

# Jacobi Fields in Optimal Control I: Morse and Maslov Indices

Andrei Agrachev\*      Ivan Beschastnyi†

## Abstract

In this paper we discuss a general framework based on symplectic geometry for the study of second order conditions in optimal control problems. Using the notion of  $L$ -derivatives we construct Jacobi curves, which represent a generalization of Jacobi fields from the classical calculus of variations. This construction includes in particular the previously known constructions for specific types of extremals. We state and prove Morse-type theorems that connect the negative inertia index of the Hessian of the problem to some symplectic invariants of Jacobi curves.

## Introduction

Consider a standard variational problem of the form

$$J[q(t)] = \int_0^T L(q, \dot{q}) dt \rightarrow \min, \quad q \in \mathbb{R}^n, \quad (1)$$

with fixed boundary conditions  $q(0) = q_0$ ,  $q(T) = q_T$ . We assume that our Lagrangian is sufficiently smooth and that the strong Legendre condition holds

$$L_{\dot{q}\dot{q}} > 0.$$

We also use the usual convention of summation over the repeating indices whenever it is possible.

If  $\tilde{q}(t)$  is a minimum, then it must be a critical point of  $J$ , i.e.

$$dJ[\tilde{q}(t)](\phi(t)) = 0, \quad \forall \phi \in C^2 : \phi(0) = \phi(T) = 0.$$

Equivalently one can say that  $\tilde{q}(t)$  satisfies the Euler-Lagrange equation which can be written in a Hamiltonian form as

$$\begin{aligned} \dot{q}^i &= \frac{\partial H}{\partial p_i}, \\ \dot{p}_i &= -\frac{\partial H}{\partial q^i} \end{aligned} \quad (2)$$

---

\*Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy, agrachevaa@gmail.com

†Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy, i.beschastnyi@gmail.com

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.

After that, one usually proceeds to studying the necessary second order conditions like  $d^2 J[\tilde{q}(t)] \geq 0$ . But one can ask a more general question of calculating the negative inertia index  $\text{ind}^- d^2 J[\tilde{q}(t)]$ . It can be done via the notion of conjugate points due to Jacobi. Namely we linearize system (2) to get the Jacobi equation

$$\begin{aligned} \dot{x}^i &= \frac{\partial^2 H}{\partial p_i \partial q^j} x^j + \frac{\partial^2 H}{\partial p_i \partial p_j} y_j, \\ \dot{y}_i &= -\frac{\partial^2 H}{\partial q^i \partial q^j} x^j - \frac{\partial^2 H}{\partial q^i \partial p_j} y_j. \end{aligned} \quad (3)$$

A moment of time  $t_c$  is called *conjugate* if there exists a non-trivial solution of the Jacobi equation (3) with boundary conditions  $x(0) = 0, x(t_c) = 0$ . The corresponding point  $\tilde{q}(t_c)$  of our extremal is called a *conjugate point* and the number of linearly independent solutions of the Jacobi boundary value problem is called the *multiplicity* of  $\tilde{q}(t_c)$ .

A famous theorem due to Morse states

**Theorem 0.1** (Morse). *The index  $\text{ind}^- d^2 J[\tilde{q}(t)]$  is equal to the number of conjugate points counted with multiplicities.*

Due to the strong Legendre condition conjugate points can not accumulate and the index must be finite.

We can give a more geometric interpretation of conjugate points using symplectic geometry. Consider the standard symplectic form on  $\mathbb{R}^{2n}$

$$\sigma((x_1, y_1), (x_2, y_2)) = \begin{pmatrix} x_1^T & y_1^T \\ -\text{id} & 0 \end{pmatrix} \begin{pmatrix} 0 & \text{id} \\ \text{id} & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

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

Notice that the Jacobi equation (3) is a Hamiltonian system on  $\mathbb{R}^{2n}$ . So its flow  $\Phi_t$  is a linear symplectic transformation and hence it maps Lagrangian planes to Lagrangian planes. Thus by fixing an initial Lagrangian plane  $L_0 = \Pi$  we get a curve  $L_t = \Phi_t(\Pi)$  in the Lagrangian Grassmanian that is known as the *Jacobi curve*. Then a moment of time  $t$  is conjugate if and only if  $L_t \cap \Pi \neq \{0\}$  and the multiplicity of the corresponding conjugate point is given by  $\dim(L_t \cap \Pi)$ .

The set of all Lagrangian planes that have a non zero intersection with a fixed Lagrangian plane  $\Pi$  is called the *Maslov train*  $\mathcal{M}_\Pi$ . We have a conjugate point whenever our curve  $L_t$  crosses it. The set  $\mathcal{M}_\Pi$  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, see also [16, 11, 19]. Thus we can reformulate the Morse index theorem as follows.

**Theorem 0.2.** *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  $L_t|_{[\varepsilon, T]}$  for some  $\varepsilon$  small enough.*

So we see that a study of the functional  $J[q]$  on an infinite dimensional space can be reduced to the study of behaviour of some curves in finite dimensional spaces. In the case of the first variation to the study of the curves defined as solutions of the Hamiltonian system (2) and in the case of the second variation to the study of the Jacobi curve  $L_t$ . Moreover the Maslov index turns out to be a very flexible tool for computing the index of Hessians in variational problems (see for example [18, 23]), due to its homotopy invariance.

