on Optimization*, 9 :466–492, 1999.

[17] J.F. Bonnans and A. Hermant. Second-order analysis for optimal control problems with pure state constraints and mixed control-state constraints. INRIA Research Report 6199, to appear in Annales de l'Institut Henri Poincaré (C) Analyse Non Linéaire.

[18] J.F. Bonnans and A. Hermant. Conditions d'optimalité du second ordre nécessaires ou suffisantes pour les problèmes de commande optimale avec une contrainte sur l'état et une commande scalaires. *C. R. Math. Acad. Sci. Paris*, 343(7) :473–478, 2006.

[19] J.F. Bonnans and A. Hermant. Well-posedness of the shooting algorithm for state constrained optimal control problems with a single constraint and control. *SIAM J. on Control and Optimization*, 46(4) :1398–1430, 2007.

[20] J.F. Bonnans and A. Hermant. Stability and sensitivity analysis for optimal control problems with a first-order state constraint and application to continuation methods. *ESAIM Control Optim. Calc. Var.*, 14(4) :825–863, 2008.

[21] J.F. Bonnans and A. Hermant. No-gap second-order optimality conditions for optimal control problems with a single state constraint and control. *Mathematical Programming*, 117 :21–50, 2009.

[22] J.F. Bonnans and G. Launay. Large scale direct optimal control applied to a re-entry problem. *AIAA J. of Guidance, Control and Dynamics*, 21 :996–1000, 1998.

[23] J.F. Bonnans and P. Rouchon. *Commande et optimisation de systèmes dynamiques*. Editions de l'Ecole Polytechnique, Palaiseau, 2005.

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

[25] J.F. Bonnans and H. Zidani. Optimal control problems with partially polyhedric constraints. *SIAM Journal on Control and Optimization*, 37 :1726–1741, 1999.

[26] B. Bonnard, L. Faubourg, G. Launay, and E. Trélat. Optimal control with state constraints and the space shuttle re-entry problem. *J. Dynam. Control Systems*, 9(2) :155–199, 2003.

[27] B. Bonnard, L. Faubourg, and E. Trelat. Optimal control of the atmospheric arc of a space shuttle and numerical simulations with multiple-shooting method. *Mathematical Models & Methods in Applied Sciences*, 15(1) :109–140, 2005.

[28] A. E. Bryson and Y.-C. Ho. *Applied optimal control*. Hemisphere Publishing, New-York, 1975.

[29] A.E. Bryson, W.F. Denham, and S.E. Dreyfus. Optimal programming problems with inequality constraints I : necessary conditions for extremal solutions. *AIAA Journal*, 1 :2544–2550, 1963.

[30] R. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. I. Necessary conditions. *J. of Optimization Theory and Applications*, 70 :1–23, 1991.

[31] R. Bulirsch, F. Montrone, and H. J. Pesch. Abort landing in the presence of windshear as a minimax optimal control problem. II. Multiple shooting and homotopy. *J. Optim. Theory Appl.*, 70(2) :223–254, 1991.

[32] C. Büskens and M. Knauer. Higher order real-time approximations in optimal control of multibody-systems for industrial robots. *Multibody System Dynamics*, 15(1) :85–106, 2006.

[33] C. Büskens and H. Maurer. Nonlinear programming methods for real-time control of an industrial robot. *J. Optim. Theory Appl.*, 107(3) :505–527, 2000.

[34] P. Cardaliaguet, M. Quincampoix, and P. Saint-Pierre. Numerical schemes for discontinuous value functions of optimal control. *Set-Valued Anal.*, 8(1-2) :111–126, 2000. Set-valued analysis in control theory.

[35] E. Casas. Control of an elliptic problem with pointwise state constraints. *SIAM Journal on Control and Optimization*, 24 :1309–1318, 1986.

[36] E. Casas and M. Mateos. Second order optimality conditions for semilinear elliptic control problems with finitely many state constraints. *SIAM J. Control Optim.*, 40(5) :1431–1454 (electronic), 2002.

[37] E. Casas and F. Tröltzsch. Second-order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM J. Optim.*, 13(2) :406–431 (electronic), 2002.

[38] E. Casas, F. Tröltzsch, and A. Unger. Second order sufficient optimality conditions for some state-constrained control problems of semilinear elliptic equations. *SIAM J. Control Optim.*, 38(5) :1369–1391 (electronic), 2000.

