



SCUOLA INTERNAZIONALE
SUPERIORE di STUDI AVANZATI
International School
for Advanced Studies

JACOBI FIELDS IN OPTIMAL CONTROL

Supervisor

Prof. Andrei A. Agrachev

Candidate

Ivan Yu. Beschastnyi

Thesis submitted for the degree of
“Doctor Philosophiæ”

Academic Year 2017/2018

Contents

Introduction	i
1 Preliminaries	3
1.1 Linear symplectic geometry	3
1.1.1 General definitions	3
1.1.2 Lagrangian Grassmanian	6
1.1.3 Maslov, Kashiwara and Leray indices	9
1.2 \mathcal{L} -derivatives	16
1.3 Chronological calculus	19
1.4 Generalities on sub-Riemannian geometry	20
2 Jacobi curves and Morse-type theory	23
2.1 \mathcal{L} -derivatives for optimal control problems	23
2.2 Approximation algorithm	32
2.3 A quick application: the bang-bang case	37
2.4 Gluing formula	38
2.5 Morse-type theorems	42
3 Jacobi differential equations	57
3.1 Jacobi DEs for good one control parameter systems	57
3.2 The Goh condition and Jacobi DEs for good singular extremals	62
4 Jacobi curves in the simplest singular case	69
4.1 Normal form for the Jacobi DE	70
4.2 Heuristics for one-dimensional variational problems	74
4.3 Kneser oscillation criteria	78
4.4 Model examples	83
4.4.1 Case $m < 2$	83
4.4.2 Case $m = 2$	85
4.4.3 Case $m > 2$	89
4.5 Jacobi curves in the general case for $m \leq 2$	92
5 Geodesic flow and singular trajectories of left-invariant Engel structures	103
5.1 Normal frame and classification of left-invariant Engel structures	103
5.2 Integrability of the geodesic flow	107
5.3 Minimality properties of singular geodesics	110
Bibliography	113

Introduction

Every high-school student knows that given a smooth function $f : \mathbb{R} \rightarrow \mathbb{R}$, in order to find its minimum, one should first find a critical point x_0 , where $f'(x_0) = 0$, and then check that $f''(x_0) > 0$. Under these two conditions we are guaranteed to have a local minimum at x_0 .

This simple fact is just a very special application of the general theory of calculus of variations which was pioneered by the mathematicians of the late 17th-early 18th century with the invention of calculus and developed continuously until the present day.

In the calculus of variations one is interested in finding a minimum of a smooth functional $J : \mathcal{U} \rightarrow \mathbb{R}$ from a possibly infinite dimensional Banach manifold \mathcal{U} to \mathbb{R} . For example, the classical problem of the calculus of variations is to minimize

$$J[q(t)] = \int_0^T L(q, \dot{q}) dt, \quad q \in \mathbb{R}^n$$

with fixed boundary conditions $q(0) = q_0$, $q(T) = q_T$.

The first step is always to ensure that the corresponding minimum exists. If we can prove the existence, the next step usually is to apply the first order necessary conditions, which follow the same idea of taking the first derivative of a functional. Namely, if $u \in \mathcal{U}$ is a minimum, then necessarily

$$dJ[u](w) = 0, \quad \forall w \in T_u\mathcal{U},$$

where $dJ[u]$ is the differential of J at u . In this case $u \in \mathcal{U}$ is called a *critical point* of J or an *extremal point*.

For the classical problem of the calculus of variations an extremal point is a curve $q(t)$. If the Lagrangian L is sufficiently smooth, then one can show that $q(t)$ must satisfy a system of second order non-linear ODEs called the Euler equations or the Euler-Lagrange equations

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^j} = \frac{\partial L}{\partial q^j} \tag{1}$$

in honour of the Leonard Euler who was the first one to derive them and Joseph-Louis Lagrange who extended them to handle constrained variational problems as well.

Once a critical point has been found, like in the case of smooth functions, we consider the second derivative $d^2J[u]$ as a quadratic form on an infinite dimensional space. If $d^2J[q(t)]$ has a negative eigenspace, then using arbitrary small variations one can construct a curve $q(t)'$, s.t. $J[q(t)'] < J[q(t)]$. And so $q(t)$ is not a local minimum. Therefore a necessary condition for a local minimum is $d^2J[u] \geq 0$.

Luckily, for the problem of the classical calculus of variations there is no need in computing the spectra of $d^2J[q(t)]$. Almost a century after Euler published his equations, Jacobi gave a relatively simple sufficient criteria for $q(t)$ not to be a local minimum. Following Jacobi we linearize (1) to obtain what is known as the Jacobi equation

$$\frac{d}{dt} \left(\frac{\partial^2 L}{\partial \dot{q}^j \partial \dot{q}^i} \dot{x}^i + \frac{\partial^2 L}{\partial \dot{q}^j \partial q^i} x^i \right) = \frac{\partial^2 L}{\partial q^j \partial \dot{q}^i} \dot{x}^i + \frac{\partial^2 L}{\partial q^j \partial q^i} x^i. \tag{2}$$

A moment of time t_c is called *conjugate*, if there exists a solution of (2) with $x(0) = x(t_c) = 0$. The point $q(t_c)$ is called a *conjugate point*, and the number of solutions to the previous boundary-value problem is called the *multiplicity of $q(t_c)$* . Jacobi proved that whenever along an extremal curve quadratic form $L_{\dot{q}\dot{q}}$ is positive, if $q(t)$ contains a conjugate point, then it is not a local minimum. Inequality $L_{\dot{q}\dot{q}} > 0$ is known as the *strict Legendre condition*. A precise relation between the theory of conjugate points and the second differential of J was given later by Morse.

Theorem (Morse, [38]). *Under the strong Legendre conditions*

$$\text{ind}^- d^2 J[q(t)] = \#\{\text{conjugate points along } q(t) \text{ counted with multiplicities}\},$$

$$\text{ker } d^2 J[q(t)] = \text{multiplicity of } q(T).$$

The importance of the second variations comes not just from the fact that it allows to obtain necessary conditions for local minima. These conditions can be often reinterpreted as stability conditions. To see how it works, let us consider a classical problem from the theory for elasticity:

“Given an elastic rod of certain length what shape can it take?”

First successful attempts to solve this problem were done by the Bernoulli family [33]. James Bernoulli has noted that if the rod is confined to lie in a plane, then the equation for the elastic curve can be formulated in terms of its curvature. He was able to integrate those equations in quadratures, but due to difficulty of his method he was not able to analyse properly the solutions. This became possible after his nephew Daniel Bernoulli formulated the problem as a minimization of bending energy, that can be stated as follows:

“Among all curves of the same length l that not only pass through points A and B , but are also tangent to given straight lines at these points, find the one minimizing the value of

$$J[\kappa(s)] = \int_0^l \kappa(s)^2 ds$$

where $\kappa(s)$ is the curvature of the curve.”

He proposed this problem to Euler [18], who used his equation to derive the equations of elastic rods in a nice form, to solve them in quadratures and to describe qualitatively different types of curves that can occur (see Figure 1). Today we call such curves Euler Elasticas.

In his book [27], Euler gives a qualitative picture of all the solutions of the Euler-Lagrange equations. Interestingly enough, there are some configurations that can not be observed in nature, like one in Figure 2 on the left. This is because most of the solutions of the Euler-Lagrange equation correspond to non-stable profiles of the rod. For example, if we constrain an elastic rod to have a form like in Figure 2 on the left, as soon as we remove

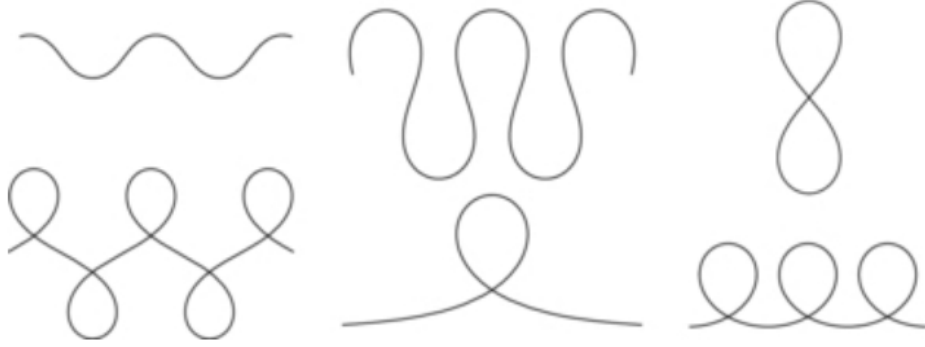


Figure 1: Different types of Euler's elastica curves. The picture is taken from [50]

the constraints it will take a different form, similar to the picture on the right. From the variational point the reason for this is simple. The left profile contains a conjugate point, while the right one does not (see [44] for the proof of this fact).

In the second part of the 20th century there has been a renewed interest in this classical methods due to the emergence of symplectic geometry [14, 25]. It turns out that the whole theory mentioned so far can be naturally reinterpreted using the language of symplectic geometry and extended much further. For example, it was known for a long time that under the strict Legendre condition the Euler-Lagrange equation is equivalent to a Hamiltonian system

$$\begin{aligned}\dot{q}^i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q^i}\end{aligned}$$

with a Hamiltonian

$$H = p_i \dot{q}^i - L(q, \dot{q}), \quad p_i = L_{\dot{q}^i}.$$

The Legendre condition ensures that we can resolve at least locally $p_i = L_{\dot{q}^i}$ with respect to \dot{q} and obtain a Hamiltonian $H(p, q)$ that depends only on the phase variables.

If we introduce a *complex structure* given by a matrix

$$J = \begin{pmatrix} 0 & \text{id}_n \\ -\text{id}_n & 0 \end{pmatrix},$$

then we can rewrite the Hamiltonian system as

$$\frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = -J \nabla H(p, q). \quad (3)$$

Given a solution $(p(t), q(t))$ of this system, we can rewrite the Jacobi equation along the extremal curve $q(t)$, by simply linearizing (3) along $(p(t), q(t))$:

$$\frac{d}{dt} \begin{pmatrix} y \\ x \end{pmatrix} = -J \left(\nabla_{(p(t), q(t))}^2 H \right) \begin{pmatrix} y \\ x \end{pmatrix}, \quad (4)$$

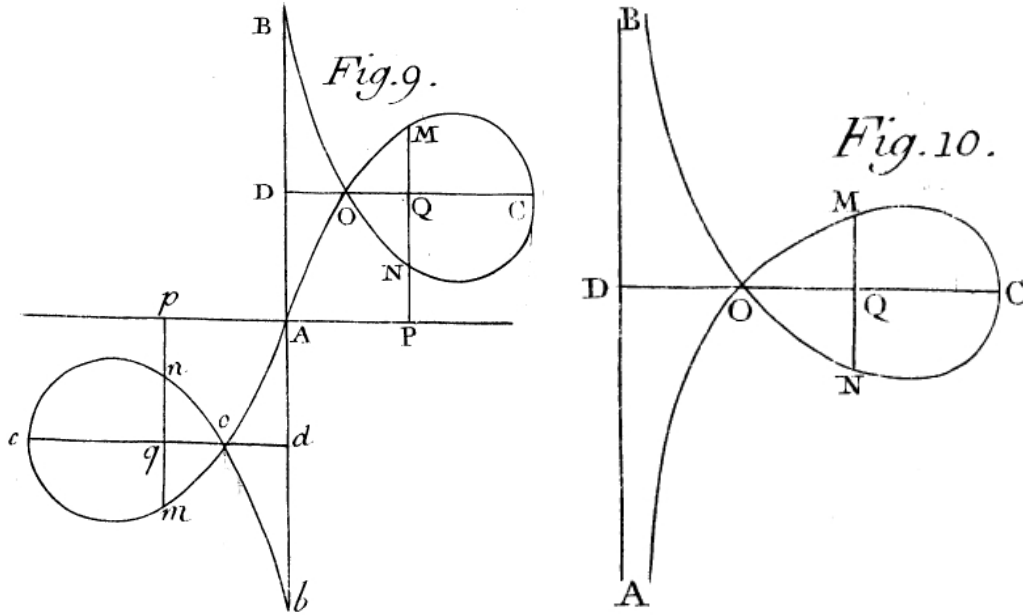


Figure 2: Original pictures from Euler's book [27]

which is again a Hamiltonian system with a Hamiltonian given by the Hessian along $(p(t), q(t))$.

Complex structure J gives rise to a symplectic form

$$\sigma(\lambda_1, \lambda_2) = \lambda_1^T J \lambda_2.$$

A plane $L \in \mathbb{R}^{2n}$ is called *isotropic* if the restriction $\sigma|_L$ is zero. A *Lagrangian* subspace L is the maximal isotropic subspace, which means that $\sigma|_L = 0$ and $\dim L = n$. For example it is easy to see that the $\Pi = \{(y, 0) \in \mathbb{R}^{2n}\}$ is Lagrangian. The set of Lagrangian planes is called the *Lagrangian Grassmanian*.

Since Jacobi equation (4) is symplectic, its flow Φ_t preserves the symplectic form and maps Lagrangian planes to Lagrangian planes. Thus by fixing an initial Lagrangian plane $\mathcal{L}_0 = \Pi$ we get a curve $\mathcal{L}_t = \Phi_t(\Pi)$ in the Lagrangian Grassmanian that is known as the *Jacobi curve*. Then a moment of time t is conjugate if and only if $\mathcal{L}_t \cap \Pi \neq \{0\}$ and the multiplicity of the corresponding conjugate point is given by $\dim(\mathcal{L}_t \cap \Pi)$.

The set of all Lagrangian planes that have a non zero intersection with a fixed Lagrangian plane Π is called the *Maslov train* \mathcal{M}_Π . We have a conjugate point whenever our curve \mathcal{L}_t crosses it. The set \mathcal{M}_Π is an algebraic hypersurface in the Lagrange Grassmanian with codimension three singularities and a coorientation. Therefore there is a well defined intersection index with curves, which is called the *Maslov index*. We will give precise definitions later. Thus we can reformulate the Morse index theorem as follows.

Theorem 0.1. *If q_T is not a conjugate point, then $\text{ind}^- d^2 J[\tilde{q}(t)]$ is equal to the Maslov index of the curve $\mathcal{L}_t|_{[\varepsilon, T]}$ for some ε small enough.*

It may seem just a reformulation of known results, but it actually gives us new tools to handle old problems. For example, Maslov index is a homotopy invariant. This fact was used quite recently to give a description of conjugate points and therefore stability criteria for the Euler elastica problem mentioned earlier [44].

We can take the last theorem and use it to formulate a guiding principle

"Morse index of the Hessian at an extremal is equal to the Maslov index of the corresponding Jacobi curve"

The goal of this thesis is to explore how general this principle is. And it turns out that it is surprisingly very general. We just need the right definition for "Jacobi curve" and "Maslov index". The need for an extension of the classical theory comes from constrained variational problems and especially control theory.

To illustrate this, let us consider the following problem. We would like to minimize

$$\frac{1}{2} \int_0^T (q^1)^2 dt \rightarrow \min,$$

subject to

$$\dot{q}^1 = q^2, \quad \dot{q}^2 = u, \quad |u| \leq 1,$$

In this case extremal curves satisfy a Hamiltonian system as well, but the Hamiltonian now is discontinuous. If we define the two Hamiltonians

$$H_{\pm}(p, q) = p_1 q^2 \pm p_2 - \frac{(q^1)^2}{2},$$

then if $p_2 \neq 0$ the Hamiltonian, that we need to use, is given by the max of two, i.e.

$$H = \max\{H_+, H_-\}.$$

It turns out that for T sufficiently large, the optimal solution is the following. One should alternate between controls $u = \pm 1$ that correspond to H_{\pm} infinite number of times in a finite period of time until we reach the origin. Then we stay for some time in the origin, and then we should alternate between controls $u = \pm 1$ an infinite number of times in order to exit from the origin.

Here we immediately encounter a number of problems. Even if we suppose that the linearization is the right way to define the Jacobi curve, there is clearly an uncertainty at the accumulation time when we hit the origin. And even if we can resolve it, clearly the Jacobi curve is going to be discontinuous, but the Maslov index is a homotopy invariant of continuous curves. Is it possible to define it for discontinuous curves as well? All of these problems can be resolved if we find the right definitions for all the objects involved.

There have been a number of results in specific cases. For example, in article [42] the authors study the bang-bang case, in [31] singular curves and in [17] bang-singular arcs.

The theory presented in this thesis allows to unify all the previous results and derive them as special cases of this general framework.

To be a little bit more precise, in this thesis we consider an optimal control problem of the form

$$\dot{q} = f(u, q), \quad u \in U \subset \mathbb{R}^k, \quad q \in M,$$
$$\int_0^T L(u, q) dt \rightarrow \min.$$

where M is a smooth manifold, $f(u, q)$ and $L(u, q)$ are smooth in both variables, final time T is fixed, $q(0)$ lies in some submanifold N_0 , and $q(T) = q_T$ is fixed. We assume that the controls $u(t)$ lie in $L_k^\infty[0, T]$ as functions of time. We show how to define the Jacobi curves for extremals of this kind of problems and how to extract the information about the Hessian using symplectic invariants.

The thesis has the following structure. In Chapter 1 some background material is given. First we discuss linear symplectic geometry and various invariants of the linear symplectic group, including the Maslov index. Some good materials on the subject include [23, 34] and [10] for the positive Maslov index.

Then we define and discuss \mathcal{L} -derivatives, which will be our principal tool. They were introduced and studied in [1]. A \mathcal{L} -derivative is a map that assigns to a space of variations in a specific constrained variational problem a Lagrangian subspace in some symplectic space. By comparing the relative positions of those Lagrangian subspaces we obtain information about how the Morse index of the Hessian changes as we add new variations. Given a one-parametric family of variations, we obtain a one-parametric family of \mathcal{L} -derivatives that will be Jacobi curves that we seek. This will be the right definition and we will see in subsequent chapters that the Jacobi equation should be thought just as a tool to compute them. Then we briefly discuss chronological calculus and basic definitions from sub-Riemannian geometry.

In Chapter 2 we give a general definition of Jacobi curves that can be applied to any optimal control problem and any type of extremals. We show then how approximately compute them with arbitrary good precision and as an application we compute them for bang-bang extremals. Second order necessary and sufficient optimality conditions for bang-bang extremals were previously studied in a number of works of different groups [42, 45, 11]. The novelty of our approach is that it gives a unified way to construct Jacobi curves for any type of extremals. The algorithm that we present works for any possible type of extremals.

We then prove the glueing formula that allows us to reconstruct the Jacobi curves defined on an interval from separate Jacobi curves on the subintervals. This technique can be used to construct Jacobi curves in the presence of Fuller phenomena. In the last section of that chapter we prove Morse-type theorems that relate the index or the nullity of the Hessian to some symplectic invariants of the Jacobi curve or its approximation, which extends greatly the results (see [14, 25]). Some results in this direction for specific types of extremals in optimal control theory were given in [31, 17, 16, 45, 8] and other articles of the same authors. Most of them can be derived as a special case of the Morse

theorem that we prove.

In Chapter 3 we look how one can derive the usual Jacobi equations for both singular and regular extremals directly from the definitions. In that chapter we give a geometric interpretation of the classical Goh and generalized Legendre-Clebsch conditions. We give a completely new proof of the Goh conditions that works well for high-order singular extremals. Some of the results of this chapter were previously derived in other works (see for example [19, 8]). The goal of this chapter is to show how one can use the technique of \mathcal{L} -derivatives to derive classical results, simplify proofs and generalize them to more difficult cases.

In Chapter 4 we compute Jacobi curves for the simplest singular case when the strict Legendre-Clebsch conditions fails at a single point. Variational problems with similar singularities were studied mainly by Morse himself and some of his students [40, 39, 32]. They considered the usual problems of calculus of variations and used essentially functional analytic techniques. Recently some singular situations were considered in [21, 4]. In that chapter we consider optimal control problems with a single control parameter, which seems to not have been considered previously in the literature. We show that using the definition of \mathcal{L} -derivative, one can give a dynamical characterization of the Jacobi curve as a certain boundary value problem.

In the last Chapter 5 we study geodesics of Engel structures, which are 4-dimensional sub-Riemannian manifolds endowed with a rank two distribution. We show in which cases the normal flow is integrable and as application of the previous results we study C^0 -local optimality of abnormal geodesics using Jacobi equations. In particular we give many new examples of sub-Riemannian structures with integrable geodesic flows and strictly abnormal geodesics (the only previous known example was given in [37]). Later as an application of the previous results we derive some simple comparison theorems for conjugate times along abnormal extremals.

The results presented in this PhD thesis appear in the following works:

- A.A. Agrachev, I. Yu. Beschastnyi. *'Symplectic Geometry of Constrained Optimization'*, Regular and Chaotic Dynamics, vol 22, no. 6, 2017, p. 750–770;
- I. Beschastnyi, A. Medvedev. *'Left-invariant sub-Riemannian structures'*, SIAM J. on Control and Optimization, accepted. arXiv:1611.03634.
- A.A. Agrachev, I. Yu. Beschastnyi. *'Jacobi fields in optimal control I: Morse and Maslov indices'*, in preparation.
- A.A. Agrachev, I. Yu. Beschastnyi. *'Jacobi fields in optimal control II: One-dimensional variations'*, in preparation.

One more article that is not included in the current thesis, but was prepared during the PhD is

- D. Barilari, I. Beschastnyi, A. Lerario. *'Volume of small balls and sub-Riemannian curvature in 3D contact case'*, Journal of Symplectic Geometry, accepted. arXiv:1802.10155.

Acknowledgments

I would like to first of all thank my science adviser prof. Andrey Agrachev for his patience and knowledge that he shared with me during these four years and my dear collaborators and friends: Alexander Medvedev, Antonio Lerario and Davide Barilari and prof. Sachkov without whom this thesis would be impossible.

I would like to give a special thanks to prof. Ugo Boscain for many fruitful discussions that we had, and the opportunity to meet the wonderful mathematical community of Paris.

I would also like to say thank you to all the great people I met in Trieste and especially to Gabriele Bogo, Andrea Maiorana, the Khozasov family, Giordanno Cotti, Paolo Boanicato, Elio Marconi, Elena Facco, Elena Tartaglia, Veronica Fantini, Carlo Scarpa, Ulisse Bocchero, Fiamma Romagnoli, Guglielmo Costa, Paolo Campetti and to all of my dear Berimbau friends.

I am very grateful to my father and my sister, and to Stella Valentina Paronuzzi Ticco for their constant support and encouragement.

Preliminaries

1.1 Linear symplectic geometry

In this section we give a some basic definitions from symplectic geometry. For further results and proofs see [23, 36, 29].

1.1.1 General definitions

A *symplectic space* is a pair (Σ, σ) that consists of an even-dimensional real vector space and a non-degenerate skew-symmetric bilinear form σ . One can always choose a basis in Σ , s.t. σ is of the form

$$\sigma(\lambda_1, \lambda_2) = \lambda_1^T J \lambda_2, \quad \forall \lambda_i \in \Sigma,$$

where J is the standard complex structure

$$J = \begin{pmatrix} 0 & \text{id}_n \\ -\text{id}_n & 0 \end{pmatrix}.$$

In particular $J^2 = -\text{id}_{2n}$. Coordinates in which σ has such a form are called *Darboux coordinates*. We use usual position-momenta notations in this case, i.e. we write $\lambda = (p, q) = (p_1, \dots, p_n, q^1, \dots, q^n)$.

In Darboux coordinates a Hamiltonian system with a perhaps time-dependent Hamiltonian $H : \Sigma \times \mathbb{R} \rightarrow \mathbb{R}$ is a system of ODEs

$$\dot{\lambda} = -J \nabla H(t, \lambda),$$

where ∇H is the gradient of H . In particular, if H is a quadratic form

$$H(t, \lambda) = p^T C(t) p - 2q^T A(t) p - q^T B(t) q,$$

where $B(t), C(t)$ are symmetric matrices, then we obtain a linear Hamiltonian system

$$\frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} A(t) & B(t) \\ C(t) & -A^T(t) \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}. \quad (1.1)$$

Given J we can define the *symplectic group* $\text{Sp}(2n)$ and the corresponding *symplectic algebra* $\text{sp}(2n)$ as

$$\text{Sp}(2n) = \{ F \in \text{Mat}(2n \times 2n, \mathbb{R}) : F^T J F = J \},$$

$$\mathfrak{sp}(2n) = \{X \in \text{Mat}(2n \times 2n, \mathbb{R}) : X^T J + JX = 0\}.$$

If we write down $X \in \mathfrak{sp}(2n)$ as a block matrix, we will see that it has the same form as the matrix in the Hamiltonian system (1.1). Therefore we immediately can see that the flow $\Phi(t)$ of (1.1) is symplectic.

We define the *skew-orthogonal complement* of a subspace Γ in a symplectic space Σ as a subspace

$$\Gamma^\perp = \{\lambda \in \Sigma : \sigma(\lambda, \mu) = 0, \forall \mu \in \Gamma\}.$$

One has the following special situations

- If $\Gamma \subset \Gamma^\perp$, then Γ is called *isotropic*;
- If $\Gamma \supset \Gamma^\perp$, then Γ is called *coisotropic*;
- If $\Gamma = \Gamma^\perp$, then Γ is called *Lagrangian*.

From the definition we can see, that Γ is isotropic if and only if the restriction $\sigma|_\Gamma$ vanishes. Since σ is non-degenerate, we have

$$\dim \Gamma + \dim \Gamma^\perp = \dim \Sigma.$$

Therefore a subspace Γ is Lagrangian if and only if Γ is isotropic and has dimension $(\dim \Sigma)/2$.

Since σ is skew-symmetric, any one-dimensional direction $\mathbb{R}v$, $v \in \Sigma$ is isotropic. For the same reasons any codimension one subspace is coisotropic. Two main examples of Lagrangian subspaces are the *horizontal subspace* Ξ and the *vertical subspace* Π defined as

$$\begin{aligned} \Pi &= \{(p, q) \in \Sigma : q = 0\}, \\ \Xi &= \{(p, q) \in \Sigma : p = 0\}. \end{aligned}$$

We can construct other examples as follows. Let $S = S^T$ be a symmetric matrix. Then

$$\Lambda_S = \{(p, Sp) \in \Sigma : p \in \mathbb{R}^n\}$$

is a Lagrangian subspace transversal to Ξ . Conversely to any Lagrangian subspace Λ transversal to Ξ (we denote this by $\Lambda \pitchfork \Xi$) we can associate a symmetric operator S from Π to Ξ .

There exists a close relation between symplectic maps and Lagrangian subspaces. Given (Σ, σ) we can construct a new symplectic space $(\Sigma \times \Sigma, -\sigma \oplus \sigma)$ of double dimension. It can be used to give an alternative definition of a symplectic map.

Proposition 1.1. *Let $F : \Sigma \rightarrow \Sigma$ be a linear map. F is symplectic if and only if the graph of F in $(\Sigma \times \Sigma, -\sigma \oplus \sigma)$ is Lagrangian.*

The square of a symplectic space will play later an important role in this work. So it makes sense to characterize all Lagrangian planes in $(\Sigma \times \Sigma, -\sigma \oplus \sigma)$. We have seen that a graph of a symplectic space gives an example of a Lagrangian subspace. Another example is just a direct product of two Lagrangian subspaces $\Lambda_1 \times \Lambda_2$. It turns out that the general situation is an interpolation between those two.

Lemma 1.2. *Any Lagrangian subspace Λ in $(\Sigma \times \Sigma, -\sigma \times \sigma)$ is a direct sum of three subspaces $(\Gamma_1, 0) \oplus \text{graph}(\Phi_{12}) \oplus (0, \Gamma_2)$, where $\Gamma_i \subset \Sigma$ are isotropic of the same dimension and $\text{graph}(\Phi_{12})$ is the graph of some symplectic map $\Phi_{12} : \Gamma_1^\perp/\Gamma_1 \rightarrow \Gamma_2^\perp/\Gamma_2$. Conversely, given two isotropic spaces $\Gamma_i \subset \Sigma$ and a symplectic map $\Phi_{12} : \Gamma_1^\perp/\Gamma_1 \rightarrow \Gamma_2^\perp/\Gamma_2$, the space indicated above is going to be Lagrangian.*

Proof. The second part is a straightforward computation. Let us prove the first part. We denote by π_i projections into each factor. It is clear from the definitions that $\Gamma_1 = \pi_1(\ker \pi_2|_\Lambda)$ and $\Gamma_2 = \pi_2(\ker \pi_1|_\Lambda)$ are isotropic subspaces of Σ . We can naturally identify the quotient $((\Gamma_1, 0) \oplus (0, \Gamma_2))^\perp / ((\Gamma_1, 0) \oplus (0, \Gamma_2))$ with $\Gamma_1^\perp/\Gamma_1 \times \Gamma_2^\perp/\Gamma_2$. But since $((\Gamma_1, 0) \oplus (0, \Gamma_2)) \subset \Lambda$, we have that $\Lambda' = \Lambda / ((\Gamma_1, 0) \oplus (0, \Gamma_2))$ can be identified with a Lagrangian subspace in $\Gamma_1^\perp/\Gamma_1 \times \Gamma_2^\perp/\Gamma_2$. Moreover Λ' is a linear subspace that projects onto Γ_i^\perp/Γ_i . But this can happen if and only if $\dim \Gamma_1^\perp/\Gamma_1 = \dim \Gamma_2^\perp/\Gamma_2$. Therefore Λ' must be a graph of a symplectic mapping and $\dim \Gamma_1 = \dim \Gamma_2$. \square

We can extended all these definitions to the non-linear setting. A *symplectic manifold* is a pair (N, σ) , where N is a smooth manifold and σ is a closed non-degenerate differential two-form. Similar to the linear case, one can show that locally all symplectic manifolds have the same structure.

Theorem 1.3 (Darboux). *For any point x of a symplectic manifold (N, σ) one can find a neighbourhood U and a local diffeomorphism $\psi : U \rightarrow \mathbb{R}^{2n}$, s.t.*

$$\sigma = \psi^*(dp_i \wedge dq^i),$$

where (p, q) are coordinates in \mathbb{R}^{2n} .

Note that a tangent space $T_x N$ has naturally a structure of a symplectic space. Therefore we can say that a submanifold $M \subset N$ is *isotropic/coisotropic/Lagrangian* if the same property is true for each subspace $T_x M \subset T_x N$ for all $x \in M$. Similarly to the linear case a submanifold M is isotropic if and only if $\sigma|_M = 0$ and Lagrangian if additionally $\dim M = (\dim N)/2$.

A *symplectomorphism* of (N, σ) is a smooth map $f : N \rightarrow N$, that preserves the symplectic structure, i.e.

$$f^* \sigma = \sigma.$$

Given a smooth function $h : N \rightarrow \mathbb{R}$, a *Hamiltonian vector field* \vec{h} is defined by the identity $dh = \sigma(\cdot, \vec{h})$. The flow generated by the *Hamiltonian system* $\dot{x} = \vec{h}(x)$ preserves the symplectic structure. In Darboux coordinates, Hamiltonian system has the form:

$$\dot{p} = -\frac{\partial h}{\partial q}, \quad \dot{q} = \frac{\partial h}{\partial p}.$$

The non-linear analogue of Proposition 1.1 holds as well

Proposition 1.4. *A diffeomorphism $f : N \rightarrow N$ of a symplectic manifold (Σ, σ) is a symplectomorphism if and only if the graph of f in $(N \times N, -\sigma \oplus \sigma)$ is a Lagrangian submanifold.*

The most basic and important examples of symplectic manifolds are the cotangent bundles T^*M . To define invariantly the symplectic form we use the projection map $\pi : T^*M \rightarrow M$. Its differential is a well defined map $\pi_* : T(T^*M) \rightarrow TM$. We can define the *Liouville one-form* $s \in \Lambda^1(T^*M)$ at $\lambda \in T^*M$ as

$$s_\lambda = \lambda \circ \pi_*.$$

Then the canonical symplectic form on T^*M is simply given by the differential $\sigma = ds$.

In local coordinates T^*M is locally diffeomorphic to $\mathbb{R}^n \times \mathbb{R}^n$ with coordinates (p, q) , where q are coordinates on the base and p are coordinates on the fibre. In these coordinates the Liouville form s is written as $s = p_i dq^i$. Thus (p, q) are actually Darboux coordinates. We can use this fact to construct many Lagrangian manifolds. Namely

Proposition 1.5. *Let $S : M \rightarrow \mathbb{R}$ be a smooth function. Then the graph of the differential $d_q S$ is a Lagrangian submanifold in T^*M .*

The proof is a straightforward computation in the Darboux coordinates and follows from the commutativity of the second derivative of S .

1.1.2 Lagrangian Grassmanian

The set of all Lagrangian planes of a symplectic space (Σ, σ) is called the *Lagrangian Grassmanian*, and we denote it by $L(\Sigma)$. We also use notation Π^\natural for the set of all Lagrangian planes transversal to a given Lagrangian plane $\Pi \in L(\Sigma)$.

The set $L(\Sigma)$ is a manifold, whose atlas is given by Λ^\natural . Coordinate charts are maps from Λ^\natural to the space of symmetric matrices constructed like in the previous subsection. Let $\dim(\Sigma) = 2n$. Throughout this thesis we will also use another representation of a Lagrangian plane $\Lambda \in L(\Sigma)$ as a span of n independent vectors v_i . It is clear that such a representation is not unique. We can replace v_i by any linear span of the same vectors as long as they remain independent. This means that in general we need to quotient a natural $\mathrm{GL}(n)$ action. We can arrange v_i in a single $n \times 2n$ matrix and we write

$$\Lambda = [v_1 \quad \dots \quad v_n],$$

where the square brackets indicate the equivalence class under the $\mathrm{GL}(n)$ action. We denote this action by

$$g [v_1 \quad \dots \quad v_n] := [gv_1 \quad \dots \quad gv_n], \quad g \in \mathrm{GL}(n).$$

For example, if $\Lambda \in \Sigma^\natural$ we can write

$$\Lambda = \begin{bmatrix} \mathrm{id}_n \\ S \end{bmatrix},$$

where S is a symmetric matrix like in the example above. Or we can assume that v_i form an orthonormal basis of Λ in \mathbb{R}^{2n} . Then

$$\Lambda = \begin{bmatrix} X \\ Y \end{bmatrix},$$

where $X^T X + Y^T Y = \text{id}_n$ (orthonormality property) and $X^T Y - Y^T X = 0$ (Lagrangian property) are satisfied. A matrix $X + iY$ that satisfies these properties is unitary and the converse is true as well. We can choose v_i in such a way up to a $O(n)$ -action, which is given by

$$\begin{pmatrix} O & 0 \\ 0 & O \end{pmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}, \quad O \in O(n).$$

This gives the usual identification of $L(\Sigma) \simeq U(n)/O(n)$.

We will use this idea many times when we will consider the singular case, so at this point it makes sense to consider a simple example that will be useful for us later.

Example 1.6. Suppose that we would like to find a simple representation of a Lagrangian plane $\Lambda \in L(\mathbb{R}^4)$ knowing that $\dim(\Lambda \cap \Xi) = 1$. in this case it must be of the form

$$\Lambda = [v_1 \ v_2] = \begin{bmatrix} x_1 & 0 \\ y_1 & 0 \\ z_1 & z_2 \\ w_1 & w_2 \end{bmatrix}, \quad x_1^2 + y_1^2 \neq 0.$$

We can assume that v_1 and v_2 are orthonormal. We then apply a rotation $O \in O(n)$, so that y component of v_1 becomes zero. Then

$$\Lambda = [Ov_1 \ Ov_2] = \begin{bmatrix} \sqrt{x_1^2 + y_1^2} & 0 \\ 0 & 0 \\ \tilde{z}_1 & \tilde{z}_2 \\ \tilde{w}_1 & \tilde{w}_2 \end{bmatrix},$$

but since Λ is Lagrangian we must have $\tilde{z}_2 = 0$. Changing the basis we then find

$$\Lambda = \begin{bmatrix} Ov_1 - (\tilde{w}_1/\tilde{w}_2)Ov_2 & Ov_2 \\ \sqrt{x_1^2 + y_1^2} & \tilde{w}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ z & 0 \\ 0 & 1 \end{bmatrix}.$$

Given an isotropic subspace Γ and a Lagrangian plane Λ , we can construct a new Lagrangian plane Λ^Γ , which is a Lagrangian plane that contains Γ and the dimension of $\Lambda \cap \Lambda^\Gamma$ is maximal. It is defined as

$$\Lambda^\Gamma = (\Lambda \cap \Gamma^\perp) + \Gamma = (\Lambda + \Gamma) \cap \Gamma^\perp.$$

If $\Gamma = \mathbb{R}X$ for some vector $X \in \mathbb{R}^{2n}$ we will simply write Λ^X instead of $\Lambda^{\mathbb{R}X}$.

Let us have a look at another example that will be useful in future.

Example 1.7. Assume that $X = (1 \ 0 \ 0 \ 0)^T$ and we would like to construct Λ^X for $\Lambda \in L(\mathbb{R}^4)$. We have that

$$\Lambda = [v_1 \ v_2] = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \\ z_1 & z_2 \\ w_1 & w_2 \end{bmatrix}.$$

Subspace X^\perp consists of vectors $v \in \mathbb{R}^4$ whose z -component is zero. So assume first that $z_1 = z_2 = 0$. Then $\sigma(X, v_1) = \sigma(X, v_2) = 0$. But since it is a Lagrangian subspace, it means that $X \in \Lambda$ and by definition $\Lambda^X = \Lambda$. Thus we can take $v_1 = X$. In this case we obtain

$$\Lambda^X = [X \ v_2] = [X \ v_2 - x_2 X] = \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix}.$$

Suppose that $X \notin \Lambda$. Then $z_1^2 + z_2^2 \neq 0$ and as a result $\sigma(X, z_1 v_1 + z_2 v_2) \neq 0$, but $\sigma(X, z_1 v_2 - z_2 v_1) = 0$. So $\Lambda \cap X^\perp = z_1 v_2 - z_2 v_1$ and by definition

$$\Lambda^X = [X \ z_1 v_2 - z_2 v_1] = \begin{bmatrix} 1 & z_1 x_2 - z_2 x_1 \\ 0 & z_1 y_2 - z_2 y_1 \\ 0 & 0 \\ 0 & z_1 w_2 - z_2 w_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & z_1 y_2 - z_2 y_1 \\ 0 & 0 \\ 0 & z_1 w_2 - z_2 w_1 \end{bmatrix}.$$

We will often use the following simple, but important lemma.

Lemma 1.8. *For any countable set $S \in L(\Sigma)$ the set of Lagrangian planes, that intersect transversally any Lagrangian plane from S , is dense in $L(\Sigma)$.*

Proof. The proof is a simple consequence of the Baire category theorem. Indeed, we know that $\Lambda^\#$ is an open set in $L(\Sigma)$. One should just prove that those sets are dense, then the intersection

$$\bigcap_{\Lambda \in S} \Lambda^\#$$

must be dense and therefore non-empty. □

The tangent space $T_\Lambda L(\Sigma)$ can be identified with the space $\text{Sym}(\Lambda)$ of all symmetric quadratic forms on Λ . Indeed, given Λ take any curve $\Lambda(\varepsilon) \in L(\Sigma)$, s.t. $\Lambda(0) = \Lambda$ and fix $\lambda(0) \in \Lambda(0)$ and take any curve $\lambda(\varepsilon) \in \Lambda(\varepsilon)$. Then we identify $\Lambda(0)$ with the quadratic form $\sigma(\lambda(0), \dot{\lambda}(0))$. An easy calculation shows that the definition does not depend on the choice of the curve $\lambda(\varepsilon)$.

Definition 1.9. We say that a C^1 -curve $\Lambda(t) \in L(\Sigma)$ is *monotone increasing* if the corresponding matrix $\dot{\Lambda}(t) \geq 0$ as a quadratic form on $\Lambda(t)$ for every t . We say that it is *strictly monotone* if this inequality is strict.

Monotone decreasing curves are defined in a similar way, but it turns out that C^1 -smooth Jacobi curves of minimum problems, that we will define in Chapter 2, are always monotone increasing.

In some particular cases the Jacobi curves will come from a flow $\Phi(t)$ of a linear Hamiltonian system (1.1). We will simply take a point $\Lambda \in L(\Sigma)$ and consider a curve $\Lambda(t) = \Phi(t)\Lambda$. More generally a linear Hamiltonian system induces a dynamical system on $L(\Sigma)$. We can write down an ODE for that system using local charts. Indeed, let $(p(t), q(t))$ be a solution of (1.1) and $S(t)$ be a curve of symmetric matrices that correspond to $\Lambda(t)$. Then $q(t) = S(t)p(t)$ and we differentiate this expression. This way we obtain a Riccati equation of the form

$$\dot{S} + SA + A^T S + SBS - C = 0. \quad (1.2)$$

Since a coordinate chart Ξ^{th} is dense in $L(\Sigma)$ the opposite is also true: a Riccati equation of the form above gives rise to a Hamiltonian system and a well defined flow on $L(\Sigma)$. In order to write down a Riccati equation in a different chart we can apply a symplectic transformation to the corresponding Hamiltonian, s.t. a given Lagrangian plane Λ is mapped to Ξ . Then we simply insert the new expressions for A, B and C in (1.2).

1.1.3 Maslov, Kashiwara and Leray indices

The *Maslov train* \mathcal{M}_Π is the set of all Lagrangian planes non-transversal to $\Pi \in L(\Sigma)$, i.e. $\mathcal{M}_\Pi = L(\Sigma) \setminus \Pi^{\text{th}}$. This is a stratified manifold where each strata \mathcal{M}_Π^k is the set of all $\Lambda \in L(\Sigma)$ s.t. $\dim(\Lambda \cap \Pi) = k$. The dimension of each strata is

$$\dim \mathcal{M}_\Pi^k = \frac{n(n+1)}{2} - \frac{k(k+1)}{2}.$$

We can see that the highest dimensional strata \mathcal{M}_Π^1 has codimension one in $L(\Sigma)$. To define an intersection index we need to define a co-orientation on \mathcal{M}_Π^1 . Suppose that $\Lambda_\varepsilon \in L(\Sigma)$ intersects \mathcal{M}_Π^1 transversally at $\varepsilon = 0$, i.e. there exists unique up to a scalar factor $\lambda \in \Lambda(0) \cap \Pi$. We define a positive co-orientation when

$$\dot{\Lambda}_0(\lambda) > 0.$$

Similarly one defines a negative co-orientation.

Definition 1.10. *The Maslov index* $\text{Mi}_\Pi(\Lambda(t))$ of a curve $\Lambda(t)$ is the intersection number of $\Lambda(t)$ with \mathcal{M}_Π^1 . For a curve in general position this is just a number of intersections $\Lambda(t) \cap \Pi$ counted with signs.

Since \mathcal{M}_Π^2 has codimension three, the Maslov index is well defined and it is a homotopy invariant. It turns out, that it is a complete homotopic invariant for loops in $L(\Sigma)$. More precisely we have two very important properties of the Maslov index related to a change of the reference plane.

Theorem 1.11 ([6]). *If a curve $\Lambda(t) \subset L(\Sigma)$ is closed, then its Maslov index does not depend on the choice of the reference plane, i.e.*

$$\text{Mi}_{\Delta_1} \Lambda(t) = \text{Mi}_{\Delta_2} \Lambda(t) = \text{Mi} \Lambda(t), \quad \forall \Delta_i \in L(\Sigma).$$

Moreover Mi induces an isomorphism $\pi_1(L(\Sigma)) \rightarrow \mathbb{Z}$.

If $\Lambda(t)$ is not closed we have the following estimate

$$|\text{Mi}_{\Delta_1} \Lambda(t) - \text{Mi}_{\Delta_2} \Lambda(t)| \leq n.$$

In the cases when a curve in $L(\Sigma)$ comes from a flow of a linear Hamiltonian system, its Maslov index can be computed directly from the Hamiltonian.

Theorem 1.12. *Let $H(t)$ be a quadratic non-autonomous Hamiltonian and let $\Phi(t)$ be a flow of the corresponding Hamiltonian system. Fix two transversal Lagrangian planes Δ and Λ and assume that $\Delta \pitchfork \Phi(T)\Lambda$. If $H(t)|_{\Delta} \geq 0$ for $t \in [0, T]$, then*

$$\text{Mi}_{\Delta} \Phi(t)\Lambda = \sum_{t \in [0, T]} \dim(\Delta \cap \Phi(t)\Lambda).$$

If $H(t)|_{\Delta} \leq 0$, then the same formula holds with a minus in front of the sum.

The proof of this theorem is a straightforward application of definitions. These last two theorems will be our main tool in proving oscillation results in Section 4.3.

The given definition is very useful in many theoretical studies, but not very convenient for computations, since one needs to put the curve in a general position and verify that the boundary points are not in \mathcal{M}_{Π} . To overcome this, one usually uses some other symplectic invariants of Lagrangian planes and curves. We will need the Kashiwara index $\text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3)$ of a triple of Lagrangian planes $\Lambda_i \in L(\Sigma)$ and the Leray index $\text{Li}(\tilde{\Lambda}_1, \tilde{\Lambda}_2)$ of two points $\tilde{\Lambda}_i$ in the universal cover $\widetilde{L(\Sigma)}$ to state and prove the main Morse index theorem, but for many intermediate steps it is more useful to use an index introduced in [10].

To define it we take three Lagrangian planes $\Lambda_1, \Lambda_2, \Pi \in L(\Sigma)$ and define a quadratic form q on $((\Lambda_1 + \Lambda_2) \cap \Pi) / (\Lambda_1 \cap \Pi \cap \Lambda_2)$ as

$$q(\lambda) = \sigma(\lambda_1, \lambda_2), \quad \lambda = \lambda_1 + \lambda_2, \quad \lambda_i \in \Lambda_i. \quad (1.3)$$

Definition 1.13. The *positive Maslov index* of a triple $(\Lambda_1, \Pi, \Lambda_2)$ is a half-integer number

$$\begin{aligned} \text{ind}_{\Pi}(\Lambda_1, \Lambda_2) &= \text{ind}^+ q + \frac{1}{2} \dim \ker q = \\ &= \text{ind}^+ q + \frac{1}{2} (\dim(\Lambda_1 \cap \Pi) + \dim(\Lambda_2 \cap \Pi)) - \dim(\Lambda_1 \cap \Lambda_2 \cap \Pi). \end{aligned}$$

The positive Maslov index has many important properties. We list just a few and refer to [10] for some others and the proves. We note that in [10] a different sign convention was used and therefore in the cited paper the negative Maslov index played the central role.