A natural question is whether or not this theory generalizes to constrained variational problems, like the problems of optimal control. In the case of the first variation the answer was given by Pontryagin and his students, who showed, that under very weak conditions minimal curves satisfy a Hamiltonian system with a Hamiltonian defined by a maximum condition, that is now widely known as the Pontryagin's maximum principle (PMP).

To motivate such a generalization let us look at some possible applications of this theory to engineering and pure mathematics. First of all, using those techniques it is possible to answer certain stability questions. Indeed, many physical problems are formulated using variational principles. For example, given an elastic rod, what shapes can it take? One knows that those curves must be local minimizers of the bending energy and the overall problem can be formalized as a constrained variational problem [23]. The PMP allows us to characterize critical curves, but most of them are not going to give a local minima of the functional. This means that even using infinitesimal perturbation of a given solution, one can produce curves with a smaller value of the bending energy. Hence such solutions are not stable. Studying the second variation allows us separate stable solutions from the unstable ones. For an example of this approach see [20].

Another closely related application is motion planning. Given a control system one wants to transfer it from one state to another. Often there is an infinite number of ways of doing this (think of parking a car) and simple algorithms that allow to associate to two states a control function that steers the system between those two states are strongly desirable. One way to approach this problem is to look for minimum curves of some simple functional. Usually simplicity in this case means that the functional and the system has as many symmetries as possible which simplify its study. Then one can apply numerical algorithms or try to solve the problem by hand. Usually the last one is too difficult and a mixture of both is applied. Nevertheless with each new optimality condition we improve the speed of the convergence of the numerical solution, since we restrict to a much smaller set of candidates. This way we can obtain a better initial guess or even obtain a finite number of possibilities that can be checked by a computer. A classical example of this approach is the Reeds-Shepp car. Reeds and Shepp took the simplest model for a mobile robot and studied its time-optimal solutions [22]. They were very able to reduce the number of candidate trajectories to a finite number of configurations. This number was

later improved by Sussman and Tang using the technique of generalized envelopes [25]. There is a deep connection between this technique and the theory we develop here. All optimality conditions from [25] can be derived using our methods.

The second variation has also a purely geometric application. We can consider a variational problem on the space of curves of a differentiable manifold. For example, if the minimized functional is the length functional of a Riemannian metric, the theory of second variation allows to arrive naturally at the notion of sectional curvature. This approach was successfully applied in sub-Riemannian geometry, where curvature invariants that are directly related to the behavior of the geodesic flow were found [3]. It turns out that good models for constant curvature spaces are linear quadratic problems that are not sub-Riemannian manifolds, but which constitute a very natural class of optimal control problems. Using this class of systems in [14] the authors were able to prove comparison theorems, and one can hope that a similar theory can be developed for much more general constrained variational problems on manifolds.

Finally, many applications of PMP have produced various examples of bad behaving minimizers, which do not need to be smooth, which can accumulate an infinite number of discontinuities in finite time [26] or even be chaotic [27]. Thus it may not be even possible to write down a Jacobi equation like in the situation above. Nevertheless surprisingly it was proven in [6, 1] that there exists a Jacobi curve  $L_t$  in the Lagrangian Grassmanian and a Maslov type theory that allows to codify all the information about the second variation. The basis of this general theory is the notion of  $\mathcal{L}$ -derivatives. The goal of this paper is to explain this theory, to demonstrate some old and new ways to compute  $\mathcal{L}$ -derivatives, how to construct the corresponding Jacobi curves and how to extract from them information about the Hessian of the functional along an extremal curve. We note that the notion of a conjugate point and Morse index theorems in optimal control were previously known for some specific cases of extremals like regular, singular or bang-bang (see, for example, [15, 24, 13, 12] and references there in). In this work we develop a general unified framework based on symplectic geometry, that can be applied to all the previously known cases and beyond.

## 1 $\mathcal{L}$ -derivatives for optimal control problems

### 1.1 General definitions

A  $\mathcal{L}$ -derivative is a rule that for a given constrained 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 bigger and bigger spaces of variations. Using this theory one can recover the classical theory of Jacobi and much more.

We consider the following setting. Let  $J : \mathcal{U} \rightarrow \mathbb{R}$  be a functional and  $F : \mathcal{U} \rightarrow M$  be a map with smooth finite-dimensional restrictions, where  $\mathcal{U}$  is a smooth Banach manifold and  $M$  is a finite-dimensional manifold. Given a point  $q \in M$ , we are interested in finding

$\tilde{\omega} \in F^{-1}(q)$  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 the next subsection.

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

$$\langle \lambda, dF[\tilde{\omega}](w) \rangle - \nu dJ[\tilde{\omega}](w) = 0, \quad \forall w \in T_{\tilde{\omega}}\mathcal{U}. \quad (4)$$

A pair  $(\tilde{\omega}, \lambda)$  that satisfies the equation above is called a *Lagrangian point* and  $\tilde{\omega}$  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 it *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 study the *Hessian* of a pair  $(J, F)$  at a Lagrangian point  $(\tilde{\omega}, \lambda)$  defined as

$$\text{Hess}(F, \nu J)[\tilde{\omega}, \lambda] := (\nu d^2 J[\tilde{\omega}] - \langle \lambda, d^2 F[\tilde{\omega}] \rangle) |_{\ker dF[\tilde{\omega}]}. \quad (5)$$

The index and the nullity of the Hessian are directly related to optimality of the critical point  $\tilde{\omega}$ . Namely we have