[39] F.H. Clarke. *Optimization and nonsmooth analysis*. Wiley, New York, 1983.

[40] F.H. Clarke, Yu. S. Ledyaev, R.J. Stern, and P. Wolenski. *Nonsmooth analysis and control theory*, volume 178 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[41] R. Cominetti. Metric regularity, tangent sets and second order optimality conditions. *Journal of Applied Mathematics & Optimization*, 21 :265–287, 1990.

[42] R. Cominetti and J.P. Penot. Tangent sets to unilateral convex sets. *Comptes Rendus de l'Académie des Sciences de Paris, Série I*, 321 :1631–1636, 1995.

[43] M.G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton Jacobi equations. *Bull. American Mathematical Society*, 277 :1–42, 1983.

[44] M.G. Crandall and P.-L. Lions. Two approximations of solutions of Hamilton-Jacobi equations. *Mathematics of Computation*, 43 :1–19, 1984.

[45] P. Deuflhard. *Newton methods for nonlinear problems*, volume 35 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2004. Affine invariance and adaptive algorithms.

[46] M. Diehl, R. Findeisen, S. Schwarzkopf, I. Uslu, F. Allgöwer, H.G. Bock, E.D. Gilles, and J.P. Schlöder. An efficient algorithm for nonlinear model predictive control of large-scale systems. Part I : Description of the method. *Automatisierungstechnik*, 50(12) :557–567, 2002.

[47] M. Diehl, R. Findeisen, S. Schwarzkopf, I. Uslu, F. Allgöwer, H.G. Bock, E.D. Gilles, and J.P. Schlöder. An efficient algorithm for nonlinear model predictive control of large-scale systems. Part II : Application to a distillation column. *Automatisierungstechnik*, 51(1) :22–29, 2003.

[48] V.V. Dikusar and A.A. Milyutin. *Qualitative and numerical methods in the maximum principle*. "Nauka", Moscow, 1989. (in Russian).

[49] A.V. Dmitruk. Quadratic conditions for the Pontryagin minimum in an optimal control problem linear with respect to control. I. Decoding theorem. *Izv. Akad. Nauk SSSR Ser. Mat.*, 50(2) :284–312, 1986.

[50] A.V. Dmitruk. Maximum principle for the general optimal control problem with phase and regular mixed constraints. *Computational Mathematics and Modeling*, 4(4) :364–377, 1993. Software and models of systems analysis. Optimal control of dynamical systems.

[51] A.V. Dmitruk. Quadratic order conditions of a local minimum for singular extremals in a general optimal control problem. In *Differential geometry and control (Boulder, CO, 1997)*, volume 64 of *Proc. Sympos. Pure Math.*, pages 163–198. Amer. Math. Soc., Providence, RI, 1999.

[52] A.L. Dontchev and W. Hager. Lipschitzian stability in nonlinear control and optimization. *SIAM J. Control Optim.*, 31(3) :569–603, 1993.

[53] A.L. Dontchev and W.W. Hager. Lipschitzian stability for state constrained nonlinear optimal control. *SIAM J. on Control and Optimization*, 36(2) :698–718 (electronic), 1998.

[54] A.L. Dontchev and W.W. Hager. The Euler approximation in state constrained optimal control. *Mathematics of Computation*, 70 :173–203, 2001.

[55] A.L. Dontchev, W.W. Hager, A.B. Poore, and B. Yang. Optimality, stability, and convergence in nonlinear control. *Applied Mathematics and Optimization*, 31(3) :297–326, 1995.

[56] A.L. Dontchev and I. Kolmanovsky. On regularity of optimal control. In *Recent developments in optimization (Dijon, 1994)*, volume 429 of *Lecture Notes in Econom. and Math. Systems*, pages 125–135. Springer, Berlin, 1995.

[57] A.Ya. Dubovitskiĭ and A.A. Milyutin. Theory of the principle of the maximum. In *Methods of the theory of extremal problems in economics*, pages 7–47. "Nauka", Moscow, 1981.

[58] N. Dunford and J. Schwartz. *Linear operators, Vol I and II*. Interscience, New York, 1958, 1963.