Lemma 1.14. *The positive Maslov index has the following properties for all $\Lambda_i, \Pi \in L(\Sigma)$*

1. *Explicit finite bounds*

$$0 \leq \text{ind}_{\Pi}(\Lambda_1, \Lambda_2) \leq \frac{\dim \Sigma}{2};$$

2. *If $\Gamma \subset \Lambda_1 \cap \Lambda_2$, then*

$$\text{ind}_{\Pi}(\Lambda_1, \Lambda_2) = \text{ind}_{\Pi \cap \Gamma}(\Lambda_1, \Lambda_2);$$

3. *Triangle inequality*

$$\text{ind}_{\Pi}(\Lambda_1, \Lambda_3) \leq \text{ind}_{\Pi}(\Lambda_1, \Lambda_2) + \text{ind}_{\Pi}(\Lambda_2, \Lambda_3);$$

4. *A formula*

$$\text{ind}_{\Pi}(\Lambda_1, \Pi) = \text{ind}_{\Pi}(\Pi, \Lambda_1) = \frac{1}{2} \left(\frac{\dim \Sigma}{2} - \dim(\Lambda_1 \cap \Pi) \right).$$

A similar invariant is the Kashiwara index of a triple of Lagrangian planes

Definition 1.15. Let $q(\lambda)$ be the quadratic form from (1.3), but defined on all $(\Lambda_1 + \Lambda_2) \cap \Pi$. Then the *Kashiwara index* of the triple $(\Lambda_1, \Pi, \Lambda_2)$ is the signature of the form q :

$$\text{Ki}(\Lambda_1, \Pi, \Lambda_2) = \text{sign } q.$$

Lemma 1.16. *The Kashiwara index has the following properties for all $\Lambda_i, \Pi \in L(\Sigma)$*

1. *Explicit finite bounds*

$$|\text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3)| \leq \frac{\dim \Sigma}{2};$$

2. *The cocycle property*

$$\text{Ki}(\Lambda_2, \Lambda_3, \Lambda_4) - \text{Ki}(\Lambda_1, \Lambda_3, \Lambda_4) + \text{Ki}(\Lambda_1, \Lambda_2, \Lambda_4) - \text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3) = 0;$$

3. *Antisymmetry*

$$\text{Ki}(\Lambda_{p(1)}, \Lambda_{p(2)}, \Lambda_{p(3)}) = (-1)^{\text{sign}(p)} \text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3),$$

where p is a permutation of $\{1, 2, 3\}$;

4. *Relation with the positive Maslov index*

$$-\text{Ki}(\Lambda_1, \Pi, \Lambda_2) + 2 \text{ind}_{\Pi}(\Lambda_1, \Lambda_2) + \dim(\Lambda_1 \cap \Lambda_2) = \frac{\dim(\Sigma)}{2}.$$

The proves of the first three properties can be found in [23] or [34]. The last one is proved in [10].

Let us consider what are these indices geometrically in the simplest case, when $\Sigma = \mathbb{R}^2$ (see picture 1.1). Fix some Darboux coordinates (p, q) , s.t. $\Pi = \{(p, 0)\}$. Then all the Lagrangian planes close to Π are parametrized by a single parameter S as (p, Sp) . Consider a curve of Lagrangian planes $\Lambda(t) : [-1, 1] \rightarrow L(\Sigma)$ s.t. $\Lambda(0) = \Pi$. Then we easily compute, the derivative

$$\dot{\Lambda}(0)(\lambda) = \sigma \left((p, 0), (p, \dot{S}_0 p) \right) = \dot{S}_0 p^2.$$

Thus when the curve $\Lambda(t)$ crosses Π in the clockwise direction, we add +1 to the Maslov index, and -1 in when it crosses counter-clockwise.

Consider now $\text{ind}_{\Pi}(\Lambda(-1), \Lambda(1))$ and $\text{Ki}(\Lambda(-1), \Pi, \Lambda(1))$. By working out the definitions one can check that the values of both indices depend only on the relative positions of $\Lambda(-1), \Pi, \Lambda(1)$, where we have four situations, some of which are depicted in Figure 1.1:

1. if $\Lambda(-1) = \Lambda(1) = \Pi$, then $\text{ind}_{\Pi}(\Lambda(-1), \Lambda(1)) = \text{Ki}(\Lambda(-1), \Pi, \Lambda(1)) = 0$;
2. if $\Lambda(-1) = \Pi$ or $\Lambda(1) = \Pi$, then $\text{ind}_{\Pi}(\Lambda(-1), \Lambda(1)) = 1/2$ and $\text{Ki}(\Lambda(-1), \Pi, \Lambda(1)) = 0$;
3. if by rotating Λ_{-1} in the clockwise direction we meet Λ_1 before Π , then

$$\text{ind}_{\Pi}(\Lambda(-1), \Lambda(1)) = 0 \quad \text{and} \quad \text{Ki}(\Lambda(-1), \Pi, \Lambda(1)) = -1;$$

4. if by rotating Λ_{-1} in the clockwise direction we meet Π before Λ_1 , then

$$\text{ind}_{\Pi}(\Lambda(-1), \Lambda(1)) = 1 \quad \text{and} \quad \text{Ki}(\Lambda(-1), \Pi, \Lambda(1)) = 1.$$

The fact that these indices depend only on the relative positions of the Lagrangian planes is a consequence of the following statement.

Proposition 1.17 ([23]). *The Kashiwara index $\text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3)$ and the positive Maslov index $\text{ind}_{\Lambda_2}(\Lambda_1, \Lambda_3)$ are constant on the set*

$$\{(\Lambda_1, \Lambda_2, \Lambda_3) : \dim(\Lambda_1 \cap \Lambda_2) = k_1, \dim(\Lambda_2 \cap \Lambda_3) = k_2, \dim(\Lambda_3 \cap \Lambda_1) = k_3\} \subset L(\Sigma)^3,$$

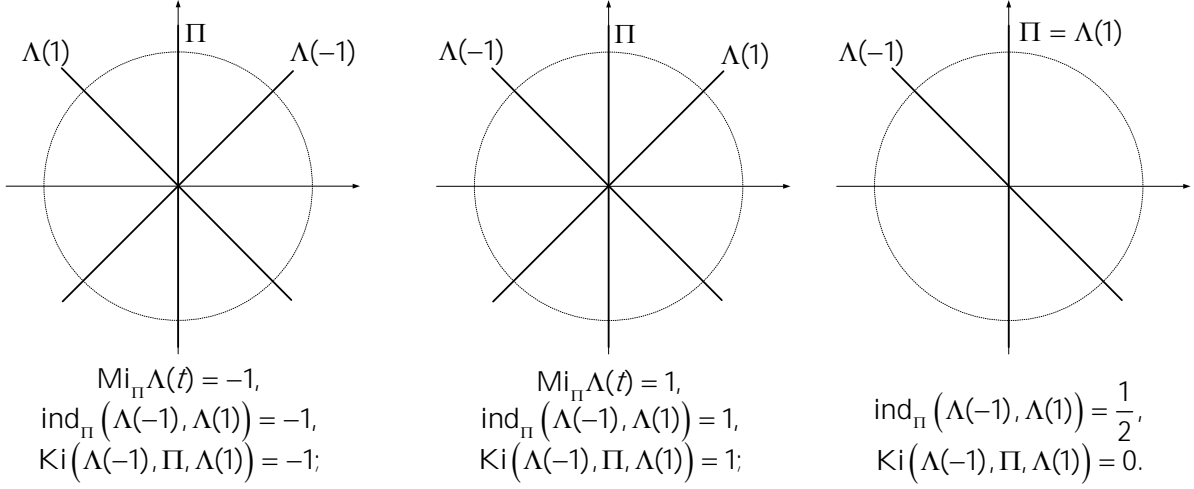
where k_i are some constants.

To state precisely what is the relation between the indices Mi , Ki and ind we need the following definition

Definition 1.18. A curve $\Lambda(t)$ is called *simple* if there exists $\Delta \in L(\Sigma)$, s.t. $\Lambda(t) \in \Delta^{\text{h}}$.

Proposition 1.19 ([10]). *Let $\Lambda(t)$, $t \in [0, 1]$ be a continuous curve, s.t. there exists $\Delta \in L(\Sigma)$, for which $\Lambda(t) \cap \Delta = \Pi \cap \Delta = \{0\}$. Then*

$$\text{Mi}_{\Pi}(\Lambda(t)) = \frac{1}{2} (\text{Ki}(\Delta, \Lambda_0, \Pi) - \text{Ki}(\Delta, \Lambda_1, \Pi)).$$

Figure 1.1: The Kashiwara and the positive Maslov indices in \mathbb{R}^2

Any two given points $\Lambda_1, \Lambda_2 \in L(\Sigma)$ can be joined by a simple monotone curve. It is easy to see this using an affine chart on the Grassmanian. So it makes sense to reformulate this result for a closed monotone curve $\Lambda(t)$.

Proposition 1.20 ([10]). *Suppose that $\Lambda(t)$, $t \in [0, 1]$ is a closed continuous monotone curve, $0 = t_0 < t_1 < t_2 < \dots < t_N = 1$ is a partition of $[0, 1]$ and $\Lambda_i = \Lambda(t_i)$. Then one has the estimate*

$$\text{Mi}(\Lambda(t)) \geq \sum_{i=0}^N \text{ind}_\Pi(\Lambda_i, \Lambda_{i+1}),$$

where $\Lambda_{N+1} = \Lambda_0$. Moreover if all pieces $\Lambda(t)|_{[t_i, t_{i+1}]}$ are simple, i.e. there exist $\Delta_i \in L(\Sigma)$, s.t. $\Delta \cap \Lambda(t)|_{[t_i, t_{i+1}]} = \{0\}$, then we have an equality

$$\text{Mi}(\Lambda(t)) = \sum_{i=0}^N \text{ind}_\Pi(\Lambda_i, \Lambda_{i+1}) = \frac{1}{2} \sum_{i=0}^N (\text{Ki}(\Delta_i, \Lambda_i, \Pi) - \text{Ki}(\Delta_i, \Lambda_{i+1}, \Pi)).$$

This motivates the following definition, that extends the notions of Maslov index and monotonicity from continuous curves to general curves in the Lagrangian Grassmanian. This extension is important, since even in the relatively simple case of bang-bang trajectories Jacobi curves are discontinuous.

Definition 1.21. Let $\Lambda(t) : [0, T] \rightarrow L(\Sigma)$ be a curve in the Lagrangian Grassmanian. Given a partition $D = \{0 = t_0 < t_1 < \dots < t_N = T\}$ we define

$$\text{ind}_\Pi^D \Lambda(t) = \sum_{i=0}^{N-1} \text{ind}_\Pi(\Lambda_i, \Lambda_{i+1}).$$

where Λ_i are as in the Proposition 1.20. We say that $\Lambda(t)$ is *monotone increasing*, if

$$\text{ind}_{\Pi} \Lambda(t) = \sup_D \text{ind}_{\Pi}^D \Lambda(t) < +\infty.$$

The quantity $\text{ind}_{\Pi} \Lambda(t)$ we call the *Maslov index of a monotone increasing curve*.

Note that if $D_1 \subset D_2$, then $\text{ind}_{\Pi}^{D_1} \Lambda(t) \leq \text{ind}_{\Pi}^{D_2} \Lambda(t)$ by the triangle inequality. We need to check that this definition is well defined, i.e. that it coincides with the previous one in the case of differentiable curves.

Theorem 1.22. *A differentiable curve $\Lambda(t) \in L(\Sigma)$ is monotone increasing if and only if $\text{ind}_{\Pi} \Lambda(t) < +\infty$ for some $\Pi \in L(\Sigma)$.*

Proof. Suppose that $\Lambda(t) \in C^1([0, T], L(\Sigma))$ is a monotone curve. Without any loss of generality we can assume that the curve is closed, since by Lemma 1.14, adding a monotone piece can increase the index by no more than $\dim \Sigma / 2$. Then it does not matter which Lagrangian plane Π we take. We split the whole curve into small pieces. Suppose that on some piece $[t_1, t_2] \subset [0, T]$ we have $\Lambda(t) \cap \Gamma \neq \{0\}$ for any $t \in [t_1, t_2]$ and some isotropic plane Γ . Assume that Γ is the maximal isotropic subspace for which this is true. Then this subspace does not give any contribution to the overall index, because of the definition of the positive Maslov index and the property 2 from Lemma 1.14. It allows us to reduce the symplectic space to $\Sigma = \Gamma^{\perp} / \Gamma$ and the reduced curve will be strictly increasing. That is why it is enough to prove the result for strictly increasing curves.

The first observation is that a strictly increasing curve can intersect \mathcal{M}_{Π} only at isolated instances of time. Indeed suppose that times of intersection have an accumulation point t . Then it is not hard to see that there exists $\lambda \in \Lambda(t)$, s.t. $\dot{\Lambda}(t)(\lambda) = 0$, which gives a contradiction with strict monotonicity. Therefore by taking small enough time intervals we can ensure that only finitely many pieces will intersect the Maslov train. The statement in one direction now follows from Proposition 1.20.

Now we prove the converse. Suppose that we have $\Lambda(t) \in C^1([0, T], L(\Sigma))$ with $\text{ind}_{\Pi} \Lambda(t) < +\infty$. We claim that $\Lambda(t)$ is monotone increasing in the sense of Definition 1.9. Let us suppose the contrary. Then for some t

$$\sigma(\lambda(t), \dot{\lambda}(t)) < 0,$$

and for some $\varepsilon > 0$ small enough

$$\frac{1}{\varepsilon} \sigma(\lambda(t), \lambda(t + \varepsilon)) < 0.$$

and since the curve is smooth, we can assume without loss of generality that $\Lambda(t) \cap \Pi = \Lambda(t + \varepsilon) \cap \Pi = \{0\}$ for all ε small enough. But this implies, that q has a negative subspace and $\text{ind}_{\Pi}(\Lambda(t), \Lambda(t + \varepsilon)) \geq 1$. And the same is true for every t', ε' sufficiently close to t, ε , s.t. $[t', t' + \varepsilon'] \subset [t, t + \varepsilon]$. So we can take a finer and finer subdivision of $[t, t + \varepsilon]$ and the index will go to $+\infty$. \square

Although these invariants were already successfully applied in [10, 6] to the study of the second variation of some classes of optimal control problem, in order to formulate the main Morse theorem we need one more symplectic invariant.

Definition 1.23. Let $\widetilde{L(\Sigma)}$ be the universal covering of $L(\Sigma)$. The Leray index is the unique mapping

$$\text{Li} : \widetilde{L(\Sigma)} \times \widetilde{L(\Sigma)} \rightarrow \mathbb{Z}$$

that satisfies the following two properties:

1. Li is locally constant on the set $\{(\tilde{\Lambda}_1, \tilde{\Lambda}_2) : \Lambda_1 \cap \Lambda_2 = \{0\}\}$;
2. $\text{Li}(\tilde{\Lambda}_2, \tilde{\Lambda}_3) - \text{Li}(\tilde{\Lambda}_1, \tilde{\Lambda}_3) + \text{Li}(\tilde{\Lambda}_1, \tilde{\Lambda}_2) = \text{Ki}(\Lambda_1, \Lambda_2, \Lambda_3)$.

An explicit construction of the Leray index using matrix logarithms can be found in [23] or [29]. We only list its main properties, that are going to be useful for the computations.

Lemma 1.24. *The Leray index Li has the following properties*

1. *Antisymmetry*

$$\text{Li}(\tilde{\Lambda}_1, \tilde{\Lambda}_2) = -\text{Li}(\tilde{\Lambda}_2, \tilde{\Lambda}_1),$$

2. *If $\tilde{\Lambda}(t)$ is a lift of a closed continuous curve $\Lambda(t) : [0, T] \rightarrow L(\Sigma)$ to $\widetilde{L(\Sigma)}$, then*

$$\text{Li}(\tilde{\Lambda}(0), \tilde{\Lambda}) - \text{Li}(\tilde{\Lambda}(T), \tilde{\Lambda}) = 2 \text{Mi}(\Lambda(t)), \quad \forall \tilde{\Lambda} \in \widetilde{L(\Sigma)}.$$

The Leray index allows to define the Maslov index and other intersection indices for curves in the Lagrangian Grassmanian and symplectic group in an abstract way. But one of its most important applications is that it can be used to construct an explicit model for the universal covering space $\widetilde{L(\Sigma)}$.

Theorem 1.25. *Let $\tilde{\Lambda}_\alpha$ be a lift of an arbitrary Lagrangian plane Λ_α to the universal covering $\widetilde{L(\Sigma)}$. Define a mapping $\Phi_\alpha : \widetilde{L(\Sigma)} \rightarrow L(\Sigma) \times \mathbb{Z}$ by*

$$\Phi_\alpha(\tilde{\Lambda}) = \left(\Lambda, \frac{1}{2} \text{Li}(\tilde{\Lambda}, \tilde{\Lambda}_\alpha) \right).$$

Then

1. *The mapping Φ_α is a bijection, whose restrictions to the subset $\{\tilde{\Lambda} \in \widetilde{L(\Sigma)} : \Lambda \cap \Lambda_\alpha = \{0\}\}$ is a homeomorphism onto $\{\Lambda \in L(\Sigma) : \Lambda \cap \Lambda_\alpha = \{0\}\}$.*
2. *The set of all bijections Φ_α forms a system of local charts of $\widetilde{L(\Sigma)}$ whose transitions $\Phi_{\alpha\beta} = \Phi_\alpha \Phi_\beta^{-1}$ are the functions*

$$\Phi_{\alpha\beta}(\Lambda, k) = \left(\Lambda, k + \frac{\text{Ki}(\Lambda, \Lambda_\alpha, \Lambda_\beta) - \text{Li}(\tilde{\Lambda}_\alpha, \tilde{\Lambda}_\beta)}{2} \right)$$

The proof of this theorem and the last lemma, as well as many other applications of the Leray index can be found in [23].

We need a couple of lemmas related to curves in the Lagrangian Grassmanian and its universal covering. We will use them only to prove the main Morse Theorem 2.23. So we just sketch the proofs.

Lemma 1.26. *Let $\Lambda(t)$ be a closed monotone curve in the sense of Definition 1.21. Then*

$$\text{ind}_{\Pi} \Lambda(t) = \text{ind}_{\Delta} \Lambda(t), \quad \forall \Delta, \Pi \in L(\Sigma).$$

Proof. Since the curve is monotone, the supremum in the definition is finite. But since it can take only discrete values, it must be attained by some partition D , i.e.

$$\text{ind}_{\Pi} \Lambda(t) = \sum_{i=0}^N \text{ind}_{\Pi}(\Lambda(t_i), \Lambda(t_{i+1})),$$

where $\Lambda(t_{N+1}) = \Lambda(t_0)$, $t_i \in D$. At the same time we can join $\Lambda(t_i)$ with simple monotone curves and construct this way a closed curve $\hat{\Lambda}(t)$. Then by Theorem 1.11 and Proposition 1.20

$$\text{ind}_{\Pi} \Lambda(t) = \text{Mi}_{\Pi}(\hat{\Lambda}(t)) = \text{Mi}_{\Delta}(\hat{\Lambda}(t)) = \sum_{i=1}^n \text{ind}_{\Delta}(\Lambda(t_i), \Lambda(t_{i+1})) = \text{ind}_{\Delta} \Lambda(t).$$

□

Lemma 1.27. *Any two simple monotone curves connecting $a, b \in L(\Sigma)$ are homotopic.*

This lemma is a direct consequence of Propositions 1.17 and 1.20. It shows Maslov index of monotone curves depends only on the relative position of its end-points.

Remark 1.28. This result has an important application that we will use later. Let $\Lambda(t)$ be a curve in $L(\Sigma)$ with a finite number of discontinuities. Then there is a canonical way of lifting a curve to the universal covering $\widetilde{L(\Sigma)}$. One has to glue all the discontinuities with simple monotone curves and lift it to the universal covering and then delete the lifts of the glued in monotone parts. The result will not depend on the way of gluing. Indeed, the previous lemma shows that two monotone curves are homotopic and therefore their lifts starting at the same point will also end at the same point.

1.2 \mathcal{L} -derivatives

A \mathcal{L} -derivative is a rule that in a given variational problem assigns to an admissible space of variations a Lagrangian plane in some symplectic space. As we add variations we can compare the relative positions of the corresponding \mathcal{L} -derivatives and deduce from that how the inertia indices and nullity of the Hessian change as we consider a bigger and bigger space of variations. As a result one can recover the classical theory of Jacobi and

much more. This theory is applicable in a great variety of cases, even when there is no Jacobi equation at all.

Assume that we have following constrained variational problem. Let $J : \mathcal{U} \rightarrow \mathbb{R}$ be a smooth functional and $F : \mathcal{U} \rightarrow M$ be a smooth map, where \mathcal{U} is a Banach manifold and M is a finite-dimensional manifold. Given a point $q \in M$, we are interested in finding $\tilde{u} \in F^{-1}(q)$ that minimize J among all other points $u \in F^{-1}(q)$. In the case of optimal control problems \mathcal{U} is the space of admissible controls. The map F is usually taken to be the end-point map, which we will introduce in Chapter 2.

The first step is to apply the Lagrange multiplier rule that says that if \tilde{u} is a minimal point then there exists a covector $\lambda \in T_q^*M$ and a number $\nu \in \{0, 1\}$, s.t.

$$\langle \lambda, dF[\tilde{u}](w) \rangle - \nu dJ[\tilde{u}](w) = 0, \quad \forall w \in T_{\tilde{u}}\mathcal{U}. \quad (1.4)$$

A pair (\tilde{u}, λ) that satisfies the equation above is called a *Lagrangian point* and \tilde{u} is called a critical point of (F, J) . If $\nu = 0$ we say that the critical point is *abnormal*, and if $\nu = 1$ we call *normal*. There are of course many critical points that are not minimal. So in order to find the minimal ones we have to apply high order conditions for minimality. For example, we can look at the Hessian $\text{Hess}(F, \nu J)[\tilde{u}, \lambda]$ at a Lagrangian point (\tilde{u}, λ) that we define as

$$\text{Hess}(F, \nu J)[\tilde{u}, \lambda] := (\nu d^2 J[\tilde{u}] - \langle \lambda, d^2 F[\tilde{u}] \rangle) |_{\ker dF[\tilde{u}]}. \quad (1.5)$$

The index and the nullity of the Hessian are directly related to optimality of the critical point \tilde{u} [8].

In order to motivate the definition let us compute the second derivative of J on the level set of $F^{-1}(q)$ at a Lagrangian point (\tilde{u}, λ) . Let $u(s)$ be a curve in $F^{-1}(q)$, s.t. $u(0) = \tilde{u}$. Then by differentiating twice $F(u(s)) = q$ at $s = 0$ we find that

$$\begin{aligned} dF[\tilde{u}](\dot{u}) &= 0, \\ d^2 F[\tilde{u}](\dot{u}, \dot{u}) + dF[\tilde{u}](\ddot{u}) &= 0. \end{aligned}$$

Similarly we find that

$$\begin{aligned} \frac{\partial^2}{\partial s^2} \Big|_{s=0} J(u(s)) &= d^2 J[\tilde{u}](\dot{u}, \dot{u}) + dJ[\tilde{u}](\ddot{u}) = d^2 J[\tilde{u}](\dot{u}, \dot{u}) + \langle \lambda, dF[\tilde{u}](\ddot{u}) \rangle = \\ &= d^2 J[\tilde{u}](\dot{u}, \dot{u}) - \langle \lambda, dF[\tilde{u}](\dot{u}, \dot{u}) \rangle \end{aligned}$$

where in the second equality we have used that $(\tilde{u}, \lambda(t))$ is a Lagrange point. From here we can see that this expression is equal exactly to $\text{Hess}(F, \nu J)[\tilde{u}, \lambda](\dot{u}, \dot{u})$.

We are now ready to define \mathcal{L} -derivatives. We linearise (1.4) with respect to λ and u , and obtain the following equation

$$\langle \xi, dF[\tilde{u}](w) \rangle + \langle \lambda, d^2 F[\tilde{u}](v, w) \rangle - \nu d^2 J[\tilde{u}](v, w) = 0.$$

Or if we define $Q(v, w) := \langle \lambda, d^2 F[\tilde{u}](v, w) \rangle + \nu d^2 J[\tilde{u}](v, w)$, we can rewrite this as

$$\langle \xi, dF[\tilde{u}](w) \rangle + Q(v, w) = 0. \quad (1.6)$$

We note that $\text{Hess}(F, \nu J)[\tilde{u}, \lambda] = -Q|_{\ker dF[\tilde{u}]}$.

Definition 1.29. A \mathcal{L} -derivative of a pair (F, J) at a Lagrangian point (\tilde{u}, λ) constructed over a finite-dimensional space of variations $V \subset T_{\tilde{u}}\mathcal{U}$ that we denote as $\mathcal{L}(F, \nu J)[\tilde{u}, \lambda](V)$ is the set of vectors $(\xi, dF[\tilde{u}](v)) \in T_{\lambda}(T^*M)$, s.t. $(\xi, v) \in (T_{\lambda}(T_q^*M), V)$ solve (1.6) for all $w \in V$.

This set is a Lagrangian plane [2] if V is finite-dimensional. The reason why we do not take directly $T_{\tilde{u}}\mathcal{U}$ instead of V is that it is a linear equation defined on an infinite-dimensional space and it might be ill-posed. In this case $\mathcal{L}(F, \nu J)[\tilde{u}, \lambda](V)$ is just isotropic. But if we have chosen the right topology for our space of variations, we are going to get exactly $\dim M$ independent solutions.

Lagrangian subspace $\mathcal{L}(F, \nu J)[u, \lambda](V)$ contains information about the second variation restricted to the subspace V . To obtain a Lagrangian subspace that encodes the information about all the possible variations one has to use generalized sequences.

Definition 1.30. A directed set (I, \preceq) is a set I with a preorder \preceq , s.t. for any two elements $\alpha, \beta \in I$ there exists an element γ , s.t. $\alpha \preceq \gamma$ and $\beta \preceq \gamma$.

Definition 1.31. Given a directed set (I, \preceq) a *generalized sequence* or a net is a function from the set of indices I to a topological space X . A generalized sequence $\{x_{\alpha}\}_{\alpha \in I} \in X$ converges to a limit $x \in X$, if for any open neighbourhood $O_x \ni x$ there exists an element $\beta \in I$, s.t. for all $\alpha \succ \beta$ one has $x_{\alpha} \in O_x$.

Finite dimensional subspaces of $T_u\mathcal{U}$ form a directed set with a partial ordering given by the inclusion $V \subset W$. This motivates the following definition

Definition 1.32. A \mathcal{L} -derivative of (F, J) at a Lagrange point (u, λ) constructed over a subspace $V \subset T_u\mathcal{U}$ is the generalized limit

$$\mathcal{L}(F, \nu J)[u, \lambda](V) = \lim_{W \nearrow V} \mathcal{L}(F, \nu J)[u, \lambda](W).$$

taken over increasing finite-dimensional subspaces $W \subset V$.

When V is the whole space of available variations, we simply write $\mathcal{L}(F, \nu J)[\tilde{u}, \lambda]$ for the corresponding \mathcal{L} -derivative.

We have the following important theorem proved in [1], that gives the existence of this limit and a way to compute it.

Theorem 1.33. Let (\tilde{u}, λ) be a Lagrangian point of (F, J) .

1. If either the positive or the negative inertia index of $\text{Hess}(F, \nu J)[\tilde{u}, \lambda]$ is finite, then $\mathcal{L}(F, \nu J)[\tilde{u}, \lambda]$ exists;
2. $\mathcal{L}(F, \nu J)[\tilde{u}, \lambda] = \mathcal{L}(F, \nu J)[\tilde{u}, \lambda](V)$ for any V dense in $T_{\tilde{u}}\mathcal{U}$.

\mathcal{L} -derivatives contain information about the inertia indices and nullity of the Hessian (1.5) restricted to some space of variations. By comparing two \mathcal{L} -derivatives constructed over two subspaces $V \subset W$, we can see how the inertia indices change as we add variations to our variations space [1].

1.3 Chronological calculus

We have seen in the previous subsection that the definition of \mathcal{L} -derivatives involves first and second derivatives of maps F and J . In the case of optimal control problems, usually the end-point map E_t plays the role of F . One of the ways to obtain explicit formulas for the first and second differential of E_t is to use results from the chronological calculus, that we explain next.

The idea of the chronological calculus is to reinterpret all geometric objects on a manifold M as linear maps on $C^\infty(M)$. For example, a point q can be seen as a linear operator $\hat{q} : C^\infty(M) \rightarrow \mathbb{R}$ defined in a natural way

$$\hat{q}(a) = a(q), \quad \forall a \in C^\infty(M).$$

Similarly one defines an operator analogue of a diffeomorphism P :

$$(\hat{q} \circ \hat{P})(a) = P(a(q)), \quad \forall a \in C^\infty(M), \forall q \in M.$$

Here $\hat{P} : C^\infty(M) \rightarrow C^\infty(M)$ is an algebra automorphism, that geometrically is just a change of variables. A vector field V is represented by a differentiation \hat{V} of the algebra $C^\infty(M)$.

In [8] one can find the proof of the fact, that any algebra homomorphism/automorphism/differentiation can be represented by a point/diffeomorphism/vector field. A one-parametric family of these objects can be integrated and differentiated with the usual properties like, for example, the Leibnitz rule.

Consider a non-autonomous vector field $V(t)$ and the corresponding differential equation

$$\dot{q}(t) = V(t)(q(t))$$

that can be rewritten in the operator form as

$$\dot{\hat{q}}(t) = \hat{q}(t) \circ \hat{V}(t).$$

From here we omit the "hat" in the operator notation, since we will always speak about operators unless it is stated otherwise. If the Cauchy problem for this ODE is well posed, we have a well defined flow P^t that must be a unique solution to the operator equation

$$\dot{P}^t = P^t \circ V(t). \tag{1.7}$$

A solution to this equation is called the *right chronological exponent* and is denoted by

$$P^t = \overrightarrow{\exp} \int_0^t V(\tau) d\tau.$$

Since we know that $P^0 = \text{id}$, we can rewrite equation (1.7) in the integral form

$$P^t = \text{id} + \int_0^t P^\tau \circ V(\tau) d\tau.$$

Iterating this expression gives us *the Volterra expansion* for the right chronological exponent

$$P^t = \text{id} + \int_0^t V(\tau) d\tau + \int_0^t \int_0^\tau V(\theta) \circ V(\tau) d\theta d\tau + \dots \quad (1.8)$$

The last thing that we need is the variation formula for the right chronological exponent. Suppose that $V(t), W(t)$ are non-autonomous vector fields and P^t satisfies (1.7). Then the following formula are true

$$\overrightarrow{\text{exp}} \int_0^t (V(\tau) + W(\tau)) d\tau = \overrightarrow{\text{exp}} \int_0^t (P_*^\tau)^{-1} W(\tau) d\tau \circ P^t. \quad (1.9)$$

$$\overrightarrow{\text{exp}} \int_0^t (V(\tau) + W(\tau)) d\tau = P^t \circ \overrightarrow{\text{exp}} \int_0^t P_*^t (P_*^\tau)^{-1} W(\tau) d\tau. \quad (1.10)$$

Here P_*^t should be understood as a pushforward map, i.e. in the expressions above $(P_*^t)^{-1} W$ should be read as $\widehat{(P_*^t)^{-1} W}$. The proof can be found in the book [8].

1.4 Generalities on sub-Riemannian geometry

Definition 1.34. A *sub-Riemannian manifold* is a triple (M, \mathcal{D}, g) , where M is a connected smooth manifold, \mathcal{D} is a distribution of planes in TM and g is a Riemannian metric on \mathcal{D} . The dimension of \mathcal{D} at $q \in M$ is the *rank* of the distribution at that point.

Definition 1.35. A curve $q(t) : [0, T] \rightarrow M$ is said to be *admissible* or *horizontal* if it is Lipschitz continuous and for almost every $t \in [0, T]$ satisfies

$$\dot{q}(t) \in \mathcal{D}_{q(t)}.$$

Using the Riemannian metric g we can define length of admissible curves in the usual way as

$$l[q(t)] = \int_0^T \sqrt{g(\dot{q}, \dot{q})} dt.$$

This transforms a sub-Riemannian manifold into a metric length space with the usual distance

$$d(q_0, q_T) = \inf\{l(q(t)) : q(0) = q_0, q(T) = q_T, q(t) \text{ horizontal}\},$$

which in the context of sub-Riemannian geometry is known as the *Carnot-Caratheodory distance*.

The following theorem plays a central role in sub-Riemannian geometry.

Theorem 1.36 (Rashevsky-Chow). *Suppose that*

$$\text{span}\{[X_1, [\dots[X_{j-1}, X_j]]](q) : X_i(q) \in \mathcal{D}_q, j \in \mathbb{N}\} = T_q M, \quad \forall q \in M \quad (1.11)$$

where X_j are horizontal vector fields, which are sections of Δ . Then any two points of M can be connected by an admissible curves and the Carnot-Caratheodory distance induces a topology equivalent to the manifold topology.

Condition 1.11 is also known as the *Hörmander condition*.

Definition 1.37. An admissible curves $q(t)$ is said to be a *geodesic*, if its restriction to any sufficiently small interval $[t_1, t_2] \subset [0, T]$ is a length minimizer between $q(t_1)$ and $q(t_2)$.

One can check that the length functional is invariant under reparameterization of curves. A simple application of the Cauchy-Schwarz inequality shows that constant speed geodesics are also extremal curves of the action functional

$$J[q(t)] = \frac{1}{2} \int_0^T g(\dot{q}, \dot{q}) dt$$

and the other way around. Any extremal curve of the action functional is a constant speed geodesics.

We can assume that the distribution at least locally is spanned by a set of orthonormal vector fields

$$\mathcal{D}_q = \text{span}\{X_1, \dots, X_k\}, \quad g(X_i, X_j) = \delta_{ij}.$$

Then we can reformulate the problem of finding minimal curves between two given points q_0 and q_T as an optimal control problem of the form

$$\dot{q}(t) = \sum_{i=1}^k u_i(t) X_i(q(t)), \quad u_i(t) \in L^\infty[0, T],$$

$$q(0) = q_0, \quad q(T) = q_T,$$

$$J[u(t)] = \frac{1}{2} \int_0^T \sum_{i=1}^k u_i^2(t) dt \rightarrow \min.$$

It turns out that the Rashevsky-Chow theorem is enough to guarantee existence of solutions to this variational problem [3].

Jacobi curves and Morse-type theory

The goal of this chapter is to define the notion of Jacobi curves and show how to extract the information of the Hessian from it. First we derive explicit formulas for (1.6) using the language of symplectic geometry. Then we show how to calculate this curve approximately and using those approximations we show a very general Morse theorem that applies to any type of extremals.

2.1 \mathcal{L} -derivatives for optimal control problems

We consider the following optimal control problem

$$\dot{q} = f(u, q), \quad u \in U \subset \mathbb{R}^k, \quad q \in M, \quad (2.1)$$

$$J_T(u) = \int_0^T L(u, q) dt \rightarrow \min. \quad (2.2)$$

We assume that we look for a minimum control in $L_k^\infty[0, T]$, that $f(u, q)$ and $L(u, q)$ are smooth in both variables and that necessary growth conditions are satisfied to ensure uniqueness of the Cauchy problem (see, for example, Caratheodory theorem in [22]). The set $U \subset \mathbb{R}^k$ is a union

$$U = \bigcup_{i=1}^N U_i$$

of closed embedded submanifolds $U_i \subset \mathbb{R}^k$ without boundary. A typical example in control theory is a curve-linear polytope in \mathbb{R}^k defined by a number of inequalities

$$p_i(u) \leq 0$$

that satisfy

$$p_i(u) = 0, \quad \Leftrightarrow \quad d_u p_i = 0.$$

Then U is union of the interior of the polytope and faces of different dimensions. Moreover we assume that the final time T is fixed and that $q(0)$ can be free and lies in some submanifold N_0 , and $q(T) = q_T$ is fixed. We note that if $q(0)$ is fixed and $q(T)$ is not, then we transform our problem to a free starting point by making a change of time variable $t \mapsto T - t$.

We would like to have a definition of the Jacobi curve that is invariant under feedback transformations and diffeomorphisms of M and that does not use any additional constructions on the manifold. To do this we need to reformulate our problem in an invariant manner using the language of vector bundles. We begin with an invariant reformulation of (2.1).

Definition 2.1. $V = M \times \mathbb{R}^k$ be a trivial Euclidean bundle. Then a control system on M is a smooth map $f : V \rightarrow TM$, s.t. $f(\mathbb{R}^k) \subset T_q M$.

Definition 2.2. An admissible pair is a bounded measurable map $\omega : [0, T] \rightarrow V$, s.t. $q(t) = \pi(\omega(t))$ is a Lipschitz curve that satisfies for almost every t the control equation

$$\dot{q} = f(\omega(t)).$$

The curve $q(t)$ is called an admissible trajectory.

An admissible pair $\omega(t)$ can be seen as a pair $(q(t), u(t))$ which consists of a Lipschitz trajectory $q(t)$ that satisfies (2.1) for almost every t and the corresponding control $u(t)$.

In [2] the following result was proven

Proposition 2.3. If $\dim V_q = k$ and $\dim M = n$, then the set of admissible pairs Ω has a structure of a smooth Banach manifold modelled over $\mathbb{R}^n \times L_k^\infty[0, T]$.

We don't give here a complete proof of this result, but we explain how one can construct an open neighbourhood of some admissible pair $\tilde{\omega}(t) = (\tilde{q}(t), \tilde{u}(t))$. Fix a moment of time $\tau \in [0, T]$ and consider an open neighbourhood U of $\tilde{q}(\tau) = \pi(\tilde{\omega}(\tau))$ that is diffeomorphic to an open set in \mathbb{R}^n . Then if we fix $u(t)$, s.t. $\|u(t) - \tilde{u}(t)\|_\infty < \varepsilon$, through each point $q \in U$ passes a unique solution of (2.1) at a moment of time τ , because of the well-posedness of the Cauchy problem. So one can see that locally a neighbourhood of $\tilde{\omega}(t)$ is a product of a small open neighbourhood of $\tilde{q}(\tau)$ in M and an open neighbourhood of \tilde{u} in $L_k^\infty[0, T]$.

In order to construct the \mathcal{L} -derivative we must specify a map that corresponds to our constraint.

Definition 2.4. The evaluation map $F_t : \Omega \rightarrow M$ is a map, that is defined as

$$F_t(\omega) = \pi(\omega(t)) = q(t).$$

This map is actually smooth because of the smooth dependence on parameters of the solutions of the Cauchy problem. Also from the construction of a neighbourhood of $\omega \in \Omega$ it is easy to see that F_t is a submersion. Moreover the classical end-point map can be characterized as

$$E_T = F_T|_{F_0^{-1}(q_0)}, \quad q_0 \in M.$$

The end-point map is the basic object in the study of time optimal control problems. It takes a control $u(t)$ and maps it to the end of the corresponding trajectory that begins at $q_0 \in M$. In our problem we have that q_0 is not fixed, but lies in a manifold N_0 . So we will define an analog of the classical end-point map as

$$E_{N_0, T} = F_T|_{F_0^{-1}(N_0)}.$$

Remark 2.5. It is important to note that the definition of the end-point map, evaluation map, admissible curves etc. are all invariant. This means that the \mathcal{L} -derivatives that we are going to construct are invariant as well and thus we can exploit the local structure of the space of admissible curves to simplify explicit computations. Previously we have discussed that the space of admissible curves is locally equivalent to $L_k^\infty[0, T] \times \mathbb{R}^n$, which simply means that we look for the solutions of (2.1) with some control $u(t) \in L_k^\infty[0, T]$ passing through a point $q(\tau) \in M$, $\tau \in [0, T]$. But we can choose this τ as we want, the corresponding \mathcal{L} -derivative will be the same for all τ . This simplifies many things. For example, the inclusion of the space of admissible curves defined on an interval $[0, t] \subset [0, T]$ into the space of admissible curves defined on $[0, T]$ is simply given by taking the controls from $L_k^\infty[0, t] \subset L_k^\infty[0, T]$. Or by identifying a neighborhood of $q(\tau)$ with \mathbb{R}^n we can find coordinates s.t. $F_\tau(q(\tau), u(t)) = q(\tau)$. This implies that in this coordinate chart kernel of the differential of F_τ is exactly $L_k^\infty[0, T]$ and that the second derivative is zero. Finally we note that the space of variations $F_0^{-1}(N_0)$ for the very same reasons locally can be identified with $L_k^\infty[0, T] \times \mathbb{R}^{\dim N_0}$.

If $\tilde{\omega}$ is a critical point of $(E_{N_0, T}, J_T)$, then we have

$$\langle \lambda(T), dE_{N_0, T}[\tilde{\omega}](w) \rangle - \nu dJ_T[\tilde{\omega}](w) = 0, \quad \forall w \in L_k^\infty[0, T] \times T_{\tilde{q}(0)}N_0.$$

Here $(\lambda(T), -\nu) \in T_{E_{N_0, T}(\tilde{\omega})}^* \times \mathbb{R}$ are the Lagrange multipliers, where ν is normalized to take values 0 or 1, $\tilde{\omega} \in F_0^{-1}(N_0)$. If we introduce the *extended end-point map* $\hat{E}_{N_0, t} = (E_{N_0, t}, J_t)$, we can rewrite this equation as

$$\langle \hat{\lambda}(t), d\hat{E}_{N_0, t}[\tilde{\omega}](w) \rangle = 0, \quad (2.3)$$

where $\hat{\lambda}(t) = (\lambda(t), -\nu)$. The extended end-point map $\hat{E}_{N_0, t}$ can be seen as the end point-map of the following control system

$$\begin{aligned} \dot{q} &= f(u, q) \\ \dot{y} &= L(u, q) \end{aligned} \iff \begin{aligned} \dot{\hat{q}} &= \hat{f}(u, \hat{q}). \end{aligned}$$

Let us denote the flow of this system with $u = \tilde{u}$ from time t_1 till time t_2 by $\hat{P}_{t_1}^{t_2}$. We also write \hat{P}^t for \hat{P}_0^t and Ω_{N_0} for $\Omega \cap F_0^{-1}(N_0)$. We use the non-hatted notation P^t for the flow of the original control system (2.1) with the same control.

Since $d\hat{E}_{N_0, T}[\tilde{\omega}]|_{\Omega_{N_0} \cap L_k^\infty[0, T]} = (\hat{P}_T^T)^* d\hat{E}_T[\tilde{\omega}]$, by restricting (1.9) to $L_k^\infty[0, t]$ we find that

$$\langle \hat{\lambda}(t), d\hat{E}_{N_0, t}[\tilde{\omega}](w) \rangle = \langle \lambda(t), dE_{N_0, t}[\tilde{\omega}](w) \rangle - \nu dJ_t[\tilde{\omega}](w) = 0, \quad \forall w \in T_{\tilde{\omega}}\Omega_{N_0}, \quad (2.4)$$

where $\hat{\lambda}(t) = (\hat{P}_t^T)^* \hat{\lambda}(T)$ and $\lambda(t)$ is the projection of $\hat{\lambda}(t)$ to $T_{\tilde{q}(t)}^* M$. Note that in the first inequality we have used the fact that the differential flow \hat{P}_t^T leaves the subspace $(0, -\nu)$ invariant, since \dot{y} does not depend on y .

By the variation formula (1.9) we then find

$$\begin{aligned} \hat{E}_{N_0, t}(u, \hat{q}(0)) &= \hat{q}(0) \circ \overrightarrow{\text{exp}} \int_0^t (\hat{P}_\tau^T)^{-1} \left(\hat{f}_{u(\tau)} - \hat{f}_{\tilde{u}(\tau)} \right) d\tau \circ \hat{P}^t = \\ &= \hat{q}(0) \circ \overrightarrow{\text{exp}} \int_0^t \hat{g}_{\tau, u(\tau)} d\tau \circ \hat{P}^t. \end{aligned}$$

Note that $\hat{g}_{\tau, \tilde{u}(\tau)} \equiv 0$. Using the Volterra expansion (1.8) and differentiating w.r.t. to $\hat{q}(0)$ at $(\tilde{q}(0), 0)$ and $u(\tau)$ at $\tilde{u}(\tau)$, we obtain for the first variation the following expression

$$d\hat{E}_{N_0, t}[\tilde{\omega}, \hat{q}(0)](v, \hat{\zeta}) = \hat{P}_*^t \hat{\zeta} + \left(\hat{P}_*^t \int_0^t \left(\frac{\partial}{\partial u} \Big|_{u=\tilde{u}(\tau)} \hat{g}_{\tau, u} \right) v(\tau) d\tau \right) (\hat{q}(t)). \quad (2.5)$$

We define the Hamiltonian

$$h(u, \lambda) = \langle \lambda, f(u, q) \rangle - \nu L(u, q).$$

One can show [8] that $\lambda(t)$ satisfies the Hamiltonian system

$$\dot{\lambda}(t) = \vec{h}(\tilde{u}, \lambda(t)).$$

Moreover if we restrict the equation (2.4) to w in $L_k^\infty[0, t]$, we obtain

$$\begin{aligned} 0 &= \left\langle \hat{\lambda}(t), \left(\hat{P}_*^t \int_0^t \frac{\partial}{\partial u} \Big|_{u=\tilde{u}(\tau)} \hat{g}_{\tau, u} \cdot v(\tau) d\tau \right) (\hat{q}(0)) \right\rangle = \\ &= \int_0^t \frac{\partial}{\partial u} \Big|_{u=\tilde{u}(\tau)} \left\langle \hat{\lambda}(t), \hat{P}_*^t (\hat{P}_*^\tau)^{-1} (\hat{f}_u - \hat{f}_{\tilde{u}(\tau)}) (\hat{q}(t)) \cdot v(\tau) \right\rangle d\tau = \\ &= \int_0^t \frac{\partial h(u, \lambda(t))}{\partial u} \Big|_{u=\tilde{u}(\tau)} v(\tau) d\tau. \end{aligned}$$

Since this equality holds for any $v(t) \in L_k^\infty[0, t]$, we obtain this way the extremality condition

$$\frac{\partial h(u, \lambda(t))}{\partial u} \Big|_{u=\tilde{u}} = 0, \quad (2.6)$$

which is a weak form of the maximum principle.