**Theorem 1.1** ([8]). *Let  $(F, J) : \mathcal{U} \rightarrow M \times \mathbb{R}$  be a pair of smooth maps and  $\tilde{\omega}$  be a critical point with parameter  $\nu$ . Then if for any Lagrangian point  $(\tilde{\omega}, \lambda)$*

$$\text{ind}^- \text{Hess}(F, \nu J)[\tilde{\omega}, \lambda] \geq \text{codim im } d(F, J)[\tilde{\omega}],$$

*the critical point  $\tilde{\omega}$  is not optimal.*

In the normal case the Hessian of  $(J, F)$  coincides with the second derivative of  $J$  on the level set of  $F^{-1}(q)$  at a Lagrangian point  $(\tilde{\omega}, \lambda)$ . Indeed, let  $\omega(s)$  be a curve in  $F^{-1}(q)$ , s.t.  $\omega(0) = \tilde{\omega}$ . Then by differentiating twice  $F(\omega(s)) = q$  at  $s = 0$  we find that

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

Similarly we find that

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

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

We are now ready to define  $\mathcal{L}$ -derivatives. We linearise (4) with respect to  $\lambda$  and  $\omega$ , and obtain the following equation

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

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

$$\langle \xi, dF[\tilde{\omega}](w) \rangle + Q(v, w) = 0. \quad (6)$$

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

**Definition 1.1.** Let  $(F, J)$  be maps with smooth finite-dimensional restrictions. A  $\mathcal{L}$ -derivative of  $(F, J)$  at a Lagrangian point  $(\tilde{\omega}, \lambda)$  constructed over a finite-dimensional space of variations  $V \subset T_{\tilde{\omega}}\mathcal{U}$  that we denote as  $\mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda](V)$  is the set of vectors  $(\xi, dF[\tilde{\omega}](v)) \in T_{\lambda}(T^*M)$ , s.t.  $(\xi, v) \in (T_{\lambda}(T_q^*M), V)$  solve (6) for all  $w \in V$ .

This set is a Lagrangian plane if  $V$  is finite-dimensional [2]. The reason why we do not take directly  $T_{\tilde{\omega}}\mathcal{U}$  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  $\mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda](V)$  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  $\mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda](V)$  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. For reader's convenience we recall some definitions.

**Definition 1.2.** 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  $\gamma$ , s.t.  $\alpha \preceq \gamma$  and  $\beta \preceq \gamma$ .

**Definition 1.3.** 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  $T_{\tilde{\omega}}\mathcal{U}$  form a directed set with a partial ordering given by the inclusion  $V \subset W$ . This motivates the following definition

**Definition 1.4.** A  $\mathcal{L}$ -derivative of  $(F, J)$  at a Lagrange point  $(\tilde{\omega}, \lambda)$  constructed over a subspace  $V \subset T_{\tilde{\omega}}\mathcal{U}$  is the generalized limit

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

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

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

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

**Theorem 1.2.** Let  $(\tilde{\omega}, \lambda)$  be a Lagrangian point of  $(F, J)$ .

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

The second point is especially important, since it actually allows to compute the  $\mathcal{L}$ -derivative. For example, if  $\mathcal{U}$  is modelled over a separable Hilbert space we can take a dense subspace spanned by vectors  $e_1, e_2, \dots$  and define subspaces  $V_i = \text{span}\{e_1, \dots, e_i\}$ ,  $i \in \mathbb{N}$ . Then we can compute the  $\mathcal{L}$ -derivative as a limit

$$\mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda] = \lim_{i \rightarrow \infty} \mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda](V_i)$$

and there is no need for a generalized sequence in this case.

**Remark 1.1.** Another simple but important property is the following. Suppose that we have a smooth map  $G : \mathcal{V} \rightarrow \mathcal{U}$  which is a submersion. And let  $\tilde{v}$  be any preimage of a critical point  $\tilde{\omega}$  under  $G$ . Then it is easy to check that  $(\tilde{v}, \lambda)$  is a Lagrange point of  $(F \circ G, J \circ G)$  and that

$$\mathcal{L}(F, \nu J)[\tilde{\omega}, \lambda] = \mathcal{L}(F \circ G, \nu J \circ G)[\tilde{v}, \lambda].$$

$\mathcal{L}$ -derivatives contain information about the inertia indices and nullity of the Hessian (5) restricted to some space of variations. The strength of this technique lies in the fact, that by comparing relative position of two  $\mathcal{L}$ -derivatives constructed over two subspaces  $V \subset W$ , we can see how the inertia indices of the Hessian change as we add variations to our variations space. We have gathered some simple facts of this kind together with a proof of Theorem 1.2 in Appendix B.

## 1.2 Optimal control problems with no constraints on the control

Let us consider the following optimal control problem

$$\dot{q} = f(u, q), \quad u \in U \subset \mathbb{R}^k, \quad q \in M, \quad (7)$$

$$J_T(u) = \int_0^T L(u, q) dt \rightarrow \min. \quad (8)$$

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. Moreover we assume that the final time  $T$  is fixed and that  $q(0)$  can be free and lie in some submanifold  $N_0 \subset M$ , 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 moving starting point by making a change of time variable  $t \mapsto T - t$ . For now we simply assume that  $U = \mathbb{R}^k$ . We will see in the next subsection how the general case can be reduced to this one.

**Definition 1.5.** An admissible pair  $\omega(t)$  is a pair  $(q(t), u(t))$  which consists of a Lipschitz trajectory  $q(t)$  that satisfies (7) for almost every  $t$  and the corresponding control  $u(t)$ .