[59] I. Ekeland. Nonconvex minimization problems. *Bulletin of the American Mathematical Society*, 1(New series) :443–474, 1979.

[60] A.V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Academic Press, New York, 1983.

[61] A.V. Fiacco and G.P. McCormick. *Nonlinear Programming : Sequential Unconstrained Minimization Techniques*. Wiley, New York, 1968.

[62] H. Frankowska and B. Kaŝkosz. A maximum principle for differential inclusion problems with state constraints. *Systems Control Lett.*, 11 :189–194, 1988.

[63] J. Gergaud and T. Haberkorn. Homotopy method for minimum consumption orbit transfer problem. *ESAIM Control Optim. Calc. Var.*, 12(2) :294–310 (electronic), 2006.

[64] H. Goldberg and F. Tröltzsch. Second-order sufficient optimality conditions for a class of nonlinear parabolic boundary control problems. *SIAM J. Control Optim.*, 31(4) :1007–1025, 1993.

[65] W.W. Hager. Lipschitz continuity for constrained processes. *SIAM J. Control Optim.*, 17(3) :321–338, 1979.

[66] H. Halkin. Necessary conditions for optimal control problems with infinite horizons. *Econometrica*, 42 :267–272, 1974.

[67] A. Haraux. How to differentiate the projection on a convex set in Hilbert space. Some applications to variational inequalities. *Journal Mathematical Society of Japan*, 29 :615–631, 1977.

[68] R.F. Hartl, S.P. Sethi, and R.G. Vickson. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Review*, 37 :181–218, 1995.

[69] A. Hermant. Homotopy algorithm for optimal control problems with a second-order state constraint. INRIA Research Report RR-6626, 2008. http ://hal.inria.fr/inria-00316281/fr/.

[70] A. Hermant. Optimal control of the atmospheric reentry of a space shuttle by an homotopy method. INRIA Research Report RR-6627, 2008. http ://hal.inria.fr/inria-00317722/fr/.

[71] A. Hermant. Stability analysis of optimal control problems with a second-order state constraint. INRIA Research Report, 2007. http ://hal.inria.fr/inria-00186968/fr/, accepted in SIAM Journal on Optimization.

[72] R.P. Hettich and H.T. Jongen. Semi-infinite programming : conditions of optimality and applications. In J. Stoer, editor, *Optimization Techniques*. Proc. 8th IFIP Conf. on Optimization Techniques, Würzburg. Part 2, Springer-Verlag, New York, 1978.

[73] A.D. Ioffe. Necessary and sufficient conditions for a local minimum I : A reduction theorem and first order conditions, II : Conditions of Levitin-Miljutin-Osmolovskii type, III : Second order conditions and augmented duality. *SIAM Journal on Control Optimization*, 17 :245–250, 251–265 and 266–288, 1979.

[74] A.D. Ioffe and V.M. Tihomirov. *Theory of Extremal Problems*. North-Holland Publishing Company, Amsterdam, 1979. Russian Edition : Nauka, Moscow, 1974.

[75] D.H. Jacobson, M.M. Lele, and J.L. Speyer. New necessary conditions of optimality for control problems with state-variable inequality contraints. *J. of Mathematical Analysis and Applications*, 35 :255–284, 1971.

[76] K. Jittorntrum. Solution point differentiability without strict complementarity in nonlinear programming. *Mathematical Programming*, 21 :127–138, 1984.

[77] H. Kawasaki. An envelope-like effect of infinitely many inequality constraints on second order necessary conditions for minimization problems. *Mathematical Programming*, 41 :73–96, 1988.

[78] H. Kawasaki. The upper and lower second order directional derivatives of a sup-type function. *Mathematical Programming*, 41 :327–339, 1988.

[79] H. Kawasaki. Second order necessary optimality conditions for minimizing a sup-type function. *Mathematical Programming (Ser. A)*, 49 :213–229, 1990/91.

[80] H. Kawasaki and V. Zeidan. Conjugate points for variational problems with equality and inequality state constraints. *SIAM Journal on Control and Optimization*, 39(2) :433–456 (electronic), 2000.

[81] O.I. Kostyukova. Properties of solutions of a parametric linear-quadratic optimal control problem in a neighborhood of an irregular point. *Zh. Vychisl. Mat. Mat. Fiz.*, 43(9) :1364–1373, 2003.