Note that since we do not vary the initial value functional (which is zero), we have $\hat{\zeta} = (\zeta, 0) \in T_{\tilde{q}(0)} M \times \mathbb{R}$. Thus if we restrict the equation (2.4) to $w \in T_{\tilde{q}(0)} N_0$, we find

$$0 = \langle \hat{\lambda}(t), \hat{P}_*^t \hat{\zeta} \rangle = \langle (\hat{P}^t)^* \hat{\lambda}(t), \hat{\zeta} \rangle = \langle \hat{\lambda}(0), \hat{\zeta} \rangle = \langle \lambda(0), \zeta \rangle.$$

This way we obtain the *transversality conditions*

$$\lambda(0) \perp T_{\tilde{q}(0)} N_0. \quad (2.7)$$

To define the \mathcal{L} -derivatives we have to consider the solutions of the following linear equation

$$\langle \xi(t), dE_{N_0, t}[\tilde{\omega}](w) \rangle + \langle \lambda(t), d^2 E_{N_0, t}[\tilde{\omega}](v, w) \rangle = \nu d^2 J_t[\tilde{\omega}](v, w) \quad (2.8)$$

Definition 2.6. A *Jacobi curve* of an optimal control problem (2.1)-(2.2) along an extremal $\tilde{q}(t)$ is the parametrized by t curve of \mathcal{L} -derivatives $\mathcal{L}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]$.

Definition 2.6 is quite natural in the light of the previous discussion, but it also has two problems. Firstly, the extremal pair $\tilde{\omega}(t)$ might be on the boundary of the space of admissible curves. This happens, for example, in the bang-bang case, when the optimal control takes values in vertices of a polytope U . In this case the second derivative is not well defined, since on the boundary not all of the variations are two sided. The second problem that we have, is that the \mathcal{L} -derivatives will be Lagrangian subspaces in different symplectic spaces $T_{\lambda(t)}(T^*M)$. So in order to study their relative positions, we need a way to identify them.

We start by considering the first problem. A solution that we propose here consists of two steps. First we enlarge the space of variations in order to cover all the known situations (like regular, singular, bang-bang extremals and their concatenations) and then we will simply construct the corresponding \mathcal{L} -derivative using only the subspace of available two-sided variations.

Variation that we need are called time variations. Basically we introduce a new time variable τ , given by

$$t(\tau) = \int_0^\tau (1 + u_0(s)) ds.$$

We assume that $u_0(s) > -1$ and that

$$\int_0^{t_1} u_0(s) ds = 0,$$

since the final time is fixed. Or instead of the last condition, we can take the time variable as a new state variable

$$\dot{t} = 1$$

and reduce this way the problem to a free time problem.

Under these assumptions function $t(\tau)$ is strictly increasing and therefore invertible. Then our control system is transformed to

$$\frac{dq}{d\tau} = (1 + u_0(\tau))f(q, u(t(\tau)))$$

and the functional to

$$\int_0^{\tau(T)} (1 + u_0(s))L(q, u(t(s))) ds \rightarrow \min .$$

If \tilde{u} was an optimal control for (2.1)-(2.2), then $(\tilde{u}, 0)$ will be optimal for the new problem. Thus after we have included time variations, we just construct the \mathcal{L} -derivative at $((\tilde{u}, 0), \lambda(t))$ over the set of all available two sided-variations, which is now non-empty.

Time variations were previously used to derive necessary and sufficient condition of bang-bang arcs [11, 5] and with small modification of their definition, one can even prove a version of the maximum principle [24]. These variations actually do not give contribution to the index of the second variation if the considered control $\tilde{u}(s)$ has at least C^2 -regularity

as can be seen from the explicit form of the second derivative, because in this case any time variation can be realized as a control variation. But if the control has less regularity like in the bang-bang case, time variations allow to find necessary optimality conditions even when there are not enough two-sided variations. So from now on we assume that the time variations are included in the formulation of the problem and that consequently the space of two-sided variations is non-empty.

The next step is to write down explicitly equation that defines $\mathcal{L}(E_{N_0,t}, \nu J_t)$. We have seen that the first order conditions are equivalent to the maximum principle with transversality conditions. Thus in order to obtain an explicit form for the equations (2.4) it is enough to linearize the Hamiltonian system, the maximum condition and the transversality conditions w.r.t. both phase variables and control variables. The equations in this case can be quite involved but there is a small trick to make them simpler and also to fix the second problem mentioned earlier. Let Φ_t be the Hamiltonian flow of $\vec{h}_{\tilde{u}(t)}$. We simply apply a time dependent change of variables $\mu = \Phi_t^{-1}(\lambda)$ on T^*M . Then the Hamiltonian system of PMP before the maximization is transformed to

$$\dot{\mu}(t) = \vec{H}(t, u, \mu(t)),$$

where

$$H(t, u, \mu) = (\Phi_t)_*^{-1}(h(u, \cdot) - h(\tilde{u}, \cdot))(\mu).$$

The maximum principle now says that any extremal control u must satisfy the weak maximum condition (2.4), which in the new coordinates has the same form as before

$$\frac{\partial H(t, u, \mu(t))}{\partial u} = 0.$$

Since $\Phi_0 = \text{id}$, the transversality conditions have the same form as in (2.7):

$$\mu(0) \perp T_{\tilde{q}(0)}N_0.$$

We note that under this change of variables, the Lagrange point $(\tilde{\omega}, \lambda(t))$ is transformed to $(\tilde{q}(0), \lambda(0))$ and from the formula for the new Hamiltonian we obtain $\vec{H}(t, \tilde{u}, \lambda(0)) = 0$. Thus linearization at $(\tilde{q}(0), \lambda(0))$ will take a simple form. Secondly, these new formulas correspond to the pull-back $\Phi_t^* \mathcal{L}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)](V)$. For brevity we denote it by $\mathcal{L}_t(V)$. We will see that the information about the Hessian is encoded in some symplectic invariants of the Jacobi curve. Since Φ_t is a symplectomorphism, all the results about $\mathcal{L}_t(V)$ will transfer automatically to the original invariant curve and we also gain the advantage that $\mathcal{L}_t(V)$ stays in a fixed symplectic space $T_{\lambda(0)}(T^*M)$. This way we solve completely the second problem from before.

Now we can finally write down explicit equations that define $\mathcal{L}_t(V)$. First we linearize the Hamiltonian system at $(\tilde{q}(0), \lambda(0))$. We get

$$\dot{\eta}(t) = X(t)v(t) \quad \iff \quad \eta(t) = \eta_0 + \int_0^t X(\tau)v(\tau)d\tau, \quad X(t) := \left. \frac{\partial \vec{H}(t, u, \lambda(0))}{\partial u} \right|_{u=\tilde{u}}. \quad (2.9)$$

By identifying $T_{\lambda(0)}(T^*M)$ with $T_{\bar{q}(0)}^*M \times T_{\bar{q}(0)}M$, we obtain that the linearization of the transversality conditions gives

$$\eta_0 \in T_{\bar{q}(0)}^\perp N_0 \times T_{\bar{q}(0)} N_0,$$

where $T_{\bar{q}(0)}^\perp N_0 \subset T_{\bar{q}(0)}^*M$ is just the annihilator of $T_{\bar{q}(0)} N_0$.

Finally we linearize the maximum condition to obtain

$$\frac{\partial^2 H(t, u, \lambda(0))}{\partial u^2} \Big|_{u=\bar{u}} (v(t), w) + \left\langle d_{\mu(t)} \frac{\partial H}{\partial u} \Big|_{u=\bar{u}} w, \eta(t) \right\rangle = 0, \quad \forall w \in \mathbb{R}^k$$

Using the definitions we gave before, we can write

$$b(t)(v(t), w) := \frac{\partial^2 H(t, u, \lambda(0))}{\partial u^2} \Big|_{u=\bar{u}} (v(t), w) = \frac{\partial^2 h(u, \lambda(t))}{\partial u^2} \Big|_{u=\bar{u}} (v(t), w), \quad \forall w \in \mathbb{R}^k,$$

$$\left\langle d_{\mu(t)} \frac{\partial H}{\partial u} \Big|_{u=\bar{u}} w, \eta(t) \right\rangle = \sigma(\eta(t), X(t)w), \quad \forall w \in \mathbb{R}^k.$$

Collecting all the formulas proves the following result.

Proposition 2.7. *An \mathcal{L} -derivative $\mathcal{L}_t(V)$ over a subspace $V \in L_k^\infty[0, t]$ consists of vectors of the form*

$$\eta(t) = \eta_0 + \int_0^t X(\tau)v(\tau)d\tau. \quad (2.10)$$

where $\eta_0 \in T_{\bar{q}(0)}^\perp N_0 \times T_{\bar{q}(0)} N_0$ and $v \in V$ satisfy

$$\int_0^t \left(\sigma \left(\eta_0 + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)(v(\tau), w(\tau)) \right) d\tau = 0, \quad \forall w(t) \in V. \quad (2.11)$$

As we have discussed before the full \mathcal{L} -derivative is defined as $\mathcal{L}_t = \lim \mathcal{L}_t(V)$. The next step is to compute this limit.

We will need later an explicit expression of $Q|_{\ker dE_{N_0, t}[\tilde{\omega}]}$. So let us derive it as a final application of chronological calculus.

Proposition 2.8. *The Hessian $\text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]$ has the following form*

$$\begin{aligned} & \text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]((\zeta_1, v_1), (\zeta_2, v_2)) = \\ & = - \int_0^t \sigma \left(\zeta_1 + \int_0^\tau X(\theta)v_1(\theta)d\theta, X(\tau)v_2(\tau) \right) + b(\tau)(v_1(\tau), v_2(\tau))d\tau. \end{aligned} \quad (2.12)$$

Note that there is no ζ_2 due to the fact that $(\zeta, v) \in \ker dE_{N_0, T}$ are not independent.

Proof. We introduce a map $\hat{G}_{N_0,t} = (\hat{P}^t)^{-1} \hat{E}_{N_0,t}$. Then we can write equivalently

$$Q(v, w) = \langle \hat{\lambda}(t), d^2 \hat{E}_{N_0,t}[\tilde{\omega}](v, w) \rangle = \langle \hat{\lambda}(0), d^2 \hat{G}_{N_0,t}[\tilde{\omega}](v, w) \rangle.$$

To simplify the notations we define

$$g'(\tau) = \left. \frac{\partial}{\partial u} \right|_{u=\tilde{u}(\tau)} g_{\tau,u},$$

and similarly the “hatted” \hat{g}'_τ for the extended system. We also define the “hatted” version of $X(t)$, which is defined in the same way, using the same Hamiltonian $h(u, \lambda)$, but viewed as a Hamiltonian on $T^*M \times \mathbb{R}^2$ and the corresponding extended Hamiltonian flow $\hat{\Phi}_t$. We note that the projection of $\hat{X}(t)$ to $T(T^*M)$ is exactly $X(t)$ and that the projections of $\hat{X}(t)$ and $X(t)$ to $TM \times \mathbb{R}$ and TM are $\hat{g}'(t)$ and $g'(t)$ for all $t \in [0, T]$. Another important point is that the standard symplectic form $\hat{\sigma}$ on the extended phase space is equal to

$$\hat{\sigma} = \sigma - d\nu \wedge dy,$$

where σ is the standard symplectic form on T^*M .

Using the Volterra expansion once more we obtain an explicit formula

$$\begin{aligned} & d^2 \hat{G}_t[\tilde{\omega}]((\hat{\zeta}_1, v_1), (\hat{\zeta}_2, v_2)) = \\ &= \int_0^t d_{\hat{q}} \hat{g}'(\tau) (\hat{\zeta}_1 v_2(\tau) + \hat{\zeta}_2 v_1(\tau)) d\tau + \int_0^t \hat{g}''(\tau) (v_1(\tau), v_2(\tau)) d\tau + \\ &+ \int_0^t \left(\int_0^\tau \hat{g}'(\theta) v_1(\theta) d\theta \circ \hat{g}'(\tau) v_2(\tau) + \int_0^\tau \hat{g}'(\theta) v_2(\theta) d\theta \circ \hat{g}'(\tau) v_1(\tau) \right) d\tau. \end{aligned}$$

By exchanging the order of integration we have

$$\int_0^t \left(\int_0^\tau \hat{g}'(\theta) v_2(\theta) d\theta \circ \hat{g}'(\tau) v_1(\tau) \right) d\tau = \int_0^t \left(\hat{g}'(\tau) v_2(\tau) \circ \int_\tau^t \hat{g}'(\theta) v_1(\theta) d\theta \right) d\tau.$$

By adding and subtracting

$$\int_0^t \hat{g}'(\tau) v_2(\tau) d\tau \circ \int_0^t \hat{g}'(\tau) v_1(\tau) d\tau$$

we find

$$\begin{aligned} d^2 \hat{G}_t[\tilde{\omega}]((\hat{\zeta}_1, v_1), (\hat{\zeta}_2, v_2)) &= \int_0^t d_{\hat{q}} \hat{g}'(\tau) \zeta_1 v_2(\tau) d\tau + \int_0^t \hat{g}''(\tau) (v_1(\tau), v_2(\tau)) d\tau + \\ &+ \int_0^t d_{\hat{q}} \hat{g}'(\tau) \hat{\zeta}_2 v_1(\tau) d\tau + \int_0^t \hat{g}'(\tau) v_2(\tau) d\tau \circ \int_0^t \hat{g}'(\tau) v_1(\tau) d\tau + \\ &+ \int_0^t \left[\int_0^\tau \hat{g}'(\theta) v_1(\theta) d\theta, \hat{g}'(\tau) v_2(\tau) \right] d\tau. \end{aligned}$$

We need to reinterpret each summand in terms of symplectic geometry. One can check (or see [8]), that

$$\int_0^t \langle \hat{\lambda}(0), \hat{g}''(\tau)(v_1(\tau), v_2(\tau)) \rangle d\tau = \int_0^t b(\tau)(v_1(\tau), v_2(\tau)) d\tau$$

and

$$\left\langle \hat{\lambda}(0), \int_0^t \left[\int_0^\tau \hat{g}'(\theta)v_1(\theta)d\theta, \hat{g}'(\tau)v_2(\tau) \right] d\tau \right\rangle = \int_0^t \hat{\sigma} \left(\int_0^\tau \hat{X}(\theta)v_1(\theta)d\theta, \hat{X}(\tau)v_2(\tau) \right) d\tau.$$

To give an interpretation to the first term let us choose Darboux coordinates in $T_{\hat{\lambda}(0)}(T^*M)$ subordinate to the Lagrangian splitting $T_{\hat{\lambda}(0)}(T^*(M \times \mathbb{R})) = T_{(\hat{q}(0),0)}(M \times \mathbb{R}) \times T_{(\hat{q}(0),0)}^*(M \times \mathbb{R})$. We note that in this case $-d_{\hat{q}}\hat{g}'(\tau)v_2(\tau) \cdot$ can be associated with a covector that is nothing but the projection of $\hat{X}(\tau)v_2(\tau)$ to the fibre. Therefore we have

$$\int_0^t d_{\hat{q}}\hat{g}'(\tau)\zeta_1 v_2(\tau) d\tau = \int_0^t \hat{\sigma}(\hat{\zeta}_1, \hat{X}(\tau)v_2(\tau)) d\tau.$$

Finally for the term in the middle, we use the fact that we are restricting to the kernel of $dE_{N_0,t}$. On it we have

$$\zeta_2 = - \int_0^t g'(\tau)v_2(\tau) d\tau.$$

Since $\hat{\zeta}_2 = (\zeta_2, 0)$, we find that

$$\int_0^t d_{\hat{q}}\hat{g}'(\tau)\zeta_1 v_2(\tau) d\tau = \int_0^t g'(\tau)v_2(\tau) d\tau \circ \int_0^t \hat{g}'(\tau)v_1(\tau) d\tau$$

We can write $\hat{g}'(\tau) = g'(\tau) + (g^0)'(\tau)\partial_y$, then

$$\begin{aligned} & \int_0^t d_{\hat{q}}\hat{g}'(\tau)\hat{\zeta}_2 v_1(\tau) d\tau + \int_0^t \hat{g}'(\tau)v_2(\tau) d\tau \circ \int_0^t \hat{g}'(\tau)v_1(\tau) d\tau = \\ & = \int_0^t (g^0)'(\tau)v_2(\tau) d\tau \partial_y \circ \int_0^t \hat{g}'(\tau)v_1(\tau) d\tau, \end{aligned}$$

but as we have seen the horizontal part is independent of the y variable. So this terms is just zero.

Collecting everything we find an explicit formula for the Hessian

$$\begin{aligned} & \text{Hess}(E_{N_0,t}, \nu J_t)[\omega, \lambda(t)]((\zeta_1, v_1), (\zeta_2, v_2)) = \\ & = \int_0^t \hat{\sigma} \left(\hat{\zeta}_1 + \int_0^\tau \hat{X}(\theta)v_1(\theta)d\theta, \hat{X}(\tau)v_2(\tau) \right) + b(\tau)(v_1(\tau), v_2(\tau)) d\tau. \end{aligned}$$

This can be simplified even further, if we note that the ν component of $\hat{X}(\tau)$ is equal to zero, which can be easily seen from the definitions. Therefore from the explicit form of $\hat{\sigma}$ we derive that

$$\hat{\sigma}(\hat{X}(\theta), \hat{X}(\tau)) = \sigma(X(\theta), X(\tau))$$

and

$$\hat{\sigma}((\zeta_1, 0), \hat{X}(\tau)) = \sigma(\zeta_1, X(\tau)),$$

which proves the proposition. □

We note that although we assumed initially that the space of admissible controls is in L^∞ , the differential $dE_{N_0,t}[\tilde{\omega}]$ and the quadratic form Q from (1.6) are continuous in the L^2 -topology of controls. Thus we can extend them by continuity to a weaker topology. Theorem 1.33 implies that \mathcal{L} -derivatives will not change. This allow us to prove a simple, but important lemma.

Lemma 2.9. *The Jacobi curve \mathcal{L}_t is left continuous.*

Proof. Without any loss of generality we assume that all variations are two-sided. We compute \mathcal{L}_t over the space of piecewise constant functions with zero on the last interval. This space is dense in $L^2([0, t], \mathbb{R}^k)$ and therefore \mathcal{L}_t does not change. Fix a neighborhood $O_{\mathcal{L}_t} \subset L(T_{\lambda(0)}(T^*M))$ of \mathcal{L}_t . Then by definition of a generalized sequence there exists a finite-dimensional subspace V of simple functions, s.t. for all $W \supset V$ one has $\mathcal{L}_t(W) \in O_{\mathcal{L}_t}$. Let $\alpha = \{0 = t_0 < t_1 < \dots < t_N = t\}$ be the set of jump points of all variations $v(t) \in V$. By construction $v(t) = 0$ for $t \in [t_{N-1}, t_N]$. We define $V_\beta \supset V$ to be the space of simple functions $v^\beta(t)$ with possible discontinuities in $\beta \supset \alpha$, s.t. $v^\beta(t) = 0$ for $t \in [t_{N-1}, t_N]$. Then by definition $\mathcal{L}_t(V_\beta) \in O_{\mathcal{L}_t}$ for any $\beta \supset \alpha$. By refining the partition β on $[t_0, t_{N-1}]$ we obtain

$$\lim \mathcal{L}_t(V_\beta) = \mathcal{L}_t(L^2([0, t_{N-1}], \hat{U})) = \mathcal{L}_{t_{N-1}} \in O_{\mathcal{L}_t}.$$

Since t_{N-1} can be arbitrary close to t , the result follows. □

2.2 Approximation algorithm

As we have mentioned before we will construct the \mathcal{L} -derivatives using only two-sided variations. Thus we must restrict the operator $X(t)$ to the tangent spaces of U_i . But by assumptions each U_i is embedded in \mathbb{R}^k . So let us choose any metric in the ambient space and for each point $u \in U_i$ we define an orthogonal projection $\pi_u^i : \mathbb{R}^k \rightarrow T_u U_i$ that depends on a point. Then we use this to define a projection of a given variation to the subspaces of two-sided variations as

$$\pi_\tau v(\tau) = \sum_{i=1}^n \chi_{U_i}(\tilde{u}(\tau)) \pi_{\tilde{u}(\tau)}^i v(\tau),$$

where χ_{U_i} is the indicator function of U_i .

We can see that finite-dimensional approximations to the \mathcal{L} -derivative will depend on the choice of the metric in \mathbb{R}^k , but the limit \mathcal{L} -derivative itself will not, since we approximate the same space of variations in two different ways.

For our algorithm we will need variations “constant” in time. We will define them as projection of constant sections $v \in \mathbb{R}^k$. Equivalently we can replace $X(\tau)$ by $X(\tau)\pi_\tau$ considered as a time dependent map from \mathbb{R}^k to $T_{\lambda(0)}(T_{\tilde{q}(0)}^*M)$. This way we have reduced our problem to a problem without the constraints in the control and from now on we assume that the variations $v(t)$ can take any value in \mathbb{R}^k .

The \mathcal{L} -derivative for optimal control problems enjoys several useful properties. First of all we have seen that the \mathcal{L} -derivative \mathcal{L}_T exists if $\text{ind}^\pm \text{Hess}(E_{N_0,T}, \nu J_T)[\tilde{\omega}, \lambda(T)] < +\infty$. But since $E_{N_0,T}|_{\Omega \cap L_k^\infty[0,t]} = E_{N_0,t}$ for $t \leq T$ and $L_k^\infty[0,t] \subset L_k^\infty[0,T]$ is an isometrical embedding, we have that the existence at a moment of time T implies the existence for all $t \leq T$.

Secondly we can compute \mathcal{L}_t using a dense subspace of $L_k^\infty[0,t]$. But one can also do the contrary and expand $L_k^\infty[0,t]$ to some weaker space. The \mathcal{L} -derivative will not change if the first and the second differential are continuous in a weaker norm. One can note from formulas (2.5) and (2.12), that the first and the second derivatives of $(E_{N_0,t}, J_t)$ are actually continuous in $L_k^2[0,t]$. That is why from now on we use the space of square-integrable functions as our space of variations.

Next we prove one more useful property that greatly simplifies the computation of \mathcal{L}_t .

Lemma 2.10 (Additivity). *Take $0 < t_1 < t_2$ and suppose that the index of the Hessian of the extremal curve on $[0, t_2]$ is finite. We denote by V_2 some finite dimensional subspace of $L_k^2[t_1, t_2]$ and we consider the following equation*

$$\int_{t_1}^{t_2} \left[\sigma \left(\lambda + \int_{t_1}^\tau X(\theta)v_2(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau) (v_2(\tau), w(\tau)) \right] d\tau = 0, \quad \forall w(\tau) \in V_2, \quad (2.13)$$

where $v_2(\tau) \in V_2$, and $\lambda \in \mathcal{L}_{t_1}$.

Then \mathcal{L}_{t_2} is a generalized limit of Lagrangian subspaces

$$\left\{ \lambda + \int_{t_1}^{t_2} X(\tau)v(\tau)d\tau : \lambda \in \mathcal{L}_{t_1}, v(\tau) \in V_2 \text{ satisfies (2.13) for any } w(\tau) \in V_2 \right\}.$$

Proof. By the existence theorem and the remark above we know that \mathcal{L}_{t_2} and \mathcal{L}_{t_1} exist and \mathcal{L}_{t_2} can be computed over any dense subspace of the variation space. So we compute it over $V_1 \oplus V_2 = V \subset L_k^2[0, t_2]$, where V_1 is a span of a countable dense subset in $L_k^2[0, t_1]$. Denote by π_i the projection onto V_i .

Now fix a neighborhood $O_{\mathcal{L}_{t_2}}$ in the Lagrangian Grassmanian and consider a finite-dimensional subspace $W \subset V$, s.t. for any finite dimensional $U \supset W$ we have $\mathcal{L}_t(U) \in O_{\mathcal{L}_{t_2}}$. Then we can construct a countable sequence of nested subspaces

$$U_1 \subset U_2 \subset \dots$$

by adding vectors from the basis of V_1 . As a result we get a sequence $\mathcal{L}_t(U_i)$ which converges to $\mathcal{L}_t(V_1 \oplus \pi_2(W))$, since the index over the Hessian on this subspace must be finite as well. Note that in this case by construction $\mathcal{L}_t(V_1 \oplus \pi_2(W)) \subset O_{\mathcal{L}_{t_2}}$. By taking finer and finer $O_{\mathcal{L}_{t_2}}$ we realize \mathcal{L}_{t_2} as a limit of vectors from $\mathcal{L}_t(V_1 \oplus \pi_2(W))$.

It remains to show that (2.13) holds. And indeed, any element of $\mathcal{L}_t(U_i)$ is of the form

$$\eta + \int_0^{t_1} X(\tau)v_1^i(\tau)d\tau + \int_{t_1}^{t_2} X(\tau)v_2(\tau)d\tau, \quad v_1^i \in V_1 \cap U_i \quad v_2 \in \pi_2(W)$$

s. t.

$$\int_0^{t_1} \left[\sigma\left(\eta + \int_0^\tau X(\theta)v_1^i(\theta)d\theta, X(\tau)w_1(\tau)\right) + b(\tau)(v_1^i(\tau), w_1(\tau)) \right] d\tau = 0$$

$$\int_{t_1}^{t_2} \left[\sigma\left(\eta + \int_0^t X(\theta)v_1^i(\theta)d\theta + \int_{t_1}^t X(\theta)v_2(\theta)d\theta, X(\tau)w_2(\tau)\right) + b(\tau)(v_2(\tau), w_2(\tau)) \right] d\tau = 0$$

for any $w_1 \in V_1 \cap U_i$, $w_2 \in \pi_2(W)$. Therefore as we take the limit, the vectors

$$\eta + \int_0^{t_1} X(\tau)v_1^i(\tau)d\tau$$

will converge to vectors from $\mathcal{L}_t(V_1)$. □

These properties are enough to have an algorithm for computing \mathcal{L}_t at each moment of time t with arbitrary good precision. In fact, if the index of the Hessian is finite, we have existence. Since we can replace $L_k^2[0, t]$ with any dense subset, we compute \mathcal{L}_t over the space of piecewise constant functions. To construct an approximation of \mathcal{L}_t we just have to take some partition $D = \{0 < t_1 < t_2 < \dots < t_N = t\}$ of $[0, t]$ and construct $\mathcal{L}_t(V_D)$, where $V_D \subset \Omega_{N_0}$ is the space of variations of the initial point and piece-wise constant variations of the control with jumps at D . Then we can use the additivity lemma to iteratively construct an approximation to $\mathcal{L}_t(V_D)$, given by $T_{\bar{q}(0)}^\perp N_0 \times T_{\bar{q}(0)} N_0 = \mathcal{L}_t(V^{\{0\}})$, $\mathcal{L}_t(V^{\{0, t_1\}})$, $\mathcal{L}_t(V^{\{0, t_1, t_2\}})$ and so on. So at the end we just need to understand how \mathcal{L}_t changes when we add constant variations $\mathbb{R}^k \chi_{[t, t+\varepsilon]}$. In this case at each step we need to solve an over-determined finite-dimensional linear system. A convenient machinery for such type of equations is the notion of pseudo-inverses. We recall their basic definition.

Definition 2.11. Let $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be a linear map between two Euclidean spaces and A^* be its adjoint. Then the Moore-Penrose pseudoinverse A^+ can be defined as

$$A^+ = \lim_{\varepsilon \rightarrow 0} (\varepsilon \text{id} + A^* A)^{-1} A^*.$$

The Moore-Penrose inverse has many interesting properties. The most useful one for us will be the following one.

Proposition 2.12. *If the linear solution $Ax = b$ admits at least one solution, then $y = A^+b$ is the minimal norm solution of this equation.*

Now we are ready to describe our algorithm.

Theorem 2.13. *Suppose that we know $\mathcal{L}_t(V)$, where V is some space of variations defined on $[0, t]$. We identify $\mathcal{L}_t(V)$ with \mathbb{R}^n and the space of control parameters with \mathbb{R}^k , and put an arbitrary Euclidean metric on both of them. Let E be the space of all $v \in \mathbb{R}^k$ for which*

$$\sigma \left(\eta, \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \right) = 0, \quad \forall \eta \in \mathcal{L}_t(V)$$

and let $L \subset \mathcal{L}_t(V)$ consisting of all $\eta \in \mathcal{L}_t(V)$, s.t.

$$\sigma \left(\eta, \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot w \right) = 0, \quad \forall w \in \mathbb{R}^k.$$

We define two bilinear maps $A_R : \mathcal{L}_t(V) \times E^\perp \rightarrow \mathbb{R}$, $Q_R : E^\perp \times E^\perp \rightarrow \mathbb{R}$:

$$\begin{aligned} A_R : (\eta, w) &\mapsto \sigma \left(\eta, \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot w \right), \\ Q_R : (v, w) &\mapsto \frac{1}{\varepsilon} \int_t^{t+\varepsilon} \sigma \left(\int_t^\tau X(\theta) d\theta \cdot v, X(\tau) w \right) + b(\tau)(v, w) d\tau, \end{aligned}$$

and we use the same symbols for the corresponding matrices.

Then the new \mathcal{L} -derivative $\mathcal{L}_{t+\varepsilon}(V \oplus \mathbb{R}^k \chi_{[t, t+\varepsilon]})$ is a span of vectors from the subspace L and vectors

$$\eta_i + \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v_i,$$

where v_i is an arbitrary basis of E^\perp and η_i are defined as

$$\eta_i = -A_R^+ Q_R v_i.$$

Proof. From the additivity lemma it follows that it is sufficient to construct n independent solutions of the equation

$$\int_t^{t+\varepsilon} \sigma \left(\eta + \frac{1}{\varepsilon} \int_t^\tau X(\theta) d\theta \cdot v, X(\tau) w \right) + \frac{b(\tau)(v, w)}{\varepsilon} d\tau = 0, \quad \forall w \in \mathbb{R}^k, \quad (2.14)$$

where $\eta \in \mathcal{L}_t(V)$, $v \in \mathbb{R}^k$. The idea of the prove can be easily seen from the statement. One has to show that the subspaces L and E don't give non-trivial contributions to the new \mathcal{L} -derivative. Meaning that (2.14) is well defined on the corresponding quotients.

Consider subspace

$$L = \mathcal{L}_t(V) \cap \mathcal{L}_{t+\varepsilon}(V \oplus \mathbb{R}^k \chi_{[t, t+\varepsilon]}) = \left\{ \eta \in \mathcal{L}_t(V) : \sigma \left(\eta, \int_t^{t+\varepsilon} X(\tau) d\tau \cdot w \right) = 0, \forall w \in \mathbb{R}^k \right\}.$$

Suppose that the dimension of this space is equal to $n - l$, where $l \leq \min\{k, n\}$. But since this is just a homogeneous system of k linear equations with n variables, it means that there must exist $k - l$ vectors $v \in \mathbb{R}^k$ for which

$$\sigma \left(\eta, \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \right) = 0, \quad \forall \eta \in \mathcal{L}_t(V) \quad \Rightarrow \quad \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \in \mathcal{L}_t(V) \quad (2.15)$$

since $\mathcal{L}_t(V)$ is a Lagrangian subspace.

We note that vectors (2.15) are just linear combinations of $\eta \in \mathcal{L}_t(V)$. Since we are looking for solutions of the form

$$\eta + \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v, \quad \eta \in \mathcal{L}_t(V), v \in \mathbb{R}^k$$

we can just take $v \in E^\perp$, $\dim E^\perp = l$, or else we would have replaced the part from E with the corresponding $\eta \in \mathcal{L}_t(V)$. This basically means that we have reduced our system (2.14) of k linear equations with $n + k$ variables to a system with just $n + \dim E^\perp$ variables. But then the only way that we can have n independent solutions, if there exist $k - \dim E^\perp$ dependent relations in (2.14) with $v \in E^\perp$. This means that there must exist $k - \dim E^\perp = \dim E$ independent $w_i \in \mathbb{R}^k$ for which

$$\int_t^{t+\varepsilon} \sigma \left(\eta + \frac{1}{\varepsilon} \int_t^\tau X(\theta) d\theta \cdot v, X(\tau) w_i \right) + \frac{b(\tau)(v, w_i)}{\varepsilon} d\tau = 0, \quad \forall v \in E^\perp, \forall \eta \in \mathcal{L}_t(V).$$

In particular

$$\int_t^{t+\varepsilon} \sigma(\eta, X(\tau) w_i) d\tau = 0, \quad \forall \eta \in \mathcal{L}_t(V)$$

i.e. w_i form a basis of E . So we see that (2.14) is reduced to

$$A_R \eta = -Q_R v.$$

The solution of this equation necessarily exists for any $v \in E^\perp$. Indeed, the \mathcal{L} -derivative consists of vectors from L and some vectors constructed from solutions of this equations. Since by definition $L = \ker A_R$, we have that these solutions are unique modulo elements of L . Moreover we have $\dim L + \dim E^\perp = n$. So if assume, that there exists $v \in E^\perp$ for which there is no solution η , then we would have arrived at a contradiction with the fact that \mathcal{L} -derivative is a Lagrangian plane.

This implies that we can apply the pseudo-inverse A_R^+ to find $n - \dim L$ independent vectors

$$\eta_i + \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot e_i.$$

These are indeed independent, because by the definition E contains all the vectors v s.t.

$$\int_t^{t+\varepsilon} X(\tau) d\tau \cdot v = 0.$$

This means that

$$v \mapsto \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \tag{2.16}$$

is a bijection between the image of this map and E^\perp , and therefore independent $v_i \in E^\perp$ are mapped to independent vectors in the image. This finishes the proof. \square

We stress once again that the \mathcal{L} -derivative itself is invariant and does not depend on the choices we make. The proven theorem is going to play an essential role in the Morse-type theorems that we are going to state and prove in Section 2.5.

2.3 A quick application: the bang-bang case

The algorithm becomes particularly nice if we have a single control parameter. In this case $b(\tau)$ is just a function and $X(\tau)$ is a \mathbb{R}^{2n} -valued vector function.

Proposition 2.14. *Consider a single control parameter system. Given a \mathcal{L} -derivative $\mathcal{L}_t(V)$, where V is some space of variations defined on $[0, t]$, we have $\mathcal{L}_t(V \oplus \mathbb{R}\chi_{[t, t+\varepsilon]}) = \mathcal{L}_t(V)^{\eta(t+\varepsilon)}$, where $\chi_{[t, t+\varepsilon]}$ is the characteristic function of the corresponding interval and $\eta(t + \varepsilon)$ is determined by one of the two alternatives*

1. If

$$\int_t^{t+\varepsilon} X(\tau) d\tau \in \mathcal{L}_t(V)$$

then $\mathcal{L}_t(V \oplus \mathbb{R}\chi_{[t, t+\varepsilon]}) = \mathcal{L}_t(V)$ and we can take $\eta(t + \varepsilon)$ to be any vector from $\mathcal{L}_t(V)$

2. Else we fix any $\eta(t) \in \mathcal{L}_t(V)$ satisfying

$$\sigma\left(\eta(t), \int_t^{t+\varepsilon} X(\tau) d\tau\right) \neq 0$$

and take

$$\eta(t + \varepsilon) = K\eta(t) + \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau,$$

where

$$K = -\frac{\frac{1}{\varepsilon} \int_t^{t+\varepsilon} \left[\sigma\left(\int_t^\tau X(\theta) d\theta, X(\tau)\right) + b(\tau) \right] d\tau}{\sigma\left(\eta(t), \int_t^{t+\varepsilon} X(\tau) d\tau\right)}.$$

One can apply this immediately to bang-bang extremals of fixed boundary problems. Let us assume that we have a time optimal control problem, i.e. $L \equiv 1$, and consider an extremal pair $(\tilde{q}(t), \tilde{u}(t))$ which is bang-bang. This means that the control $u(t)$ is piece-wise constant and takes values on the boundary of U . In this case we only have time variations as admissible variations.

Let

$$f(q, \tilde{u}(t)) =: f_i(q), \quad t \in [t_i, t_{i+1}].$$

be the controlled system and the minimized functional on an interval of constancy $[t_i, t_{i+1}]$ of $u(t)$. Since the system under consideration was autonomous we have that $X(\tau)$ is also a piece-wise constant function. We define $X_i = X(\tau)$, $\tau \in (t_i, t_{i+1}]$. We have

$$(P_0^\tau)_*^{-1} f(q, \tilde{u}(\tau)) = e_*^{-t_1 f_1} (e_*^{(t_1 - t_2) f_2} (\dots (e_*^{(t_{i+1} - t_i) f_{i-1}} f_i) \dots)), \quad \tau \in (t_i, t_{i+1}]$$

So if we let $h_i(\lambda) = \langle \lambda, (P^\tau)_*^{-1} f(q, \tilde{u}(\tau)) \rangle$, $t \in (t_i, t_{i+1}]$, then $X_i = \vec{h}_i$.

We can now apply Proposition 2.14 to find an approximation of the Jacobi curve. We take V_j to be the space of variations constant on the intervals $[t_i, t_{i+1}]$ and which are zero

for $t \geq t_j$. We have $\mathcal{L}_t(\{0\}) = T_{\lambda(0)}(T_{\tilde{q}(0)}^*M)$ and it is possible now to apply inductively Proposition 2.14.

Since the new system and the Lagrangian are linear in controls, we obtain $b(\tau) \equiv 0$. Then

$$\sigma \left(\int_{t_i}^{\tau} X(\theta) d\theta, X(\tau) \right) + b(\tau) = (\tau - t_i) \sigma(X_i, X_i) = 0, \quad \forall \tau \in [t_i, t_{i+1})$$

and so $K = 0$ on each step. Then

$$\eta(t_{i+1}) = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} X(\tau) d\tau = X_i.$$

This way we obtain a sequence of Lagrangian subspace $\mathcal{L}_t(V_j)$, defined inductively as

$$\mathcal{L}_t(V_0) = \mathcal{L}_t(\{0\}) = T_{\lambda(0)}(T_{\tilde{q}(0)}^*M), \quad \mathcal{L}_t(V_{i+1}) = \mathcal{L}_t(V_i)^{X_i}.$$

If we take a finer splitting of the interval the corresponding approximation to the Jacobi curve is the same as above, because $(\mathcal{L}_t(V_i)^{X_i})^{X_i} = \mathcal{L}_t(V_i)^{X_i}$.

The final algorithm of constructing the Jacobi curve goes as follows. One defines $\mathcal{L}_0 = T_{\lambda(0)}(T_{\tilde{q}(0)}^*M)$. The Jacobi curves \mathcal{L}_τ is constant for $\tau \in (t_i, t_{i+1}]$ and after a switching it jumps to $\mathcal{L}(t_i+) = \mathcal{L}(t_i)^{X_i}$. This is the same algorithm that was obtained in [11].

2.4 Gluing formula

In the previous sections in order to construct an approximation we have used the additivity Lemma 2.10 which essentially exploits the direction of time. We could have used it in the other direction by extending the support of variations on the left instead of right.

This is very useful, for example, in the case of the Fuller phenomena. We have already computed Jacobi curves of bang-bang arcs. If we have a bang-singular arc, we can use and then we apply the additivity Lemma 2.10 and after this we can compute separately the Jacobi curve of the singular arc using Jacobi differential equations, but with the correct boundary conditions (see Chapter 3).

If the number of switches in the bang-arc is finite, we do not really care from which of the two endpoints to start. We could have easily reversed the direction of time and first used the Jacobi equation and only after the algorithm for a bang-bang extremal. But if the number of switches is infinite, we can only apply the known algorithms in one direction, namely bang-singular, because in the other direction we have a singularity that must be resolved. On the other hand if we have a singular-bang arc, we can simply reverse the direction of time by taking $s = T - t$ as the new time variable.

So a natural question arises that can be roughly stated as follows: if we have already computed two different Jacobi curves using two spaces of variations with non-intersecting support, is it possible from this information to compute the Jacobi curve constructed over

the sum of the two spaces of variations? In the case of bang-singular-bang extremals this corresponds to computing separately Jacobi curves of bang-singular and singular-bang arcs and gluing the two.

Let us consider a fixed end-point problem. In this case in order to give an affirmative answer to this question, we must consider a different \mathcal{L} -derivative and here the definition of the evaluation maps F_t is essential. What we need is the \mathcal{L} -derivative $\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[(\tilde{\omega}, (-\lambda(t_0), \lambda(t_1)))]$ defined like in the Definition 2.6. In order to compute it, we can already apply the algorithm from the previous section. Indeed, we can consider the following optimal control problem for an extended control system

$$\begin{aligned} \dot{x} &= 0, \\ \dot{q} &= f(u, q), \\ J_{t_0}^{t_1} &= \int_{t_0}^{t_1} L(u, q) dt \rightarrow \min \end{aligned}$$

with the boundary conditions

$$(x_{t_0}, q_{t_0}) \in \{(x, x) \in M\} \subset M \times M.$$

Following Remark 2.5, we have simply introduced some special coordinates, such that F_{t_0} is linear and the space of variations splits into variations of control and variations of the initial point (we can move it freely on the diagonal). If we apply now the Lagrange multiplier rule we will find that $(\tilde{\omega}, (-\lambda(t_0), \lambda(t_1)))$ is a Lagrange point if and only if $\hat{\lambda}(t_0) = (\hat{P}_{t_0}^{t_1})^* \hat{\lambda}(t_1)$, allowing us to recover the Hamiltonian system.

We can now compute $\mathcal{L}(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})[(\tilde{\omega}, (-\lambda(t_0), \lambda(t_1)))]$ using our algorithm or some other method. This object contains all information about the \mathcal{L} -derivatives of end-point maps, that we have considered so far. More precisely, we can reconstruct the latter one from the former one.

Lemma 2.15. *Let $(\omega, (-\lambda(t_0), \lambda(t_1)))$ be a Lagrange point of the map $(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})$. We consider two projections*

$$\begin{aligned} \pi_0 &: T_{-\lambda(t_0)}(T^*M) \times T_{\lambda(t_1)}(T^*M) \rightarrow T_{-\lambda(t_0)}(T^*M), \\ \pi_1 &: T_{-\lambda(t_0)}(T^*M) \times T_{\lambda(t_1)}(T^*M) \rightarrow T_{\lambda(t_1)}(T^*M). \end{aligned}$$

If $\omega \in F_{t_0}^{-1}(q_0)$ with $q_0 \in M$, then $(\omega, \lambda(t_1))$ is a Lagrange point of $F_{t_1}|_{F_{t_0}^{-1}(q_0)}$ and the corresponding \mathcal{L} -derivative can be computed as

$$\begin{aligned} \mathcal{L}(F_{t_1}, \nu J_{t_0}^{t_1})[(\omega, \lambda(t_1))](F_{t_0}^{-1}(q_0)) &= \\ &= \pi_1 \left(\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[(\omega, (-\lambda(t_0), \lambda(t_1)))] \cap \pi_0^{-1}(T_{-\lambda(t_0)}(T_{F_{t_0}(u)}M)) \right). \end{aligned}$$

Proof. Since the formulation only involves projection maps, it is clear that there will be no problem in taking the generalized limits and we can assume that the space of variations is finite. We use the special coordinates mentioned in Remark 2.5. Since F_{t_0}

is a submersion, its differential has maximum rank and we can always find coordinates in which $d^2F_0[\omega] = 0$. We split the space of variations into a direct sum $V_0 \oplus V_1$, where V_0 is a subspace isomorphic to $T_{F_0(\omega)}M$ and V_1 is in the closure of the orthogonal complement of V_0 with respect to the quadratic form $Q = \langle \lambda(t_1), d^2F_{t_1}[\omega] \rangle - \nu dJ_{t_0}^{t_1}[\omega]$, i.e.

$$\langle \lambda(t_1), d^2F_{t_1}[\omega](v_0, w_1) \rangle - \nu dJ_{t_0}^{t_1}[\omega](v_0, w_1) = 0, \quad \forall v_0 \in V_0, w_1 \in V_1. \quad (2.17)$$

The fact that $(\omega, \lambda(t_1))$ is a Lagrange point now follows easily. Indeed, we have that $(\omega, (-\lambda(t_0), \lambda(t_1)))$ is a Lagrange point if for any $w = w_0 + w_1$ the following equation is satisfied

$$\begin{aligned} & \langle \lambda(t_1), dF_{t_1}[\omega](w_0) \rangle + \langle \lambda(t_1), dF_{t_1}[\omega](w_1) \rangle - \langle \lambda(t_0), dF_{t_0}[\omega](w_0) \rangle = \\ & = \nu dJ_{t_0}^{t_1}[\omega](w_0) + \nu dJ_{t_0}^{t_1}[\omega](w_1). \end{aligned}$$

If we restrict all maps to $T_\omega(F_{t_0}^{-1}(q_0))$, or equivalently we take $w_0 = 0$, then we obtain exactly conditions for $(\omega, \lambda(t_1))$ being a Lagrange point of $F_{t_1}|_{F_{t_0}^{-1}(q_0)}$.