In [2] the following result was proven

**Proposition 1.1.** If  $\dim M = n$ , then the set of admissible pairs  $\Omega$  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  $\tilde{q}(\tau) = \pi(\tilde{\omega}(\tau))$  which 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 (7) at a moment of time  $\tau$ , 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 1.6.** *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 optimal control problems. It takes a control  $u(t)$  and maps it to the end of the corresponding trajectory that begins at  $q_0 \in M$ . Similarly if  $q_0$  is not fixed, but lies in a manifold  $N_0$ , we define an analog of the classical end-point map as

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

We can then proceed to the construction of the  $\mathcal{L}$ -derivative of  $(E_{N_0, T}, J_T)$ , that will contain information about the Hessian of this problem.

In the most well studied case of regular extremals  $\mathcal{L}(E_T, \nu J_T)[\tilde{\omega}, \lambda(T)]$  is the set of values of all Jacobi fields at a moment of time  $T$  that have a zero projection on  $TM$  at time 0. If we assume that  $q_0$  is free, then we need to modify accordingly the boundary condition at zero, but the meaning of the  $\mathcal{L}$ -derivative is essentially the same. By constructing these  $\mathcal{L}$ -derivatives for all times  $t \in [0, T]$ , and not just for the final time  $T$ , we reconstruct the whole set of Jacobi fields.

**Remark 1.2.** It is important to note that the definition of the end-point map, evaluation map, admissible curves etc. are all intrinsic, i.e. they do not depend on the choice of coordinates in  $\mathcal{U}$  and  $M$ . This means that the  $\mathcal{L}$ -derivatives that we will construct are intrinsic 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 (7) 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  $\tau$  as we want, the corresponding  $\mathcal{L}$ -derivative will be the same for all  $\tau$ . 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 neighbourhood 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  $F_\tau$  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  $L_k^\infty[0, T] \times \mathbb{R}^{\dim N_0}$ .

Let  $\Omega_{N_0}^T = F_0^{-1}(N_0)$  be the space of all solutions of (7) starting at  $N_0$  up to time  $T$ . If  $\tilde{\omega}$  is a critical point of  $(E_{N_0, T}, J_T)$ , then we have

$$\langle \lambda(T), dE_{N_0, T}[\tilde{\omega}](w) \rangle - \nu dJ_T[\tilde{\omega}](w) = 0, \quad \forall w \in \Omega_{N_0}^T. \quad (9)$$

Here  $(\lambda(T), -\nu) \in T_{E_{N_0, T}(\tilde{\omega})}^* \times \mathbb{R}$  are the Lagrange multipliers, where  $\nu$  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

$$\langle \hat{\lambda}(t), d\hat{E}_{N_0, t}[\tilde{\omega}](w) \rangle = 0, \quad (10)$$

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

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

Let us denote the flow of this system with  $u = \tilde{u}$  from time  $t_0$  till time  $t_1$  by  $\hat{P}_{t_0}^{t_1}$ . We also write  $\hat{P}^t$  for  $\hat{P}_0^t$ . We use the non-hatted notation  $P^t$  for the flow of the original control system (7) with the same control.

Since  $d\hat{E}_{N_0, T}[\tilde{\omega}]|_{\Omega_{N_0}^T \cap L_k^\infty[0, T]} = (\hat{P}_T^T)^* d\hat{E}_{N_0, t}[\tilde{\omega}]$ , by restricting (9) to  $L_k^\infty[0, t]$  we find that

$$\langle \hat{\lambda}(t), d\hat{E}_{N_0, t}[\tilde{\omega}](w) \rangle = \langle \lambda(t), dE_{N_0, t}[\tilde{\omega}](w) \rangle - \nu dJ_t[\tilde{\omega}](w) = 0, \quad \forall w \in T_{\tilde{\omega}}\Omega_{N_0}^T, \quad (11)$$

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  $y$ .

If we define a Hamiltonian

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

then 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 (11) to  $w \in L_k^\infty[0, t]$  then we get a condition of the form

$$\left. \frac{\partial h(u, \lambda(t))}{\partial u} \right|_{u=\tilde{u}} = 0, \quad (12)$$

which is a weak form of the maximum principle. If we restrict the equation (11) to  $w \in T_{\tilde{q}(0)}N_0$  we obtain instead the transversality conditions

$$\lambda(0) \perp T_{\tilde{q}(0)}N_0. \quad (13)$$

For completeness we give a derivation of these two conditions in Appendix A.

**Definition 1.7.** A *Jacobi curve* of an optimal control problem (7)-(8) is the family of time parameterized Lagrangian subspace  $\mathcal{L}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)]$ .

It is important to note that Jacobi curves are feedback invariant. Indeed,  $\mathcal{L}$ -derivatives remain unchanged under diffeomorphism of  $\Omega_{N_0}$  as discussed in Remark 1.1 and feedback transformations just constitute a special case.

Definition 1.7 is quite natural in the light of the previous discussion, but it also has a small problem, namely  $\mathcal{L}$ -derivatives from the definition are Lagrangian subspaces in different symplectic spaces  $T_{\lambda(t)}(T^*M)$ , and if we want to study their relative positions, we need a way to identify them. In order to fix this problem, let us consider the Hamiltonian flow of  $\vec{h}_{\tilde{u}(t)}$ , that we denote as  $\Phi_t$ . We are going to compute  $\Phi_t^* \mathcal{L}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)](V)$ . For brevity we denote this Lagrange subspace 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  $\Phi_t^*$  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  $T_{\lambda(0)}(T^*M)$ .