[82] O.I. Kostyukova. Sensitivity analysis for optimal control problems with phase constraints and changing index. *Vestsī Nats. Akad. Navuk Belarusī Ser. Fīz.-Mat. Navuk*, (1) :12–16, 2004.

[83] E. Kreindler. Additional necessary conditions for optimal control with state-variable inequality constraints. *J. Optim. Theory Appl.*, 38(2) :241–250, 1982.

[84] J. Laurent-Varin, F. Bonnans, N. Bérend, M. Haddou, and C. Talbot. Interior-point approach to trajectory optimization. *Journal of Guidance, Control, and Dynamics*, 30(5) :1228–1238, 2007.

[85] X.J. Li and J.M. Yong. *Optimal control theory for infinite-dimensional systems*. Systems & Control : Foundations & Applications. Birkhäuser Boston Inc., Boston, MA, 1995.

[86] K. Malanowski. Second order conditions and constraint qualifications in stability and sensitivity analysis of solutions to optimization problems in Hilbert spaces. *Journal of Applied Mathematics & Optimization*, 25 :51–79, 1992.

[87] K. Malanowski. Two-norm approach in stability and sensitivity analysis of optimization and optimal control problems. *Advances in Mathematical Sciences and Applications*, 2 :397–443, 1993.

[88] K. Malanowski. Stability and sensitivity of solutions to nonlinear optimal control problems. *Journal of Applied Mathematics & Optimization*, 32 :111–141, 1995.

[89] K. Malanowski. Sufficient optimality conditions for optimal control subject to state constraints. *SIAM J. on Control and Optimization*, 35 :205–227, 1997.

[90] K. Malanowski. Stability and sensitivity analysis for optimal control problems with control-state constraints. *Dissertationes Math. (Rozprawy Mat.)*, 394 :51, 2001.

[91] K. Malanowski. Stability analysis for nonlinear optimal control problems subject to state constraints. *SIAM J. on Optimization*, 18(3) :926–945, 2007.

[92] K. Malanowski. Sufficient optimality conditions in stability analysis for state-constrained optimal control. *Applied Mathematics and Optimization*, 55(2) :255–271, 2007.

[93] K. Malanowski and H. Maurer. Sensitivity analysis for state constrained optimal control problems. *Discrete and Continuous Dynamical Systems*, 4 :241–272, 1998.

[94] K. Malanowski and H. Maurer. Sensitivity analysis for optimal control problems subject to higher order state constraints. *Annals of Operations Research*, 101 :43–73, 2001. Optimization with data perturbations, II.

[95] K. Malanowski, H. Maurer, and S. Pickenhain. Second-order sufficient conditions for state-constrained optimal control problems. *J. of Optimization Theory and Applications*, 123 :595–617, 2004.

[96] O.L. Mangasarian. Sufficient conditions for the optimal control of nonlinear systems. *SIAM J. Control*, 4 :139–152, 1966.

[97] P. Martinon and J. Gergaud. An application of PL continuation methods to singular arcs problems. In *Recent advances in optimization*, volume 563 of *Lecture Notes in Econom. and Math. Systems*, pages 163–186. Springer, Berlin, 2006.

[98] H. Maurer. On the minimum principle for optimal control problems with state constraints. Schriftenreihe des Rechenzentrum 41, Universität Münster, 1979.

[99] H. Maurer. First and second order sufficient optimality conditions in mathematical programming and optimal control. *Math. Programming Stud.*, (14) :163–177, 1981.

[100] H. Maurer and N.P. Osmolovskii. Second order sufficient conditions for time-optimal bang-bang control. *SIAM J. Control Optim.*, 42(6) :2239–2263 (electronic), 2004.

[101] H. Maurer and H.J. Pesch. Solution differentiability for nonlinear parametric control problems. *SIAM Journal on Control and Optimization*, 32 :1542–1554, 1994.

[102] H. Maurer and J. Zowe. First and second-order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, 16 :98–110, 1979.

[103] F. Mignot. Contrôle dans les inéquations variationnelles elliptiques. *Journal of Functional Analysis*, 22 :130–185, 1976.

[104] A.A. Milyutin. *The maximum principle in the general problem of optimal control*. Fizmatlit, Moscow, 2001.