The second part also is just a consequence of the basic definitions. In these coordinates by (2.17) we have the following equation for \mathcal{L} -derivative of $(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})$:

$$\begin{aligned} & \langle \xi(t_1), dF_{t_1}[\omega](w_0) \rangle + \langle \xi(t_1), dF_{t_1}[\omega](w_1) \rangle - \\ & - \langle \xi(t_0), dF_{t_0}[\omega](w_0) \rangle + \langle \lambda(t_1), d^2F_{t_1}[\omega](v_0, w_0) \rangle + \\ & + \langle \lambda(t_1), d^2F_{t_1}[\omega](v_1, w_1) \rangle = \nu d^2J_{t_1}^{t_2}[\omega](v_0, w_0) + \nu d^2J_{t_1}^{t_2}[\omega](v_1, w_1). \end{aligned} \quad (2.18)$$

Similarly the equation for the \mathcal{L} -derivative of $(F_{t_1}, J_{t_0}^{t_1})|_{F_{t_0}^{-1}(q_0)}$ can be written as

$$\langle \xi(t_1), dF_{t_1}[\omega](w_1) \rangle + \langle \lambda(t_1), d^2F_{t_1}[\omega](v_1, w_1) \rangle = \nu d^2J_{t_1}^{t_2}[\omega](v_1, w_1). \quad (2.19)$$

Let $(v_1, \xi(t_1))$ be a solution of the last equation. Then clearly the quadruple

$$(v_1, \xi(t_1), v_0 = 0, \xi(t_0) = \langle \xi(t_1), dF_{t_1}[\omega](dF_{t_0}[\omega])^{-1} \rangle)$$

is a solution of (2.18), where $(dF_{t_0}[\omega])^{-1} : T_{F_{t_0}(\omega)}M \rightarrow V_0$. But, since $v_0 = 0$, all of those solutions will indeed lie in $\pi_0^{-1}(T_{-\lambda(t_0)}(T_{F_{t_0}(\omega)}M))$, since the corresponding \mathcal{L} -derivative consists of vectors of the form

$$((\xi(t_0), dF_{t_0}[\omega](v_0)), (\xi(t_1), dF_{t_1}[\omega](v_0 + v_1))) \in T_{-\lambda(t_0)}(T^*M) \times T_{\lambda(t_1)}(T^*M).$$

So we get the the left inclusion. On the other hand if we take an element from

$$\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[\omega, (-\lambda(t_0), \lambda(t_1))] \cap \pi_0^{-1}(T_{-\lambda(t_0)}(T_{F_0(u)}M)),$$

then it corresponds to a solution of (2.18) with $v_0 = 0$ for any $w = w_0 + w_1$. In particular, (2.18) is satisfied for $w_0 = 0$. But in this case (2.18) reduces to (2.19). \square

Using the last lemma and Lemma 1.2, we can prove the following result.

Theorem 2.16. *Let $(\omega, (-\lambda(t_0), \lambda(t_1)))$ and $(\omega, (-\lambda(t_1), \lambda(t_2)))$ be a Lagrange point of the map $(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})$ and $(F_{t_1}, F_{t_2}, \nu J_{t_1}^{t_2})$. We assume that \mathcal{L} -derivatives of $(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})$ and $(F_{t_1}, F_{t_2}, \nu J_{t_1}^{t_2})$ are decomposed like in Lemma 1.2 as*

$$\begin{aligned}\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[(\omega, (-\lambda(t_0), \lambda(t_1)))] &= (\Gamma_0, 0) \oplus \text{graph } \Phi_{01} \oplus (0, \Gamma_1), \\ \mathcal{L}(F_{t_1}, F_{t_2}, \nu J_{t_1}^{t_2})[(\omega, (-\lambda(t_1), \lambda(t_2)))] &= (\tilde{\Gamma}_1, 0) \oplus \text{graph } \Phi_{12} \oplus (0, \Gamma_2).\end{aligned}$$

Let $\Gamma_{\ker} \subset \Gamma_1$ be a subspace isomorphic to $(\ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1} \cap \Gamma_1) / (\Gamma_1 \cap \tilde{\Gamma}_1)$ and $\tilde{\Gamma}_{\ker} \subset \tilde{\Gamma}_1$ be isomorphic to $(\ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1} \cap \tilde{\Gamma}_1) / (\Gamma_1 \cap \tilde{\Gamma}_1)$.

Then $(\omega, (-\lambda(t_0), \lambda(t_2)))$ is a Lagrange point of $(F_{t_0}, F_{t_2}, J_{t_0}^{t_2})$ and

$$\mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})[(\omega, (-\lambda(t_0), \lambda(t_2)))] = (\Gamma_0 \oplus \Phi_{01}^{-1}(\tilde{\Gamma}_{\ker}), 0) \oplus \text{graph } \Phi_{02} \oplus (0, \Phi_{12}(\Gamma_{\ker}) \oplus \Gamma_2),$$

where $\text{graph } \Phi_{02}$ is a graph of the symplectic map

$$\Phi_{02} = \Phi_{12} \circ \Phi_{01} : \Phi_{01}^{-1}((\Gamma_1 + \tilde{\Gamma}_1)^\angle / \ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1}) \rightarrow \Phi_{12}((\Gamma_1 + \tilde{\Gamma}_1)^\angle / \ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1}).$$

Proof. In this proof for simplicity we do not indicate explicitly the Lagrange points assuming that they are the same as indicated in the statement of the theorem.

The fact that $(\omega, (-\lambda(t_0), \lambda(t_2)))$ is a Lagrange point of $(F_{t_0}, F_{t_2}, J_{t_0}^{t_2})$ follows immediately from the definition.

To prove the statement we first prove *the chain rule* for \mathcal{L} -derivatives that can be stated as follows. Assume that $x_{01} \in \mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})$ and $x_{12} \in \mathcal{L}(F_{t_1}, F_{t_2}, \nu J_{t_1}^{t_2})$ are such that $\pi_1(x_{01}) = \pi_1(x_{12})$. Then $(\pi_0(x_{01}), \pi_2(x_{12})) \in \mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$. To prove this we use local coordinates and split the space of variations into $V_0 \oplus V_1 \oplus V_2$, where V_1 is isomorphic to $T_{F_{t_1}(u)}M$, V_0 is isomorphic to $L_k^\infty[t_0, t_1]$ and V_2 to $L_k^\infty[t_1, t_2]$. In other words V_1 are variations of the mid point of the curve and V_0, V_2 are variations of the control on intervals $[t_0, t_1]$ and $[t_1, t_2]$ correspondingly, then we have $d^2 F_{t_1}[\omega] = 0$, $dF_{t_0}[\omega](w_2) = d^2 F_{t_0}[\omega](\cdot, w_2) = 0$, $dF_{t_2}[\omega](w_0) = d^2 F_{t_2}[\omega](\cdot, w_0) = 0$ and by additivity of the functional

$$d^2 J_{t_0}^{t_2}[\omega](v_0 + v_1 + v_2, w_0 + w_1 + w_2) = d^2 J_{t_0}^{t_1}[\omega](v_0 + v_1, w_0 + w_1) + d^2 J_{t_1}^{t_2}[\omega](v_1 + v_2, w_1 + w_2).$$

The chain rule for the finite-dimensional approximations follows now easily. We can see this by writing the equation for the \mathcal{L} -derivative, and adding and subtracting $\langle \xi(t_1), dF_{t_1}[\tilde{\omega}](w_1) \rangle$. For the infinite dimensional we can assume that we construct the \mathcal{L} -derivatives using piecewise constant functions. Fix three neighbourhoods $O_{ij} \ni \mathcal{L}(F_{t_i}, F_{t_j}, \nu J_{t_i}^{t_j})$. Then by the definition of \mathcal{L} -derivative there must exist three subspaces $V_{01} \subset V_0 \oplus V_1$, $V_{12} \subset V_1 \oplus V_2$ and $V_{02} \subset V_0 \oplus V_1 \oplus V_2$, such that the \mathcal{L} -derivatives constructed over spaces of variations containing V_{ij} will remain in O_{ij} . Since those are just piecewise constant functions, we can simply refine the partition and use the space of variations $U_0 \oplus V_1 \oplus U_2$, where variations U_0, U_2 have the same discontinuities as all variations from V_{ij} . We can then apply the chain rule to this new space of variations. Since elements of \mathcal{L} -derivatives are limits of vectors of finite-dimensional approximations we get that the chain rule holds in infinite dimensions as well. So it only remains to exploit this rule to construct enough independent vectors of $\mathcal{L}(F_{t_0}, F_{t_2}, J_{t_0}^{t_2})$.

Since the zero vector always lies in a \mathcal{L} -derivative, it is clear from the chain rule that $(\Gamma_0, 0) \oplus (0, \Gamma_2)$ is a subspace of $\mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$. We note that by construction Γ_{\ker} can be identified with a subspace in $\tilde{\Gamma}_1^\zeta/\tilde{\Gamma}_1$. Thus, by the chain rule $(0, \Phi_{12}(\Gamma_{\ker})) \in \mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$. For the same reason we also have $(\Phi_{01}^{-1}(\tilde{\Gamma}_{\ker}), 0) \in \mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$.

Finally we have that $(\Gamma_1 + \tilde{\Gamma}_1)^\zeta/(\ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1})$ can be identified with a subspace in Γ_1^ζ/Γ_1 and at the same time with a subspace in $\tilde{\Gamma}_1^\zeta/\tilde{\Gamma}_1$. Thus we see that a graph of the map Φ_{02} is going to be a Lagrangian subspace in $\Phi_{01}^{-1}(\Gamma_1^\zeta/\Gamma_1) \times \Phi_{12}(\tilde{\Gamma}_1^\zeta/\tilde{\Gamma}_1)$, and by the chain rule it will be a subspace of $\mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$. A simple dimensional count now shows that the resulting space is indeed Lagrangian. \square

2.5 Morse-type theorems

Now we are ready to state some Morse-type theorems. The simplest one allows us to compute the dimension of the kernel of the Hessian.

Lemma 2.17. *Let $F : \mathcal{U} \rightarrow M$ be a smooth map from a finite dimensional manifold \mathcal{U} to a finite dimensional manifold M , $J : \mathcal{U} \rightarrow R$ be a smooth functional, and let (u, λ) be a Lagrange point of (F, J) . Then to any vector in $\mathcal{L}(F, \nu J)[u, \lambda] \cap \Pi$ we can associate a unique up to an element of $\ker Q \cap \ker dF[u]$ variation $v \in \ker \text{Hess}(F, \nu J)[u, \lambda]$. Consequently*

$$\dim(\ker \text{Hess}(F, \nu J)[u, \lambda]) - \dim(\ker Q \cap \ker dF[u]) = \dim(\mathcal{L}(F, \nu J)[u, \lambda] \cap \Pi).$$

Proof. The uniqueness part is proved easily. So we can assume that $\ker Q \cap \ker dF[u] = \{0\}$ by factoring out this intersection if necessary.

If $(\xi, 0) \in \mathcal{L}(F, \nu J)[u, \lambda] \cap \Pi$, then by definition there must exist $v \in \ker dF[u]$, that solves the \mathcal{L} -derivative equation

$$\langle \xi, dF[u](w) \rangle + Q(v, w) = 0, \quad \forall w \in T_u \mathcal{U}. \quad (2.20)$$

Restricting w to $\ker dF[u]$ shows that

$$\dim(\ker \text{Hess}(F, \nu J)[u, \lambda]) \geq \dim(\mathcal{L}(F, \nu J)[u, \lambda] \cap \Pi).$$

To prove the other inequality and that the map described in the statement is a bijection, let us assume that $v \in \ker \text{Hess}(F, \nu J)[u, \lambda]$. Then

$$Q(v, w) = 0, \quad \forall w \in \ker dF[u].$$

But this implies that $Q(v, \cdot)$ is linear combination of rows of $dF[u]$. Therefore there must exist ξ , s.t. (2.20) holds. \square

The next step is to extract the information about the index of the Hessian from the Jacobi curve. We will rely heavily on the following lemma from linear algebra

Lemma 2.18. *Suppose that Q is a quadratic form defined on \mathbb{R}^N and let $V \subset \mathbb{R}^N$ be some subspace. If we define*

$$V^\perp = \{x \in \mathbb{R}^N : Q(x, y) = 0, \forall y \in V\},$$

then

$$\text{ind}^+ Q = \text{ind}^+ Q|_V + \text{ind}^+ Q|_{V^\perp} + \dim(V \cap V^\perp) - \dim(V \cap \ker Q) \quad (2.21)$$

Remark 2.19. Note that this lemma holds in a more general situation of a continuous quadratic form Q on a Hilbert space with finite positive inertia index and a closed subspace V .

We will also need the following result. We begin the following lemma that we have used several times in the text.

Lemma 2.20. *Let Q be a quadratic form defined on a finite-dimensional space V_2 , $A : V_2 \rightarrow \mathbb{R}^n$ be a linear map and $N \subset \mathbb{R}^n$ be a linear subspace. Take any subspace $V_1 \subset V_2$ and write $V_i^N = V_i \cap A^{-1}(N)$. Then the orthogonal complement of V_1^N in V_2^N with respect to Q consists of vectors $v \in V_2^N$, for which there exists ξ in the annihilator $N^\perp \subset (\mathbb{R}^n)^*$, s.t.*

$$\langle \xi, Aw \rangle + Q(v, w) = 0, \quad \forall w \in V_1. \quad (2.22)$$

Similarly $\ker Q \cap V_1^N$ consists of vectors $v \in V_1^N$, for which there exists ξ in the annihilator $N^\perp \subset (\mathbb{R}^n)^$, s.t. the quality above holds for all $w \in V_2$.*

Proof. We denote by W the subspace defined in the statement. By restricting (2.22) to $w \in V_1^N$, we can easily see that $W \subset (V_1^N)^\perp_Q$.

To prove the other inclusion we identify the annihilator of N^\perp with the orthogonal complement of N in \mathbb{R}^n . Let us then take a complement of $A^{-1}(N)$ in V_2 which will be isomorphic to $\text{im } A \cap N^\perp$ and a basis e_i in this subspace such that the images Ae_i form an orthonormal basis of $\text{im } A \cap N^\perp$. Then if $v \in (V_1^N)^\perp_Q$, by identifying $(\mathbb{R}^n)^*$ with \mathbb{R}^n using the Euclidean inner product, we find that (2.22) is satisfied if we take

$$\xi = - \sum_{i=1}^d \frac{Q(v, e_i)}{|Ae_i|^2} Ae_i,$$

where $d = \dim(\text{im } A \cap N^\perp)$. Thus $W \supset (V_1^N)^\perp_Q$.

The rest of the statement is proved using exactly the same argument. \square

The following analog of the Morse theorem was proven in [1].

Theorem 2.21. *Let $F : \mathcal{U} \rightarrow M$ be a map from a possibly infinite dimensional Banach manifold \mathcal{U} to a finite dimensional manifold M , $J : \mathcal{U} \rightarrow \mathbb{R}$ and let (u, λ) be a Lagrange point of (F, J) . We denote by $V_1 \subset V_2$ two finite dimensional subspaces of $T_u \mathcal{U}$ and $V_i^0 = V_i \cap \ker dF[u]$. If we choose V_1, V_2 be such that $\text{rank } dF[u]|_{V_1} = \text{rank } dF[u]|_{V_2}$ and $\ker \text{Hess}(F, \nu J)[u, \lambda]|_{V_1^0} = \ker \text{Hess}(F, \nu J)[u, \lambda]|_{V_2^0}$, then*

$$\begin{aligned} & \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{V_2^0} - \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{V_1^0} \geq \\ & \geq \text{ind}_\Pi(\mathcal{L}(F, \nu J)[u, \lambda](V_1), \mathcal{L}(F, \nu J)[u, \lambda](V_2)). \end{aligned}$$

It is clear that an equality can not hold in general. The right hand side of those inequalities is limited by the dimension of the manifold M , while the jump in the index can be arbitrary large. Nevertheless when we take piece-wise constant functions we can reconstruct exact formulas. The idea is that when we add some constant variations, using our algorithm (see Theorem 2.13) we can track exactly how the \mathcal{L} -derivative changes and use this to obtain an exact formula for the index. This will be the main building block in the general Morse index theorem, that we will prove immediately after.

Theorem 2.22. *Let $D = \{0 = t_0 < t_1 < \dots < t_N = t\}$ be a partition of the interval $[0, t]$ and let V_D be a direct product of the space of variations of the initial point and piece-wise constant variations with jumps at moments of time t_i . We denote by $V_i \subset V_D$ the subspace of V_D of functions that are zero for $t > t_i$ and $V_i^0 = V_i \cap \ker dE_{N_0, t}[\tilde{\omega}]$. Then the following formula is true*

$$\text{ind}^- \text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]|_{V_D^0} = \sum_{i=-1}^N \text{ind}_\Pi(\mathcal{L}_i, \mathcal{L}_{i+1}) + \dim \left(\bigcap_{i=-1}^N \mathcal{L}_i \right) - n, \quad (2.23)$$

where for simplicity we wrote $\mathcal{L}_i = \mathcal{L}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]|_{\Omega_{N_0} \cap V_i}$, $\mathcal{L}_{-1} = \mathcal{L}_{N+1} = \Pi$ and $\mathcal{L}_0 = T_{\tilde{q}(0)}^\perp N_0 \times T_{\tilde{q}(0)} N_0$.

Proof. As before we write

$$Q = \lambda d^2 E_{N_0, t}[\tilde{\omega}] - \nu d^2 J_t[\tilde{\omega}].$$

As we have already mentioned, $\text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]$ is equal to $-Q|_{\ker dE_{N_0, t}}$ as a quadratic form. So it is enough to prove the formula with $\text{ind}^+ Q|_{V_D^0}$ on the left-hand side.

We prove it by a recursive computation of $\text{ind}^+ Q|_{V_{i+1}^0}$ in terms of \mathcal{L}_m , $m \leq i$. The main tool will be the formula (2.21). We denote by Q_i the restriction of Q to V_i^0 and by $(V_i^0)^\perp$ the orthogonal complement of V_i^0 with respect to Q_{i+1} . First we establish the formula for $\text{ind}^+ Q_{i+1}|_{(V_i^0)^\perp}$ and then for $\dim(V_i^0 \cap (V_i^0)^\perp) - \dim(\ker Q_{i+1} \cap V_i^0)$ in terms of \mathcal{L}_i .

Step 1. We prove the following statement. Given two subspace

$$T_{\tilde{q}(0)} N_0 \subset U_1 \subset U_2 \subset T_{\tilde{q}(0)} N_0 \times L_k^2[0, t],$$

we claim that the subspace $(U_1^0)^\perp$ is equal to a subspace W_2 which consists of $(\zeta, v_2(\tau)) \in U_2^0$, s.t. there exists $\eta \in \mathcal{L}_0$ for which the following conditions are satisfied

$$\pi(\eta) = \zeta,$$

$$\int_0^t \sigma \left(\eta + \int_0^\tau X(\theta) v_2(\theta) d\theta, X(\tau) v_1(\tau) \right) + b(\tau)(v_2(\tau), v_1(\tau)) d\tau = 0, \quad \forall v_1(\tau) \in U_1 \quad (2.24)$$

This is a consequence of Lemma 2.20. Indeed, as we have discussed before vector fields $X(\tau)$ are lifts of $g'(\tau)$. Therefore from formula (??) it follows that we can characterize the kernel in terms of the Hamiltonian vector field $X(\tau)$ as

$$\ker dE_{N_0,t}[\tilde{\omega}] = \left\{ (v(\tau), \zeta) \in \Omega_{N_0} : \eta + \int_0^t X(\tau)v(\tau)d\tau \in \Pi, \forall \eta \in \mathcal{L}_0, \pi(\eta) = \zeta \right\}.$$

We apply Lemma 2.20 with A being equal to the operator

$$A : (\zeta, v) \mapsto \int_0^t X(\tau)v(\tau)d\tau$$

and $N = \Pi \times T_{\tilde{q}(0)}N_0 \subset T_{\lambda(0)}(TM)$. Then $\xi \in (T_{\lambda(0)}(TM))^*$ from that lemma must annihilate N . Since σ is a non-degenerate symplectic form we can use it to identify $(T_{\lambda(0)}(TM))^*$ with $(T_{\lambda(0)}(TM))$ and in this case $\xi \in N^\perp \simeq N^\angle \simeq T_{q_0}^\perp N_0$ and

$$\langle \xi, Av_1 \rangle = \sigma \left(\xi, \int_0^t X_\tau v_1(\tau)d\tau \right)$$

from which the statement follows.

Step 2. We have by definition and formula (2.12) that the subspace $(V_i^0)^\perp$ is equal to the space of variations $(\zeta, v(\tau), \alpha) \in V_{i+1}^0$, s.t.

$$\int_0^{t_i} \sigma \left(\zeta + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)(v(\tau), w(\tau))d\tau = 0, \quad \forall w(\tau) \in V_i^0.$$

Then from the step 1 it follows that $(V_i^0)^\perp$ is actually equal to the space W_i of vectors $(\zeta, v(\tau), \alpha) \in V_i \times \mathbb{R}^k$, s.t. there exists $\eta \in \mathcal{L}_0$ for which $\pi(\eta) = \zeta$ and

$$\int_0^{t_i} \sigma \left(\eta + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)(v(\tau), w(\tau))d\tau = 0, \quad \forall w(\tau) \in V_i, \quad (2.25)$$

$$\eta + \int_0^{t_i} X(\tau)v(\tau)d\tau + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \in \Pi. \quad (2.26)$$

We denote by

$$\lambda = \eta + \int_0^{t_i} X(\tau)v(\tau)d\tau.$$

Then the first condition just tells us that $\lambda \in \mathcal{L}_i$.

Let $(\zeta, v(\tau), \alpha) \in W_i$, then we obtain

$$\begin{aligned}
Q(\zeta, v(\tau), \alpha) &= \int_0^{t_i} \sigma \left(\zeta + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)v(\tau) \right) + b(\tau)(v(\tau), v(\tau))d\tau + \\
&+ \int_{t_i}^{t_{i+1}} \sigma \left(\zeta + \int_0^{t_i} X(\theta)v(\theta)d\theta + \int_{t_i}^\tau X(\theta)d\theta \cdot \alpha, X(\tau)\alpha \right) + b(\tau)(\alpha, \alpha)d\tau = \\
&= -\sigma \left(\mu, \int_0^{t_i} X(\tau)v(\tau)d\tau \right) + \\
&+ \int_{t_i}^{t_{i+1}} \sigma \left(\zeta + \int_0^{t_i} X(\theta)v(\theta)d\theta + \int_{t_i}^\tau X(\theta)d\theta \cdot \alpha, X(\tau)\alpha \right) + b(\tau)(\alpha, \alpha)d\tau = \\
&= \int_{t_i}^{t_{i+1}} \sigma \left(\lambda + \int_{t_i}^\tau X(\theta)d\theta \cdot \alpha, X(\tau)\alpha \right) + b(\tau)(\alpha, \alpha)d\tau
\end{aligned}$$

where in the last line we have used a consequence of (2.26)

$$-\sigma \left(\mu, \int_0^{t_i} X(\tau)v(\tau)d\tau \right) = -\sigma \left(\mu, \eta + \int_0^{t_i} X(\tau)v(\tau)d\tau \right) = \sigma \left(\mu, \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \right)$$

since $\mu, \eta \in \mathcal{L}_0$.

Thus we have shown that $Q|_{W_i}$ is equal to the form

$$P(\lambda, \alpha) = \int_{t_i}^{t_{i+1}} \sigma \left(\lambda + \int_{t_i}^\tau X(\theta)d\theta \cdot \alpha, X(\tau)\alpha \right) + b(\tau)(\alpha, \alpha)d\tau$$

defined on a finite-dimensional space

$$S = \left\{ (\lambda, \alpha) \in \mathcal{L}_i \times \mathbb{R}^k : \lambda + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \in \Pi \right\}.$$

Now we consider the quadratic form q from the definition of the positive Maslov index defined on $(\mathcal{L}_i + \mathcal{L}_{i+1}) \cap \Pi$. Let

$$\lambda_1, \lambda_2 \in \mathcal{L}_i, \quad \lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \in \mathcal{L}_{i+1}, \quad \lambda_1 + \lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \in \Pi. \quad (2.27)$$

Then from the definition of \mathcal{L}_{i+1} we obtain

$$\begin{aligned}
q(\lambda_1, (\lambda_2, \alpha)) &= \sigma \left(\lambda_1, \lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \right) = \\
&= \sigma \left(\lambda_1 + \lambda_2, \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \right) - \sigma \left(\lambda_2, \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \right) = \\
&= \int_{t_i}^{t_{i+1}} \sigma \left(\lambda_1 + \lambda_2 + \int_{t_i}^\tau X(\theta)d\theta \cdot \alpha, X(\tau)\alpha \right) + b(\tau)(\alpha, \alpha)d\tau.
\end{aligned}$$

If we denote

$$\tilde{S} = \{(\lambda_1 + \lambda_2, \alpha) \in \mathcal{L}_i \times \mathbb{R}^k : \lambda_j, \alpha \text{ satisfy (2.27)}\},$$

then we have that $\tilde{S} \subset S$ and $P|_{\tilde{S}} = q|_{\tilde{S}}$. So

$$\text{ind}^+ Q_{i+1}|_{(V_i^0)^\perp} \geq \text{ind}^+ q.$$

Step 3. The reason why we didn't apply immediately the Theorem 2.21 is that we needed these specific expressions for $Q|_{W_i}$ and q to prove the other inequality. We do it by demonstrating that all λ, α , that actually give a contribution to the index of P , lie in \tilde{S} . And indeed, this is just a consequence of our algorithm.

Take $(\lambda, \alpha) \in S$. In Theorem 2.13 we have defined subspaces $L \subset \mathcal{L}_i$ and $E \subset \mathbb{R}^k$, which from their definition can be seen to lie in the kernel of P . Thus it is enough consider P on any complementary subspace L^\perp and E^\perp . But from Theorem 2.13 we know, that for any $\alpha \in E^\perp$ there exists a unique $\lambda_2 \in L^\perp \subset \mathcal{L}_i$, s.t.

$$\lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau) d\tau \cdot \alpha \in \mathcal{L}_{i+1}.$$

Thus we can take $\lambda_1 = \lambda - \lambda_2$ and then it follows that $(\lambda, \alpha) \in \tilde{S}$, which proves that $S \subset \tilde{S}$ and

$$\text{ind}^+ Q_{i+1}|_{(V_i^0)^\perp} \leq \text{ind}^+ q.$$

Then from the definition of the positive Maslov index, we have

$$\begin{aligned} \text{ind}^+ Q_{i+1}|_{(V_i^0)^\perp} &= \text{ind}_\Pi(\mathcal{L}_i, \mathcal{L}_{i+1}) - \frac{1}{2} (\dim(\mathcal{L}_i \cap \Pi) + \dim(\mathcal{L}_{i+1} \cap \Pi)) \\ &\quad + \dim(\mathcal{L}_i \cap \mathcal{L}_{i+1} \cap \Pi) \end{aligned} \quad (2.28)$$

Using exactly the same arguments one can prove the formula for Q_1 , that gives the base of the induction

$$\text{ind}^+ Q_1 = \text{ind}_\Pi(\mathcal{L}_0, \mathcal{L}_1) - \frac{1}{2} (\dim(\mathcal{L}_0 \cap \Pi) + \dim(\mathcal{L}_1 \cap \Pi)) + \dim(\Pi \cap \mathcal{L}_0 \cap \mathcal{L}_1). \quad (2.29)$$

Step 4. Now we obtain an expression for $\dim(V_i^0 \cap (V_i^0)^\perp) - \dim(\ker Q_{i+1} \cap V_i^0)$ in terms of \mathcal{L}_m , $m \leq i+1$. Here again our algorithm plays the central role. It gives us a sequence of maps P_i

$$T_{\tilde{q}(0)}^\perp N_0 \times T_{\tilde{q}(0)} N_0 = \mathcal{L}_0 \xrightarrow{P_0} \mathcal{L}_1 \xrightarrow{P_1} \dots \xrightarrow{P_{N-1}} \mathcal{L}_N = \mathcal{L}_t(V_D).$$

We want to reconstruct all $v \in V_i$, s.t.

$$\eta + \int_0^t X(\tau) v(\tau) d\tau \in \mathcal{L}_i \cap \Pi$$

by inverting P_i and going backwards from \mathcal{L}_N to \mathcal{L}_{N-1} , then to \mathcal{L}_{N-2} and so on. Maps P_i are indeed invertible, since they are surjective linear maps between spaces of the same

dimension. Thus by fixing $\lambda \in \mathcal{L}_i \cap \Pi$, we get a sequence $P_{i-1}^{-1}(\lambda)$, $P_{i-2}^{-1} \circ P_{i-1}^{-1}(\lambda)$ and so on, that can be seen as a sort of a solution of the Jacobi equation passing through λ .

We need to understand how many different variations $v \in V_i$ correspond to the same sequence. First note that all

$$\lambda \in \bigcap_{i=0}^m \mathcal{L}_i \cap \Pi$$

correspond to the same variation $(\zeta, v) \equiv (0, 0)$. Thus we must extract $\dim(\bigcap \mathcal{L}_i)$ from the overall expression. All the other $\lambda \in \mathcal{L}_i \cap \Pi$ correspond to some non-zero $v \in V_i$.

Let $L_m \in \mathcal{L}_m$, $E_m \in \mathbb{R}^k$ be the subspace L , E from the Theorem 2.13 for $V = V_m$ and L_m^\perp , E_m^\perp be the orthogonal complements in \mathcal{L}_m and \mathbb{R}^k correspondingly. Note that to each $\lambda \in L_m^\perp$ corresponds a unique variation $\alpha \chi_{[t_m, t_{m+1}]}$, $\alpha \in E_m^\perp$. But if $\beta \in E_m$ the variation $(\alpha + \beta) \chi_{[t_m, t_{m+1}]}$ corresponds to the same vector in the \mathcal{L} -derivative. Therefore we have $\sum \dim E_m$ of variations that correspond to the same $\lambda \in \mathcal{L}_i \cap \Pi$, and so we get a formula

$$\dim(V_i^0 \cap (V_i^0)^\perp) = \dim(\mathcal{L}_i \cap \Pi) + \sum_{m=0}^{i-1} \dim E_m - \dim \left(\bigcap_{m=-1}^i \mathcal{L}_m \right). \quad (2.30)$$

Now we compute $\dim(\ker Q_{i+1} \cap V_i^0)$. Using Lemma 2.20, the same proof as in the step 1 shows, that $v \in \ker Q_{i+1} \cap V_i^0$ if and only if there exists $\eta \in \mathcal{L}_0$ for which $\pi(\eta) = \zeta$, s.t.

$$\begin{aligned} \int_0^{t_i} \sigma \left(\eta + \int_0^\tau X(\theta) v(\theta) d\theta, X(\tau) w(\tau) \right) + b(\tau)(v(\tau), w(\tau)) d\tau &= 0, \quad \forall w(\tau) \in V_i, \\ \sigma \left(\eta + \int_0^{t_i} X(\tau) v(\tau) d\tau, \int_{t_i}^{t_{i+1}} X(\tau) d\tau \cdot \alpha \right) &= 0, \quad \forall \alpha \in \mathbb{R}^k, \\ \eta + \int_0^{t_i} X(\tau) v(\tau) d\tau &\in \Pi. \end{aligned}$$

If denote

$$\lambda = \eta + \int_0^{t_i} X(\tau) v(\tau) d\tau,$$

then equivalently we can write $\lambda \in \mathcal{L}_i \cap \mathcal{L}_{i+1} \cap \Pi$.

Using same argument as for $\dim(V_i^0 \cap (V_i^0)^\perp)$ we get

$$\dim(\ker Q_{i+1} \cap V_i^0) = \dim(\mathcal{L}_i \cap \mathcal{L}_{i+1} \cap \Pi) + \sum_{m=0}^{i-1} \dim E_m - \dim \left(\bigcap_{m=-1}^{i+1} \mathcal{L}_m \right). \quad (2.31)$$

So we sum over all i the formulas (2.28)-(2.31) to obtain

$$\text{ind}^+ Q|_{V_D} = \sum_{i=0}^{N-1} \text{ind}_\Pi(\mathcal{L}_i \cap \mathcal{L}_{i+1}) - \frac{1}{2} \dim(\mathcal{L}_0 \cap \Pi) - \frac{1}{2} \dim(\Pi \cap \mathcal{L}_N) + \dim \left(\bigcap_{i=-1}^N \mathcal{L}_i \right).$$

The final formula follows from $\mathcal{L}_{-1} = \mathcal{L}_{N+1} = \Pi$ and property 4 in Lemma 1.14. \square

This approximation lemma can now be used to prove a very general Morse theorem, that establishes relation between some symplectic invariants of the Jacobi curve and index of the Hessian. After fixing some partition D , we introduce the following curves using the notations of the previous theorem

$$\Lambda^D(t) = \begin{cases} \Pi & \text{if } -1 \leq t < 0 = t_0, \\ \mathcal{L}_i & \text{if } t_i < t \leq t_{i+1}, \\ \Pi & \text{if } t_N < t \leq t_N + 1. \end{cases}$$

We extend the Jacobi curve \mathcal{L}_t by assuming that $\mathcal{L}_t = \Pi$ for $t \in [-1, 0) \cup (T, T + 1]$. Then by definition $\Lambda^D(t) \rightarrow \mathcal{L}_t$ pointwise as a generalized limit. To shorten the notations we also write $\Sigma = T_{\lambda(0)}(T^*M)$.

Theorem 2.23. *Suppose that $\text{ind}^- \text{Hess}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)] < \infty$ at a Lagrange point $(\tilde{\omega}, \lambda(t))$. Let $\tilde{\Pi}$ be a point in the universal covering $\tilde{L}(\Sigma)$, that projects to $\Pi \in L(\Sigma)$. Let $\Lambda^D : [-1, T + 1] \rightarrow L(\Sigma)$ be the extended Jacobi curve built over the space of piece-wise constant variations with discontinuities in D as defined above, and $\tilde{\Lambda}^D(s)$ be the corresponding curves in the universal covering with the same initial point $\tilde{\Lambda}^D(-1) = \tilde{\Lambda}_{-1}$, s.t. $\Lambda_{-1} = \Pi$.*

Then there exists a point-wise generalized limit $\tilde{\Lambda}^D(s) \rightarrow \tilde{\mathcal{L}}_s$, such that $\tilde{\mathcal{L}}_s$ is the lift of the Jacobi curve \mathcal{L}_s and

$$\text{ind}^- \text{Hess}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)] = \frac{1}{2} \left(\text{Li}(\tilde{\mathcal{L}}_{T+1}, \tilde{\Pi}) - \text{Li}(\tilde{\mathcal{L}}_{-1}, \tilde{\Pi}) \right) + \dim \left(\bigcap_{s=0}^T \mathcal{L}_s \cap \Pi \right) - n.$$

Proof. Step 1. We are going to show that index of the Hessian restricted to the dense sub-space of piece-wise constant functions coincides with the index of the Hessian on the whole kernel. This will allow us to apply directly Morse Theorem 2.22.

Map $dE_{N_0,T}$ is a continuous finite rank operator between an infinite dimensional Hilbert manifold that is locally isomorphic to $T_{\tilde{q}(0)}N_0 \times L_k^2[0, T]$ and $T_{\tilde{q}(0)}M$. We have that the intersection of $\ker dE_{N_0,T}[\tilde{\omega}]$ with the space of piece-wise constant functions is dense in $\ker dE_{N_0,T}[\tilde{\omega}]$. Indeed, by continuity of $dE_{N_0,T}[\tilde{\omega}]$ we have that its restriction to the subspace of piece-wise constant functions must have the same rank. Therefore the subspace of linear piecewise constant functions splits into two disjoint subspaces: the intersection with the kernel of $dE_{N_0,T}[\tilde{\omega}]$ and a finite-dimensional complement that is isomorphic to its image. If P_{ker} and P_{fin} are two projections to these subspace, s.t. $P_{ker}P_{fin} = P_{ker}P_{fin} = 0$, then given a sequence of piecewise constant function f_n converging to $f \in \ker dE_{N_0,T}[\tilde{\omega}]$, the projections of $P_{fin}f_n$ must converge to zero and $P_{ker}f_n$ converge to f .

At the same time the quadratic form Q is continuous in $T_{\tilde{q}(0)}N_0 \times L_k^2[0, t]$, therefore by restricting to a dense subspace we will get the same index. This implies that we can from the beginning compute the index of Q restricted to the intersection of $\ker dE_{N_0,T}[\tilde{\omega}]$ with piece-wise constant functions.

Step 2. We apply Theorem 2.22 to a special sequence of spaces V_D . We take a finite number of piece-wise constant functions v_i , s.t. they span a negative subspace of maximal

dimension of the Hessian. Let D_0 be a splitting $0 = t_0 < t_1 < \dots < t_{N_0} = T$, where t_i are the discontinuity points of v_i . Then we can consider any sequence $\{0 = t_0^m, \dots, t_{N_m}^m = T\} = D_m \supset D_0$, s.t. $\max |t_{i+1}^m - t_i^m| \rightarrow 0$ and the corresponding subspace of piecewise constant variations $V^m = V^{D_m}$ as in Morse Theorem 2.22. We also use notations analogous to Theorem 2.22 to define a subspace $V_i^m \subset V^m$ of functions that are zero for $t > t_i$ and $(V_i^m)^0 = V_i^m \cap \ker dE_{N_0, T}[\tilde{\omega}]$.

For what follows we will need the following sequence of curves:

$$\Lambda_t^m(s) = \begin{cases} \Pi & \text{if } -1 \leq s < 0, \\ \Lambda^m(s) & \text{if } 0 \leq s \leq t, \\ \Pi & \text{if } t < s \leq t+1; \end{cases}$$

which are just closed extensions of the restrictions $\Lambda^m(s)|_{[0, t]}$, where we have shortened the notation for $\Lambda^{D_m}(s)$ just to $\Lambda^m(s)$.

By the additivity Lemma 2.10 and Morse Theorem 2.22 for the piece-wise constant approximations, we obtain

$$\text{ind}^- \text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)] + n = \text{ind}_\Pi \Lambda_t^m(s) + \dim \left(\bigcap_{s=-1}^t \Lambda_t^m(s) \right). \quad (2.32)$$

It only remains to study the limit of the right hand-side when $m \rightarrow \infty$.

Step 3. We start by considering the second term containing the dimension of the intersections. Since $\mathcal{L}_{-1} = \Pi$ and $\Lambda_t^m(s) = \Lambda^m(s)$ for $s \in [0, t]$, we have that

$$\bigcap_{s=-1}^t \Lambda_t^m(s) = \bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)).$$

From the Theorem 2.13 and the definition of $\Lambda_t^m(s)$ it follows that

$$\bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)) = \left\{ \mu \in T_{\tilde{q}(0)}^\perp N_0 : \sigma \left(\mu, \int_0^t X(\tau) w(\tau) d\tau \right) = 0 : \forall w(\tau) \in V^m \cap L_k^2[0, t] \right\}.$$

Therefore since $D_i \subset D_{i+1}$, we have that

$$\bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)) \subset \bigcap_{s=0}^t (\Pi \cap \Lambda^l(s)), \quad \forall l \leq m.$$

Thus we get a sequence of nested subspace, and since $\Lambda^m(s)$ converge pointwise, this sequence must stabilize for m large enough.

We claim that

$$\bigcap_{s=0}^t (\Pi \cap \mathcal{L}_s) = \bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)) \quad (2.33)$$

for m large enough. Again, from the point-wise convergence it is obvious that

$$\bigcap_{s=0}^t (\Pi \cap \mathcal{L}_s) \subset \bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)).$$

The other inclusion holds true as well. Given $\mu \in \bigcap (\Pi \cap \mathcal{L}_s)$ we can find a sequence

$$\mu_m \in \bigcap_{s=0}^t (\Pi \cap \Lambda^m(s))$$

s.t. $\mu_m \rightarrow \mu$. But then for any $w \in V^m \cap L_k^2[0, t]$, we have:

$$\sigma \left(\mu, \int_0^t X(\tau)w(\tau)d\tau \right) = \lim_{m \rightarrow \infty} \left(\mu_m, \int_0^t X(\tau)w(\tau)d\tau \right) = 0.$$

Thus $\eta \in \Lambda^m(s)$ by definition for m large enough and (2.33) holds.

This way we have shown that

$$\bigcap_{s=0}^t (\mathcal{L}_s \cap \Pi) = \bigcap_{s=0}^t (\Lambda_t^m(s) \cap \Pi), \quad \forall t \in [0, T].$$

Step 4. To arrive at the final result we need to express $\text{ind}_{\Pi} \Lambda^m(t)$ in terms of the Leray index.

Fix some m and sequence of Lagrangian planes Δ_i^m , s.t. there exist monotone curves which connect $\Lambda_t^m(t_i)$ with $\Lambda_t^m(t_{i+1})$ and do not intersect the corresponding Δ_i^m . Let Δ be any Lagrangian plane. Then from the Maslov index formula in Proposition 1.20 and the definition of the Leray index we get

$$\begin{aligned} \text{ind}_{\Delta} \Lambda_t^m(s) &= \frac{1}{2} \sum_{i=-1}^{N_m} (\text{Ki}(\Delta_i^m, \Lambda_t^m(t_i), \Delta) - \text{Ki}(\Delta_i^m, \Lambda_t^m(t_{i+1}), \Delta)) = \\ &= \frac{1}{2} \sum_{i=-1}^{N_m} \left(\text{Li}(\tilde{\Lambda}_t^m(t_i), \tilde{\Delta}) + \text{Li}(\tilde{\Delta}_i^m, \tilde{\Lambda}_t^m(t_i)) - \text{Li}(\tilde{\Lambda}_t^m(t_{i+1}), \tilde{\Delta}) - \text{Li}(\tilde{\Delta}_i^m, \tilde{\Lambda}_t^m(t_{i+1})) \right). \end{aligned}$$

By definition Leray index $\text{Li}(\tilde{\Lambda}_1, \tilde{\Lambda}_2)$ is locally constant on the set $\{(\tilde{\Lambda}_1, \tilde{\Lambda}_2) : \Lambda_1 \cap \Lambda_2 = \{0\}\}$. Since $\Lambda_t^m(t_i)$ and $\Lambda_t^m(t_{i+1})$ can be connected by a curve that does not pass through Δ_i^m , we obtain by Lemma 1.24

$$\text{Li}(\Delta_i^m, \Lambda_t^m(t_i)) - \text{Li}(\Delta_i^m, \Lambda_t^m(t_{i+1})) = 0.$$

This way we get

$$\text{ind}_{\Delta} \Lambda_t^m(s) = \frac{1}{2} \left(\text{Li}(\tilde{\Lambda}_{-1}, \tilde{\Delta}) - \text{Li}(\tilde{\Lambda}_t^m(t+1), \tilde{\Delta}) \right).$$

Step 5. Assume for now that $\tilde{\Lambda}^m(s)$ converges pointwise to a curve $\tilde{\mathcal{L}}_s$. The previous formula then implies the final result. Indeed, we put $\Delta = \Pi$ and choose any Lagrangian plane Δ' , s.t. $\Pi \cap \Delta' = \{0\}$. Then we obtain from the properties of Leray index

$$\begin{aligned} & \text{ind}^-(\text{Hess}(E_{N_0, T}, \nu J_T)[\tilde{\omega}, \lambda(T)]) + n - \dim \left(\bigcap_{s=0}^T \mathcal{L}_s \cap \Pi \right) \\ &= \frac{1}{2} \left(\text{Li}(\tilde{\Lambda}^m(-1), \tilde{\Pi}) - \text{Li}(\tilde{\Lambda}^m(T+1), \tilde{\Pi}) \right) = \frac{1}{2} \left(\text{Li}(\tilde{\Lambda}^m(-1), \tilde{\Delta}') - \right. \\ & \left. - \text{Li}(\tilde{\Lambda}^m(T+1), \tilde{\Delta}') + \text{Ki}(\Lambda^m(-1), \Pi, \Delta') - \text{Ki}(\Lambda^m(T+1), \Pi, \Delta') \right). \end{aligned}$$

By construction $\Lambda^m(T+1) = \Lambda^m(-1) = \Pi$. Therefore the Kashiwara indexes in the expression are zero and we can take limit as $m \rightarrow \infty$, since the Leray index is locally constant. So we see that the result indeed holds if the pointwise convergence is true.

Step 6. Fix a moment of time $t \in [0, T]$. To prove that the sequence $\Lambda^m(t)$ converges point-wise, we are going fix a special Lagrangian plane Δ , s.t. it does not intersect any $\Lambda^m(t)$ or \mathcal{L}_t and moreover

$$\text{Li}(\tilde{\Lambda}_t^m(t+1), \tilde{\Delta}) = \text{Li}(\tilde{\Lambda}^m(t), \tilde{\Delta}). \quad (2.34)$$

Then by Lemma 1.26, formula (2.32) and steps 3 and 4 we have

$$\text{Li}(\tilde{\Lambda}^m(t), \tilde{\Delta}) = \text{Li}(\tilde{\Lambda}_{-1}, \tilde{\Delta}) - 2 \text{ind}^-(\text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]) - 2n + 2 \dim \left(\bigcap_{s=0}^t \mathcal{L}_s \cap \Pi \right) \quad (2.35)$$

for m sufficiently large. Therefore the limit on the left hand side exists. But we recall that by the Theorem 1.25 an open subset in $\{\tilde{\Lambda} \in \widetilde{L}(\Sigma) : \Lambda \cap \Delta = \{0\}\}$ can be identified with $\Delta^{\text{h}} \times \mathbb{Z}$. And therefore we can take

$$\tilde{\mathcal{L}}_t = \left(\mathcal{L}_t, \lim_{m \rightarrow \infty} \frac{1}{2} \text{Li}(\mathcal{L}_t^m, \Delta) \right).$$

To prove that the Lagrangian plane Δ with the desired properties exists, we follow until some point the proof of the existence of the \mathcal{L} -derivative. The idea is that we expect that the \mathcal{L} -derivative encodes all the information about the index and the nullity of the Hessian. So we construct a \mathcal{L} -derivative over a finite-dimensional subspace, which contains already the kernel and a negative subspace of maximal dimension. Then adding up variations should not change the \mathcal{L} -derivative to much, at least we can hope that it is not going to produce any contribution to the Maslov index in the process. This procedure was done in [1] and we will just discuss the first steps.

We are going to use the formulas from the definition of a \mathcal{L} -derivative as a linearisation of the Lagrange multiplier rule. As before we write

$$Av = dE_{N_0, t}[\tilde{\omega}]v.$$

First of all, we note that directly from definition variations from $\ker Q \cap \ker A$ do not give any contribution to the \mathcal{L} -derivative. This is a simple consequence of the definition. Next we refine our initial partition D_0 . We assume that D_0 is such that the space V_{D_0} satisfies the following three conditions for any $D \supset D_0$:

1. $\text{rank } A|_{V_D} = \text{rank } A|_{V_{D_0}}$;
2. $\text{ind}^+ Q|_{V_D} = \text{ind}^+ Q|_{V_{D_0}}$;
3. $v \in \ker Q|_{(V_D)^0}$ and $v \notin (V_{D_0})^0$ hold at the same time if and only if $v \in \ker Q \cap \ker A$.