The next step is to write down explicitly the equation that defines  $\mathcal{L}_t(V)$ . 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 (11) it is enough to linearize the Hamiltonian system, the maximum condition and the transversality conditions w.r.t. both phase variables and control variables. Since we are interested in the pull-back of the Jacobi curve under the pullback  $\Phi_t^*$ , we 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

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

where

$$H(t, u, \cdot) = (h(u, \cdot) - h(\tilde{u}, \cdot)) \circ \Phi_t.$$

The maximum principle now says that any extremal control  $u$  must satisfy the weak maximum condition (11), 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 (13):

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

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.

Now we can finally write down explicit equations that define  $\mathcal{L}_t(V)$ . Let us define a time-dependent vector field

$$X(t) := \left. \frac{\partial \vec{H}(t, u, \lambda(0))}{\partial u} \right|_{u=\tilde{u}}$$

and a quadratic form

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

First we linearise the Hamiltonian system at  $(\tilde{q}(0), \lambda(0))$ . We get

$$\dot{\eta}(t) = X(t)v(t) \iff \eta(t) = \eta_0 + \int_0^t X(\tau)v(\tau)d\tau. \quad (14)$$

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

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

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

Finally we linearize the maximum condition to obtain

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

Using the definitions we gave before, we can write

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

Collecting all the formulas proves the following result.

**Proposition 1.2.** *An  $\mathcal{L}$ -derivative  $\mathcal{L}_t(V)$  over a subspace  $V \in L_k^\infty[0, t]$  consists of vectors of the form*

$$\eta(t) = \eta_0 + \int_0^t X(\tau)v(\tau)d\tau. \quad (15)$$

where  $\eta_0 \in T_{\tilde{q}_0}^\perp N_0 \times T_{\tilde{q}_0} N_0$  and  $v \in V$  satisfy

$$\int_0^t \left( \sigma \left( \eta_0 + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)(v(\tau), w(\tau)) \right) d\tau = 0, \quad \forall w(t) \in V. \quad (16)$$

As we have discussed before in the case when  $U = \mathbb{R}^k$ , the full  $\mathcal{L}$ -derivative is defined as  $\mathcal{L}_t = \lim \mathcal{L}_t(V)$ .

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 (42) and (44) in Appendix A, 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. This allow us to prove a simple, but important lemma.

**Lemma 1.1.** *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  $\mathcal{L}_t(W \cap \Omega_{N_0}^t) \in O_{\mathcal{L}_t}$ . 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  $\mathcal{L}_t(V_\beta \cap \Omega_{N_0}^t) \in O_{\mathcal{L}_t}$  for any  $\beta \supset \alpha$ . By refining the partition  $\beta$  on  $[t_0, t_{N-1}]$  we obtain

$$\lim \mathcal{L}_t(V_\beta \cap \Omega_{N_0}^t) = \mathcal{L}_t(\Omega_{N_0}^{t_{N-1}}) = \mathcal{L}_{t_{N-1}} \in O_{\mathcal{L}_t}.$$

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

### 1.3 Optimal control problems with constraints

Let us now consider the case when set  $U \subset \mathbb{R}^k$  is a disjoint 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.

In order to compute the  $\mathcal{L}$ -derivatives we need to take derivatives of the functional and the end-point map, which implies that we use two sided variations. But when we are

on the boundary of the manifold of admissible pairs, we can not variate in all directions. Definition 1.7 still makes sense if we just compute  $\mathcal{L}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)](V)$ , where  $V$  is the set of admissible two-sided variations. However, it does not fix the problem entirely that can be seen very well from the example, when  $U$  is a polytope and the extremal curve is bang-bang. This is a type of trajectory for which the extremal control  $\tilde{u}$  takes values in the vertices of the polytope. Therefore there are no two-sided variations at all. To solve this issue, we introduce additional variations in our problem so that the set of two sided variations is never empty and define Jacobi curves as  $\mathcal{L}$ -derivatives constructed over the it. This will be enough to enclose all the known results and to generalize them.

Variations that we need are called time variations. Basically we introduce a new time variable  $\tau$ , given by

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

We assume that  $u_0(s) > -1$  and we take the time variable as a new state variable

$$\dot{t} = 1 + u_0(\tau)$$

since the final time is fixed.

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

$$\begin{aligned} \frac{dq}{d\tau} &= (1 + u_0(\tau))f(q, u(t(\tau))), \\ \dot{t} &= 1 + u_0(\tau) \end{aligned}$$

and the functional to

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

If  $\tilde{u}$  was an optimal control for (7)-(8), 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.

**Remark 1.3.** Time variations were previously used to derive necessary and sufficient condition of bang-bang arcs [10, 5] and with small modification of their definition, one can even prove a version of the maximum principle [17]. 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 (see the discussion in [4]). 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.

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.

Now we construct the  $\mathcal{L}$ -derivatives. Note that since  $U$  is embedded in  $\mathbb{R}^k$ , the operators  $X(t) : \mathbb{R}^k \rightarrow T_{\lambda(0)}(T^*M)$  and quadratic forms  $b(t) : \mathbb{R}^k \times \mathbb{R}^k \rightarrow \mathbb{R}$  are well defined.