[105] A.A. Milyutin and N. N. Osmolovskii. *Calculus of Variations and Optimal Control*. American Mathematical Society, Providence, 1998.

[106] A.A. Milyutin and N.P. Osmolovskii. *Calculus of variations and optimal control*, volume 180 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1998. Translated from the Russian manuscript by Dimitrii Chibisov.

[107] H.J. Oberle and W. Grimm. Bndsco – a program for the numerical solution of optimal control problems. Technical report, Report No. 515, Institut for Flight Systems Dynamics, Oberpfaffenhofen, German Aerospace Research Establishment DLR, 1989.

[108] N.P. Osmolovskii. Higher-order necessary and sufficient conditions for Pontryagin and restricted-strong minima in an optimal control problem. *Doklady Akademii Nauk SSSR*, 303(5) :1052–1056, 1988. translation in Soviet Phys. Dokl. 33 (1988), no. 12, 883–885 (1989).

[109] N.P. Osmolovskii. Quadratic conditions for nonsingular extremals in optimal control (a theoretical treatment). *Russian Journal of Mathematical Physics*, 2(4) :487–516, 1995.

[110] N.P. Osmolovskii and H. Maurer. Equivalence of second order optimality conditions for bang-bang control problems. I. Main results. *Control Cybernet.*, 34(3) :927–950, 2005.

[111] Z. Páles and V. Zeidan. First- and second-order necessary conditions for control problems with constraints. *Transactions of the American Mathematical Society*, 346(2) :421–453, 1994.

[112] Z. Páles and V. Zeidan. Optimal control problems with set-valued control and state constraints. *SIAM J. on Optimization*, 14 :334–358 (electronic), 2003.

[113] Z. Páles and V. Zeidan. Strong local optimality conditions for state constrained control problems. *J. Global Optim.*, 28(3-4) :363–377, 2004.

[114] Zs. Páles and V. M. Zeidan. Nonsmooth optimum problems with constraints. *SIAM J. Control Optim.*, 32(5) :1476–1502, 1994.

[115] H. J. Pesch. A practical guide to the solution of real-life optimal control problems. *Control and Cybernetics*, 23 :7–60, 1994.

[116] L.S. Pontryagin, V.G. Boltyanskii, R.V. Gamkrelidze, and E. F. Mishchenko. *The mathematical theory of optimal processes*. Translated from the Russian by K. N. Trirogoff ; edited by L. W. Neustadt. Interscience Publishers John Wiley & Sons, Inc. New York-London, 1962.

[117] J.-P. Raymond and F. Tröltzsch. Second order sufficient optimality conditions for nonlinear parabolic control problems with state constraints. *Discrete Contin. Dynam. Systems*, 6(2) :431–450, 2000.

[118] H.M. Robbins. Junction phenomena for optimal control with state-variable inequality constraints of third order. *J. of Optimization Theory and Applications*, 31 :85–99, 1980.

[119] S.M. Robinson. First order conditions for general nonlinear optimization. *SIAM Journal on Applied Mathematics*, 30 :597–607, 1976.

[120] S.M. Robinson. Stability theorems for systems of inequalities, part II : Differentiable nonlinear systems. *SIAM Journal on Numerical Analysis*, 13 :497–513, 1976.

[121] S.M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5 :43–62, 1980.

[122] A. Seierstad and K. Sydsaeter. Sufficient conditions in optimal control theory. *Internat. Econom. Rev.*, 18(2) :367–391, 1977.

[123] J. Sokolowski. Sensitivity analysis of control constrained optimal control problems for distributed parameter systems. *SIAM Journal on Control and Optimization*, 25 :1542–1556, 1987.

[124] P.E. Souganidis. Approximation schemes for viscosity solutions of Hamilton-Jacobi equations. *J. Differential Equations*, 59(1) :1–43, 1985.

[125] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, New-York, 1993.

[126] R. Vinter. *Optimal control*. Systems & Control : Foundations & Applications. Birkhäuser Boston Inc., Boston, MA, 2000.

[127] V. Zeidan. The Riccati equation for optimal control problems with mixed state-control constraints : necessity and sufficiency. *SIAM J. on Control and Optimization*, 32 :1297–1321, 1994.