To see that such a partition exists we note that $\ker Q \cap \ker A \subset \ker Q|_{\ker A}$ and that it has a finite codimension. Indeed if $v \in \ker Q|_{\ker A}$ then it satisfies $Q(v, w) = 0$, for all $w \in \ker A$. On the other hand all $v \in \ker Q \cap \ker A$ satisfy the same equation, but for all variations w and not just those in the kernel. Since $\dim \text{rank } A < \infty$ we conclude that indeed $\ker Q \cap \ker A$ has finite codimension in $\ker Q|_{\ker A}$. Therefore to construct V_{D_0} it is enough to take a direct sum of any complement of $\ker Q \cap \ker A$ in $\ker Q|_{\ker A}$, any negative subspace in $\ker A$ of maximal dimension and any subspace isomorphic to $\text{im } A$ via A .

Given V_{D_0} by Lemma 2.17 for any $D \supset D_0$ one has $\mathcal{L}_t(V_D) \cap \Pi = \mathcal{L}_t(V_{D_0}) \cap \Pi$. This allows us to search for Δ in $(\Pi \cap \mathcal{L}_t(V_{D_0}))^\perp / (\Pi \cap \mathcal{L}_t(V_{D_0}))$, i.e. we can assume that $\mathcal{L}_t(V_D) \cap \Pi = \{0\}$. Geometrically this means that we look for Δ that contains $\mathcal{L}_t(V_{D_0}) \cap \Pi$. Indeed as a result we will get monotone curves that have constant intersection with Δ , so it is going to be enough to replace it with Δ^Γ , where Γ is any isotropic subspace s.t. $\sigma|_{\Gamma \times (\Pi \cap \mathcal{L}_t(V_{D_0}))}$ is symplectic.

We return now to our sequence of partitions $D_m \supset D_0$. All V^m satisfy the assumptions of Theorem 2.21. Therefore

$$\text{ind}_\Pi(\Lambda^i(t), \Lambda^j(t)) = 0, \quad \forall i, j \in \mathbb{N}$$

We note that from definition it follows, that when $\text{ind}_{\Lambda_1}(\Lambda_2, \Lambda_3)$ is equal to zero or n , it implies that $\Lambda_1 \cap \Lambda_2 = \Lambda_1 \cap \Lambda_3 = \{0\}$. So $\Lambda^m(t) \cap \Pi = \{0\}$ for all $m \in \mathbb{N}$.

We choose any $\Delta \in L(\Sigma)$, s.t. $\text{ind}_\Pi(\Lambda^0(t), \Delta) = n$. Then by the triangle inequality we get

$$\text{ind}_\Pi(\Lambda^0(t), \Delta) \leq \text{ind}_\Pi(\Lambda^0(t), \Lambda^m(t)) + \text{ind}_\Pi(\Lambda^m(t), \Delta), \quad \forall m \in \mathbb{N}.$$

From the bounds on the indices it follows that $\text{ind}_\Pi(\Lambda^m(t), \Delta) = n$ and therefore $\Lambda^m(t) \in \Delta^\#$. From the relations between the positive Maslov index and the Kashiwara index, we get

$$\text{Ki}(\Lambda^m(t), \Pi, \Delta) = 2 \text{ind}_\Pi(\Lambda^m(t), \Delta) + \dim(\Lambda^m(t) \cap \Delta) - n = n + \dim(\Lambda^m(t) \cap \Delta).$$

Thus

$$\dim(\Lambda^m(t) \cap \Delta) = 0.$$

This establishes the existence of Δ . Using the same formula and antisymmetry of the Kashiwara index, we obtain.

$$\text{ind}_\Delta(\Lambda^m(t), \Pi) = 0.$$

Therefore all $\Lambda^m(t)$ can be connected to Π by a monotone curve that does not intersect Δ . By properties of the Leray index it now follows that (2.34) holds. \square

We have seen that if we add an arbitrary subspace of variations like in Theorem 2.21, we can only expect inequality. But the proof of Theorem 2.22 relies heavily on the fact that the inequality becomes an equality if we use piece-wise constant functions. So it is natural to ask what other type of variations we can add to have an equality. The next theorem gives another sufficient condition for this.

Theorem 2.24. *Assume that index of $\text{Hess}(F, \nu J)[u, \lambda]|_U$ is finite and that we can find a splitting $U_1 \oplus U_2$ of a possibly infinite-dimensional U , s.t.*

1. U_1 and U_2 are orthogonal with respect to Q ;
2. $\text{Hess}(F, \nu J)[u, \lambda]|_{U_2^0} > 0$;
3. $\dim \mathcal{L}(F, \nu J)[u, \lambda](U_1) \cap \Pi = 0$.

Then

$$\begin{aligned} & \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{U^0} - \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{U_1^0} = \\ & = \text{ind}_\Pi(\mathcal{L}(F, \nu J)[u, \lambda](U_1), \mathcal{L}(F, \nu J)[u, \lambda](U)). \end{aligned}$$

Proof. We are going to apply twice the Lemma 2.18: first time to the subspace U_1^0 in U^0 and the second time to U_2^0 in $(U_1^0)^\perp$. Assume for now, that $\dim U < \infty$.

As in Theorem 2.22 we can show that $(U_1^0)^\perp$ under the assumptions is actually equal to the subspace

$$\{v_1 + v_2 \in U : dF[u](v_1 + v_2) = 0, \exists \xi_1 \in T_{F(u)}^* M, \langle \xi_1, dF[u](U_1) \rangle + Q(v_1, U_1) = 0\}$$

From the orthogonality assumption it follows that $U_2^0 \in (U_1^0)^\perp$. The orthogonal complement of U_2^0 in $(U_1^0)^\perp$ is equal to $(U_1^0 + U_2^0)^\perp$. Like before, by using Lemma 2.20, we can show that this subspace consists of $v_1 + v_2 \in U^0$, s.t. there exist $\xi_1, \xi_2 \in T_{F(u)}^* M$ for which

$$\begin{aligned} & \langle \xi_1, dF[u](U_1) \rangle + Q(v_1, U_1) = 0, \\ & -\langle \xi_2, dF[u](U_2) \rangle + Q(v_2, U_2) = 0. \end{aligned}$$

We can now compute the quadratic form Q restricted to $(U_1^0 + U_2^0)^\perp$. Again we use the orthogonality assumption and the equivalent definition of $(U_1^0 + U_2^0)^\perp$ above. Assume that $v = v_1 + v_2 \in (U_1^0 + U_2^0)^\perp$ and $\xi = \xi_1 + \xi_2$. Then

$$\begin{aligned} & \langle Q(v_1 + v_2), v_1 + v_2 \rangle = \langle Qv_1, v_1 \rangle + \langle Qv_2, v_2 \rangle = \\ & = -\langle \xi_1, dF[u](v_1) \rangle + \langle \xi_2, dF[u](v_2) \rangle = -\langle \xi, dF[u](v_1) \rangle. \end{aligned}$$

Now we would like to write down the expression for the matrix S from the definition of the positive Maslov index. First we write down the definition of the two \mathcal{L} -derivatives:

$$\mathcal{L}(F, \nu J)(U_1) = \{(\eta_1, dF[u](v_1^1)) : \langle \eta_1, dF[u](U_1) \rangle + Q(v_1^1, U_1) = 0\};$$

$$\begin{aligned} \mathcal{L}(F, \nu J)(U) = \{(\eta_2, dF[u](v_1^2 + v_2)) : &\langle \eta_2, dF[u](U_1) \rangle + Q(v_1^2, U_1) = 0, \\ &\langle \eta_2, dF[u](U_2) \rangle + Q(v_2, U_2) = 0\}. \end{aligned}$$

The quadratic form from the Maslov index is defined on

$$(\mathcal{L}(F, \nu J)(U_1) + \mathcal{L}(F, \nu J)(U)) \cap \Pi.$$

We write $v_1^1 + v_1^2 = v_1$, $\xi_1 = \eta_1 + \eta_2$, $\xi_2 = -\eta_2$ and suppose that $dF[u](v_1) + dF[u](v_2) = 0$. Then for the quadratic form q we have

$$\begin{aligned} q &= \sigma((\eta_1, dF[u](v_1^1)), (\eta_2, dF[u](v_1^2 + v_2))) = \sigma((\eta_1, dF[u](v_1^1)), (0, dF[u](v_2))) = \\ &= \langle \eta_1, dF[u](v_2) \rangle = -\langle \eta_1, dF[u](v_1) \rangle = -\langle \xi, dF[u](v_1) \rangle. \end{aligned}$$

In the second equality we have used that $(\eta_1, dF[u](v_1^1))$ and $(\eta_2, dF[u](v_1^2))$ belong to $\mathcal{L}(F, \nu J)(U_1)$ by definition.

We see that this gives the same expression as for $Q|_{(U_2^0)^\perp}$. But moreover both quadratic forms are actually defined on the same space. Indeed, we have

$$\begin{aligned} (\mathcal{L}(F, \nu J)(U_1) + \mathcal{L}(F, \nu J)(U)) \cap \Pi &= \{(\xi_1 + \xi_2, 0) : \exists v_i \in V_i, dF[u](v_1 + v_2) = 0, \\ &(\xi_1, dF[u]v_1) \in \mathcal{L}(F, \nu J)(U_1), (-\xi_2, dF[u]v_2) \in \mathcal{L}(F, \nu J)(U_2)\} = \\ &= (\mathcal{L}(F, \nu J)(U_1) + \mathcal{L}(F, \nu J)(U_2)) \cap \Pi \end{aligned}$$

But if we add to $(\xi_1 + \xi_2) \in (\mathcal{L}(F, \nu J)(U_1) + \mathcal{L}(F, \nu J)(U_2)) \cap \Pi$ the corresponding v_i and to $v_i \in U_i \cap (U_1^0 + U_2^0)^\perp$ the corresponding $\xi_1 + \xi_2$, we obtain the same space.

Now we compute the other terms from the formula in Lemma 2.18. We have

$$U_1^0 \cap (U_1^0)^\perp = \{v_1 \in U_1^0 : Q(v_1, U_1^0) = 0\}.$$

Similarly as before, we can show that

$$U_1^0 \cap (U_1^0)^\perp = \{v_1 \in U_1^0 : \exists \xi \in T_{F(u)}^* M, \langle \xi, dF[u](U_1) \rangle + Q(v_1, U_1) = 0\}.$$

We do now the same for $\ker Q|_{U^0 \cap U_1^0}$:

$$\begin{aligned} \ker Q|_{U^0 \cap U_1^0} &= \{v_1 \in U_1^0 : Q(v_1, U^0) = 0\} = \\ &= \{v_1 \in U_1^0 : \exists \xi \in T_{F(u)}^* M, \langle \xi, dF[u](U) \rangle + Q(v_1, U) = 0\} \end{aligned}$$

To understand the dimensions, we look carefully at the equation

$$dF[u](U_1) + Q(v_1, U_1) = 0$$

If there are two solutions (ξ, v_1) and (ξ, v'_1) of this equation, then by linearity $(0, v_1 - v'_1)$ is a solution as well and thus all solutions are uniquely defined by different ξ modulo $\ker Q|_{U_1} \cap U_1^0$. By Lemma 2.17 we know, that these ξ lie in $\mathcal{L}(F, \nu J)(U_1) \cap \Pi$. Therefore

$$\dim(U_1^0 \cap (U_1^0)^\perp) = \dim(\mathcal{L}(F, \nu J)(U_1) \cap \Pi) + \dim(\ker Q|_{U_1} \cap U_1^0)$$

Now we do the same for

$$\langle \xi, dF[u](U) \rangle + Q(v_1, U) = \langle \xi, dF[u](U_1) \rangle + Q(v_1, U_1) + \langle \xi, dF[u](U_2) \rangle = 0$$

Again ξ are defined uniquely modulo $\ker Q|_{U_1} \cap U_1^0$, but now they lie in $\mathcal{L}(F, \nu J)(U) \cap \Pi$. Therefore

$$\dim(\ker Q|_{U^0} \cap U_1^0) = \dim(\mathcal{L}(F, \nu J)(U) \cap \Pi) + \dim(\ker Q|_{U_1} \cap U_1^0).$$

Since Q is positive on U_2^0 , we have $(U_2^0)^\perp \cap U_2^0 = \{0\}$ and so we can collect all the formulas using the fact that $(\mathcal{L}(F, \nu J)(U) \cap \Pi) \subset (\mathcal{L}(F, \nu J)(U_1) \cap \Pi)$:

$$\begin{aligned} \operatorname{ind}^+ Q|_{U^0} - \operatorname{ind}^+ Q|_{U_1^0} &= \operatorname{ind}_\Pi(\mathcal{L}(F, \nu J)(U_1), \mathcal{L}(F, \nu J)(U)) + \\ &+ \frac{1}{2} \dim(\mathcal{L}(F, \nu J)(U_1) \cap \Pi) - \frac{1}{2} \dim(\mathcal{L}(F, \nu J)(U) \cap \Pi) \end{aligned}$$

Under the assumption three the formula is valid also in the infinite dimensional case. We know that the \mathcal{L} -derivatives constructed over finite-dimensional subspace will converge and that the quadratic form from the positive Maslov index is continuous. The only possibly discontinuous term are the dimensions of various intersections, but they are zero now for all \mathcal{L} -derivatives constructed over sufficiently large subspace of variations. \square

Jacobi differential equations

In this Chapter we discuss some situations when it is possible to characterize the Jacobi curve as a solution of some boundary value problem for an ODE known as the Jacobi equation. We start with the case of a single control parameter system which contains all the conceptual difficulties but lacks of technical difficulties present in the multidimensional case. Then we proceed to studying the multidimensional case. Using the techniques of \mathcal{L} -derivatives we will give a geometric meaning to the famous Goh condition optimal control. This will allow us to generalize this condition to very non-trivial singular extremals without any difficulties at all.

In this and the next chapters we make the following assumptions:

Assumption 1. Functions $b(t)$, $X(t)$ and the extremal control $\tilde{u}(t)$ are piece-wise analytic as functions of t .

Assumption 2. The initial point $q_0 \in M$ is fixed.

3.1 Jacobi DEs for good one control parameter systems

Let us assume that the control $\tilde{u}(t)$ takes values in the interior of the set U . We consider a sequence of functions $b^i(\tau)$ that we define as

$$b^i(\tau) = \begin{cases} b(\tau), & \text{if } i = 0, \\ \sigma(X^{(i+1)}(\tau), X^{(i)}(\tau)), & \text{if } i \geq 1. \end{cases}$$

For the sake of simplicity we will often to drop in the future the explicit dependence on time τ and simply write b^i or $X^{(i)}$, when there is no confusion.

The *strengthened Legendre condition of order m* is a series of identities of the form

$$b^m \leq \beta < 0, \quad b^j \equiv 0, \quad j < m$$

for some $m \in \mathbb{Z}_{\geq 0}$, where β is just a constant. We say that an extremal curve $\tilde{q}(\tau)$ is a *singular curve of order m* , if along it the strengthened Legendre condition of order m is satisfied. If along a trajectory $b^i \equiv 0$ for all $i \in \mathbb{Z}_{\geq 0}$, we say that the trajectory has order infinity.

We define the *Goh subspaces* as

$$\Gamma^i(\tau) = \text{span}\{X^{(j)}(\tau) : j \leq i\}.$$

Lemma 3.1. *Assume that the strengthened Legendre condition of order m is satisfied along an extremal trajectory $\tilde{q}(\tau)$. Then $\Gamma^{m-1}(\tau)$ is an isotropic subspace. Moreover $X^m(\tau) \in \Gamma^{m-2}(\tau)^\perp$*

Proof. The proof is a simple inductive argument. For $i = 1$ the statement is obvious since $\Gamma^1 = X$. Assume that the statement is true for $i < m - 1$. Then in particular we have

$$\sigma(X^{(i)}, X^{(j)}) = 0, \quad \forall j < i.$$

Differentiating this identity and using the induction assumption we find that

$$\sigma(X^{(i+1)}, X^{(j)}) = 0, \quad \forall j < i.$$

The equality for $j = i$ is obviously true, because $i < m - 1$ and the Legendre condition has order m .

The fact that $X^m(\tau)$ lies in $\Gamma^{m-2}(\tau)^\perp$ now follows from the differentiation of

$$\sigma(X^{(m-1)}, X^{(i)}) = 0.$$

□

We are going to prove the following characterization of the Jacobi curve.

Theorem 3.2. *Let $\tilde{q}(\tau)$ be a regular or a singular extremal of order m . Then \mathcal{L}_t for $t > 0$ is a linear span of $\Gamma^{m-1}(t)$ and the solutions of the following linear ODE*

$$\dot{\mu} = \frac{\sigma(X^{(m)}, \mu)}{b^m} X^{(m)},$$

with boundary conditions $\mu(0) \in T_{\lambda(0)}(T_{\tilde{q}(0)}^ M) \cap (\Gamma^{m-1}(0))^\perp$.*

If the trajectory has infinite order, then we can define

$$\Gamma(\tau) = \bigcup_{i=1}^{\infty} \Gamma^i(\tau)$$

and $\mathcal{L}_t = (T_{\lambda(0)}(T_{\tilde{q}(0)}^ M))^{\Gamma(0+)}$.*

Proof. The proof is based on the technique called the Goh transformations. The idea of that technique is that if an extremal is singular, then the first and second differentials in the definition of the \mathcal{L} -derivative remain continuous in a much weaker topology. So we can extend this map by continuity to a bigger space, in which the original space is dense. From Theorem 1.33 we know that this will not change the \mathcal{L} -derivative.

Let us assume first that the extremal trajectory is regular. If a vector

$$\eta(t) = \eta + \int_0^t X(\tau)v(\tau)d\tau, \quad \eta \in \mathcal{L}_0$$

is in \mathcal{L}_t , then it satisfies

$$\int_0^t \sigma \left(\eta + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)v(\tau)w(\tau)d\tau = 0, \quad \forall w \in L^2[0, t] \quad (3.1)$$

From (3.1) we get that $v(\tau)$ must satisfy

$$\sigma(\eta(\tau), X(\tau)) + b(\tau)v(\tau) = 0 \iff v(\tau) = -b(\tau)^{-1}\sigma(\eta(\tau), X(\tau)), \quad \text{a.e. } \tau \in [0, t]$$

But on the other hand from the definition of $\eta(\tau)$ we have

$$\dot{\eta}(\tau) = X(\tau)v(\tau) \quad \Rightarrow \quad \dot{\eta}(\tau) = -X(\tau)b(\tau)^{-1}\sigma(\eta(\tau), X(\tau))$$

which gives us the classical Jacobi equation [8].

We assume now that the extremal is singular of order m . It is clear that this derivation is not going to work anymore, since $b(\tau) \equiv 0$, so we modify it in the following way. We denote

$$P^m v(t) = \int_0^t \int_0^{\tau_1} \dots \int_0^{\tau_{m-1}} v(\tau_m) d\tau_m \dots d\tau_1, \quad P^m w(t) = \int_0^t \int_0^{\tau_1} \dots \int_0^{\tau_{m-1}} w(\tau_m) d\tau_m \dots d\tau_1$$

the m -th primitives of v and w . We integrate by parts m times the first summand of (3.1)

$$\begin{aligned} & \sigma \left(\eta, \int_0^t X(\tau)w(\tau)d\tau \right) = \\ & = \sigma \left(\eta, \sum_{i=0}^{m-1} (-1)^i X^{(i)}(t)(P^{i+1}w(t)) + (-1)^m \int_0^t X^{(m)}(\tau)(P^m w(\tau))d\tau \right) \end{aligned}$$

Now we integrate by parts the other summand of (3.1). Exchanging the order of integration, using Lemma 3.1 and the assumption on the order of our extremal curve:

$$\begin{aligned} & \int_0^t \sigma \left(\int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) d\tau = (\text{integration by parts}) = \\ & = \int_0^t \sigma \left(X(\tau)(Pv(\tau)) - \int_0^\tau \dot{X}(\theta)(Pv(\theta))d\theta, X(\tau)w(\tau) \right) d\tau = (\text{exchanging order}) = \\ & = \int_0^t \sigma \left(-\dot{X}(\tau)(Pv(\tau)), \int_\tau^t X(\theta)w(\theta)d\theta \right) d\tau = (\text{integration by parts}) = \\ & = \int_0^t \sigma \left(-\dot{X}(\tau)(Pv(\tau)), X(t)(Pw(t)) - X(\tau)(Pw(\tau)) - \int_\tau^t \dot{X}(\theta)(Pw(\theta))d\theta \right) d\tau = \\ & = (\text{assumption}) = \sigma \left(- \int_0^t \dot{X}(\tau)(Pv(\tau))d\tau, X(t)(Pw(t)) \right) + \\ & + \int_0^t \sigma \left(\int_0^\tau \dot{X}(\theta)(Pv(\theta))d\theta, \dot{X}(\tau)(Pw(\tau)) \right) d\tau \end{aligned}$$

We continue to integrate by parts both summands and use Lemma 3.1, until the Legendre term $\sigma(X^{(m)}(t), X^{(m-1)}(t))$ will not appear explicitly. At the end we get

$$\begin{aligned} & \int_0^t \sigma \left(\int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) d\tau = \\ & = \sigma \left((-1)^m \int_0^t X^{(m)}(\tau)(P^m v(\tau))d\tau, \sum_{i=0}^{m-1} (-1)^i X^{(i)}(t)(P^{i+1}w(t)) \right) + \\ & + \int_0^t \sigma \left(X^{(m)}(\tau)(P^m v(\tau)), X^{(m-1)}(\tau)(P^m w(\tau)) \right) d\tau + \\ & + \int_0^t \sigma \left(\int_0^\tau X^{(m)}(\theta)(P^m v(\theta))d\theta, X^{(m)}(\tau)(P^m w(\tau)) \right) d\tau \end{aligned}$$

Thus (3.1) is transformed into

$$\begin{aligned} & \sum_{i=0}^{m-1} \sigma \left(\eta + (-1)^m \int_0^t X^{(m)}(\tau)(P^m v(\tau))d\tau, (-1)^i X^{(i)}(t)(P^{i+1}w(t)) \right) + \\ & + \int_0^t \sigma \left((-1)^m \eta + \int_0^\tau X^{(m)}(\theta)(P^m v(\theta))d\theta, X^{(m)}(\tau)(P^m w(\tau)) \right) d\tau + \quad (3.2) \\ & + \int_0^t \sigma \left(X^{(m)}(\tau)(P^m v(\tau)), X^{(m-1)}(\tau)(P^m w(\tau)) \right) d\tau = 0 \end{aligned}$$

We also integrate by parts the integral representation of $\eta(t)$, to get

$$\eta(t) = \eta + \sum_{i=0}^{m-1} (-1)^i X^{(i)}(t)(P^{i+1}v(t)) + (-1)^m \int_0^t X^{(m)}(\tau)(P^m v(\tau))d\tau$$

We can see that the right hand side of this expression and quadratic form in (3.2) are continuous in the topology $\hat{H}^{-m}[0, t]$ given by the norm

$$\|v\|_{-m} = \sqrt{\sum_{i=0}^{m-1} (P^i v(t))^2 + \|P^m v\|_{L^2}^2}.$$

So we extend by continuity on $\hat{H}^{-m}[0, t]$. It is important to note that in the $\hat{H}^{-m}[0, t]$ the end-points $P^i v(t)$ represent separate variables. This implies immediately that $\Gamma^{m-1}(t) \subset \mathcal{L}_t$. Indeed, we can see that the right-hand side of (3.2) does not depend on $P^i v(t)$ at all. So if we take $\eta = 0$ and $P^m v(\tau) \equiv 0$, then (3.2) will be satisfied automatically. But then $\eta(t) \in \Gamma^{m-1}(t)$ and every vector of $\Gamma^{m-1}(t)$ can be realized this way.

This means that \mathcal{L}_t actually consists of vectors

$$\mu(t) = \eta + (-1)^m \int_0^t X^{(m)}(\tau)(P^m v(\tau))d\tau$$

and vectors from $\Gamma^{m-1}(t)$.

The derivative of $\mu(\tau)$ is given by

$$\dot{\mu}(\tau) = (-1)^m X^{(m)}(\tau)(P^m v(\tau)).$$

We only need to find $P^m v(\tau)$. We do this by solving (3.2), which gives us a system of equations

$$\sigma(\mu(t), X^{(i)}(t)) = 0, \quad 0 \leq i \leq m-1, \quad (3.3)$$

$$(-1)^m \sigma(\mu(\tau), X^{(m)}(\tau)) + b^m(\tau) P^m v(\tau) = 0, \quad \text{a.e. } \tau \in [0, t]. \quad (3.4)$$

Then from the last equation we recover

$$P^m v(\tau) = (-1)^{(m+1)} (b^m(\tau))^{-1} \sigma(\mu(\tau), X^{(m)}(\tau))$$

and so

$$\dot{\mu}(\tau) = -X^{(m)}(\tau) (b^m(\tau))^{-1} \sigma(\mu(\tau), X^{(m)}(\tau)).$$

From (3.3) we recover boundary conditions

$$\mu(t) \in \Gamma^{m-1}(t)^\perp. \quad (3.5)$$

We can prove that this identity is true not only for the chosen time t , but for any $\tau \in [0, t]$. Indeed, from the explicit form of the Jacobi DE we can see that $X^{(m-1)}(\tau)$ is a particular solution. But since all solutions lie in \mathcal{L}_τ we have $\sigma(\mu(\tau), X^{(m-1)}(\tau)) = 0$ for any solution $\mu(\tau)$. Assume that the same is true for $X^{(i)}(\tau)$, $i \leq m-1$. Then for $X^{(i-1)}(\tau)$ we have

$$\frac{d}{d\tau} \sigma(\mu(\tau), X^{(i-1)}(\tau)) = -\frac{\sigma(X^{(m)}(\tau), \mu(\tau))}{b^m(\tau)} \sigma(X^{(m)}(\tau), X^{(i-1)}(\tau)) + \sigma(\mu(\tau), X^{(i)}(\tau)) \equiv 0$$

by Lemma 3.1 and the induction assumption.

So we see that $\mu(\tau) \in \Gamma^{m-1}(\tau)^\perp$ is satisfied automatically if

$$\mu(0+) = (T_{\lambda(0)}(T_{\tilde{q}(0)}^* M))^{\Gamma^{m-1}(0)}.$$

It means that by fixing the appropriate boundary conditions all n independent solutions will lie in \mathcal{L}_t .

Let us now look at what can happen if the singularity is of an infinite order. Since we have already established that $\Gamma(\tau)$ is an isotropic subspace, its dimension is limited. This can happen only if higher derivatives of X become dependent from the lower derivatives. Let us assume that the first l derivatives are generically independent and the $(l+1)$ -th is not. Then $\Gamma^l(\tau)$ is a fixed subspace. Indeed, we can represent $\Gamma^l(\tau)$ as an element of $\wedge^l \mathbb{R}^{2n}$

$$\Gamma^l(\tau) = X(\tau) \wedge \dot{X}(\tau) \wedge \dots \wedge X^{(l)}(\tau).$$

Then $\dot{\Gamma}^l(\tau) = \kappa(\tau)\Gamma^l(\tau)$ for some function $\kappa(\tau)$. Since $\wedge^l \mathbb{R}^{2n}$ is a linear space, the solution of this equation is simply

$$\Gamma^l(\tau) = e^{\int_s^\tau \kappa(\theta) d\theta} \Gamma^l(s).$$

So we see that $\Gamma(\tau) = \Gamma$ is constant except maybe a finite number of points, where the first l derivatives of X can become dependent. Therefore Γ can be taken to be equal to $\Gamma(0+)$.

By assumption on the infinite order we have that the elements of \mathcal{L}_t must satisfy

$$\int_0^t \sigma \left(\eta + \int_0^\tau X^{(l)}(\theta) P^l v(\theta), X^{(l)}(\tau) w(\tau) \right) d\tau = 0.$$

We look for a solution with $P^l v(\tau) = 0$, then the equation above is transformed to

$$\sigma \left(\eta, \int_0^t X^{(l)}(\tau) w(\tau) d\tau \right) = 0.$$

But we have seen that $X^{(l)}(\tau)$ never leaves Γ . Therefore all $\eta \in (T_{\lambda(0)}(T_{\tilde{q}(0)}^* M)) \cap \Gamma^\perp$ satisfy the equation above as well as the boundary conditions (3.3). Therefore those vectors together with vectors from $\Gamma(\tau)$ give n independent solutions whenever $\dim(\Gamma(\tau))$ is maximal. At the isolated points where the dimension of this space drops we simply use the left-continuity property from Lemma 2.9. \square

3.2 The Goh condition and Jacobi DEs for good singular extremals

Let us now look at what happens in the multidimensional case. In this case $X(\tau)$ is a map from $\mathbb{R}^k \rightarrow \mathbb{R}^{2n}$ and $b^i(\tau)$ are quadratic forms which may have kernels.

Let us consider first as an example how to derive the Goh condition if $b^0(\tau)$ vanishes completely on an interval. The classical proof of the Goh condition usually relies on the assumption that the negative or positive index of the Hessian is finite. This is equivalent to the existence of the \mathcal{L} -derivative. So the Goh condition should follow from some argument involving the \mathcal{L} -derivative. And this is indeed so, since the geometric meaning of the Goh condition is that the corresponding Goh subspace is a subspace of the \mathcal{L} -derivative \mathcal{L}_t .

Theorem 3.3. *Suppose that $b(\tau) \equiv 0, \forall \tau \in [0, t]$ along an extremal trajectory $\tilde{q}(\tau)$. Then $X(t)v \in \mathcal{L}_t, \forall v \in \mathbb{R}^k$.*

Proof. By the previous lemma we know that for any arbitrary small neighborhood $O_{\mathcal{L}_t}$ of \mathcal{L}_t , there exists $\varepsilon > 0$, s.t. $\mathcal{L}_{t-\varepsilon} \in O_{\mathcal{L}_t}$. To prove the theorem, we are going to show that $\mathcal{L}_{t-\varepsilon}$ always contains a vector, that is close to $X(\tau)v$. To simplify the notations we write $\mathcal{L}_{t_1}^{t_2}[V] = \mathcal{L}[L_k^2[0, t_1] \oplus V]$, where $t_1 < t_2$ and $V \subset L_k^2[t_1, t_2]$.

We use Lemma 2.10 and compute $\mathcal{L}_{t-\varepsilon}^t[\mathbb{R}(v\phi_\varepsilon(t))]$, where $v \in \mathbb{R}^k$ and $\phi_\varepsilon(t)$ is a smooth approximation of a delta function $\delta(t)$ with support in $(t - \varepsilon, t)$, i.e. we have

$$\int_{t-\varepsilon}^t X(\tau)v \phi_\varepsilon(\tau) d\tau \xrightarrow{\varepsilon \rightarrow 0} X(t)v.$$

By the definition of the \mathcal{L} -derivative, we get $\mathcal{L}_{t-\varepsilon}^t [\mathbb{R}(v\phi_\varepsilon(t))] \in U$. So if we can prove that the vectors

$$\eta_\varepsilon + \int_{t-\varepsilon}^t X(\tau)v\phi_\varepsilon(\tau)d\tau, \quad \eta_\varepsilon \in \mathcal{L}_{t-\varepsilon} \quad (3.6)$$

belong to $\mathcal{L}_{t-\varepsilon}^t [\mathbb{R}(v\phi_\varepsilon(t))]$ and $\eta_\varepsilon \rightarrow 0$, then we are done.

The condition that our candidate vector (3.6) lies in $\mathcal{L}_{t-\varepsilon}^t [\mathbb{R}(v\phi_\varepsilon(t))]$ is determined by a single equation of the form

$$\sigma \left(\eta_\varepsilon, \int_{t-\varepsilon}^t X(\tau)v\phi_\varepsilon(\tau)d\tau \right) + \int_{t-\varepsilon}^t \sigma \left(\int_{t-\varepsilon}^\tau X(\theta)v\phi_\varepsilon(\theta)d\theta, X(\tau)v\phi_\varepsilon(\tau) \right) d\tau = 0.$$

If $X(t)v = 0$ or

$$\int_{t-\varepsilon}^t X(\tau)v\phi_\varepsilon(\tau)d\tau \in \mathcal{L}_{t-\varepsilon},$$

for all ε small enough, then we can take $\eta_\varepsilon = 0$ and we are done, because $X(t)v \in \mathcal{L}_t$ by the left-continuity of the Jacobi curve. So suppose that there exists a sequence $\varepsilon_j \rightarrow 0$ and $\eta_j \in \mathcal{L}_{t-\varepsilon_j}$, s.t.

$$\sigma \left(\eta_j, \int_{t-\varepsilon_j}^t X(\tau)v\phi_{\varepsilon_j}(\tau)d\tau \right) \neq 0.$$

Then we take $\eta_{\varepsilon_j} = k_j\eta_j$ and solve for k_j the equation

$$\sigma \left(k_j\eta_j, \int_{t-\varepsilon_j}^t X(\tau)v\phi_{\varepsilon_j}(\tau)d\tau \right) + \int_{t-\varepsilon_j}^t \sigma \left(\int_{t-\varepsilon_j}^\tau X(\theta)v\phi_{\varepsilon_j}(\theta)d\theta, X(\tau)v\phi_{\varepsilon_j}(\tau) \right) d\tau = 0.$$

obtaining

$$k_j = - \frac{\int_{t-\varepsilon_j}^t \sigma \left(\int_{t-\varepsilon_j}^\tau X(\theta)v\phi_{\varepsilon_j}(\theta)d\theta, X(\tau)v\phi_{\varepsilon_j}(\tau) \right) d\tau}{\sigma \left(\eta_j, \int_{t-\varepsilon_j}^t X(\tau)v\phi_{\varepsilon_j}(\tau)d\tau \right)}$$

If $\eta_j \rightarrow X(t)v$ then we are done, because in this case $X(t)v \in \mathcal{L}_t$ by left continuity of \mathcal{L}_τ . So we assume that this is not true. Then without any loss of generality we can normalize for j big enough

$$\sigma \left(\eta_j, \int_{t-\varepsilon_j}^t X(\tau)v\phi_{\varepsilon_j}(\tau)d\tau \right) = 1.$$

Then clearly $|\eta_j| \rightarrow C < \infty$ and so it is sufficient to prove that $k_j \rightarrow 0$.

Suppose that we have a smooth function $\phi(x)$ with support in $[0, 1]$ and such that

$$\int_0^1 \phi(x)dx = 1.$$

Then we use as $\phi_\varepsilon(x)$ the standard mollifiers

$$\phi_\varepsilon(t) = \frac{1}{\varepsilon} \phi\left(\frac{x - (t - \varepsilon)}{\varepsilon}\right).$$

Thus after an obvious change of variables we find

$$\begin{aligned} k_j &= - \int_0^1 \sigma \left(\int_0^\tau X(\varepsilon_j \theta + (t - \varepsilon_j))v \phi(\theta) d\theta, X(\varepsilon_j \tau + (t - \varepsilon_j))v \phi(\tau) \right) d\tau = \\ &= -\sigma(X(t)v, X(t)v) \int_0^1 \int_0^\tau \phi(\theta)\phi(\tau) d\theta d\tau + O(\varepsilon_j) = O(\varepsilon_j) \end{aligned}$$

which proves the result. \square

The same argument can be used for singular extremals of higher order.

Let us now go back to our main problem of writing down the Jacobi equation. If $b(\tau)$ is analytic, then $b(\tau)$ may have a kernel that changes analytically in space except of a finite number of points, where the dimension of the kernel can jump. On an interval where the dimension of the kernel is constant we can actually write down the Jacobi equation that determines the \mathcal{L} -derivative. At singular points the Jacobi curve may have jump discontinuities, and we consider a simple example of the sort in the next chapter.

We assume that the dimension of the kernel of $b(\tau)$ is constant. The first step is to choose a fixed subspace E_1 of the dimension of the kernel. For example we can assume that E_1 is the span of the last $\dim \ker b(\tau)$ vectors. Then we can find an analytic family of orthogonal transformations $R^1(\tau)$, s.t. $R^1(\tau)E_0 = \ker b(\tau)$. Note that the map $v(\tau) \mapsto R^1(\tau)v(\tau)$ is an isometry of $L_k^2[0, t]$ to itself and therefore we do not change the space of variations.

As in the previous section we study the equation (3.1) for the \mathcal{L} -derivative. Let $\pi_1 : \mathbb{R}^k \rightarrow E_1$ be a projection map and $Pf(t)$ the primitive of the function $f(t)$. We define the map

$$\gamma^1 v(\tau) = (\text{id} - \pi_1)(R^1(\tau))^{-1}v(\tau) + \pi_1 P((R^1(\tau))^{-1}v(\tau)).$$

Since the kernel is now fixed π_1 commutes with P and derivatives with respect to time. Using this and the fact, that π_1 is idempotent, i.e. $\pi_1^2 = \pi_1$, we derive

$$(R^1(\tau))^{-1}v(\tau) = (\text{id} - \pi_1)\gamma^1 v(t) + \pi_1 \frac{d}{dt}(\gamma^1 v(t)).$$

In the equation (3.1) we replace $v(\tau)$ with $R^1(\tau)(R^1(\tau))^{-1}v(\tau)$ and plug in the previous expression. Then using exactly the same argument as in the proof of the Goh condition, we find that $\text{span}\{(X(\tau)R^1(\tau))\pi_1 v, \forall v \in \mathbb{R}^k\}$ is an isotropic subspace and lies in the \mathcal{L} -derivative. Having that in mind we integrate by parts (3.1) just like in the previous subsection. To simplify the final result we introduce the following notations

$$X^1(\tau) = (X(\tau)R^1(\tau))(\text{id} - \pi_1) - \frac{d}{d\tau}(X(\tau)R^1(\tau))\pi_1$$

$$\begin{aligned}
b^1(\tau) &= (R^1(\tau))^T b(\tau) R^1(\tau) + \sigma \left((X(\tau)R^1(\tau))\pi_1 \cdot, (X(\tau)R^1(\tau))(\text{id} - \pi_1) \cdot \right) - \\
&\quad - \sigma \left((X(\tau)R^1(\tau))(\text{id} - \pi_1) \cdot, (X(\tau)R^1(\tau))\pi_1 \cdot \right) + \\
&\quad + \sigma \left(\frac{d}{d\tau} (X(\tau)R^1(\tau))\pi_1 \cdot, (X(\tau)R^1(\tau))\pi_1 \cdot \right) \\
&\quad + \sigma \left(\eta + \int_0^t X^1(\tau)(\gamma^1 v(\tau)) d\tau, (X(t)R^1(t))\pi_1(\gamma^1 w(t)) \right) + \\
&\quad + \int_0^t \sigma \left(\eta + X^1(\theta)(\gamma^1 v(\theta)) d\theta, X^1(\tau)(\gamma^1 w(\tau)) \right) d\tau + \\
&\quad + \int_0^t b^1(\tau)(\gamma^1 v(\tau), \gamma^1 w(\tau)) d\tau = 0
\end{aligned} \tag{3.7}$$

We integrate the same way the expression of $\eta(t)$ to get

$$\eta(t) = \eta + (X(t)R^1(t))\pi_1(\gamma^1 v(t)) + \int_0^t X^1(\tau)(\gamma^1 v(\tau)) d\tau.$$

We introduce the space $\hat{H}_{\gamma^1}[0, t]$ as the completion of $L_k^2[0, t]$ in the norm

$$\|v\|_{\gamma^1} = \sqrt{\|\gamma^1 v\|_{L_k^2}^2 + \|(\gamma^1 v(t))\|_{\mathbb{R}^k}^2}.$$

One can check again that the quadratic form in (3.7) is continuous in that topology. Therefore we can extend it by continuity and do the same for the map

$$v(\tau) \mapsto (X(t)R^1(t))\pi_1(\gamma^1 v(t)) + \int_0^t X^1(\tau)(\gamma^1 v(\tau)) d\tau.$$

Then once again the quadratic form $b^1(\tau)$ is analytic and can have a kernel, that has a constant dimension everywhere except a finite number of points. Restricting to a smaller interval once again we can assume that the dimension of $\ker b^1(\tau)$ is constant. We claim that $\dim \ker b^1(\tau) \leq \dim \ker b(\tau)$. Indeed, we can write

$$b^1(\tau) = \begin{pmatrix} A(\tau) & B(\tau) \\ B^T(\tau) & C(\tau) \end{pmatrix},$$

where

$$\begin{aligned}
A(\tau) &= ((R^1(\tau))^T b(\tau) R^1(\tau)) |_{\text{im}(\text{id} - \pi_1)}, \\
B(\tau) &= \sigma \left((X(\tau)R^1(\tau))\pi_1 \cdot, (X(\tau)R^1(\tau))(\text{id} - \pi_1) \cdot \right), \\
C(\tau) &= \sigma \left(\frac{d}{d\tau} (X(\tau)R^1(\tau))\pi_1 \cdot, (X(\tau)R^1(\tau))\pi_1 \cdot \right).
\end{aligned}$$

By construction $A(\tau)$ is invertible. So we apply a linear change of variables

$$\begin{aligned} & \begin{pmatrix} \text{id} & -A^{-1}(\tau)B(\tau) \\ 0 & \text{id} \end{pmatrix}^T \begin{pmatrix} A(\tau) & B(\tau) \\ B^T(\tau) & C(\tau) \end{pmatrix} \begin{pmatrix} \text{id} & -A^{-1}(\tau)B(\tau) \\ 0 & \text{id} \end{pmatrix} = \\ & = \begin{pmatrix} A(\tau) & 0 \\ 0 & C(\tau) - B^T(\tau)A^{-1}(\tau)B(\tau) \end{pmatrix} \end{aligned}$$

This way we obtain that $\dim \ker b^1(\tau) = \dim \ker(C(\tau) - B^T(\tau)A^{-1}(\tau)B(\tau)) \leq \dim \ker b^0(\tau)$.

This fact allows us to repeat inductively this integration by parts procedure several times. To be more precise, we have the following proposition.

Proposition 3.4. *Let $(\gamma^0 v(\tau), X^0(\tau), b^0(\tau), E_0, R^0(\tau)) = (v(\tau), X(\tau), b(\tau), \mathbb{R}^k, \text{id})$. Then we define inductively $(\gamma^i v(\tau), X^i(\tau), b^i(\tau), E_i, R^i(\tau))$ in the following way. We take E_i to be any subspace of E_{i-1} , s.t. $\dim E_i = \dim \ker b^{i-1}(\tau)$ and $R^i(\tau)$ be any rotation matrix, s.t. $R^i(\tau)E_i = \ker b^{i-1}(\tau)$ and*

$$\gamma^{i+1}v(\tau) = (\text{id} - \pi_{i+1})(R^{i+1}(\tau))^{-1}\gamma^i v(\tau) + \pi_{i+1}P((R^{i+1}(\tau))^{-1}\gamma^i v(\tau)).$$

$$X^{i+1}(\tau) = (X^i(\tau)R^{i+1}(\tau))(\text{id} - \pi_{i+1}) - \frac{d}{d\tau}(X^i(\tau)R^{i+1}(\tau))\pi_{i+1}$$

$$\begin{aligned} b^{i+1}(\tau) &= (R^{i+1}(\tau))^T b(\tau) R^{i+1}(\tau) + \\ &+ \sigma((X^i(\tau)R^{i+1}(\tau))\pi_{i+1}, (X^i(\tau)R^{i+1}(\tau))(\text{id} - \pi_{i+1})) - \\ &- \sigma((X^i(\tau)R^{i+1}(\tau))(\text{id} - \pi_{i+1}), (X^i(\tau)R^{i+1}(\tau))\pi_{i+1}) + \\ &+ \sigma\left(\frac{d}{d\tau}(X^i(\tau)R^{i+1}(\tau))\pi_{i+1}, (X^i(\tau)R^{i+1}(\tau))\pi_{i+1}\right) \end{aligned}$$

The \mathcal{L} -derivative at moment of time t then consists of the vectors

$$\eta(t) = \eta + \sum_{j=0}^{i-1} (X^j(t)R^{j+1}(t))\pi_{j+1}(\gamma^j v(t)) + \int_0^t X^i(\tau)(\gamma^i v(\tau))d\tau.$$

where $\eta \in \Pi$ and $v(\tau)$ satisfy a system of equations

$$\begin{aligned} & \int_0^t \sigma\left(\eta + \int_0^\tau X^i(\theta)(\gamma^i v(\theta))d\theta, X^i(\tau)(\gamma^i w(\tau))\right) d\tau + \int_0^t b^i(\tau)(\gamma^i v(\tau), \gamma^i w(\tau))d\tau = 0, \\ & \sigma\left(\eta + \int_0^t X^i(\tau)(\gamma^i v(\tau))d\tau, \sum_{j=1}^i (X^{j-1}(t)R^j(t))\pi_j(\gamma^j w(t))\right) = 0, \end{aligned} \tag{3.8}$$

and $v(\tau)$ is in $\hat{H}_{\gamma^i}[0, t]$, that is defined as the completion of $L_k^2[0, t]$ in the norm

$$\|v(\tau)\|_{\gamma^i} = \sqrt{\| \gamma^i v(\tau) \|_{L_k^2}^2 + \sum_{j=1}^{i-1} |\gamma^j v(t)|^2}.$$

Moreover the \mathcal{L} -derivative contains the Goh subspace

$$\Gamma^{i-1}(\tau) = \text{span}\{(X^j(\tau)R^{j+1}(\tau))\pi_{j+1}v, v \in \mathbb{R}^k, j \leq i-1\} \subset \mathcal{L}_\tau.$$

We now have three possibilities:

1. At some point $b^i(\tau)$ becomes non degenerate;
2. $b^i(\tau)$ is degenerate for all $i \in \mathbb{N}$, but at some point $\dim \Gamma^i(\tau) = n$ and therefore $\Gamma^i(\tau) = \mathcal{L}_\tau$.
3. $b^i(\tau)$ is degenerate for all $i \in \mathbb{N}$ and $\dim \Gamma^i(\tau) < n$.