Now we must enlarge the space of variations by including time variations. After doing this, we remain in the same class of problems as before. So we keep the notations  $X_t$  and  $b_t$  for the Hamiltonian vector field and the corresponding Hessian for the new system defined above.

Since we are interested only in two-sided variations, we must restrict the operator  $\tilde{X}(t)$  to the tangent spaces of  $U_i$ . Since by assumptions each  $U_i$  is embedded in  $\mathbb{R}^k$ , 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  $\chi_{U_i}$  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. In order to reduce the problem to  $U = \mathbb{R}^k$ , we simply replace  $X(t)$  and  $b(t)$  by  $X(t)\pi_t$  and  $b(t)(\pi_t, \pi_t)$ . By doing this, we essentially exploit Remark 1.1.

This way we have reduced our problem to a problem without the constraints on the control and from now on we assume that the variations  $v(t)$  can take any value in  $\mathbb{R}^k$ . To simplify the notations we continue to write  $X(t)$  and  $b(t)$  instead of  $X(t)\pi_t$  and  $b(t)\pi_t$ .

**Remark 1.4.** We note that Jacobi curves will depend on the decomposition of  $U$  into a disjoint union of manifolds  $U_i$ . For examples, if  $U$  is a union of two intersecting lines, we can assume that the intersection point belongs either to one line or the other. But for the most common choice of  $U$  as a polytope, the most natural decomposition would be just to take as  $U_i$  different faces of various dimensions.

## 1.4 An algorithm for computing the $\mathcal{L}$ -derivative

Equations (16) can be used to construct approximations of  $\mathcal{L}$ -derivatives, which is morally the Galerkin method. In this subsection we will see that if we choose piece-wise constant functions as our basis elements, then the finite dimensional approximation of (16) takes a nice block triangular form, which can be solved explicitly, yielding an effective algorithm for constructing approximations of the Jacobi curve.

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  $E_{N_0, T}|_{\Omega_{N_0}^t} = E_{N_0, t}$  for  $t \leq T$  and  $L_k^2[0, t] \subset L_k^2[0, T]$  is an isometrical embedding, we have that the existence at a moment of time  $T$  implies the existence for all  $t \leq T$ .

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

**Lemma 1.2** (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

$$\int_{t_1}^{t_2} \left[ \sigma \left( \lambda + \int_{t_1}^{\tau} X(\theta) v_2(\theta) d\theta, X(\tau) w(\tau) \right) + b(\tau) (v_2(\tau), w(\tau)) \right] d\tau = 0, \quad \forall w(\tau) \in V_2, \quad (17)$$

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

$$\left\{ \lambda + \int_{t_1}^{t_2} X(\tau) v(\tau) d\tau : \lambda \in \mathcal{L}_{t_1}, v(t) \in V_2 \text{ satisfies (17) for any } w(t) \in V_2 \right\}.$$

*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  $V_1 \oplus V_2 = V \subset \Omega_{N_0} t_2$ , where  $V_1$  is a span of a countable dense subset in  $\Omega_{N_0}^{t_1}$  that includes variations of the initial point. Denote by  $\pi_i$  the projection onto  $V_i$ .

Now fix a neighbourhood  $O_{L_2}$  in the Lagrangian Grassmanian and consider a finite-dimensional subspace  $W \subset V$ , s.t. for any finite dimensional  $U \supset W$  we have  $\mathcal{L}_t(U) \in O_{L_2}$ . 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  $\mathcal{L}_t(V_1 \oplus \pi_2(W)) \subset O_{L_2}$ . By taking finer and finer  $O_{L_2}$  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 (17) 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.

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

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

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, and thus 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  $V_D \subset \Omega_{N_0}^t$  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 1.8.** 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 1.3.** *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 1.3.** *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  $\eta_i$  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

$$\int_t^{t+\varepsilon} \sigma \left( \eta + \frac{1}{\varepsilon} \int_t^\tau X(\theta) d\theta \cdot v, X(\tau) w \right) + \frac{b(\tau)(v, w)}{\varepsilon} d\tau = 0, \quad \forall w \in \mathbb{R}^k, \quad (18)$$

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 (18) is well defined on the corresponding quotients.

By definition we have

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

Suppose that the dimension of this space is equal to  $n-l$ , where  $l \leq \min\{k, n\}$ . But since it is defined as the solution space of 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

$$\sigma \left( \eta, \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \right) = 0, \quad \forall \eta \in \mathcal{L}_t(V) \quad \Rightarrow \quad \frac{1}{\varepsilon} \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \in \mathcal{L}_t(V) \quad (19)$$

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

We note that vectors (19) 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 (18) 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 (18) with  $v \in E^\perp$ . Then 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 (18) 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  $\eta$ , 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  $E$  already contains all the vectors  $v$  s.t.

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

This means that

$$v \mapsto \int_t^{t+\varepsilon} X(\tau) d\tau \cdot v \tag{20}$$

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.  $\square$

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 3.

## 2 Jacobi fields and the glueing formula

In the previous subsection in order to construct an approximation we have used the additivity Lemma 1.2 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. The control function of Fuller extremals exhibits a countable number of jump discontinuities in finite time. A typical example, is when a bang-bang arc with an infinite number of switches of increasing frequency is followed by a singular arc and the followed by another bang-bang arc with an infinite number of switches of decreasing frequency.

An algorithm for computation of  $\mathcal{L}$ -derivatives in the bang-bang case is known (see [7]). If we have a bang-singular arc, we can use it to construct the  $\mathcal{L}$ -derivative of the bang-bang arc, then we apply the additivity Lemma 1.2 and after this we can compute separately the Jacobi curve of the singular arc using Jacobi differential equations (see [1]) with the correct boundary conditions.