Let us see what happens in the first case, i.e. there exists an integer m , s.t. $b^m(\tau)$ is non-degenerate for all τ . We define like in the one dimensional-case

$$\mu(t) = \eta + \int_0^t X^m(\tau)(\gamma^m v(\tau))d\tau.$$

Then we have from the definition of the \mathcal{L} -derivative, that

$$\int_0^t \sigma(\mu(\tau), X^m(\tau)\gamma^m w(\tau)) + b_\tau^m(\gamma^m v(\tau), \gamma^m w(\tau))d\tau = 0 \quad (3.9)$$

$$\sigma\left(\mu(t), \sum_{i=0}^{m-1} (X^i(t)R^{i+1}(t))\pi_{i+1}w\right) = 0, \quad \forall w \in \mathbb{R}^k.$$

Then as in the previous subsection we get the Jacobi ODE:

$$\dot{\mu}(\tau) = X^m(\tau)(b^m(\tau))^{-1}\sigma(X^m(\tau)\cdot, \mu(\tau))$$

with the boundary conditions $\mu(t) \in (\Gamma^m(t))^\perp$

As in the one-dimensional case it is now enough to show that the flow of the Jacobi ODE preserves the constraint $\mu_\tau \in (\Gamma^{m-1}(\tau))^\perp$ for all $\tau \in (0, t]$. If we simply differentiate like in the previous subsection, the formulas that we obtain are quiet involved. So to overcome this difficulty, we are going to use (3.8) directly by taking $\gamma^m w(\tau)$ to be some smooth functions on $[0, t)$ and the endpoints $\pi_i \gamma^i w(t) = 0$ for all $i = 0, \dots, m-1$. This way we are going to prove that $\mu(\tau) \in (\Gamma^i(\tau)/\Gamma^{i-1}(\tau))^\perp$ for all $i \leq m-1$.

To see what variations we should use, we integrate by parts back the formula (3.8) assuming $i = m$ with respect to all $\pi_m \gamma^m w(\tau)$, expand the necessary definitions and use the Goh condition until in the expressions we are left only with $\pi_i \gamma^i w(\tau)$ for some fixed i . It can be still rather difficult to keep track of all the steps, but we can use a simple trick and instead integrate by parts (3.9) with respect to $(R^{i+1}(\tau))^{-1}\pi_{i+1}\gamma^i v(\tau)$ several times. Since $X^m(\tau)$ is analytic and $b^m(\tau)$ is analytic and non-degenerate, $\gamma^m v(\tau)$ will be analytic as well and this procedure will give the same result.

Since we take $\pi_j \gamma^j w(t) = 0$ for all $0 \leq j \leq m$ we omit the corresponding terms and obtain from (3.9)

$$\int_0^t \sigma(\mu(\tau), X^i(\tau)(\gamma^i w(\tau))) + b^i(\tau)(\gamma^i v(\tau), \gamma^i w(\tau)) + \sum_{j=i}^m \sigma(X^j(\tau)(R^{j+1}(\tau))\pi_{j+1}(\gamma^{j+1} v(\tau)), X^i(\tau)(\gamma^i w(\tau))) d\tau = 0.$$

We are going to take such variations $\gamma^m w(\tau)$, that $(R^{i+1}(\tau))^{-1} \gamma^i w(\tau) \in E_{i+1}$ and its components are trigonometric functions. Then by definition of the space E_{i+1} we have $b^i(\tau)(\cdot, \gamma^i w(\tau)) = 0$ and from the expression above, we then obtain using the Goh condition:

$$\int_0^t \sigma(\mu(\tau), (X^i(\tau)R^{i+1}(\tau))\pi_{i+1}\hat{w}(\tau))d\tau = 0$$

where $\hat{w}(\tau)$ is any trigonometric function. Thus

$$\sigma(\mu(\tau), (X^i(\tau)R^{i+1}(\tau))\pi_{i+1}\cdot) = 0, \quad \forall \tau \in [0, T]$$

and the \mathcal{L} -derivative consists of the sum of $\Gamma^m(t)$ and all the solutions of the Jacobi equation with $\mu(0) \in T_{\lambda(0)}(T_{\hat{q}(0)}^* M)^{\Gamma^m(0)}$ like in the one-dimensional case.

Jacobi curves in the simplest singular case

In this chapter we consider the simplest singularity when $b(\tau) = 0$ for a single-control system at some moment of time τ . Due to analyticity assumption such a point must be isolated on the extremal. We would like to construct the Jacobi curve after we pass the singularity. In the following sections we are going to prove the following result.

Theorem 4.1. *Let $\tilde{q}(\tau)$ be an extremal curve for which $b(\tau) = 0$ and $b(\tau + \varepsilon) = b_m \varepsilon^m + \dots$ is negative for all $\varepsilon > 0$ sufficiently small. Assume that at τ the following conditions are satisfied*

1. $\sigma(X(\tau), \dot{X}(\tau)) \neq 0$ and $4\sigma(X(\tau), \dot{X}(\tau)) + b_2 \neq 0$ if $m = 2$;
2. $\dim \text{span}\{\ddot{X}(s), \dot{X}(s), X(s)\} = \text{const}$ for $s \in [\tau, \tau + \varepsilon]$.

Then if the right limit of the Jacobi curve $\mathcal{L}(s)$ at $s = \tau$ exists, it is equal to

$$\mathcal{L}(\tau+) = \mathcal{L}_\tau^{X(\tau)}. \quad (4.1)$$

Before we proceed we would like to make some remarks about the statement of the theorem.

1. The assumptions we make allow us to have a simplest possible singularity that in some sense is generic. It is possible to replace those conditions with different ones and to study even more singular cases;
2. The Jacobi curve may not be well defined, due to an infinite inertia index of the Hessian. We will give sufficient conditions for existence and non-existence using oscillation theorems for Hamiltonian systems in Section 4.3;
3. It is not true that after the singularity the Jacobi curve is determined only by its jump. Indeed, it must satisfy the Jacobi equation, but at the same time the right hand-side of the Jacobi equation is not even continuous. So we do not have neither existence nor uniqueness of solutions and we need more information to isolate the right solution. In Section 4.5 for $m = 1, 2$ we will prove that Jacobi curve can be uniquely characterized by a one-jet.

Due to analyticity singular points can not cluster. That is why without any loss of generality from now on we assume that $b(0) = 0$ and that $b(\tau) < 0$ for τ sufficiently small. To give a characterization of the Jacobi curve after a singularity of the considered type, we use once again the theory of \mathcal{L} -derivatives.

Even in this case we still have a Jacobi equation of the form

$$\dot{\eta} = \frac{\sigma(X, \eta)}{b} X,$$

and it governs the behavior of the Jacobi curve away from singularity. In order to understand how to proceed at the singularity, we must recall that by definition \mathcal{L} -derivatives are constructed by adding more and more variations and in the limit we get pointwise convergence to the Jacobi curve. Assume that we use only variations whose support does not intersect $[0, \varepsilon]$. Then using the same argument as in the previous section we obtain a slightly different Jacobi equation of the form

$$\dot{\eta} = \begin{cases} \frac{\sigma(X, \eta)}{b} X, & \text{if } \tau \notin [0, \varepsilon], \\ 0, & \text{if } \tau \in [0, \varepsilon]. \end{cases}$$

It is equivalent to the following construction. We will denote by $\Lambda^\varepsilon(\tau)$ a solution of the induced Jacobi equation in the Lagrangian Grassmanian. We use the Jacobi flow to determine the Jacobi curve until time 0. We assume that the left limit exist and is equal to the corresponding \mathcal{L} -derivative $\Lambda^\varepsilon(0) = \mathcal{L}_0$. Then the flow does nothing for a while, meaning that $\Lambda^\varepsilon(\tau) = \mathcal{L}_0$ for $\tau \in [0, \varepsilon]$ and then we continue with the flow after the moment of time ε , where the dynamics is non-singular until the next zero of $b(\tau)$. This means that the Jacobi curve is going to be a point-wise limit of solutions of the Jacobi equation on the Lagrangian Grassmanian that satisfy $\Lambda^\varepsilon(\varepsilon) = \mathcal{L}_0$. Since outside of the singularity we have uniqueness and existence, for each ε we obtain a unique curve in $L(T_{\lambda(0)}(T * M))$. The pointwise limit of these curves is the Jacobi curve we seek. In the last section we will see how to describe this curve using singular boundary value problems for which we can prove existence and uniqueness.

4.1 Normal form for the Jacobi DE

First we try to simplify the Jacobi DE and reduce the dimension of the considered problem by separating singular and regular dynamics of the Jacobi equation. Let J be the complex structure associated to the symplectic form σ . Then we can rewrite the Jacobi DE as

$$\dot{\eta} = \frac{X X^T J}{b} \eta.$$

We make a time-dependent change of variables $\mu(\tau) = M^{-1}(\tau)\eta(\tau)$ and get

$$\dot{\mu} = -M^{-1}\dot{M}\mu + \frac{M^{-1}X X^T J M}{b} \mu.$$

First we look carefully at the second term. We assume that the matrix M is symplectic. Then $M^{-1} = -JM^T J$ and we obtain

$$M^{-1}XX^TJM = (M^{-1}X)(M^TJ^TX)^T = (M^{-1}X)(J^TJM^TJ^TX)^T = (M^{-1}X)(M^{-1}X)^TJ.$$

We also make a choice for the first column of M by assuming

$$M^{-1}(\tau)X(\tau) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

All this implies that we get an equation of the form

$$\dot{\mu} = -M^{-1}\dot{M}\mu + \frac{1}{b} \left(\begin{array}{c|ccc} & 1 & 0 & \cdots & 0 \\ & 0 & 0 & \cdots & 0 \\ & \vdots & \vdots & \ddots & \vdots \\ & 0 & 0 & \cdots & 0 \\ \hline & & & & \\ 0 & & & & 0 \end{array} \right) \mu. \quad (4.2)$$

Now we work with the first term $M^{-1}\dot{M}$. Since we choose M to be symplectic, its columns form a Darboux basis. Denote the first n columns by e_i and the last by f_i . Then we can write

$$-M^{-1}\dot{M} = \begin{pmatrix} \sigma(f, \dot{e}) & \sigma(f, \dot{f}) \\ \sigma(e, \dot{e}) & \sigma(e, \dot{f}) \end{pmatrix} = \begin{pmatrix} \sigma(f, \dot{e}) & \sigma(f, \dot{f}) \\ \sigma(e, \dot{e}) & -\sigma(f, \dot{e}) \end{pmatrix},$$

where $\sigma(x, y)$ means a matrix whose elements are $\sigma(x_i, y_j)$, for two n -tuples of vectors $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_n)$. The last equality follows from the fact that the basis is Darboux, i.e.

$$\sigma(e, f) = \text{id}_n \Rightarrow \sigma(\dot{e}, f) + \sigma(e, \dot{f}) = 0.$$

Let us first assume that $n = 1$. Since by assumption of Theorem 4.1 $\sigma(X(0), \dot{X}(0)) \neq 0$, we can choose

$$e(\tau) = e_1(\tau) = X(\tau), \quad f(\tau) = f_1(\tau) = \frac{\dot{X}(\tau)}{\sigma(X(\tau), \dot{X}(\tau))}, \quad (4.3)$$

The Jacobi DE reduces then to the following normal form

$$\frac{d}{d\tau} \begin{pmatrix} \mu_1 \\ \mu_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & \frac{\sigma(\dot{X}, \ddot{X})}{\sigma(X, \dot{X})^2} + \frac{1}{b} \\ \sigma(X, \dot{X}) & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_{n+1} \end{pmatrix} \quad (4.4)$$

Let us now assume, that $n \geq 2$. We want to separate the singular dynamics from the regular dynamics. If we look at the singular part in (4.2), then we see that the only non-zero element is in the first row and $(n + 1)$ column. So in the new coordinates we want at least some of the expressions for $\dot{\mu}^i$ for $i \neq 1, n + 1$ to be independent of μ^1, μ^{n+1} . Moreover the assumption, that $M(\tau)$ is symplectic, is going to imply in addition that we will have two invariant symplectic subspaces: a subspace containing μ_1, μ_{n+1} coordinates, where the singular dynamics happens and its complement where the dynamics is smooth.

So we look for e_i, f_i such that

$$\sigma(e_i, e_1) = 0, \quad \sigma(f_i, e_1) = 0, \quad \sigma(e_i, f_1) = 0, \quad \sigma(f_i, f_1) = 0. \quad (4.5)$$

From the assumption 2) in Theorem 4.1, that we have made, it follows that the dimension of $\text{span}\{\ddot{X}(\tau), \dot{X}(\tau), X(\tau)\}$ must be equal either to two or three. In the first case $\ddot{X}(\tau)$ is simply in the span of $X(\tau), \dot{X}(\tau)$ for small $\tau \geq 0$. So we make the same choice (4.3) for e_1, f_1 and the rest of the columns we take to be a smooth Darboux basis for the symplectic space $\text{span}\{X(\tau), \dot{X}(\tau)\}^\perp$. Then the conditions (4.5) are indeed satisfied and we obtain exactly the equation (4.4) for the singular part.

In the second case we can not guarantee that conditions (4.5) are satisfied for $i \geq 2$, since now $\ddot{X}(\tau)$ has to be accounted for. To isolate the singular dynamics we choose a Darboux basis e_1, e_2, f_1, f_2 as follows. The vectors e_1, f_1 are as before, e_2 is defined as

$$e_2 = \ddot{X} - \frac{\sigma(\ddot{X}, X)}{\sigma(\dot{X}, X)}\dot{X} + \frac{\sigma(\ddot{X}, \dot{X})}{\sigma(\dot{X}, X)}X.$$

and f_2 is chosen to be any vector such that we get a Darboux basis. In this case it just means that

$$\sigma(X, f_2) = \sigma(\dot{X}, f_2) = \sigma(\ddot{X}, f_2) - 1 = 0.$$

The rest of the columns of M_τ are chosen to be a smooth Darboux basis of the symplectic space $(\text{span}\{e_1, e_2, f_1, f_2\})^\perp$. Again, the derivatives of e_1, f_1 are contained in $\text{span}\{e_1, e_2, f_1, f_2\}$, so the dynamics splits. The singular dynamics takes place in the plane with $(\mu_1, \mu_2, \mu_{n+1}, \mu_{n+2})$ coordinates. Thus we get an invariant subsystem containing the singular part

$$\frac{d}{dt} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{n+1} \\ \mu_{n+2} \end{pmatrix} = \begin{pmatrix} \sigma(f_1, \dot{e}_1) & \sigma(f_1, \dot{e}_2) & \sigma(f_1, \dot{f}_1) + \frac{1}{b} & \sigma(f_1, \dot{f}_2) \\ \sigma(f_2, \dot{e}_1) & \sigma(f_2, \dot{e}_2) & \sigma(f_2, \dot{f}_1) & \sigma(f_2, \dot{f}_2) \\ \sigma(e_1, \dot{e}_1) & \sigma(e_1, \dot{e}_2) & \sigma(e_1, \dot{f}_1) & \sigma(e_1, \dot{f}_2) \\ \sigma(e_2, \dot{e}_1) & \sigma(e_2, \dot{e}_2) & \sigma(e_2, \dot{f}_1) & \sigma(e_2, \dot{f}_2) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{n+1} \\ \mu_{n+2} \end{pmatrix}.$$

and so we have proven our first result about the jump of the Jacobi curve \mathcal{L}_τ

Proposition 4.2. *If $b(\tau) = 0$, then $\dim(\mathcal{L}_\tau \cap \mathcal{L}_{\tau+}) \geq n - 2$.*

Let us simplify this equation even more. Since $M^{-1}\dot{M}$ is a matrix from the symplectic Lie algebra, not all the entries above are independent. More precisely, the first and the

last diagonal 2x2 minors are minus transpose of each other and the off diagonal 2x2 minors are symmetric. We can find explicitly

$$\begin{aligned}
\sigma(f_1, \dot{e}_1) &= \sigma(f_1, \dot{e}_2) = \sigma(f_2, \dot{e}_1) = \sigma(e_1, \dot{e}_2) = 0; \\
\sigma(e_1, \dot{f}_1) &= \sigma(e_1, \dot{f}_2) = \sigma(e_2, \dot{f}_1) = \sigma(e_2, \dot{e}_1) = 0; \\
\sigma(f_2, \dot{f}_1) &= \sigma(f_1, \dot{f}_2) = -\frac{1}{\sigma(X, \dot{X})}; \\
\sigma(f_2, \dot{e}_2) &= -\sigma(e_2, \dot{f}_2) = \sigma(f_2, \ddot{X}) + \frac{\sigma(\ddot{X}, X)}{\sigma(\dot{X}, X)}; \\
\sigma(e_1, \dot{e}_1) &= \sigma(X, \dot{X}); \\
\sigma(e_2, \dot{e}_2) &= \sigma(\ddot{X}, \ddot{X}) - \frac{\sigma(\ddot{X}, X)\sigma(\ddot{X}, \dot{X}) + \sigma(\ddot{X}, X)\sigma(\ddot{X}, \dot{X})}{\sigma(\dot{X}, X)}; \\
\sigma(f_1, \dot{f}_1) &= \frac{\sigma(\dot{X}, \ddot{X})}{\sigma(X, \dot{X})^2}.
\end{aligned}$$

So we get an equation of the form

$$\frac{d}{dt} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{n+1} \\ \mu_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \sigma(f_1, \dot{f}_1) + \frac{1}{b} & \sigma(f_2, \dot{f}_1) \\ 0 & \sigma(f_2, \dot{e}_2) & \sigma(f_2, \dot{f}_1) & \sigma(f_2, \dot{f}_2) \\ \sigma(e_1, \dot{e}_1) & 0 & 0 & 0 \\ 0 & \sigma(e_2, \dot{e}_2) & 0 & -\sigma(f_2, \dot{e}_2) \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{n+1} \\ \mu_{n+2} \end{pmatrix}. \quad (4.6)$$

Note that if $\sigma(f_2, \dot{f}_1) \equiv 0$ for τ small enough, we obtain a $n = 1$ normal form of the Jacobi DE as a subsystem. So without any loss of generality from now on we can assume that $n \geq 2$.

We can simplify the last equation even more by taking

$$Q(\tau) = \begin{pmatrix} 1 & 0 \\ 0 & \exp\left(\int_0^\tau \sigma(\dot{f}_2(s), e_2(s)) ds\right) \end{pmatrix}.$$

Note that $Q(0) = \text{id}_n$. We introduce new variables

$$\begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} Q^{-1} & 0 \\ 0 & Q \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_{n+1} \\ \mu_{n+2} \end{pmatrix}.$$

If we write $p = (p_1, p_2)$, $q = (q_1, q_2)$, we obtain a normal form

$$\frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & \frac{B(\tau)}{\tau^m} \\ C(\tau) & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} \quad (4.7)$$

where m is the first non zero coefficient of the Taylor expansion of $b(\tau)$,

$$C(\tau) = Q \begin{pmatrix} \sigma(e_1, \dot{e}_1) & 0 \\ 0 & \sigma(e_2, \dot{e}_2) \end{pmatrix} Q = \begin{pmatrix} c_{11}(\tau) & 0 \\ 0 & c_{22}(\tau) \end{pmatrix}$$

and

$$\frac{B(\tau)}{\tau^m} = \begin{pmatrix} 1/b & 0 \\ 0 & 0 \end{pmatrix} + Q^{-1} \begin{pmatrix} \sigma(f_1, \dot{f}_1) & \sigma(f_2, \dot{f}_1) \\ \sigma(f_2, \dot{f}_1) & \sigma(f_2, \dot{f}_2) \end{pmatrix} Q^{-1} = \begin{pmatrix} \frac{1}{b(\tau)} + b_{11}(\tau) & b_{12}(\tau) \\ b_{12}(\tau) & b_{22}(\tau) \end{pmatrix}$$

Note that (4.7) is still a Hamiltonian system and a normal form for the singular part of the Jacobi DE. Moreover we can choose our frame so that $B(\tau)$ is positive for small $\tau > 0$. Indeed, this is going to be true if the trace and the diagonal entries are non-negative. Since $b(\tau) = b_m \tau^m + O(\tau^{m+1}) < 0$ for $\tau > 0$ small and the frame $\{e_i, f_i\}$ was chosen to be analytic, we obviously have that the trace and the upper diagonal element are negative for $\tau > 0$ sufficiently small. We claim that f_2 can be chosen in such a way that also the lower diagonal term is negative as well for small $\tau > 0$. Indeed, the only freedom that we have is to replace $f_2(\tau)$ with $f_2(\tau) + a(\tau)e_2(\tau)$ for some analytic function $a(\tau)$. Then we have

$$\sigma(f_2 + ae_2, \dot{f}_2 + \dot{a}e_2 + a\dot{e}_2) = \sigma(f_2, \dot{f}_2) - \dot{a} + a^2\sigma(e_2, \dot{e}_2),$$

where we have used that $\sigma(e_2, \dot{f}_2) = 1$ and $\sigma(\dot{e}_2, \dot{f}_2) + \sigma(e_2, \dot{f}_2) = 0$. Then it is clear that we can simply choose $a(t) = \sin kt$, with k sufficiently large and the explicit form of Q implies that we can assume without any loss of generality that $B(\tau)$ is negative for τ small.

Before we start proving Theorem 4.1, it is very helpful to understand the idea of the proof using some simple heuristics in the case $n = 1$. Later we will make all the steps rigorous. In the next section we will see under which conditions \mathcal{L} -derivatives exist and how to characterize the Jacobi curve as a solution to the singular Jacobi equation with certain boundary conditions when $n = 1$.

4.2 Heuristics for one-dimensional variational problems

Assume that $n = 1$ and let $b_m \tau^m + \dots$ be the right series of $b(\tau)$ at $\tau = 0$, with $b_m < 0$. We would like to determine whether or not \mathcal{L}_{0+} exists at all. To do this we rewrite (4.4) as a second order ODE of the form

$$\ddot{\mu}_1 + a_1(\tau)\dot{\mu}_1 + a_0(\tau)\mu_1 = 0. \quad (4.8)$$

We say that this equation is oscillating on a given interval, if any solution $\mu_1(\tau)$ has an infinite number of zeroes on that interval. Equivalently the classical Sturm theory of second order ODEs implies that this equation is oscillating whenever any solution of (4.4) makes an infinite number of turns around the origin in the (μ_1, μ_{n+1}) plane. Recall that for $n = 1$ the Lagrange Grassmanian is nothing but a projective line \mathbb{P}^1 and that the Jacobi curve is just the line $\mathbb{R}(\mu_1, \mu_{n+1})$. Therefore Jacobi curves of oscillating equations have an infinite Maslov index.

For second order ODEs there exist various oscillation and non-oscillation criteria, but among them there is a particularly simple one called the Kneser criteria [49]. It states that a second order ODE of the form

$$\ddot{x} + a(s)x = 0 \quad (4.9)$$

is oscillating on $[p, \infty)$, for any $p > 0$ if

$$\lim_{s \rightarrow \infty} s^2 a(s) > \frac{1}{4}$$

and it is non-oscillating if

$$\lim_{s \rightarrow \infty} s^2 a(s) < \frac{1}{4}.$$

If the limit is exactly $1/4$, Kneser criteria gives us no information and we have to use a different criteria. In order to put the equation (4.8) into form (4.9), we simply make a change of the time variable $s = 1/\tau$ and then make a change of the dependent variable

$$x(s) = \mu(s) \exp \left(\int_p^s \frac{2\theta - a_1(\theta)}{2\theta^2} d\theta \right).$$

Then we obtain exactly equation (4.9).

After applying the Kneser criteria, we find that for any sufficiently small interval $[0, \varepsilon]$

- equation (4.8) is oscillating if $m = 2$ and $4\sigma(X(0), \dot{X}(0)) + b_2 > 0$ or if $m > 2$ and $\sigma(X(0), \dot{X}(0)) > 0$;
- equation (4.8) is non-oscillating if $1 \leq m < 2$ or $m = 2$ and $4\sigma(X(0), \dot{X}(0)) + b_2 < 0$, or if $m > 2$ and $\sigma(X(0), \dot{X}(0)) < 0$.

Thus we conclude that in the first case the Jacobi curve has no right limit and therefore it also has an infinite Maslov index.

Since we just want to give an idea of how the proof works, we assume that the Jacobi DE (4.4) is of the simplest form

$$\frac{d}{d\tau} \begin{pmatrix} \mu_1 \\ \mu_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{\tau^2} \\ C & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_{n+1} \end{pmatrix}, \quad (4.10)$$

where C is constant. We then make a time-dependent change of variables

$$\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \tau^{1/2} & 0 \\ 0 & \tau^{-1/2} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_{n+1} \end{pmatrix}$$

and obtain

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = \frac{1}{\tau} \begin{pmatrix} \frac{1}{2} & 1 \\ C & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -\tau^{-1} JH \begin{pmatrix} p \\ q \end{pmatrix} \quad (4.11)$$

Note that this change of variables can not change whether or not the Maslov index is finite. It is clear that τ^{-1} multiplier just scales the speed along solutions, but does not change the trajectories. We could have actually got rid off it using a change of time variable. This means that if we drop τ^{-1} in the equation above, the overall phase-portrait does not change. It will be completely determined by the structure of the matrix $-JH$.

A description of various phase portraits on the Lagrangian Grassmanian was given in [46]. We will use the results from that article to work out the general case. For example, for $n = 1$ we can only have fixed points or periodic trajectories, which depend on the eigenvalues and eigenvectors of the matrix $-JH$. Its eigenvalues are

$$\lambda_{1,2} = \pm \frac{\sqrt{1+4C}}{2}.$$

If $1+4C < 0$, then we have only a single closed trajectory and no equilibrium points. Thus the trajectory rotates on \mathbb{P}^1 , and because of the τ^{-1} multiplier in (4.11) the curve rotates faster and faster as we get closer to $\tau = 0$ and therefore we get an infinite Maslov index. If $1+4C > 0$, then we have two equilibrium points: a stable and a non-stable one, that are given by two lines spanned by the eigenvectors of $-JH$. Thus all solutions except the equilibrium ones tend to the unstable equilibrium as $\tau \rightarrow 0$ and to the stable one as $\tau \rightarrow \infty$. In this case the Maslov index is finite. Note that in our example $C = \sigma(X(0), \dot{X}(0))/b_2$ and thus we recover the classical Kneser criteria. From here we can also see very well why the case $1+4C = 0$ is excluded. It is not stable under small perturbations and corresponds to a resonant situation when the two equilibrium points merge.

Having a small dimensional example allows us to actually draw the extend-phase portrait. We introduce new variables U and V defined as $\mu_1 = U\mu_{n+1}$, $p = Vq$. It is clear from the definitions that $V = \tau U^{-1}$. Since we work in a coordinate chart of the Grassmanian we can assume that $U, V \neq 0$. We differentiate expressions in the definition to obtain a couple of related Riccati equations

$$\begin{aligned} \dot{U} &= C - \frac{U^2}{\tau^2}, \\ \dot{V} &= \frac{1+V-CV^2}{\tau}. \end{aligned}$$

In the picture 4.1 the non-oscillating extended phase portrait before and after the change of variables is depicted. We can see clearly, that the extended phase portrait is separated by two separatrix into three regions. After the blow up the two separatrix solutions that have different initial values are mapped to equilibrium solutions. The stable solution after a blow-up can be described using an initial value problem, where as the unstable one can not. There is an infinite number of solutions that start from the unstable equilibrium.

We claim that the stable separatrix is the Jacobi curve for $\tau > 0$. To see this we do as discussed in the previous subsection. Assume that \mathcal{L}_0 is given by $U_0 \neq 0$. Then the Jacobi

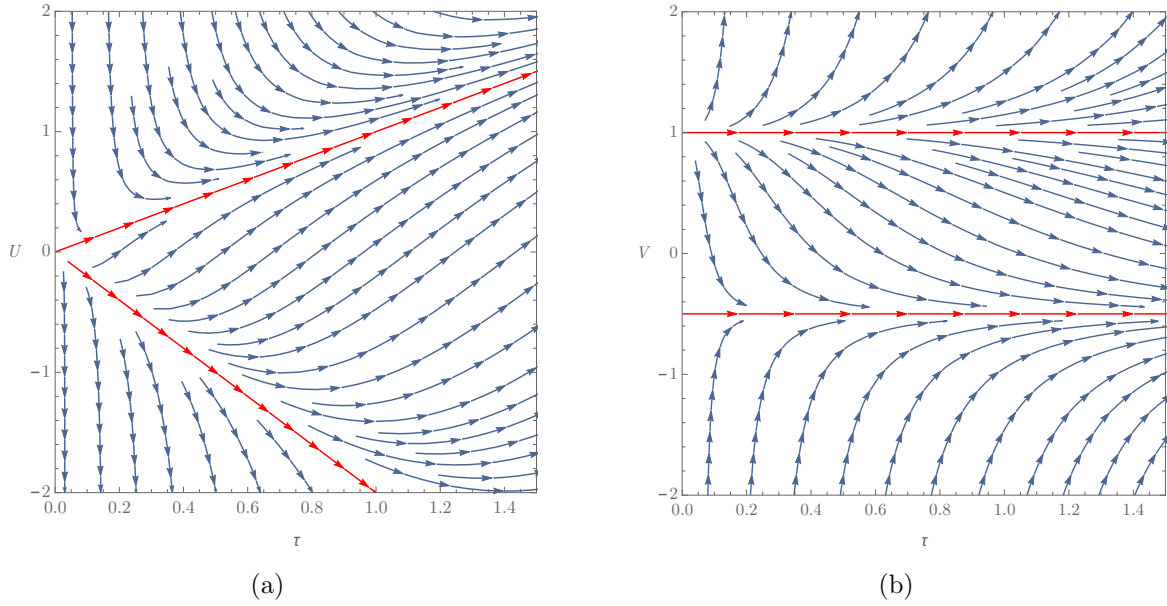


Figure 4.1: Phase portrait of a non-oscillating system: (a) before blow-up; (b) after blow-up.

curve is the limit of solutions of the Riccati equations with $U(\varepsilon) = U_0$. Similarly after a blow up it corresponds to a limit of solutions with boundary conditions $V(\varepsilon) = \varepsilon U_0^{-1}$. On the picture 4.2 we can see this convergence numerically in the original phase portrait.

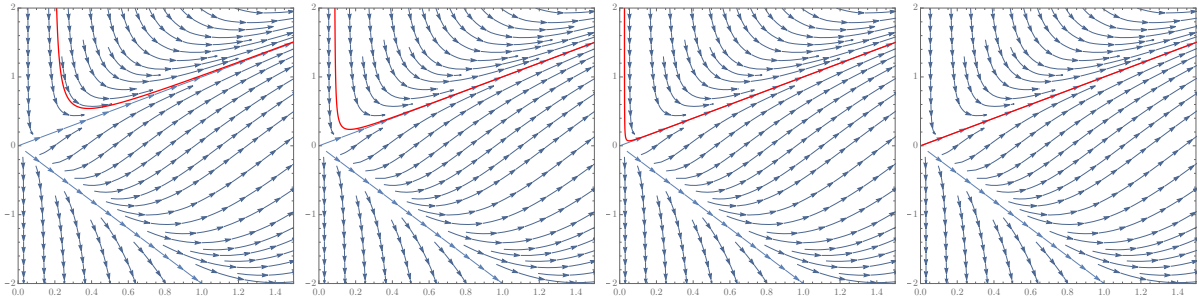


Figure 4.2: Point convergence of $\Lambda(\tau)$ to the Jacobi curve for $C = 2$.

In the next sections our goal is to make all the ideas from this section rigorous. Our proves are mostly perturbative and we will first prove them for the constant matrix case and then expand it to the general case. Following the outline of this section we first prove an analogue the Kneser criteria and identify the non-oscillating cases. Such a criteria is a necessary optimality condition on its own. Then using the general theory of ODEs with singular regular points we are going to characterize the jump of the Jacobi curve. Finally using the results from article [43] we are going to characterize the first derivative of the Jacobi curve for $m = 1, 2$ and prove that conditions on the first derivative guarantee uniqueness of the solution of the extended Jacobi equation that characterizes \mathcal{L}_t .

4.3 Kneser oscillation criteria

In the case $n = 1$, the Kneser criteria gives sufficient conditions under which a second order ODE is oscillating or non-oscillating. This result is just a consequence of the Sturm comparison theorem and an explicit solution of an Euler-type linear equation [49].

Kneser criteria can be derived as a consequence of more general integral criteria and the modern theory of oscillation of ODE systems tends to generalize those. As a result, we were not able to find in the literature a similar simple point-criteria. Thus we would like to slightly generalize the Kneser criteria to a special class of Hamiltonian systems that include system (4.7). Our main tool will be a straightforward consequence of Theorem 1 in [15, 6], that can be seen as a generalization of the Sturm comparison theorem

Theorem 4.3. *Let I be an open interval and $A_i(\tau), B_i(\tau), C_i(\tau)$, $i = 1, 2$ quadratic matrices whose elements are differentiable on I . Assume that $B_i(t)$ and $C_i(t)$ are symmetric. We consider the corresponding Hamiltonians*

$$H_i = \begin{pmatrix} C_i(\tau) & -A_i^T(\tau) \\ -A_i(\tau) & -B_i(\tau) \end{pmatrix}$$

If the Hamiltonians $H_i(\tau)$ satisfy

$$H_2(\tau) \geq H_1(\tau)$$

then for any two trajectories $\Lambda_i(t)$ whose endpoints are transversal to Λ , we have the following inequality

$$\text{Mi}_\Lambda \Lambda_1(t) - n \leq \text{Mi}_\Lambda \Lambda_2(t).$$

We are going to use a simple direct corollary of that result

Corollary 4.4. *Let I be an open interval and $A_i(\tau), B_i(\tau), C_i(\tau)$, $i = 1, 2$ quadratic matrices whose elements are differentiable on any compact subset $[a, b] \subset I$. Assume that $B_i(t)$ and $C_i(t)$ are symmetric and denote by $H_i(\tau)$ the corresponding Hamiltonians, s.t. $H_2(\tau) \geq H_1(\tau)$. Then we have the following implications:*

1. *If $\exists \Lambda \in L(\mathbb{R}^{2n})$, s.t. $H_2(\tau)|_\Lambda \leq 0$ for all $\tau \in I$ and the second system is oscillating, then the first system is oscillating as well;*
2. *If $\exists \Lambda \in L(\mathbb{R}^{2n})$, s.t. $H_1(\tau)|_\Lambda \geq 0$ for all $\tau \in I$ and the first system is oscillating, then the second system is oscillating as well;*
3. *If $\exists \Lambda \in L(\mathbb{R}^{2n})$, s.t. $H_i(\tau)|_\Lambda \leq 0$ for all $\tau \in I$ and the first system is non-oscillating, then the second system is non-oscillating as well;*
4. *If $\exists \Lambda \in L(\mathbb{R}^{2n})$, s.t. $H_i(\tau)|_\Lambda \geq 0$ for all $\tau \in I$ and the second system is non-oscillating, then the first system is oscillating as well;*

Proof. The proof is just a corollary of Theorem 4.3 and Theorem 1.12. For example, in the first case Theorem 1.12 implies that

$$\text{Mi}_\Lambda \Lambda_2(\tau) \leq 0.$$

The assumption that the corresponding system is oscillating means that the Maslov index of any solution is infinite. Then from the comparison Theorem 4.3 we obtain

$$\text{Mi}_\Lambda \Lambda_1(\tau) \leq \text{Mi}_\Lambda \Lambda_2(\tau) + n = -\infty.$$

The remaining implications are proven in the same way. \square

The goal of this section is to prove the following result

Theorem 4.5. *Consider a Hamiltonian system*

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} A(\tau) & \frac{B(\tau)}{\tau^m} \\ C(\tau) & -A^T(\tau) \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix},$$

s.t. the following assumptions are satisfied

1. $B(\tau)$ is a semi-definite smooth symmetric $n \times n$ -matrix, s.t. $B(\tau)$ is sign-definite for $\tau > 0$;
2. $C(\tau)$ is a smooth symmetric $n \times n$ -matrix;
3. $A(\tau)$ an arbitrary smooth $n \times n$ -matrix;

Then the following statements are true:

1. Let $m = 2$. If all the eigenvalues of the matrix $C(0)B(0)$ are strictly greater than $-1/4$, then the system is non-oscillating on $(0, \varepsilon)$. If at least one eigenvalue is smaller than $-1/4$, then the system is oscillating on the same interval;
2. If $0 \leq m < 2$, then the system is not oscillating on $(0, \varepsilon)$;
3. Let $m > 2$. If $C(0)$ is sign definite on the eigenspace of $B(0)$ that is transversal to the kernel of $B(0)$ with the same sign as $B(0)$, then the system is non-oscillating on $(0, \varepsilon)$. If $C(0)$ is not semi-definite on this subspace with the same sign as $B(0)$, then the system is oscillating on the same interval.

We are pretty sure that this theorem can be derived as a consequence of some existing oscillation criteria for Hamiltonian systems, but we prefer to give here a simple geometric proof of the result using Theorem 4.3, that seems to be new.

As it can be seen from the statement the matrix A plays no essential role. We make a time-dependent change of variables

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} \Phi & 0 \\ 0 & (\Phi^{-1})^T \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}$$

where $\Phi(t)$ satisfies

$$\begin{cases} \dot{\Phi} = A\Phi, \\ \Phi(0) = \text{id}_n. \end{cases}$$

Then $\Phi(t)$ is the fundamental matrix of the corresponding linear equation and it is smooth. Therefore our change of variables is a non-degenerate symplectic change of variables and it does not change oscillatory properties of the Hamiltonian systems.

Our Hamiltonian system now takes the form

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & \frac{\Phi^{-1}B(\Phi^{-1})^T}{\tau^m} \\ \Phi^T C \Phi & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix},$$

We can now simply redefine matrices B and C . This implies that without any loss of generality, we can assume that $A(\tau) \equiv 0$.

In order to apply the comparison Theorem 4.3, we need a model example, which oscillating properties we understand very well. Such model is given in the next lemma

Lemma 4.6. *Consider a Hamiltonian system of the form*

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & \frac{B}{\tau^2} \\ C & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \quad (4.12)$$

where B and C are constant symmetric matrices. This Hamiltonian system is oscillating on an interval $(0, \varepsilon)$ if and only if there exists at least one real eigenvalue λ of the matrix BC , s.t. $\lambda < -1/4$.

This result is a consequence of the following theorem proven in [7].

Theorem 4.7. *A linear autonomous Hamiltonian system*

$$\frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = -JH \begin{pmatrix} p \\ q \end{pmatrix},$$

is oscillating on an unbounded interval if and only if the matrix $-JH$ has a purely imaginary eigenvalue.

Proof of the Lemma 4.6. We do another symplectic transformation of the form

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto \begin{pmatrix} \tau^{-1/2} & 0 \\ 0 & \tau^{1/2} \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}$$

The transformation is smooth for $\tau > 0$ and therefore oscillating property is preserved. Our Hamiltonian system then becomes

$$\tau \frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \text{id}_n & B \\ C & -\frac{1}{2} \text{id}_n \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix} = -JH \begin{pmatrix} p \\ q \end{pmatrix}$$

Let us perform a change of time variable

$$s = \ln \tau.$$

Then we obtain a linear autonomous Hamiltonian system of the form

$$\frac{d}{ds} \begin{pmatrix} p \\ q \end{pmatrix} = -JH \begin{pmatrix} p \\ q \end{pmatrix}$$

Note that the change of time that we have made, maps the bounded interval $(0, \varepsilon)$ to an unbounded one. So by the Theorem 4.7 it just remains to compute the eigenvalues of the matrix $-JH$, i.e. to solve

$$\det(-JH - \lambda \text{id}_{2n}) = 0.$$

In this case the diagonal blocks are multipliers of the identity and hence commute with all the other blocks. Under this assumption it is easy to show that

$$\det(-JH - \lambda \text{id}_{2n}) = \det \left(\left(\lambda^2 - \frac{1}{4} \right) - BC \right).$$

If a matrix $-JH$ has a pair of purely complex eigenvalues $\lambda = \pm ib$, we obtain that the matrix BC has an eigenvalue $-b^2 - 1/4 < -1/4$. It is obvious that the converse holds as well. So the result follows from Theorem 4.7. \square

Finally we need the following fact proven in [30]:

Theorem 4.8. *Let B, C be two constant symmetric matrices, s.t. one of them is semidefinite. Then the spectrum of BC is real.*

Proof of the Theorem 4.5. We assume that $B(\tau) < 0$ for sufficiently small $\tau > 0$. The case $B(\tau) > 0$ is proven in a similar way. In this case the corresponding Hamiltonian H is positive semidefinite on the horizontal plane Ξ (the q -plane).

1) Let us start with the case $m = 2$. We know by the previous theorem that all the eigenvalues of $B(0)C(0)$ are real, and let us assume first that the minimum one is strictly less than $-1/4$. We define

$$\begin{aligned} B_1(\tau) &= B(0) + \varepsilon \text{id}_n, & B_2(\tau) &= B(\tau), \\ C_1(\tau) &= C(0) - \varepsilon \text{id}_n, & C_2(\tau) &= C(\tau). \end{aligned}$$

By assumption of the theorem and Lemma 4.6 the Hamiltonian system (4.12) with $B = B(0)$ and $C = C(0)$ is oscillating. This implies that a system of the form with $B = B(0) + \varepsilon \text{id}_n$ and $C = C(0) - \varepsilon \text{id}_n$ must be oscillating as well. Indeed, eigenvalues of BC are solutions of the characteristic equation whose coefficients depend continuously on the coefficients of matrices B, C . Therefore a small perturbation of matrices produces a small change in the eigenvalues of BC . But we have chosen such a perturbation in a way, that according to Theorem 4.8, the spectrum of BC remains real. So all the eigenvalues shift on the real axis, and if we choose $\varepsilon > 0$ small enough, the minimum eigenvalue of the perturbed matrix will stay strictly smaller than $-1/4$ and the corresponding Hamiltonian system (4.12) stays oscillating by Lemma 4.6.

So we can use the implication 2) of Corollary 4.4. By smoothness assumption, indeed, for sufficiently small times $H_1(\tau) \leq H_2(\tau)$.

2) The non-oscillating case for $m = 2$ is proven using exactly the same argument and matrices

$$\begin{aligned} B_1(\tau) &= B(\tau), & B_2(\tau) &= B(0) - \varepsilon \text{id}_n, \\ C_1(\tau) &= C(\tau) & C_2(\tau) &= C(0) + \varepsilon \text{id}_n. \end{aligned}$$

3) The case $0 \leq m < 2$ is now just a consequence of what we have proven so far. Indeed, we can consider the Hamiltonian as having a singularity with $m = 2$ and with a new matrix $\hat{B}(\tau) = \tau^{2-m}B(\tau)$. Then $\hat{B}(0)C(0) = 0$ and all the eigenvalues are zero. Hence the system is not oscillating.

4) Let us now assume that $m > 2$. We apply a symplectic transform

$$\begin{pmatrix} R & 0 \\ 0 & R^T \end{pmatrix},$$

where $R \in \text{SO}(n)$. Then in the new coordinates $B(\tau)$ and $C(\tau)$ will be replaced by the same matrices conjugated with R . Let us choose R , s.t. in the new coordinates $B(0)$ is diagonalized.

We take

$$\begin{aligned} B_1(\tau) &= -k\tau^{m-2} \begin{pmatrix} \text{id}_l & 0 \\ 0 & 0 \end{pmatrix} & B_2(\tau) &= B(\tau), \\ C_1(\tau) &= C(\tau), & C_2(\tau) &= C(\tau), \end{aligned}$$

where k is some constant and $l = n - \dim \ker B(0)$. If $C(0)$ is not negative semi-definite on the eigenspace of $B(0)$ that corresponds to the non-zero eigenvalue, then by taking k large enough, we find that the system one is oscillating by Lemma 4.6. We also have from the assumptions that $B_1(\tau) - B_2(\tau)$ is non-negative for sufficiently small $\tau > 0$. Thus we can use the implication 2) of Corollary 4.4 to deduce that the second system is going to be oscillating as well.

If $C(0)$ is negative definite on the eigenspace of $B(0)$, we repeat the proof with exactly the same B_1, C_1 . In this case we know from what we have proven already, that the system one is not oscillating. Then the result follows from the implication 3) from Corollary 4.4 with $\Lambda = \Pi$ being the vertical plane (the p -plane). □

Let us apply the theorem to our case. We are not oscillating for $0 \leq m < 2$ and for $m = 2$ whenever the eigenvalues of the

$$B(0)C(0) = \begin{pmatrix} \frac{1}{b_2} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sigma(e_1(0), \dot{e}_1(0)) & 0 \\ 0 & \sigma(e_2(0), \dot{e}_2(0)) \end{pmatrix} = \begin{pmatrix} \sigma(X(0), \dot{X}(0))/b_2 & 0 \\ 0 & 0 \end{pmatrix}$$

are bigger than $-1/4$, i.e. whenever the only non-zero element above is bigger than $-1/4$. If $m > 2$ then we must have $\sigma(X(0), \dot{X}(0)) < 0$.

So from now on, we assume that our Hamiltonian system is non-oscillating, ensuring that the right limit \mathcal{L}_{0+} exists.

4.4 Model examples

We are now ready to make the first step and compute the jump of the Jacobi curve. We will give three similar but separate proofs for $m = 1$, $m = 2$ and $m \geq 3$. For $m = 1$ we will do this using the general theory of linear ODEs with regular singular points. For $m = 2$ and $m \geq 3$ the strategy of the proof is going to be very similar to the proof of the Kneser theorem in the previous section. Namely we first look at some model examples and then we use the comparison theory of Riccati equations, to obtain the result in the most general case.

4.4.1 Case $m < 2$

Let $\lambda = (p, q)$. We rewrite the system (4.7) in the following form

$$\dot{\lambda} = \left(\frac{H_{-1}}{\tau} + H(\tau) \right) \lambda, \quad (4.13)$$

where $H(\tau)$ is an analytic matrix function and as can be easily seen

$$H_{-1} = \begin{pmatrix} 0 & 0 & \frac{1}{b_1} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

This matrix up to a reordering of coordinates is in its Jordan normal form, and all of its eigenvalues are zero. Therefore by a well known theorem [22], the fundamental matrix $\Phi(\tau)$ of the system (4.13) can be written as

$$\Phi(\tau) = P(\tau)\tau^{H_{-1}},$$

where $P(\tau)$ is an analytic matrix function with $P(0) = \text{id}_{2n}$. A power series expansion can be obtained by plugging this solution into (4.13) and expanding all the analytic functions into their Taylor series. It is easy to check that

$$\tau^{H_{-1}} = \begin{pmatrix} 1 & 0 & \frac{\ln \tau}{b_1} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Let \mathcal{L}_0 be the \mathcal{L} -derivative at moment of time $\tau = 0$. The flow of the Hamiltonian system $\Phi(\tau)$ induces a flow on the Lagrangian Grassmanian $L(\mathbb{R}^4)$ that we denote using the same symbol. As we have discussed previously the Jacobi curve \mathcal{L}_τ is going to be a pointwise limit of the solutions of the Jacobi DE on $L(\mathbb{R}^4)$ with boundary conditions $\Lambda(\varepsilon) = \mathcal{L}_0$. Since we know explicitly the flow, we can write the solution of this boundary problem as

$$\mathcal{L}_\tau = \lim_{\varepsilon \rightarrow 0^+} \Phi(\tau)\Phi^{-1}(\varepsilon)\mathcal{L}_0. \quad (4.14)$$

We note that $\Phi(\tau)$ is smooth and invertible for $\tau > 0$. Therefore we can exchange the limit with $\Phi(\tau)$, and we just need to compute the limit of $\Phi^{-1}(\varepsilon)\mathcal{L}_0$. To do this we use a concrete representation of Lagrangian planes as span of a couple of vectors like in Section 1.1

$$\Phi^{-1}(\varepsilon)\mathcal{L}_0 = [\lambda_1(\varepsilon) \quad \lambda_2(\varepsilon)].$$

Let us find the limits of $\lambda_i(\varepsilon)$ as $\varepsilon \rightarrow 0$. We have

$$(\varepsilon^{H-1})^{-1} = \begin{pmatrix} 1 & 0 & -\frac{\ln \varepsilon}{b_1} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Assume that $X(0) \in \mathcal{L}_0$. Then as we have seen in the Example 1.7 we can assume

$$\mathcal{L}_0 = \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix} \quad (4.15)$$

Then since $P(0) = \text{id}_{2n}$, we obtain

$$\lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon)\mathcal{L}_0 = \lim_{\varepsilon \rightarrow 0^+} (\varepsilon^{H-1})^{-1}P^{-1}(\varepsilon) \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix},$$

because $(\varepsilon^{H-1})^{-1}$ acts as the identity on \mathcal{L}_0 . For the same reason

$$\lim_{\tau \rightarrow 0^+} \Lambda(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix}$$

and so $\mathcal{L}_{0^+} = \mathcal{L}_0$ and the Jacobi curve is actually continuous.

If $X(0) \notin \mathcal{L}_0$, then again from Example 1.7 we know, that we can take

$$\mathcal{L}_0 = [\lambda_1(\varepsilon) \quad \lambda_2(\varepsilon)] = \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \\ 1 & 0 \\ w_1 & w_2 \end{bmatrix} \quad (4.16)$$

Similarly to the previous case we find that

$$\lim_{\varepsilon \rightarrow 0^+} \lambda_2(\varepsilon) = \begin{pmatrix} x_2 \\ y_2 \\ 0 \\ z_2 \end{pmatrix}.$$

Let us see what happens to the limit of the first vector. We have

$$\lim_{\varepsilon \rightarrow 0^+} \lambda_1(\varepsilon) = \lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon) \begin{pmatrix} x_1 \\ y_1 \\ 1 \\ w_1 \end{pmatrix} = \lim_{\varepsilon \rightarrow 0^+} \begin{pmatrix} x_1 - \frac{\ln \varepsilon}{b_1} \\ y_1 \\ 1 \\ w_1 \end{pmatrix}$$

which is equal to infinity. As we have said before a representation of a Lagrangian plane as a span of two vectors is not unique. We can scale them as we want as we take the limit. So we take

$$\lim_{\varepsilon \rightarrow 0^+} -\frac{b_1}{\ln \varepsilon} \lambda_1(\varepsilon) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

So

$$\lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon) \mathcal{L}_0 = \begin{bmatrix} 1 & x_2 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix}.$$

Then as before we find that

$$\lim_{\tau \rightarrow 0^+} \Lambda(\tau) = \begin{bmatrix} 1 & x_2 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & y_2 \\ 0 & 0 \\ 0 & w_2 \end{bmatrix}.$$

So summarizing everything we have done in a more invariant manner the jump can be computed as follows. Given \mathcal{L}_0 , the new \mathcal{L} -derivative \mathcal{L}_{0^+} is going to be a direct sum of $\mathcal{L}_0 \cap X(0)^\perp$ and $X(0)$. But this is by definition $\mathcal{L}_0^{X(0)}$. The goal of the following subsections is to prove the same for $m \geq 2$.

4.4.2 Case $m = 2$

For $m \geq 2$ we proceed in a different way. One can reduce by a change of variables the system (4.7) to a system with a regular singular point at $\tau = 0$. Thus all the methods from the theory of linear systems of ODE's can be used. But these techniques work well under some non-resonance conditions. In our case we can use techniques from Hamiltonian dynamics to arrive at the results even in the presence of resonances. First we prove the result for some model problems similarly as we have done in the case $m = 1$. Then we apply Riccati comparison theorems, to prove the general result.

For $m = 2$ we choose the following Hamiltonian systems as our models

$$\frac{d}{dt} \begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{b_{11}}{\tau^2} & 0 \\ 0 & 0 & 0 & b_{22} \\ c_{11} & 0 & 0 & 0 \\ 0 & c_{22} & 0 & 0 \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}, \quad (4.17)$$

where b_{ii}, c_{ii} are constants, $b_{11} \neq 0$ (or else there is no singularity) and $c_{11} \neq 0$ (because we have $\sigma(X(0), \dot{X}(0)) \neq 0$). First of all we notice that this system splits into two invariant sub-systems

$$\frac{d}{dt} \begin{pmatrix} p_1 \\ q_1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{b_{11}}{\tau^2} \\ c_{11} & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}, \quad (4.18)$$

$$\frac{d}{dt} \begin{pmatrix} p_2 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0 & b_{22} \\ c_{22} & 0 \end{pmatrix} \begin{pmatrix} p_2 \\ q_2 \end{pmatrix}. \quad (4.19)$$

We denote by $\Phi_i(t)$ the corresponding fundamental matrices. Without any loss of generality, we can assume that $\Phi_2(0) = \text{id}_2$. In order to find Φ_1 , we do a symplectic change of variables

$$\begin{pmatrix} \tilde{p}_1 \\ \tilde{q}_1 \end{pmatrix} = \begin{pmatrix} \tau^{1/2} & 0 \\ 0 & \tau^{-1/2} \end{pmatrix} \begin{pmatrix} p_1 \\ q_1 \end{pmatrix}.$$

Then the first system is transformed to

$$\frac{d}{dt} \begin{pmatrix} \tilde{p}_1 \\ \tilde{q}_1 \end{pmatrix} = \frac{1}{\tau} \begin{pmatrix} 1/2 & b_{11} \\ c_{11} & -1/2 \end{pmatrix} \begin{pmatrix} \tilde{p}_1 \\ \tilde{q}_1 \end{pmatrix} = \frac{Y}{\tau} \begin{pmatrix} \tilde{p}_1 \\ \tilde{q}_1 \end{pmatrix}$$

which is a simple linear system with a regular singular point. Therefore the fundamental solution Φ_1 is the following matrix function

$$\Phi_1(\tau) = \begin{pmatrix} \tau^{-1/2} & 0 \\ 0 & \tau^{1/2} \end{pmatrix} \tau^Y$$

or in a more detailed form

$$\Phi_1(\tau) = \begin{pmatrix} \frac{\tau^{-\frac{1-\Delta}{2}}(-1 + \Delta + (1 + \Delta)\tau^\Delta)}{2\Delta} & \frac{b_{11}\tau^{-\frac{1-\Delta}{2}}(-1 + \tau^\Delta)}{\Delta} \\ \frac{c_{11}\tau^{\frac{1-\Delta}{2}}(-1 + \tau^\Delta)}{\Delta} & \frac{\tau^{\frac{1-\Delta}{2}}(1 + \Delta + (-1 + \Delta)\tau^\Delta)}{2\Delta} \end{pmatrix},$$

where $\Delta = \sqrt{1 + 4b_{11}c_{11}}$. Under the non-oscillation assumption we have $\Delta > 0$. It is easy to check that $\det \Phi_1(\tau) = \det \Phi_2(\tau) \equiv 1$, so the inverse matrix of $\Phi(\tau)$ can be computed easily.

The Jacobi curve is given by (4.14) and as in the previous subsection the fundamental matrix $\Phi(\tau)$ is smooth for $\tau > 0$, so we can exchange it with the limit. So first of all we need to find the limit

$$\lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon) \mathcal{L}_0.$$

As in the previous case we are going to consider two situations: when $X(0) \in \mathcal{L}_0$ and when $X(0) \notin \mathcal{L}_0$.

As before, if $X(0) \in \mathcal{L}_0$ we can assume that \mathcal{L}_0 is given by (4.15). Then the first and the second vector lie in its own invariant subspace. For example since $\Phi_2(0) = \text{id}_2$, we

immediately get that

$$\lim_{\tau \rightarrow 0^+} \Phi(\tau) \lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon) \begin{pmatrix} 0 \\ y_2 \\ 0 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ y_2 \\ 0 \\ w_2 \end{pmatrix}.$$

Let us see what happens to the first vector. We have

$$\Phi^{-1}(\varepsilon) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon^{\frac{1-\Delta}{2}} (1 + \Delta + (-1 + \Delta)\varepsilon^\Delta)}{2\Delta} \\ 0 \\ -\frac{c_{11}\varepsilon^{\frac{1-\Delta}{2}} (-1 + \varepsilon^\Delta)}{\Delta} \\ 0 \end{pmatrix}.$$

Then we find that

$$\lim_{\varepsilon \rightarrow 0^+} \left[\Phi^{-1}(\varepsilon) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right] = \lim_{\varepsilon \rightarrow 0^+} \left[\varepsilon^{-\frac{1-\Delta}{2}} \Phi^{-1}(\varepsilon) \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \right] = \begin{bmatrix} \frac{1+\Delta}{2\Delta} \\ 0 \\ \frac{c_{11}}{\Delta} \\ 0 \end{bmatrix} \in \lim_{\varepsilon \rightarrow 0^+} \Phi^{-1}(\varepsilon) \mathcal{L}_0$$

and

$$\lim_{\tau \rightarrow 0^+} \left[\tau^{\frac{1+\Delta}{2}} \Phi(\tau) \begin{pmatrix} \frac{1+\Delta}{2\Delta} \\ 0 \\ \frac{c_{11}}{\Delta} \\ 0 \end{pmatrix} \right] = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad (4.20)$$

which means that the limit is up to a constant the vector $X(0)$. Thus in this case $\mathcal{L}_{0^+} = \mathcal{L}_0$ and the Jacobi curve is actually continuous as expected.

We now look at the situation when $X(0) \notin \mathcal{L}_0$. Then \mathcal{L}_0 can be assumed to be of the form (4.16). We consider the vectors $\lambda_i(\varepsilon)$ and their projection onto the first invariant subspace

$$n_1(\varepsilon) = \Phi_1^{-1}(\varepsilon) \begin{pmatrix} x_1 \\ 1 \end{pmatrix}, \quad n_2(\varepsilon) = \Phi_1^{-1}(\varepsilon) \begin{pmatrix} x_2 \\ 0 \end{pmatrix}. \quad (4.21)$$

We have

$$n_1(\varepsilon) = \varepsilon^{-\frac{1+\Delta}{2}} \begin{pmatrix} \frac{-2b_{11}(-1 + \varepsilon^\Delta) + \varepsilon(1 + \Delta + (-1 + \Delta)\varepsilon^\Delta)x_1}{2\Delta} \\ \frac{-1 + \Delta + 2c_{11}\varepsilon x_1 + \varepsilon^\Delta(1 - 2c_{11}\varepsilon x_1 + \Delta)}{2\Delta} \end{pmatrix}$$

So it is clear that

$$\lim_{\varepsilon \rightarrow 0^+} \varepsilon^{\frac{1+\Delta}{2}} \lambda_1(\varepsilon) = \begin{pmatrix} \frac{b_{11}}{\Delta} \\ 0 \\ -1 + \Delta \\ \frac{2\Delta}{0} \end{pmatrix}.$$

Then the formula (4.20) proves that $X(0) \in \mathcal{L}_{0+}$. Now we need to find an independent from $X(0)$ limit vector that would lie in \mathcal{L}_{0+} .

Vector $n_2(\varepsilon)$ has the following form

$$n_2(\varepsilon) = \varepsilon^{\frac{1-\Delta}{2}} \begin{pmatrix} \frac{(1 + \Delta + \varepsilon^\Delta(-1 + \Delta))x_2}{2\Delta} \\ \frac{c_{11}(1 - \varepsilon^\Delta)x_2}{\Delta} \end{pmatrix}.$$

If $0 < \Delta < 1$ or $x_2 = 0$, then it is clear that

$$\lim_{\varepsilon \rightarrow 0+} \lambda_2(\varepsilon) = \begin{pmatrix} 0 \\ y_2 \\ 0 \\ w_2 \end{pmatrix},$$

where we have used that $\Phi_2(0) = \text{id}_2$. For the same reason the very same vector is going to lie in \mathcal{L}_{0+} and the result follows.

If $\Delta = 1$, then either $b_{11} = 0$ or $c_{11} = 0$. Since we have excluded these possibilities it only remains to see what happens, when $\Delta > 1$ and $x_2 \neq 0$.

We can see that the expressions for $\lambda_i(\varepsilon)$ are just sums of power series of ε . Therefore it is convenient to introduce the following notation

$$a(\varepsilon) = b(\varepsilon) \pmod{\varepsilon^{>0}}$$

which means that $a(\varepsilon)$ and $b(\varepsilon)$ agree modulo terms of positive degree of ε . Then $\mathcal{L}_{0+} = \mathcal{L}_0^{X(0)}$ follows from the following lemma

Lemma 4.9. *Let $\Delta > 1$ and $x_2 \neq 0$. Then there exist constants c_0, c_1, \dots, c_{l-1} , s.t.*

$$\lambda_2(\varepsilon) - x_2 \varepsilon \lambda_1(\varepsilon) \sum_{i=0}^{l-1} c_i \varepsilon^i = \begin{pmatrix} k_1 \varepsilon^{\frac{2l+1-\Delta}{2}} \\ y_2 \\ k_2 \varepsilon^{\frac{2l+1-\Delta}{2}} \\ w_2 \end{pmatrix} \pmod{\varepsilon^{>0}},$$

where k_i are some constants.

Indeed, the vector on the left hand side is a linear span of $\lambda_1(\varepsilon)$ and $\lambda_2(\varepsilon)$. We can choose l sufficiently big so that $2l + 1 - \Delta > 0$. Then in the limit we obtain a vector $(0 \ y_2 \ 0 \ w_2)^T$, which lies in the second invariant subspace where there is no singularity at all.

Proof of the lemma. We denote by $\lambda(\varepsilon)$ the vector on the left. It is easy to see why the second and fourth components of $\lambda(\varepsilon)$ have this form. It follows from the fact that $\Phi_2(\varepsilon)$ is an analytic matrix function with $\Phi_2(0) = \text{id}_2$.

So it is enough to look on the projection of $\lambda(\varepsilon)$ to the singular invariant subspace. We can write

$$\begin{aligned} n_1(\varepsilon) &= \frac{\varepsilon^{-\frac{1+\Delta}{2}}}{\Delta} \begin{pmatrix} b_{11} \\ -\frac{1+\Delta}{2} \end{pmatrix} + \frac{\varepsilon^{\frac{1-\Delta}{2}} x_1}{\Delta} \begin{pmatrix} \frac{1+\Delta}{2} \\ c_{11} \end{pmatrix} \quad \text{mod } \varepsilon^{>0}, \\ n_2(\varepsilon) &= \frac{\varepsilon^{\frac{1-\Delta}{2}} x_2}{\Delta} \begin{pmatrix} \frac{1+\Delta}{2} \\ c_{11} \end{pmatrix} \quad \text{mod } \varepsilon^{>0}. \end{aligned}$$

Let us denote

$$\alpha(\varepsilon) = \frac{\varepsilon^{\frac{1-\Delta}{2}}}{\Delta} \begin{pmatrix} b_{11} \\ -\frac{1+\Delta}{2} \end{pmatrix}$$

From here we see that

$$x_2 \varepsilon n_1(\varepsilon) = x_2 \alpha(\varepsilon) + x_1 \varepsilon n_2(\varepsilon) \quad \text{mod } \varepsilon^{>0}.$$

We then find an expression for the projection of $\lambda(\varepsilon)$:

$$\begin{aligned} n_2(\varepsilon) - x_2 \varepsilon n_1(\varepsilon) &= \sum_{i=0}^{l-1} c_i \varepsilon^i = \\ &= n_2(\varepsilon) - x_2 \alpha(\varepsilon) c_0 - \sum_{i=1}^{l-1} (x_2 \alpha(\varepsilon) c_i + x_1 n_2(\varepsilon) c_{i-1}) \varepsilon^i - x_1 n_2(\varepsilon) c_{l-1} \varepsilon^l \quad \text{mod } \varepsilon^{>0}. \end{aligned}$$

So it is enough to choose c_i to be s.t. they solve

$$\begin{aligned} n_2(\varepsilon) - x_2 \alpha(\varepsilon) c_0 &= 0 \quad \text{mod } \varepsilon^{>0}, \\ x_2 \alpha(\varepsilon) c_{i+1} + x_1 n_2(\varepsilon) c_i &= 0 \quad \text{mod } \varepsilon^{>0}. \end{aligned}$$

The first equality is satisfied, if

$$c_0 = \frac{1 + \Delta}{2b_{11}}, \quad (\text{recall that } \Delta = \sqrt{1 + 4b_{11}c_{11}}).$$

But then we can obtain an expression for $n_2(\varepsilon)$ from the first equation and plug it into the second one. We get

$$\alpha(\varepsilon) x_2 (c_{i+1} + x_1 c_0 c_i) = 0.$$

So we simply choose recursively $c_{i+1} = -x_1 c_i c_0$. □

4.4.3 Case $m > 2$

We now consider the same model as (4.17) but with singularity of order $m > 2$. Recall that B_{ii}, C_{ii} are constants and b_{11}, c_{11} are non zero. It is convenient to define $m = 2 + \beta$. Again we have two invariant subsystems and equation (4.18) has the form

$$\begin{aligned} \dot{p}_1 &= \frac{b_{11}}{\tau^{2+\beta}} q_1, \\ \dot{q}_1 &= c_{11} p_1. \end{aligned}$$

We differentiate the second equation to obtain

$$\ddot{q}_1 - \frac{b_{11}c_{11}}{\tau^{2+\beta}}q_1 = 0.$$

If we introduce a new independent variable

$$y(\tau) = \frac{q_1(\tau)}{\sqrt{\tau}}$$

and a new dependent variable

$$s(\tau) = \frac{2\sqrt{b_{11}c_{11}}}{\beta}\tau^{-\frac{\beta}{2}},$$

we obtain a modified Bessel equation

$$s^2 \frac{d^2 y}{ds^2} + s \frac{dy}{ds} - \left(s^2 + \frac{1}{\beta^2} \right) y = 0.$$

Two independent solutions of this equation are given by the two modified Bessel functions $I_{\beta-1}(s)$, $K_{\beta-1}(s)$ [41]. Therefore the fundamental matrix $\Phi_1(\tau)$ is given by

$$\Phi_1(\tau) = \begin{pmatrix} \frac{1}{c_{11}} \frac{d}{d\tau} \sqrt{\tau} I_{\beta-1}(s(\tau)) & \frac{1}{c_{11}} \frac{d}{d\tau} \sqrt{\tau} K_{\beta-1}(s(\tau)) \\ \sqrt{\tau} I_{\beta-1}(s(\tau)) & \sqrt{\tau} K_{\beta-1}(s(\tau)) \end{pmatrix}$$

We can simplify considerably the first row using the following formulas for the derivatives of modified Bessel functions [41]

$$\begin{aligned} I'_a(x) &= \frac{a}{x} I_a(x) + I_{a+1}(x), \\ K'_a(x) &= \frac{a}{x} K_a(x) - K_{a+1}(x). \end{aligned}$$

After some simplifications we find that

$$\Phi_1(\tau) = \begin{pmatrix} -\sqrt{\frac{b_{11}}{c_{11}}} \tau^{-\frac{1+\beta}{2}} I_{\beta-1+1}(s(\tau)) & \sqrt{\frac{b_{11}}{c_{11}}} \tau^{-\frac{1+\beta}{2}} K_{\beta-1+1}(s(\tau)) \\ \sqrt{\tau} I_{\beta-1}(s(\tau)) & \sqrt{\tau} K_{\beta-1}(s(\tau)) \end{pmatrix}.$$

Since $\beta > 0$, as $\tau \rightarrow 0+$ we get $s(\tau) \rightarrow +\infty$. Therefore we need an asymptotic expansion of modified Bessel functions as the argument goes to $+\infty$. They are given by [41]

$$\begin{aligned} I_a(x) &\sim \sqrt{\frac{1}{2\pi x}} e^x, & x \rightarrow +\infty, \\ K_a(x) &\sim \sqrt{\frac{\pi}{2x}} e^{-x}, & x \rightarrow +\infty. \end{aligned}$$

In particular we see that the limit does not depend on the parameter, and therefore for any real a, b we have

$$\begin{aligned} I_a(x) &\rightarrow +\infty, & \frac{I_a(x)}{I_b(x)} &\rightarrow 1, \\ K_a(x) &\rightarrow 0, & \frac{K_a(x)}{K_b(x)} &\rightarrow 1, \end{aligned}$$

as $x \rightarrow +\infty$.

The matrix $\Phi_1(\tau)$ is invertible and smooth for $\tau > 0$. From the explicit form of the equation it follows that determinant of $\Phi_1(\tau)$ is constant. Using the asymptotics above we can find that it is actually equal to $-\beta/(2c_{11})$. The very same asymptotics and an argument similar to the one for $m = 1, 2$ implies that if $X(0) \in \mathcal{L}_0$, then $\mathcal{L}_{0+} = \mathcal{L}_0$. So we assume that $X(0) \notin \mathcal{L}_0$ and consequently that \mathcal{L}_0 is given by (4.16).

If $x_2 = 0$, then it is clear that

$$\lim_{\tau \rightarrow 0+} \Phi(\tau) \lim_{\varepsilon \rightarrow 0+} \Phi^{-1}(\varepsilon) \begin{pmatrix} 0 \\ y_2 \\ 0 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ y_2 \\ 0 \\ w_2 \end{pmatrix}.$$

So it remains to find a single independent vector in \mathcal{L}_{0+} in this case. Let us slightly abuse the notation and denote

$$\begin{pmatrix} x_1(\varepsilon) \\ y_1(\varepsilon) \\ z_1(\varepsilon) \\ w_1(\varepsilon) \end{pmatrix} = \Phi^{-1}(\varepsilon) \begin{pmatrix} x_1 \\ y_1 \\ 1 \\ w_1 \end{pmatrix}.$$

Using an explicit expression for the fundamental matrix, we find that $x_1(\varepsilon)$ is a linear combination of the modified Bessel K -functions and $z_1(\varepsilon)$ is a linear combination of I -functions. Due to the exponential behaviour of $I_a(x)$ and $K_a(x)$, we find that

$$\lim_{\varepsilon \rightarrow 0+} \frac{1}{z_1(\varepsilon)} \Phi^{-1}(\varepsilon) \begin{pmatrix} x_1 \\ y_1 \\ 1 \\ w_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

For the same reason

$$\lim_{\tau \rightarrow 0+} \frac{\sqrt{c_{11}} \tau^{\frac{1+\beta}{2}}}{\sqrt{b_{11}} K_{\beta-1+1}(s(\tau))} \Phi(\tau) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.22)$$

and we obtain that $\mathcal{L}_{0+} = \mathcal{L}_0^{X(0)}$.

Assume now that $x_2 \neq 0$. Then we obtain by the same argument as above

$$\lim_{\varepsilon \rightarrow 0^+} \frac{2c_{11}}{\beta x_2 \sqrt{\varepsilon} I_{\beta-1}(s(\varepsilon))} \Phi^{-1}(\varepsilon) \begin{pmatrix} x_2 \\ y_2 \\ 0 \\ w_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Exploiting once more the formula (4.22), we once again find that $X(0) \in \mathcal{L}_{0+}$. To find an independent vector limit let us write down explicitly the vectors $n_1(\varepsilon)$ and $n_2(\varepsilon)$ defined in (4.21) of the previous subsection. We have

$$n_1(\varepsilon) = \begin{pmatrix} x_1(\varepsilon) \\ z_1(\varepsilon) \end{pmatrix} = -\frac{2c_{11}}{\beta} \begin{pmatrix} -\frac{\sqrt{b_{11}}\varepsilon^{-\frac{1+\beta}{2}} K_{\beta-1+1}(s(\varepsilon))}{\sqrt{c_{11}}} + \sqrt{\varepsilon} x_1 K_{\beta-1}(s(\varepsilon)) \\ -\frac{\sqrt{b_{11}}\varepsilon^{-\frac{1+\beta}{2}} I_{\beta-1+1}(s(\varepsilon))}{\sqrt{c_{11}}} - \sqrt{\varepsilon} x_1 I_{\beta-1}(s(\varepsilon)) \end{pmatrix},$$

$$n_2(\varepsilon) = -\frac{2c_{11}}{\beta} \begin{pmatrix} \sqrt{\varepsilon} x_2 K_{\beta-1}(s(\varepsilon)) \\ -\sqrt{\varepsilon} x_2 I_{\beta-1}(s(\varepsilon)) \end{pmatrix}.$$

To find an independent limit vector above we consider

$$\lambda(\varepsilon) = \lambda_2(\varepsilon) + \frac{\sqrt{\varepsilon} x_2 I_{\beta-1}(s(\varepsilon))}{z_1(\varepsilon)} \lambda_1(\varepsilon).$$

The only component of $\lambda_1(\varepsilon)$ and $\lambda_2(\varepsilon)$ escaping to infinity are the z -components as can be easily seen from the explicit expression of $n_i(\varepsilon)$. But the z -component of $\lambda(\varepsilon)$ is equal to zero. Moreover the coefficient in front of $\lambda_1(\varepsilon)$ tends to zero as $\varepsilon \rightarrow 0^+$. Thus from the explicit expressions for $x_1(\varepsilon)$ and $x_2(\varepsilon)$ we obtain that

$$\lim_{\varepsilon \rightarrow 0^+} \lambda(\varepsilon) = \begin{pmatrix} 0 \\ y_2 \\ 0 \\ z_2 \end{pmatrix},$$

which is a vector that does not lie in the singular invariant subspace. Thus as in the previous sections the same vector lies in \mathcal{L}_{0+} which proves the result.

4.5 Jacobi curves in the general case for $m \leq 2$

In the previous subsections we have seen, that for the autonomous models the Jacobi curve has the right limit $\mathcal{L}_{0+} = \mathcal{L}_0^{X(0)}$. Now we are ready to prove this for a general system (4.7). We use the standard Riccati comparison result from [43].

Lemma 4.10. *Suppose that $B(\tau)$ and $C(\tau)$ are two symmetric continuous matrix functions that satisfy $B(\tau) \geq 0$ and $C(\tau) \geq 0$ for almost every τ of any closed subinterval*

$[a, b]$ of a given open interval I . Then given a symmetric matrix $S_a \geq 0$, any Cauchy solution of

$$\begin{aligned} \dot{S} + SA + A^T S + SBS - C &= 0, \\ S(a) &= S_a, \end{aligned} \quad (4.23)$$

satisfies $S(\tau) \geq 0$ for all $\tau \in [a, b]$.

We consider now the general system (4.7). Let $q = Sp$ and we write the corresponding Riccati equation like discussed in Section 1.1

$$\dot{S} + \frac{SB(\tau)S}{\tau^m} - C(\tau) = 0. \quad (4.24)$$

If the system is not oscillating, then we have existence of the Cauchy problem with the boundary data $S(t) = S$ on the interval $(0, t]$ for any fixed symmetric matrix S and for t small enough.

Assume that $B_1(\tau) \leq B(\tau) \leq B_2(\tau) \leq 0$ and $C_1(\tau) \geq C(\tau) \geq C_2(\tau)$ for small $\tau \in [0, t]$. We assume that B_i and C_i are diagonal matrices like in our models from the previous subsection satisfying the non-oscillation conditions. Then we can define S_i^ε to be solutions of the Cauchy problem

$$\dot{S} + \frac{SB_i(\tau)S}{\tau^m} - C_i(\tau) = 0, \quad S(\varepsilon) = S_{\mathcal{L}_0},$$

where $S_{\mathcal{L}_0}$ is a symmetric matrix that corresponds to \mathcal{L}_0 assuming of course that \mathcal{L}_0 is transversal to the horizontal plane $q = 0$. We denote by $S^\varepsilon(\tau)$ be a solution of (4.24) with $S(\varepsilon) = S_{\mathcal{L}_0}$.

Let us assume, for example $W^\varepsilon(\tau) = S_1^\varepsilon(\tau) - S_2^\varepsilon(\tau)$. Then we have that $W^\varepsilon(\tau)$ satisfies

$$\dot{W}^\varepsilon + W^\varepsilon \frac{B_1}{\tau^m} S_2^\varepsilon + S_2^\varepsilon \frac{B_1}{\tau^m} W^\varepsilon + W^\varepsilon \frac{B_1}{\tau^m} W^\varepsilon + \frac{S_2^\varepsilon(B_1 - B_2)S_2^\varepsilon}{\tau^m} - (C_1 - C_2) = 0$$

with $W_\varepsilon(\varepsilon) = 0$. But then by the Lemma 4.10 we obtain that

$$W^\varepsilon(\tau) \geq 0 \iff S_1^\varepsilon(\tau) \geq S_2^\varepsilon(\tau),$$

for any $\tau \geq \varepsilon$ as long as $S_2^\varepsilon(\tau)$ is defined.

By replacing $B_1(\tau)$ with $B(\tau)$ and then $B_2(\tau)$ with $B(\tau)$, we similarly obtain that

$$S_2^\varepsilon(\tau) \leq S^\varepsilon(\tau) \leq S_1^\varepsilon(\tau),$$

for any $\tau \geq \varepsilon$ sufficiently close to ε and $\varepsilon > 0$ small. By fixing τ sufficiently small and taking limits as $\varepsilon \rightarrow 0+$ we find that

$$S_1(\tau) \leq S(\tau) \leq S_2(\tau),$$

where these matrix functions are the corresponding Jacobi curves. But we have proven in the previous subsections that for our model examples we had the same right limit. Thus $S_1(0+) = S_2(0+)$ and

$$S(0+) = S_1(0+) = S_2(0+).$$

If \mathcal{L}_0 or $\mathcal{L}_0^{X(0)}$ are not transversal to the horizontal space Ξ , then this construction clearly does not work, because either $S^\varepsilon(\varepsilon)$ or $S(0+)$ do not exist. In this case we make a change of variables of the form

$$\begin{pmatrix} p \\ q \end{pmatrix} \mapsto M \begin{pmatrix} p \\ q \end{pmatrix}, \quad M = \begin{pmatrix} \alpha_1 & 0 & \beta_1 & 0 \\ 0 & \alpha_2 & 0 & \beta_2 \\ \gamma_1 & 0 & \delta_1 & 0 \\ 0 & \gamma_2 & 0 & \delta_2 \end{pmatrix},$$

s. t.

$$\begin{vmatrix} \alpha_i & \beta_i \\ \gamma_i & \delta_i \end{vmatrix} = 1$$

Matrix M is clearly symplectic and we want to choose it in such a way that $M\mathcal{L}_0$ and $M\mathcal{L}_0^{X(0)}$ are transversal to the horizontal subspace Ξ . Such matrices M are actually dense in the set of all matrices of the given form. We can prove this by an explicit computation.

If $\dim(\mathcal{L}_0 \cap \Xi) > 0$, then following along the lines of Example 1.6 we can assume that

$$\mathcal{L}_0 = \begin{bmatrix} x & 0 \\ 0 & 0 \\ z & 0 \\ 0 & 1 \end{bmatrix}$$

Then $\dim(M\mathcal{L}_0 \cap \Xi) = 0$ is equivalent to

$$\alpha_1 x + \beta_1 z \neq 0, \quad \beta_2 \neq 0.$$

Similarly from examples 1.6 and 1.7 we know that if $\dim(\mathcal{L}_0^{X(0)} \cap \Xi) > 0$, then

$$\mathcal{L}_0^{X(0)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Then $\dim(M\mathcal{L}_0^{X(0)} \cap \Xi) = 0$ can be achieved by already taking $\beta_2 \neq 0$.

In the new coordinates our Jacobi equation takes the form

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = M \begin{pmatrix} 0 & \frac{B(\tau)}{\tau^m} \\ C(\tau) & 0 \end{pmatrix} M^{-1} \begin{pmatrix} p \\ q \end{pmatrix}.$$

An explicit computation gives us

$$M \begin{pmatrix} 0 & \frac{B(\tau)}{\tau^m} \\ C(\tau) & 0 \end{pmatrix} M^{-1} = \\ = \begin{pmatrix} c_{11}\beta_1\delta_1 - (\frac{1}{b} + b_{11})\gamma_1\alpha_1 & -b_{12}\gamma_2\alpha_1 & -c_{11}\beta_1^2 + (\frac{1}{b} + b_{11})\alpha_1^2 & b_{12}\alpha_1\alpha_2 \\ -b_{12}\gamma_1\alpha_2 & c_{22}\delta_2\beta_2 - b_{22}\gamma_2\alpha_2 & b_{12}\alpha_1\alpha_2 & -c_{22}\beta_2^2 + b_{22}\alpha_2^2 \\ c_{11}\delta_1^2 - (\frac{1}{b} + b_{11})\gamma_1^2 & -b_{12}\gamma_1\gamma_2 & -c_{11}\delta_1\beta_1 + (\frac{1}{b} + b_{11})\gamma_1\alpha_1 & b_{12}\gamma_1\alpha_2 \\ -b_{12}\gamma_1\gamma_2 & c_{22}\delta_2^2 - b_{22}\gamma_2^2 & b_{12}\gamma_2\alpha_1 & -c_{22}\delta_2\beta_2 + b_{22}\alpha_2\gamma_2 \end{pmatrix}$$

Recall that our original system was such that $b < 0$ and $b_{22} < 0$ for $\tau \geq 0$ small. Thus the upper of diagonal 2x2 block will be a negative matrix function for small $\tau > 0$, if we choose α_2 big enough. For the same reason the lower of diagonal 2x2 block will be negative if we choose γ_2 big enough. Thus we can apply the comparison lemma as before with

$$B_1(\tau) = \begin{pmatrix} \frac{\alpha_1^2}{b(\tau)} - \varepsilon & 0 \\ 0 & b_{22}(0)\alpha_2^2 - c_{22}(0)\beta_2^2 - \varepsilon \end{pmatrix}, \\ B_2(\tau) = \begin{pmatrix} \frac{\alpha_1^2}{b(\tau)} + \varepsilon & 0 \\ 0 & b_{22}(0)\alpha_2^2 - c_{22}(0)\beta_2^2 + \varepsilon \end{pmatrix}, \\ C_1(\tau) = \begin{pmatrix} -\frac{\gamma_1^2}{b(\tau)} + \varepsilon & 0 \\ 0 & -b_{22}(0)\gamma_2^2 + c_{22}(0)\delta_2^2 + \varepsilon \end{pmatrix}, \\ C_2(\tau) = \begin{pmatrix} -\frac{\gamma_1^2}{b(\tau)} - \varepsilon & 0 \\ 0 & -b_{22}(0)\gamma_2^2 + c_{22}(0)\delta_2^2 - \varepsilon \end{pmatrix}$$

where $\varepsilon > 0$ is sufficiently small. This finishes the proof of Theorem 4.1.

As we have already discussed before, the jump alone does not determine the Jacobi curve, because with a singular Jacobi DE we lose uniqueness. So we need to characterize the right solution of the Jacobi equation. For $m = 1$ and $m = 2$ we can characterize the right solution using a singular boundary value problem .

Theorem 4.11. *If $m = 1$ or $m = 2$ then Jacobi curve after a singularity can be characterized as a boundary value problem of the extended Jacobi DE on the Lagrangian Grassmanian with conditions on the left end-point and the first left derivative.*

This will be proven in a number of steps:

1. We change coordinates so that \mathcal{L}_0 and \mathcal{L}_{0+} lie in the same coordinate chart and \mathcal{L}_{0+} is taken to be zero;
2. We write down the corresponding Riccati equation and perform a certain blow-up procedure;
3. After the blow-up we obtain a non-autonomous Riccati equation. We then proceed in determining the Jacobi curve for the autonomous part;

4. Using a deformation argument we prove that in the non-autonomous case the Jacobi curve is well-defined by the same jet.

For the first step we are going to have three different situations as well

1. \mathcal{L}_0 is transversal to the horizontal plane in current coordinates and in the corresponding symmetric matrix S^0 either $S_{11}^0 \neq 0$ or $S_{11}^0 = S_{12}^0 = 0$;
2. \mathcal{L}_0 is either transversal to the horizontal plane in current coordinates and in the corresponding symmetric matrix $S_{11}^0 = 0$, $S_{22}^0 \neq 0$ or \mathcal{L}_0 and the horizontal plane Ξ have a common line;
3. \mathcal{L}_0 is either transversal to the horizontal plane in current coordinates and in the corresponding symmetric matrix $S_{11}^0 = S_{22}^0 = 0$, $S_{12}^0 \neq 0$ or \mathcal{L}_0 is the horizontal plane Ξ .

Let

$$S_{22}^+ = \begin{cases} S_{22}^0, & S_{11}^0 = 0, \\ S_{22}^0 - \frac{(S_{12}^0)^2}{S_{11}^0}, & S_{11}^0 \neq 0. \end{cases}$$

Then depending on the case we apply one of the three symplectic transformations

$$M_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -S_{22}^+ & 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

After applying those transformation \mathcal{L}_{0+} will become the vertical subspace. Let us check what happens to \mathcal{L}_0 under these transforms. We have for case 1 either

$$M_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ S_{11}^0 & S_{12}^0 \\ S_{12}^0 & S_{22}^0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ S_{11}^0 & S_{12}^0 \\ S_{12}^0 & \frac{(S_{12}^0)^2}{S_{11}^0} \end{bmatrix} \quad \text{or} \quad M_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & S_{22}^0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

For case 2 either

$$M_2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & S_{12}^0 \\ S_{12}^0 & S_{22}^0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -S_{12}^0 & -S_{22}^0 \\ 0 & S_{12}^0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{(S_{12}^0)^2}{S_{22}^0} & -\frac{(S_{12}^0)}{S_{22}^0} \\ -\frac{(S_{12}^0)}{S_{22}^0} & -\frac{1}{S_{22}^0} \end{bmatrix}$$

or

$$M_2 \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ z & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ z & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ z & 0 \\ 0 & 0 \end{bmatrix}.$$

For the case 3 either

$$M_3 \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & S_{12}^0 \\ S_{12}^0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -S_{12}^0 \\ -S_{12}^0 & 0 \\ 0 & S_{12}^0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -\frac{1}{S_{12}^0} \\ -\frac{1}{S_{12}^0} & -\frac{1}{(S_{12}^0)^2} \end{bmatrix}$$

or

$$M_3 \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & 0 \end{bmatrix}.$$

And this finishes the first step.

For the second step we have to rewrite the Jacobi equation in the new coordinates. We simply have to conjugate the right-hand side of (4.7) by the corresponding matrix M_i . Then to each case corresponds its own Jacobi equation of the form

$$\frac{d}{d\tau} \begin{pmatrix} p \\ q \end{pmatrix} = M_i \begin{pmatrix} A(\tau) & B(\tau) \\ C(\tau) & -A^T(\tau) \end{pmatrix} M_i^{-1} \begin{pmatrix} p \\ q \end{pmatrix}$$

or more precisely

$$\text{Case 1 : } \frac{d}{d\tau} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0 & b_{12}(\tau)S_{22}^+ & \frac{1}{b(\tau)} + b_{11}(\tau) & b_{12}(\tau) \\ 0 & b_{22}(\tau)S_{22}^+ & b_{12}(\tau) & b_{22}(\tau) \\ c_{11}(\tau) & 0 & 0 & 0 \\ 0 & c_{22}(\tau) - b_{22}(\tau)(S_{22}^+)^2 & -b_{12}(\tau)S_{22}^+ & -b_{22}(\tau)S_{22}^+ \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix},$$

$$\text{Case 2 : } \frac{d}{d\tau} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} 0 & -b_{12}(\tau) & \frac{1}{b(\tau)} + b_{11}(\tau) & 0 \\ 0 & 0 & 0 & -c_{22}(\tau) \\ c_{11}(\tau) & 0 & 0 & 0 \\ 0 & -b_{22}(\tau) & b_{12}(\tau) & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix},$$

$$\text{Case 3 : } \frac{d}{d\tau} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} -c_{11}(\tau) & -b_{12}(\tau) & \frac{1}{b(\tau)} + b_{11}(\tau) - c_{11}(\tau) & 0 \\ 0 & 0 & 0 & -c_{22}(\tau) \\ c_{11}(\tau) & 0 & c_{11}(\tau) & 0 \\ 0 & -b_{22}(\tau) & b_{12}(\tau) & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix}.$$

We then take $q = Sp$ and obtain a Riccati equation of the form (1.2). We do a blow-up of this equation by taking

$$S(\tau) = \tau S_1(\tau).$$

Then we obtain a Riccati equation for $S_1(\tau)$. Let \tilde{S}^0 be the symmetric matrix that corresponds to \mathcal{L}_0 in the new coordinates. We denote by $S_1^\varepsilon(\tau)$ the solution of this Riccati equation with

$$S_1^\varepsilon(\varepsilon) = \frac{\tilde{S}^0}{\varepsilon}.$$

Since outside the singularity the right-hand side is analytic and we have a family of solutions converging to a solution, it is clear that

$$\dot{S}(\tau) = \lim_{\varepsilon \rightarrow 0^+} S_1^\varepsilon(\tau).$$

Since in the new coordinates $S(0^+) = 0$, we find that the discussed previously Riccati equation has the form

$$\tau \dot{S}_1 + S_1 + S_1 \begin{pmatrix} \frac{1}{b_2} & 0 \\ 0 & 0 \end{pmatrix} S_1 - C(0) = \tau R(\tau, S) \quad (4.25)$$

for $m = 2$ and

$$\tau \dot{S}_1 + S_1 - C(0) = \tau R(\tau, S) \quad (4.26)$$

for $m = 1$. That finishes the proof of the second step.

For the third step we are going to consider just the first case. For the second and the third case the argument is repeated word by word. We assume that the right-hand side of those equations is actually zero. Then we can understand very well the whole phase portrait of this Riccati equation. Indeed, we can extend the dynamics to the whole Lagrangian Grassmanian $L(\mathbb{R}^4)$ by rewriting the corresponding Hamiltonian system.

$$\tau \frac{d}{d\tau} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{b_2} & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ c_{11}(0) & 0 & -\frac{1}{2} & 0 \\ 0 & c_{22}(0) - (S_{22}^+)^2 b_{22}(0) & 0 & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix} = H \begin{pmatrix} p_1 \\ p_2 \\ q_1 \\ q_2 \end{pmatrix}$$

A complete description of the phase portrait of such a system was given in [46]. It is clear that the equilibrium points are spanned by the eigenvectors. In our case, H has eigenvalues

$$\lambda_1 = -\frac{1}{2} \sqrt{1 + \frac{4c_{11}(0)}{b_2}}, \quad \lambda_2 = -\frac{1}{2}, \quad \lambda_3 = \frac{1}{2}, \quad \lambda_4 = \frac{1}{2} \sqrt{1 + \frac{4c_{11}(0)}{b_2}}.$$

Since we consider only non-oscillating systems, we have that all four eigenvalues are real and different. Let E_i be the corresponding eigenvectors. We have

$$E_1 = \begin{pmatrix} 1 - \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \\ 0 \\ 2c_{11}(0) \\ 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

$$E_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ c_{22}(0) - b_{22}(0)(S_{22}^+)^2 \end{pmatrix}, \quad E_4 = \begin{pmatrix} 1 + \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \\ 0 \\ 2c_{11}(0) \\ 0 \end{pmatrix}.$$

We define $E_{ij} = \text{span}\{E_i, E_j\}$. It is easy to see that we have four equilibrium points on the Lagrangian Grassmanian: $E_{12}, E_{13}, E_{24}, E_{34}$. For each of these equilibrium points we can find the corresponding stable and unstable manifolds $W^s(E_{ij})$ and $W^u(E_{ij})$. Then if $S_1^\varepsilon(\varepsilon)$ lies in $W^s(E_{ij})$, the Jacobi curve is going to be just the equilibrium solution $\mathcal{L}_\tau = E_{ij}$. Indeed, the Lagrangian Grassmanian is compact and therefore any trajectory in the stable manifold has finite length. But every trajectory of our Riccati equation has speed that goes to infinity as $\tau \rightarrow 0+$. So as we take ε smaller and smaller for a fixed time $\tau > 0$ the corresponding curve $S^\varepsilon(\tau)$ is going to get closer and closer to the equilibrium point approaching it in the limit. It remains only to describe stable manifolds of our equilibrium points.

Luckily it was already done in [46] by M. Shayman. He proved that in order to find the stable manifolds we need to form a flag $\{0\} = V_0 \subset V_1 \subset \dots \subset V_4 = \mathbb{R}^4$, where

$$V_i = \bigoplus_{j=1}^i E_j,$$

and associate to each E_{ij} a sequence $l(E_{ij}) = (l_1, l_2, l_3, l_4)$, where

$$l_k = \begin{cases} 1 & \text{if } k = i, j; \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$W^s(E_{ij}) = \left\{ \Lambda \in L(\mathbb{R}^4) : \dim \Lambda \cap V_m = \sum_{k=1}^m l_k, l_k \in l(E_{ij}), m = 1, 2, 3, 4. \right\}.$$

It remains to check for which initial data $S_1^\varepsilon(\varepsilon)$ lies in which $W^s(E_{ij})$ for small $\varepsilon > 0$ and describe the corresponding $W^s(E_{ij})$.