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 glueing the two.

The answer to this question is yes, but we need to 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 1.7. What this object represents is the set of possible boundary values of all possible Jacobi fields without fixing any of the two end-points. As we vary  $t_0$  or  $t_1$ , we can reconstruct the whole Jacobi field.

In order to compute  $\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[\tilde{\omega}, (-\lambda(t_0), \lambda(t_1))]$ , 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 1.2, 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  $\lambda(t_0) = (P_{t_0}^{t_1})^* \lambda(t_1)$ , allowing us to recover the Hamiltonian system.

Since  $\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[\tilde{\omega}, (-\lambda(t_0), \lambda(t_1))]$  is a Lagrangian subspace in  $T_{-\lambda(t_0)}(T^*M) \times T_{\lambda(t_1)}(T^*M)$ , with a symplectic form  $(-\sigma_{-\lambda(t_0)} \times \sigma_{\lambda(t_1)})$ , it makes sense to characterize

explicitly Lagrangian planes in a square  $(V \times V, -\sigma \times \sigma)$ . One can construct some simple examples first. A direct product  $\Lambda_1 \times \Lambda_2$  of Lagrangian planes obviously will be a Lagrangian plane in  $(V \times V, -\sigma \times \sigma)$ . On the other hand it is easy to see that a graph of any symplectic map  $M : V \rightarrow V$  is also a Lagrangian subspace. It turns out that the general situation is an interpolation between those two.

We recall that the skew-orthogonal complement  $\Gamma^\perp$  of a linear subspace  $\Gamma$  in a symplectic space  $\Sigma$  is the set

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

For an isotropic subspace  $\Gamma$  by definition we have  $\Gamma \subset \Gamma^\perp$ . One can check that  $\Gamma^\perp/\Gamma$  is a symplectic space, where the symplectic form is simply given by the restriction of  $\sigma$ .

**Lemma 2.1.** *Any Lagrangian subspace  $\Lambda$  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  $\pi_i$  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  $\Sigma$ . 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  $\Lambda'$  is a linear subspace that projects onto each  $\Gamma_i^\perp/\Gamma_i$ . But this can happen if and only if  $\dim \Gamma_1^\perp/\Gamma_1 = \dim \Gamma_2^\perp/\Gamma_2$ . Therefore  $\Lambda'$  must be a graph of a symplectic mapping and  $\dim \Gamma_1 = \dim \Gamma_2$ , like in the example we discussed above.  $\square$

As we have already mentioned,  $\mathcal{L}(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})[\tilde{\omega}, (-\lambda(t_0), \lambda(t_1))]$  is morally the set of all Jacobi fields without any restrictions on the boundary. We can consider the set of Jacobi fields of a fixed point problem as a subset of this set. More precisely in language of  $\mathcal{L}$ -derivatives this means that  $\mathcal{L}(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})[\tilde{\omega}, (-\lambda(t_0), \lambda(t_1))]$  contains all information about the  $\mathcal{L}$ -derivative of various versions of the end-point map, that we have considered so far, and we can reconstruct the latter one from the former one.

**Lemma 2.2.** *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  $\omega \in F_{t_0}^{-1}(N_0)$  and  $\lambda(t_0) \in T_{F_{t_0}(\omega)}^\perp N$ , then  $(\omega, \lambda(t_1))$  is a Lagrange point of  $F_{t_1}|_{F_{t_0}^{-1}(N_0)}$  and the corresponding  $\mathcal{L}$ -derivative can be computed as*

$$\begin{aligned} \mathcal{L}(F_{t_1}, \nu J_{t_0}^{t_1})[\omega, \lambda(t_1)](T_\omega F_{t_0}^{-1}(N_0)) &= \\ &= \pi_1 \left( \mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[\omega, (-\lambda(t_0), \lambda(t_1))] \cap \pi_0^{-1}(T_{F_{t_0}(\omega)}^\perp N_0 \times T_{F_{t_0}(\omega)} N_0) \right). \end{aligned}$$

where we have identified  $T_{F_{t_0}(\omega)}^\perp N_0 \times T_{F_{t_0}(\omega)} N_0$  with a subset of  $T_{-\lambda(t_0)}(T_{F_{t_0}(u)}M)$ .

To better understand this formula one should read it from right to left. If we take the preimage of the "initial" Lagrangian plane  $T_{F_{t_0}(\omega)}^\perp N_0 \times T_{F_{t_0}(\omega)} N_0$ , intersect it with the  $\mathcal{L}$ -derivative of the pair and project this intersection to  $T_{\lambda(t_1)} T^*M$ , we obtain exactly the  $\mathcal{L}$ -derivative of the corresponding optimal control problem with the initial point lying in a manifold  $N_0$ .