Lemma 4.12. *Suppose that the right hand side of (4.25) is zero. Then the curves $\Lambda_1^\varepsilon(\tau)$ that correspond to $S_1^\varepsilon(\tau)$ converge pointwise to the equilibrium solution $\Lambda(\tau) \equiv E_{34}$. Or in local coordinates we get that*

$$\lim_{\varepsilon \rightarrow 0+} S_1^\varepsilon(\tau) = \begin{pmatrix} -\frac{b_2}{2} \left(1 - \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \right) & 0 \\ 0 & c_{22}(0) - b_{22}(0)(S_{22}^+)^2 \end{pmatrix} = S_1^{34}.$$

If $R(\tau, S) = 0$ in (4.25), then exists a unique solution of this equation with $S_1(0+) = S_1^{34}$.

Proof. By definition we find that

$$W^s(E_{34}) = \{ \Lambda \in L(\mathbb{R}^4) : \dim(\Lambda \cap E_{12}) = 0 \} = E_{12}^\eta,$$

which is dense in $L(\mathbb{R}^4)$. So we only need to prove that $\Lambda^\varepsilon(\varepsilon) \in E_{12}^\eta$ for $\varepsilon > 0$ small. Indeed, in this case the unstable manifold $W^u(E_{34}) = \{E_{34}\}$ and so the only solution of (4.25) with $S_1(0+) = S_1^{34}$ can be $S_1(\tau) \equiv S_1^{34}$.

We note that if $S_{11}^0 = S_{12}^0 = 0$, then $S_1^\varepsilon(\varepsilon) = 0$. In this case for small $\varepsilon > 0$ it is clear that $\dim(\Lambda^\varepsilon(\varepsilon) \cap E_{12}) = 0$. If $S_1^\varepsilon(\varepsilon) \neq 0$ for small $\varepsilon > 0$, then $\dim(\Lambda^\varepsilon(\varepsilon) \cap E_{12}) > 0$ if and only if

$$S_{12}^+ = 0 \quad \text{and} \quad \frac{S_{11}^+}{\varepsilon} = -\frac{b_2}{2} \left(1 + \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \right),$$

but this can happen only for a single value

$$\varepsilon = -\frac{2S_{11}^+}{b_2 \left(1 + \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \right)}.$$

And so for small $\varepsilon > 0$ we indeed get $\Lambda^\varepsilon(\varepsilon) \in E_{12}^\dagger$. \square

Case $m = 1$ is easier, since the principal part of the equation (4.26) is linear and has a global stable equilibrium

$$S = \begin{pmatrix} c_{11}(0) & 0 \\ 0 & c_{22}(0) - b_{22}(0)(S_{22}^+)^2 \end{pmatrix}.$$

As for $m = 2$, we have then $S_1(0+) = S$ and a unique solution to a Cauchy problem, that characterizes our Jacobi curve.

For the case 2 and 3 we have a similar result. We obtain that

$$\lim_{\varepsilon \rightarrow 0^+} S_1^\varepsilon(\tau) = \begin{pmatrix} -\frac{b_2}{2} \left(1 - \sqrt{1 + \frac{4c_{11}(0)}{b_2}} \right) & 0 \\ 0 & -b_{22}(0) \end{pmatrix} = S_1^{34}$$

for $m = 2$ and

$$\lim_{\varepsilon \rightarrow 0^+} S_1^\varepsilon(\tau) = \begin{pmatrix} c_{11}(0) & 0 \\ 0 & -b_{22}(0) \end{pmatrix}$$

for $m = 1$ and that in this case indeed the Jacobi curve is fully determined by the first jet. We keep the notation S_1^{34} because in the case 2 and 3 we obtain a Hamiltonian system whose matrix has exactly the same eigenvalues as the Hamiltonian matrix of case 1 and the same eigenvectors except E_3 , that must be replaced by

$$E_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -b_{22}(0) \end{pmatrix}$$

It remains now to do the last step and to show the general case. Let us assume

$$S_1 = \begin{pmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{pmatrix}$$

and rewrite (4.25) or (4.26) as a system on $\mathbb{R} \times L(\mathbb{R}^4)$. Namely we have

$$\begin{aligned}\dot{S} &= Q(S) + \tau R(\tau, S), \\ \dot{\tau} &= \tau;\end{aligned}$$

where $Q(S)$ is the autonomous Riccati part. It is clear that $(S_1^{34}, 0)$ is an equilibrium point of this system. Moreover, by linearising the right hand side at $(S_1^{34}, 0)$ we obtain that it is a hyperbolic equilibrium point, since the linearized operator has eigenvalues

$$\left\{ -\sqrt{1 + \frac{4c_{11}(0)}{b_2}}, -\frac{1}{2} - \frac{1}{2}\sqrt{1 + \frac{4c_{11}(0)}{b_2}}, -1, 1 \right\},$$

the same as for the autonomous system in all three cases. So by Grobman-Hartman theorem both systems are topologically conjugate in the neighbourhood of this equilibrium point. Since both of them have a single unstable direction it means that there exists a unique trajectory of the non-autonomous system that approaches $(S_1^{34}, 0)$ as $\tau \rightarrow 0+$. We claim that this trajectory must be a lift of the Jacobi curve to the extended phase-space. This result does not follow directly from the Grobman-Hartman theorem since $S^\varepsilon(\varepsilon)$ is far from the equilibrium and a priori we have no information about the behaviour orbits close to infinity.

The result follows from an application of the variation formula (1.10). We take In our case

$$V(\tau) = \begin{pmatrix} Q(S) \\ \tau \end{pmatrix}, \quad W(\tau) = \begin{pmatrix} \tau R(\tau, S) \\ 0 \end{pmatrix}.$$

Let K^τ be the flow of $V(\tau)$. Then due to smoothness of each flow the lift of the Jacobi curve will be then given by the limit curve

$$\begin{aligned}& \lim_{\varepsilon \rightarrow 0+} \begin{pmatrix} S_1^\varepsilon(\varepsilon) \\ \varepsilon \end{pmatrix} \circ \exp \int_0^{\tau-\varepsilon} V(\theta) + W(\theta) d\theta = \\ &= \lim_{\varepsilon \rightarrow 0+} \begin{pmatrix} S_1^\varepsilon(\varepsilon) \\ \varepsilon \end{pmatrix} \circ K^{\tau-\varepsilon} \circ \exp \int_0^{\tau-\varepsilon} K_*^{\tau-\varepsilon} (K_*^\theta)^{-1} W(\theta) d\theta = \\ &= \lim_{\varepsilon \rightarrow 0+} \begin{pmatrix} S_1^\varepsilon(\varepsilon) \\ \varepsilon \end{pmatrix} \circ K^{\tau-\varepsilon} \circ \lim_{\varepsilon \rightarrow 0+} \exp \int_0^{\tau-\varepsilon} K_*^{\tau-\varepsilon} (K_*^\theta)^{-1} W(\theta) d\theta = \\ &= \lim_{\varepsilon \rightarrow 0+} \begin{pmatrix} S_1^\varepsilon(\varepsilon) \\ \varepsilon \end{pmatrix} \circ K^{\tau-\varepsilon} \circ \exp \int_0^\tau K_*^\tau (K_*^\theta)^{-1} W(\theta) d\theta.\end{aligned}$$

But the first limit corresponds to the lift of the Jacobi curve in autonomous case. Thus if we take a limit of this expression as $\tau \rightarrow 0+$ we would obtain

$$\begin{pmatrix} S_1^{34} \\ 0 \end{pmatrix},$$

like in the autonomous case. This proves Theorem 4.11.

Geodesic flow and singular trajectories of left-invariant Engel structures

An Engel structure consists of an equiregular distribution of rank 2 on a four-dimensional manifold. It was proven by Engel [26], that distributions of this kind have no local invariants. He proved that locally any Engel manifold is diffeomorphic to \mathbb{R}^4 with coordinates (x, y, u, v) , s.t. the distribution is given as the annihilator of the following two differential forms

$$\omega_1 = dy - udx, \quad \omega_2 du - vdx.$$

Their metric geometry instead even on the local level is quite rich. We will see that in the first section, when we will discuss the local classification result.

Sub-Riemannian Engel manifolds are the simplest equiregular structures that admit minimizing abnormal geodesics. Abnormal geodesics still remain quite mysterious objects and there are many open problem regarding their properties. Unfortunately, there is lack of simple examples that would allow to acquire basic intuition about them. Especially, when the abnormal geodesics are strict, meaning that they are not a projection of any normal extremal. One of the main results of this chapter is a construction of various simple examples of Engel manifolds admitting many symmetries. These symmetries allow to construct models of sub-Riemannian structures with strictly abnormal geodesics and integrable geodesics flows. The only other example, that the author is aware of, is the one given by Montgomery [37], which he used as an example of a minimizing singular trajectory.

In the last section we study the Jacobi equation for singular geodesics and derive a simple comparison result for conjugate times along them.

5.1 Normal frame and classification of left-invariant Engel structures

The results of this section are mostly due to my friend and colleague Alexander Medvedev. That is why we only sketch the main ideas that explain how to arrive at the normal form for such structures.

Every sub-Riemannian Engel structure possesses a canonical global frame. To construct it we recall that an Engel manifold possesses a global characteristic one-field, whose integral lines are abnormal geodesics [20]. Let us choose a direction on this line field. This way we obtain a vector field X_2 . Let X_1 be any orthogonal complement. It is clear that there is still a $\mathbb{Z}_2 \times \mathbb{Z}_2$ ambiguity due to this choice. Then the remaining vector fields of the frame are defined as

$$X_3 = [X_1, X_2], \quad X_4 = [X_1, X_3].$$

In particular this shows that every Engel manifold must be parallelizable. The $\mathbb{Z}_2 \times \mathbb{Z}_2$ ambiguity is resolved by choosing an orientation for the manifold and for the frame.

This frame gives us a filtration on TM :

$$0 = F^0 \subset F^{-1} \subset F^{-2} \subset F^{-3} \subset F^{-4} = TM.$$

where $F^{-1} = \mathbb{R}X_1$, $F^{-2} = \mathcal{D}$, $F^{-3} = \mathcal{D}^2 = \mathcal{D} = [\mathcal{D}, \mathcal{D}]$. This filtration induces a grading. Namely $\deg(X_i) = -i$ and for the structure constants C_{ij}^k , s.t.

$$[X_i, X_j] = X_{i+j} + C_{ij}^k X_k, \quad i < j;$$

we have $\deg(C_{ij}^k) = i + j - k$. One can check that all non-zero C_{ij}^k must have a positive degree. Note that this grading is different from the standard one.

Due to the presence of the characteristic line field, classification of Engel sub-Riemannian structures is equivalent to the classification of frames. This is not always the case, as can be seen from the 3D contact case, where the frames define a sub-Riemannian structure only up to a $SO(2)$ action. Thus in the Engel case the structure constants C_{ij}^k and their covariant derivatives give a complete set of invariants [47]. But C_{ij}^k are functionally dependent due to the Jacobi identity. Therefore, we must write down the Jacobi identity and express all C_{ij}^k in terms of a set of independent structure constants, that we can choose like in the table below.

Table 5.1: Basic invariants of Sub-Riemannian Engel structures

Degree	Invariant
1	$T_1 = C_{14}^4$
2	$T_2 = C_{23}^3$ $T_3 = C_{14}^3$
3	$T_4 = C_{23}^2$ $T_5 = C_{14}^2$
4	$T_6 = C_{23}^1$

The grading on TM simplifies this calculation considerably, since we can check each equation degree by degree. By doing this we prove the following theorem.

Theorem 5.1. *For every oriented sub-Riemannian Engel structure (M, \mathcal{D}, g) with fixed orientation on \mathcal{D} there exists a canonical frame $\{X_1, X_2, X_3, X_4\}$ given by conditions*

$$[X_1, X_2] = X_3, [X_1, X_3] = X_4, [X_2, X_3] \in \text{span } X_1, X_2, X_3, \quad (5.1)$$

such that orientations of $\{X_1, X_2, X_3, X_4\}$ and $\{X_1, X_2\}$ are compatible with orientations of M and \mathcal{D} respectively.

Apart from (5.1) the structure equations of the canonical frame are:

$$\begin{aligned} [X_1, X_4] &= C_{14}^1 X_1 + T_5 X_2 + T_3 X_3 + T_1 X_4 \\ [X_2, X_3] &= T_6 X_1 + T_4 X_2 + T_2 X_3 \\ [X_2, X_4] &= X_1(T_6)X_1 + X_1(T_4)X_2 + (T_4 + X_1(T_2))X_3 + T_2 X_4 \\ [X_3, X_4] &= C_{34}^1 X_1 + C_{34}^2 X_2 + C_{34}^3 X_3 + (T_4 + 2X_1(T_2) - X_2(T_1))X_4, \end{aligned} \quad (5.2)$$

where

$$\begin{aligned} C_{14}^1 &= \frac{1}{2} (T_1 T_4 + T_1 X_1(T_2) - 3X_1(T_4) + X_2(T_3) + X_3(T_1) - X_1^2(T_2)), \\ C_{34}^3 &= -\frac{1}{2} (T_1 T_4 + T_1 X_1(T_2) - X_1(T_4) + X_2(T_3) - X_3(T_1) - X_1^2(T_2)), \\ C_{34}^2 &= T_2 T_5 - T_3 T_4 - T_1 X_1(T_4) - X_2(T_5) + X_1^2(T_4), \\ C_{34}^1 &= T_2 C_{14}^1 - T_6 T_3 - T_1 X_1(T_6) - X_2(C_{14}^1) + X_1^2(T_6). \end{aligned}$$

In particular, the structure constants depend only on T_i , $1 \leq i \leq 6$ and their derivatives along X_j .

These equations are greatly simplified if we assume that we have a left-invariant sub-Riemannian structure on a Lie group. In this case all T_i are constants and we have the following corollary.

Corollary 5.2. *Let $\{X_1, X_2, X_3, X_4\}$ be a canonical left-invariant frame for a left-invariant Engel sub-Riemannian structure. Then the structure equations of the frame are:*

$$\begin{aligned} [X_1, X_2] &= X_3, [X_1, X_3] = X_4, \\ [X_1, X_4] &= \frac{1}{2} A X_1 + T_5 X_2 + T_3 X_3 + T_1 X_4, \\ [X_2, X_3] &= T_6 X_1 + T_4 X_2 + T_2 X_3, \\ [X_2, X_4] &= T_4 X_3 + T_2 X_4, \\ [X_3, X_4] &= C X_1 + B X_2 - \frac{1}{2} A X_3 + T_4 X_4, \end{aligned} \quad (5.3)$$

where $A = T_1 T_4$, $B = T_2 T_5 - T_3 T_4$, $C = \frac{1}{2} T_1 T_2 T_4 - T_3 T_6$.

Substituting the structure constants from Corollary 5.2 into the Jacobi formula we obtain a system of restrictions on T_i :

$$\begin{aligned}
0 &= T_1 T_6 + 2T_2 T_4, \\
0 &= T_1^2 T_4 + 4T_2 T_5, \\
0 &= T_1 T_3 T_4 - T_1 T_2 T_5 + 2T_4 T_5, \\
0 &= T_1 T_4^2 - T_1^2 T_2 T_4 + 2T_1 T_3 T_6 + 2T_5 T_6, \\
0 &= T_1 T_4^2 + 4T_2^2 T_5 - 4T_2 T_3 T_4 + 2T_5 T_6, \\
0 &= T_1 T_2^2 T_4 + T_1 T_4 T_6 - 2T_2 T_3 T_6.
\end{aligned} \tag{5.4}$$

We can now solve this system of equations to obtain a full classification of left-invariant Engel structures. Note that in the article [12, 13] some of the families are missing. Nevertheless the most symmetric example that we will consider, the type III structures, that are central extensions of 3D unimodular groups, were present and classified in those articles. Summing up we have proven the following theorem

Theorem 5.3. *Any left-invariant sub-Riemannian Engel structure is uniquely locally defined by the structure constants T_i and belongs to at least one family from Table 5.2. We list in Table 5.2 restrictions on T_i that define a family as well as corresponding non-trivial structure equations.*

Table 5.2: Classification of left-invariant Sub-Riemannian Engel structures

#	Restrictions	Structure Equations Excluding $[X_1, X_2] = X_3$, $[X_1, X_3] = X_4$
I.	$T_2 = T_4 = T_6 = 0$	$[X_1, X_4] = T_5 X_2 + T_3 X_3 + T_1 X_4$
II.	$T_4 = T_6 = T_5 = 0$	$[X_1, X_4] = T_3 X_3 + T_1 X_4,$ $[X_2, X_3] = T_2 X_3,$ $[X_2, X_4] = T_2 X_4$
III.	$T_1 = T_2 = T_5 = 0$	$[X_1, X_4] = T_3 X_3,$ $[X_2, X_3] = T_6 X_1 + T_4 X_2,$ $[X_2, X_4] = T_4 X_3,$ $[X_3, X_4] = -T_6 T_3 X_1 - T_4 T_3 X_2 + T_4 X_4$
IV.	$T_1 = T_3 = 0,$ $T_4 = T_5 = 0$	$[X_2, X_3] = T_6 X_1 + T_2 X_3,$ $[X_2, X_4] = T_2 X_4.$
V.	$T_1 \neq 0,$ $T_4 = \frac{1}{2} \frac{T_2(T_1^2 + 4T_3)}{T_1},$ $T_5 = -\frac{1}{8} T_1^3 - \frac{1}{2} T_1 T_3,$ $T_6 = -\frac{T_2^2(T_1^2 + 4T_3)}{T_1^2}$	$[X_1, X_4] = T_1 X_4 - \frac{T_1^3 + 8T_5}{4T_1} X_3 + T_5 X_2 - \frac{2T_2 T_5}{T_1} X_1,$ $[X_2, X_3] = T_2 X_3 - \frac{4T_2 T_5}{T_1^2} X_2 + \frac{8T_2^2 T_5}{T_1^3} X_1,$ $[X_2, X_4] = T_2 X_4 - \frac{4T_2 T_5}{T_1^2} X_3,$ $[X_3, X_4] = \frac{2T_2 T_5}{T_1} \left(X_3 - \frac{2}{T_1} X_4 - \frac{4T_5}{T_1^2} X_2 + \frac{8T_2 T_5}{T_1^3} X_1 \right)$

5.2 Integrability of the geodesic flow

We have seen in Section 1.4, that the problem of finding minimal curves in a sub-Riemannian problem is equivalent to an optimal control problem. So we can apply the PMP to find minimal curves.

To write the Hamiltonian system of PMP it is useful to introduce linear on fibers of T^*M functions

$$h_i = \langle \lambda, X_i \rangle, \quad \lambda \in T^*M.$$

The Hamiltonian of PMP reads as

$$H_u(\lambda, \nu) = \langle \lambda, u_1 X_1 + u_2 X_2 \rangle - \frac{\nu}{2}(u_1^2 + u_2^2) = u_1 h_1 + u_2 h_2 - \frac{\nu}{2}(u_1^2 + u_2^2).$$

We can write down the corresponding Hamiltonian systems as

$$\begin{aligned} \dot{q} &= u_1 X_1(q) + u_2 X_2(q), \\ \dot{h}_i &= \{H_u, h_i\}. \end{aligned}$$

where the Lie-Poisson bracket of vertical coordinate functions h_i depends only on the structure functions of X_i :

$$\{h_i, h_j\} = \langle \lambda, [X_i, X_j] \rangle = C_{ij}^k(q) h_k.$$

Using the structure equations 5.2 and the Leibniz rule we obtain

$$\begin{aligned} \dot{q} &= u_1 X_1(q) + u_2 X_2(q), \\ \dot{h}_1 &= -u_2 h_3, \\ \dot{h}_2 &= u_1 h_3, \\ \dot{h}_3 &= u_1 h_4 + u_2 (T_6 h_1 + T_4 h_2 + T_2 h_3), \\ \dot{h}_4 &= u_1 (C_{14}^1 h_1 + T_5 h_2 + T_3 h_3 + T_1 h_4) \\ &\quad + u_2 (X_1(T_6) h_1 + X_1(T_4) h_2 + (T_4 + X_1(T_2)) h_3 + T_2 h_4). \end{aligned} \tag{5.5}$$

We first characterize abnormal geodesics. In this case $\nu = 0$ and from the weak PMP condition we obtain that along a geodesic $h_1 = h_2 \equiv 0$. This implies $\dot{h}_1 = -u_2 h_3 = 0$ and $\dot{h}_2 = u_1 h_3 = 0$. Since we are interested in curves with non-zero constant speed $\dot{q} = u_1^2 + u_2^2$ we obtain that $h_3 \equiv 0$. The forth equation of (5.5) implies that either $u_1 \equiv 0$ or $h_4 \equiv 0$. But covector $(\lambda_t, -\nu)$ can not be zero, thus $h_4 \neq 0$ if $\nu = 0$. Therefore, $u_1 \equiv 0$ and projections of abnormal extremals are integral curves of X_2 . Along these curves the last equation reduces to

$$\dot{h}_4 = u_2 T_2 h_4,$$

whose solutions are sign-definite for non-zero initial data. Therefore the non-triviality condition is satisfied for all times and $(q(t), h(t))$ is an abnormal extremal. Moreover $q(t)$ is always a length minimizer in the Engel case [3], i.e. it is always an abnormal geodesic.

Let us consider the case $\nu = 1$. Then the maximum is achieved when

$$\frac{\partial H_u}{\partial u_i} = h_i - u_i = 0 \iff u_i = h_i, \quad i = 1, 2.$$

Substituting the obtained controls in (5.5) we get

$$\begin{aligned} \dot{q} &= h_1 X_1(q) + h_2 X_2(q), \\ \dot{h}_1 &= -h_2 h_3, \\ \dot{h}_2 &= h_1 h_3, \\ \dot{h}_3 &= h_1 h_4 + h_2 (T_6 h_1 + T_4 h_2 + T_2 h_3), \\ \dot{h}_4 &= h_1 (C_{14}^1 h_1 + T_5 h_2 + T_3 h_3 + T_1 h_4) \\ &\quad + h_2 (X_1(T_6) h_1 + X_1(T_4) h_2 + (T_4 + X_1(T_2)) h_3 + T_2 h_4). \end{aligned} \tag{5.6}$$

which is a Hamiltonian system with Hamiltonian

$$H_{\bar{u}(t)} = H = \frac{h_1^2 + h_2^2}{2}.$$

Assume that an abnormal geodesic $(q(t), h(t))$ satisfies (5.6). Since it is an integral curve of X_2 we must have $h_1 \equiv 0$. Moreover, the Hamiltonian H is a first integral of the system. Therefore $2H = h_1^2 + h_2^2 = \text{const} \neq 0$ and so $h_2 = \text{const} \neq 0$. Thus from (5.6) it follows that $h_3 \equiv 0$. But the forth equation gives us $T_4 h_2^2 = 0$, which can hold if and only if $T_4 = 0$ along the curve. If this is the case, then all these conditions reduce the system to the equation

$$\dot{h}_4 = h_2 T_2 h_4, \tag{5.7}$$

which always has a solution.

On the other hand assume that along an abnormal extremal $T_4 = 0$. By substituting $h_1 \equiv 0$, $h_2 \equiv 1$, $h_3 \equiv 0$ into (5.6) we reduce the system to (5.7). This equation always has a sign-definite solution which guarantees that the abnormal extremal is normal as well.

Thus we have proven the following fact.

Theorem 5.4. *Abnormal geodesics of an Engel sub-Riemannian structure are integral curves of X_2 . An abnormal geodesic is strict if and only if $T_4 \neq 0$ along the geodesic.*

One can check from the classification in Section 5.1, that among the type III left-invariant Engel structures, there are indeed those that have $T_4 \neq 0$. The following result says that the normal geodesic flow on all these algebras is integrable.

Theorem 5.5. *Consider a left-invariant sub-Riemannian Engel structure of type III which is defined over a Lie group with a Lie algebra*

$$[X_1, X_2] = X_3, \quad [X_1, X_3] = X_4, \tag{5.8}$$

$$[X_1, X_4] = T_3 X_3, \quad [X_2, X_3] = T_6 X_1 + T_4 X_2, \tag{5.9}$$

$$[X_2, X_4] = T_4 X_3, \quad [X_3, X_4] = -T_6 T_3 X_1 - T_4 T_3 X_2 + T_4 X_4, \tag{5.10}$$

where vector fields X_1, X_2 form an orthonormal sub-Riemannian frame. The normal Hamiltonian flow of this structure is super-integrable meaning that it has four independent commuting first integrals including the Hamiltonian H and one more independent first integral that commutes with H . If $T_4 \neq 0$ then the abnormal geodesics of the structure are strict.

Before we prove the theorem, let us investigate the structure of corresponding Lie algebras. First, one can notice that any type III Lie algebra is a central extension of a 3-dimensional Lie algebra. The center element is

$$X'_4 = X_4 + T_4 X_1 - T_3 X_2.$$

The underlying Lie algebra is semi-simple if and only if $D = (T_4)^2 + T_3 T_6 \neq 0$. If $D < 0$ and $T_3 < 0$ (equivalently $T_6 < 0$) then it is $\mathfrak{so}(3, \mathbb{R})$ or $\mathfrak{sl}(2, \mathbb{R})$ otherwise.

Consider now the case $D = 0$. If $T_4 = T_6 = 0$ then we have a trivial extension either of the Lie algebra of Euclidean motions of the plane ($T_3 > 0$) or the Lie algebra of Poincare motions of the plane ($T_3 < 0$). Otherwise $T_4 \neq 0$ and we obtain a non-trivial extension of a solvable Lie algebra of dimension 3 with 2-dimensional derived algebra. The whole family already appeared in the classification of Almeida in [13] and among examples of Engel structures in [28].

Proof of Theorem 5.5. Instead of the basis for the type III family from Table 5.2, we use basis $\{X_1, X_2, X_3, X'_4\}$ in the proof. Then the only non-zero structure equations are

$$[X_1, X_2] = X_3, \tag{5.11}$$

$$[X_1, X_3] = X'_4 - T_4 X_1 + T_3 X_2, \tag{5.12}$$

$$[X_2, X_3] = T_6 X_1 + T_4 X_2. \tag{5.13}$$

The Hamiltonian function $h'_4 = \langle \lambda, X'_4 \rangle$ which corresponds to the center element X'_4 is a first integral. In the basis X_1, X_2, X_3, X'_4 the Hamiltonian system takes the form

$$\begin{aligned} \dot{h}_1 &= -h_2 h_3, \\ \dot{h}_2 &= h_1 h_3, \\ \dot{h}_3 &= h_1 h'_4 - T_4 (h_1^2 - h_2^2) + (T_3 + T_6) h_1 h_2, \end{aligned} \tag{5.14}$$

where $h'_4 = \text{const}$. It is easy to see that (5.14) has the following first integral

$$G = \frac{h_3^2}{2} - h'_4 h_2 + \frac{T_3 + T_6}{4} (h_1^2 - h_2^2) + T_4 h_1 h_2.$$

Let $I : g \mapsto g^{-1}$ be the inverse map of the Lie algebra and $X^R(g) = I_* X^L(g)$ be the right invariant fields constructed from the left-invariant ones. Let h_i^R be the right-invariant Hamiltonian functions. We know that h_i^R commute with any left-invariant Hamiltonian function and thus they commute with H, G, h'_4 . Therefore the whole family H, G, h'_4, h_1^R is commutative and $\{H, h_2^R\} = 0$.

We claim that $dH, dG, dh'_4, dh_1^R, dh_2^R$ are linearly independent almost everywhere. Actually it is enough to check that only in one, point for example, at the identity. Indeed it is known that any finite-dimensional Lie group is analytic, i.e. it admits an analytic structure as a manifold with analytic multiplication. Then the right and left-invariant Hamiltonians and their differentials are going to be analytic as well. Since the linear dependence is an algebraic condition on the components of the corresponding vectors, we get that if the differential above are linearly independent at some point, then they must be independent almost everywhere.

Assume that left-invariant Hamiltonian functions and right invariant Hamiltonian functions are related by

$$h_i^R = a_i^j(g)h_j,$$

where $X_i^R(g) = a_i^j(g)X_j^L(g)$ and $a_i^j(\text{id}) = -\delta_i^j$. In [35] it was shown that in the coordinates of the first kind $\frac{\partial a_i^k}{\partial x^j}(\text{id}) = c_{ji}^k$. Using this identity we deduce

$$\begin{pmatrix} dH \\ dG \\ dh'_4 \\ dh_1^R \\ dh_2^R \end{pmatrix} = \begin{pmatrix} h_1 & h_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{T_3+T_6}{2}h_1+T_4h_2 & -h'_4-\frac{T_3+T_6}{2}h_2+T_4h_1 & h_3 & -h_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & -h_3 & -h'_4+T_4h_1-T_3h_2 \\ 0 & -1 & 0 & 0 & 0 & h_3 & 0 & -T_6h_1-T_4h_2 \end{pmatrix}.$$

The determinant of the first, third, fourth, fifth and sixth rows is equal to $h_1h_3^3$. Therefore H, G, h'_4, h_1^R, h_2^R are almost everywhere functionally independent first integrals. \square

Remark 5.6. It is straightforward to verify that the normal Hamiltonian flow of the left-invariant Engel structure admits Casimir functions, i.e. functions that lie in the center of the Poisson algebra, only for examples of type I and type III. As follows from the structure equations, type I Lie algebras do not admit strictly abnormal geodesics, but from the integrability point of view they are simpler and could be worth considering. For example, type I algebras with structure constants

$$T_1 = n + m - 1, \quad T_3 = n + m - nm, \quad T_6 = -nm,$$

admit polynomial first integrals of order $n + 1$ and $m + 1$, with any $m > n \geq 0$, which are given by

$$F_1 = \left(\frac{h_3 + h_4 - (h_2 + h_3)n}{(1+m)(m-n)} \right) \left(\frac{h_4 + mn h_2 - (m+n)h_3}{(1+m)(1+n)} \right)^m,$$

$$F_2 = \left(\frac{m(h_2 + h_3) - h_3 - h_4}{(1+n)(m-n)} \right) \left(\frac{h_4 + mn h_2 - (m+n)h_3}{(1+m)(1+n)} \right)^n.$$

5.3 Minimality properties of singular geodesics

In differential geometry local minimality is usually understood in the sense that sufficiently short arcs of a curve are minimal. That means that for every point t_0 on an admissible curve γ there exists a sufficiently small interval $[t_1, t_2]$ containing t_0 such that $\gamma|_{[t_1, t_2]}$ is

the shortest curve among all admissible curves connecting $\gamma(t_1)$ and $\gamma(t_2)$. Such a curve is called geodesic.

However, in calculus of variations the word local in “local minimality” often refers to topology on a space of admissible curves. Consider a curve γ defined on $[0, T]$. We are interested whether the whole curve γ is shorter than any other sufficiently close admissible curve connecting $\gamma(0)$ and $\gamma(T)$. The answer to this question depends heavily on the topology we choose. Sobolev space topology $W^{1,\infty}$ was studied in [48] for Engel manifolds and [9] for the general case. Some results on the C^1 topology can be found in [20] and local optimality conditions for rank 2 distributions in the C^0 -topology can be found in Chapter 12 of [3]. We follow the last reference.

Definition 5.7. An admissible curve γ connecting $\gamma(0) = q_0$ with $\gamma(T) = q_T$ is called a C^0 -local minimizer if there exists a C^0 -neighbourhood U of γ , s.t. any other admissible curve $\hat{\gamma}$ in U with $\hat{\gamma}(0) = q_0$ and $\hat{\gamma}(T) = q_T$ is not longer than γ .

One can prove that C^0 -local minimality in a neighbourhood of every point of a curve implies that the curve is geodesic. Nevertheless C^0 -minimality is a stronger property that allows to understand if the whole geodesic is globally optimal in time and locally optimal among all other sufficiently close admissible curves.

The analysis of local minimality of abnormal curves is a subtle question in general. However, for Engel manifolds short pieces of abnormal curves are C^0 -local minimizers. This was proven in [3] by first establishing that abnormal curves on an Engel manifold are H^1 -local minimizers and then by showing that H^1 -local minimality implies C^0 -local minimality for continuously differentiable curves.

To determine whether or not an abnormal geodesic is a C^0 -local minimizer we investigate the presence of conjugate points along it. This can be done using the results of the previous chapters, but the necessary theory was already developed in [3]. And so we simply follow this source.

Definition 5.8 ([3]). Let $\gamma(t) = e^{tX_2}(q_0)$ be a unit speed abnormal geodesic on an Engel manifold. The moment of time $t > 0$ is called conjugate if

$$e_*^{tX_2} \mathcal{D}_{q_0} = \mathcal{D}_{\gamma(t)}.$$

Theorem 5.9 ([3]). *If an abnormal geodesic of an Engel manifold does not contain conjugate points, then it is a C^0 -local minimizer. Conversely, if a **strictly** abnormal geodesic is a C^0 -local minimizer, then it does not contain conjugate points.*

It is important to note that in general a presence of a conjugate point does not imply that the abnormal geodesic is not a C^0 -minimizer. The minimizing property depends on the number of lifts this geodesic has. If it has a unique lift to the cotangent bundle, then indeed a presence of at least one conjugate point is sufficient for non-optimality. For the general case see [8, Theorem 20.3].

The next theorem establishes necessary conditions for C^0 -local minimality of abnormal geodesics.

Theorem 5.10. *Let $\gamma(t) = e^{tX_2}(q_0)$ be a unit-speed abnormal geodesic on an Engel manifold and let*

$$\Delta_\gamma(t) = T_6(\gamma(t)) + \frac{1}{2}\dot{T}_2(\gamma(t)) - \frac{1}{4}T_2(\gamma(t))^2.$$

If $\Delta_\gamma \leq 0$ on $[0, T]$, then $\gamma|_{[0, T]}$ is C^0 -local minimizing. If $\gamma(t)$ is strictly abnormal and $\Delta \geq C > 0$, then $\gamma|_{[0, \tau]}$ is not a C^0 -local minimizer for $\tau \geq \pi/\sqrt{C}$.

Proof. Let us write down and analyse the corresponding Jacobi equation. Obviously $e_*^{tX_2}(X_2(\gamma(0))) = X_2(\gamma(t)) \in \mathcal{D}_{\gamma(t)}$. So we must consider the evolution of $A(t) = e_*^{tX_2}X_1$ along the abnormal curve $\gamma(t)$. A time $t_* > 0$ is conjugate if and only if $A(t_*)(\gamma(t_*)) \in \mathcal{D}_{\gamma(t_*)}$. Using the definition of Lie derivative we see that

$$\dot{A}(t) = [A(t), X_2]. \quad (5.15)$$

Let $A(t) = A_1(t)X_1 + A_2(t)X_2 + A_3(t)X_3 + A_4(t)X_4$. Using the structure constants (5.2) of the canonical frame and projecting equation (5.15) on $\{X_1, X_2, X_3, X_4\}$ we obtain

$$\begin{aligned} \dot{A}_1 &= -T_6A_3 - X_1(T_6)A_4, \\ \dot{A}_2 &= -T_4A_3 - X_1(T_4)A_4, \\ \dot{A}_3 &= A_1 - T_2A_3 - (T_4 + X_1(T_2))A_4, \\ \dot{A}_4 &= -T_2A_4, \end{aligned} \quad (5.16)$$

where all T_i as well as $X_1(T_i)$ are evaluated along the curve γ . The system (5.16) is linear with boundary conditions $A(0) = (1, 0, 0, 0)$ and $A_3(t_*) = A_4(t_*) = 0$ where t_* is the supposed conjugate time. Note that the first, the third and the fourth equations form a closed subsystem. Moreover from the last equation and the boundary conditions we obtain $A_4 \equiv 0$. This way we are left to study the non trivial solutions to the boundary value problem

$$\begin{aligned} \dot{A}_1 &= -T_6A_3, \\ \dot{A}_3 &= A_1 - T_2A_3, \\ A_1(0) &= 1, A_3(0) = 0, A_3(t_*) = 0. \end{aligned} \quad (5.17)$$

Using the fact that the abnormal curve is smooth since it is an integral curve of a smooth vector field, we rewrite (5.17) as a single second order ODE:

$$\begin{aligned} \ddot{A}_3 + T_2\dot{A}_3 + (T_6 + \dot{T}_2)A_3 &= 0, \\ A_3(0) = 0, A_3(t_*) = 0, \dot{A}_3(0) &= 1. \end{aligned}$$

This allows us to use the results from the oscillation theory of second order ODEs. After the change of variables

$$y = A_3 \exp\left(\int_0^t \frac{T_2(\tau)}{2} d\tau\right),$$

we get an equivalent formulation of the boundary value problem

$$\ddot{y} + \left(T_6 + \frac{\dot{T}_2}{2} - \frac{T_2^2}{4} \right) y = 0, \quad (5.18)$$

$$y(0) = 0, \quad y(t_*) = 0, \quad \dot{y}(0) = 1. \quad (5.19)$$

Now the statement of the theorem is a direct consequence of the Sturm comparison theorem. \square

In the case of left-invariant structures, i.e. when all T_i 's are constants, we get a sharp result.

Corollary 5.11. *Let $\gamma(t)$ be an abnormal curve of a left-invariant Engel structure and let $\Delta = T_6 - \frac{1}{4}(T_2)^2$. If $\Delta > 0$, then all the conjugate times are given by*

$$t_{conj} = \frac{\pi k}{\sqrt{\Delta}}, \quad \forall k \in \mathbb{Z}_+$$

and if, moreover, $\gamma(t)$ is strictly abnormal then the restriction $\gamma|_{[0,\tau]}$ is a C^0 -local minimizer if and only if $\tau < \pi/\sqrt{\Delta}$. If $\Delta \leq 0$, then the restriction $\gamma|_{[0,\tau]}$ is a C^0 -local minimizer for any $\tau \in (0, +\infty)$.

Proof. In the left-invariant case Δ is a constant. Therefore we can solve the boundary value problem (5.18)-(5.19) explicitly. \square

Bibliography

- [1] A. AGRACHEV, *Feedback-invariant optimal control theory and differential geometry, ii. jacobi curves for singular extremals*, J. Dynamical and Control Systems, (1998), pp. 583–604.
- [2] —, *Nonlinear and Optimal Control Theory*, Springer, 2004, ch. Geometry of Optimal Control Problems and Hamiltonian Systems, pp. 1–59.
- [3] A. AGRACHEV, D. BARILARI, AND U. BOSCAIN, *Introduction to riemannian and sub-riemannian geometry*. <https://webusers.imj-prg.fr/~davide.barilari/Notes.php>, Ver. Nov 17, 2017.
- [4] A. AGRACHEV AND C. BIOLO, *Switching in time-optimal problem with control in a ball*, SIAM J. on Control and Optimization, 56 (2018), pp. 183–200.
- [5] A. AGRACHEV AND R. GAMKRELIDZE, *Nonlinear Controllability and Optimal Control*, CRC Press Book, 1990, ch. Symplectic geometry for optimal control, pp. 263–277.
- [6] —, *Geometry of Feedback and Optimal Control*, CRC Press Book, 1998, ch. Symplectic methods in optimization and control, pp. 19–77.
- [7] A. AGRACHEV, L. RIZZI, AND P. SILVEIRA, *On conjugate times of lq optimal control problems*, J. Dynamical and Control Systems, 21 (2015), pp. 625–641.
- [8] A. AGRACHEV AND Y. SACHKOV, *Control theory from a geometric point of view*, Springer, 2004.
- [9] A. AGRACHEV AND S. SARYCHEV, *Strong minimality of abnormal geodesics for 2-distributions*, J. Dynamical and control systems, (1995), pp. 139–176.
- [10] A. AGRACHEV, G. STEFFANI, AND P. ZEZZA, *Quadratic mappings in the geometric control theory*, Itogi nauki. VINITI. Problemy geometrii, 20 (1988), pp. 111–205.
- [11] —, *Strong optimality of a bang-bang trajectory*, SIAM J. on Control and Optimization, 41 (2002), pp. 981–1014.
- [12] D. ALMEIDA, *Sub-riemannian symmetric spaces of engel type*, Mat. contemp., 17 (1998), pp. 45–58.
- [13] —, *Sub-riemannian homogeneous spaces of engel type*, J. Dynamic. and Control syst., 20 (2014), pp. 149–166.
- [14] V. ARNOLD AND A. GIVENTAL, *Dynamical systems iv: Symplectic geometry and its applications*, Springer, 2001, ch. Symplectic geometry, pp. 1–138.

- [15] V. ARNOLV, *The sturm theorems and symplectic geometry*, *Funct. Anal. Appl.*, 19 (1985), pp. 251–259.
- [16] A. ARONNA, F. BONNANS, A. DMITRUK, AND P. LOLITO, *Quadratic order conditions of a local minimum for abnormal extremals*, *Nonlinear Analysis: Theory, Methods & Applications*, 30 (1997), pp. 2439–2448.
- [17] —, *Quadratic order conditions for bang-singular extremals*, *Numerical Algebra, Control and Optimization*, 2 (2012), pp. 511–546.
- [18] D. BERNOULLI, *Correspondence Matématique et Physique*, ch. The 26th letter to Euler. October 1742.
- [19] B. BONNARD AND E. TRELAT, *On the role of abnormal minimizers in subriemannian geometry*, *Ann. Fac. Sci. Toulouse*, 6 (2001), pp. 405–491.
- [20] R. BRYANT AND L. HSU, *Rigidity of integral curves of rank 2 distribution*, *Invent. Math.*, (1993), pp. 435–461.
- [21] J.-B. CAILLAU, J. FEJOZ, M. ORIEUX, AND R. ROUSSARIE, *Singularities of min time affine control systems*. preprint, hal-01718345, version 1.
- [22] E. CODDINGTON AND N. LEVINSON, *Theory of Ordinary Differential Equations*, Krieger Pub. Co., 1984.
- [23] M. DE GOSSON, *Symplectic Geometry and Quantum Mechanics*, Birkhauser, 2000.
- [24] A. DMITRUK AND N. OSMOLOVSKII, *Variations of the v -change of time in problems with state constraints*, *Trudy Inst. Mat. i Mekh. UrO RAN*, 24 (2018), pp. 76–92.
- [25] J. DUISTERMAAT, *On the morse index in variational calculus*, *Adv. in Math.*, 21 (1976), pp. 173–195.
- [26] F. ENGEL, *Zur Invariantentheorie der Systeme von Pfaff'schen Gleichungen*, *Leipz. Sitzungsber*, 1889.
- [27] L. EULER, *Methodus inveniendi lineas curvas maximi minimive proprietate gaudentes, sive solutio problematis isoperimetrici lattissimo sensu accepti*, 1744. <http://eulerarchive.maa.org/pages/E065.html>.
- [28] V. GERSHKOVICH, *Dynamical systems and applications*, World scientific publishing company, 1995, ch. On simplest Engel structures on 4-manifolds, pp. 279–294.
- [29] V. GUILLEMIN AND S. STERNBERG, *Geometric Asymptotics*, American Mathematical Society, 1977.
- [30] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, 2 ed., 2012.

-
- [31] R. KOPP AND G. MOYER, *Necessary conditions for singular extremals*, AIAA Journal, 3 (1965), pp. 1439–1444.
- [32] W. LEIGHTON AND M. A., *Quadratic functionals with a singular end point*, Trans. Amer. Math. Soc., 78 (1955), pp. 98–128.
- [33] R. LEVIEN, *The elastica: a mathematical history*, tech. rep., EECS Department, University of California, Berkeley, Aug 2008.
- [34] G. LION AND M. VERGNE, *The Weil representation, Maslov index and Theta series*, Birkhauser, 1980.
- [35] A. MASHTAKOV AND Y. SACHKOV, *Superintegrability of left-invariant subriemannian structures on unimodular three-dimensional lie groups*, Differential equations, 51 (2015), pp. 1476–1483.
- [36] D. MCDUFF AND D. SALAMON, *Introduction to symplectic topology*, Oxford Science Publications, 2 ed., 1999.
- [37] R. MONTGOMERY, *Abnormal minimizers*, SIAM J. Control Optim., 32 (1994), pp. 1605–1620.
- [38] M. MORSE, *The Calculus of Variations in the Large*, American Mathematical Society, 1934.
- [39] —, *Singular quadratic functionals*, Math. Ann., 201 (1973), pp. 60–76.
- [40] M. MORSE AND W. LEIGHTON, *Singular quadratic functionals*, Trans. Amer. Math. Soc., 36 (1936), pp. 252–286.
- [41] F. OLVER AND L. MAXIMON, *NIST Handbook of Mathematical Functions*, Cambridge University Press, 2010, ch. Bessel Functions, pp. 215–286.
- [42] N. OSMOLOVSKII AND H. MAURER, *Advances in Mathematical Modeling, Optimization and Optimal Control*, Springer, 2016, ch. Second-Order Optimality Conditions for Broken Extremals and Bang-Bang Controls, pp. 147–201.
- [43] W. REID, *Riccati Differential Equations*, Academic Press, 1972.
- [44] Y. SACHKOV, *Conjugate points in euler’s elastic problem*, Journal of Dynamical and Control Systems, 14 (2008), pp. 409–439.
- [45] H. SHÄTTLER AND U. LEDZEWICZ, *Geometric Optimal Control*, Springer, 2012.
- [46] M. SHAYMAN, *Phase portrait of the matrix riccati equation*, SIAM J. Control and Optimization, 24 (1986), pp. 1–65.
- [47] S. STERNBERG, *Lectures on differential geometry*, Chelsea Pub. Co., 1983.

- [48] H. SUSSMANN, *Sub-Riemannian geometry*, Birkhauser, 1996, ch. A cornucopia of four-dimensional abnormal sub-Riemannian minimizers, pp. 341–364.
- [49] G. TESCHL, *Ordinary Differential Equations and Dynamical Systems*, American Mathematical Society, 2012.
- [50] V. VASSILEV, P. DJONDJOROV, AND I. MLADENOV, *Cylindrical equilibrium shapes of fluid membranes*, *Journal of Physics A: Mathematical and Theoretical*, 41 (1998).