*Proof.* Since the formulation only involves projection maps, it is clear that there will be no problem in taking the generalized limits. Thus we can essentially treat all the spaces of variations as if they were finite. We use the special coordinates mentioned in Remark 1.2. Since  $F_{t_0}$  is a submersion, its differential has maximum rank and we can always find coordinates in which  $d^2 F_0[\omega] = 0$ . We split the space of variations into a direct sum  $V_0 \oplus V_1$ , where  $V_1$  is a subspace isomorphic to  $T_{F_{t_0}(\omega)} N_0 \times L^2[t_0, t_1]$  and  $V_0$  is the orthogonal complement of  $V_1$  with respect to the quadratic form  $Q = \langle \lambda(t_1), d^2 F_{t_1}[\omega] \rangle - \nu dJ_{t_0}^{t_1}[\omega]$ , i.e.

$$\langle \lambda(t_1), d^2 F_{t_1}[\omega](v_0, w_1) \rangle - \nu dJ_{t_0}^{t_1}[\omega](v_0, w_1) = 0, \quad \forall v_0 \in V_0, w_1 \in V_1. \quad (21)$$

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

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

If we restrict all maps to  $T_\omega(F_{t_0}^{-1}(N_0))$ , or equivalently we take  $w_0 = 0$ , then we obtain exactly conditions for  $(\omega, \lambda(t_1))$  being a Lagrange point of  $(F_{t_1}, J_{t_0}^{t_1})|_{F_{t_0}^{-1}(N_0)}$ .

The second part also is just a consequence of the basic definitions. In these coordinates by (21) we have the following equation for the  $\mathcal{L}$ -derivative of  $(F_{t_0}, F_{t_1}, J_{t_0}^{t_1})$ :

$$\begin{aligned} & \langle \xi(t_1), dF_{t_1}[\omega](w_0) \rangle + \langle \xi(t_1), dF_{t_1}[\omega](w_1) \rangle - \\ & - \langle \xi(t_0), dF_{t_0}[\omega](w_0) \rangle - \langle \xi(t_0), dF_{t_0}[\omega](w_1) \rangle + \\ & + \langle \lambda(t_1), d^2 F_{t_1}[\omega](v_0, w_0) \rangle + \langle \lambda(t_1), d^2 F_{t_1}[\omega](v_1, w_1) \rangle = \\ & = \nu d^2 J_{t_0}^{t_1}[\omega](v_0, w_0) + \nu d^2 J_{t_0}^{t_1}[\omega](v_1, w_1). \end{aligned} \quad (22)$$

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

$$\langle \xi(t_1), dF_{t_1}[\omega](w_1) \rangle + \langle \lambda(t_1), d^2 F_{t_1}[\omega](v_1, w_1) \rangle = \nu d^2 J_{t_0}^{t_1}[\omega](v_1, w_1). \quad (23)$$

Let  $(v_1, \xi(t_1))$  be a solution of the last equation. If we assume that  $\xi(t_0) \in T_{\tilde{q}(0)}^\perp N_0$ , then

$$\langle \xi(t_0), dF_{t_0}[\tilde{\omega}](w_0) \rangle = K(\xi(t_0), w_0)$$

gives a non-degenerate coupling between  $T_{F_{t_0}(\omega)}^\perp N_0$  and  $V_0$ . Thus clearly the quadruple

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

is a solution of (22), where

$$(dF_{t_0}[\omega])^+ v = \begin{cases} (dF_{t_0}|_{V_0})^{-1} v & \text{if } v \in \text{im } dF_{t_0}[\omega]|_{V_0}, \\ 0 & \text{if } v \notin \text{im } dF_{t_0}|_{V_0}. \end{cases}$$

But, since  $v_0 = 0$ , all of those solutions will indeed lie in  $\pi_0^{-1}(T_{F_{t_0}(\omega)}^\perp N_0 \times T_{F_{t_0}(\omega)} N_0)$ . This way we get the left inclusion.

On the other hand if we take an element from

$$\mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})[\omega, (-\lambda(t_0), \lambda(t_1))] \cap \pi_0^{-1}(T_{F_{t_0}(\omega)}^\perp N_0 \times T_{F_{t_0}(\omega)} N_0),$$

then it corresponds to a solution of (22) with  $v_0 = 0$  for any  $w = w_0 + w_1$ . In particular, (22) is satisfied for  $w_0 = 0$ . But in this case (22) reduces to (23).  $\square$

We have the following important property of this  $\mathcal{L}$ -derivative.

**Theorem 2.1** (Chain rule). *Let  $x_{01} \in \mathcal{L}(F_{t_0}, F_{t_1}, \nu J_{t_0}^{t_1})$  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  $(\pi_0(x_{01}), \pi_2(x_{12})) \in \mathcal{L}(F_{t_0}, F_{t_2}, \nu J_{t_0}^{t_2})$ .*

**Remark 2.1.** For simplicity we do not indicate explicitly the Lagrange points assuming that they are the same as in the rest of this section.

*Proof.* To prove the statement 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  $T_{F_{t_1}(\omega)} M$ ,  $V_0$  is isomorphic to  $L_k^2[t_0, t_1]$  and  $V_2$  to  $L_k^2[t_1, t_2]$ . 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 statement we can assume that the  $\mathcal{L}$ -derivatives are constructed using piecewise constant functions. Let us fix three neighborhoods  $O_{ij} \ni \mathcal{L}(F_{t_i}, F_{t_j}, \nu J_{t_i}^{t_j})$ . 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 elements of  $\mathcal{L}$ -derivatives constructed over finite-dimensional approximations, by taking  $O_{ij}$  smaller and smaller, we get that the chain rule holds in infinite dimensions as well.  $\square$

Finally, the next corollary explains how to glue together two Jacobi curves.

**Corollary 2.1.** *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 2.1 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.* 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.

In view of the previous theorem it only remains to exploit the chain 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  $\Gamma_{\ker}$  can be identified with a subspace in  $\tilde{\Gamma}_1^\angle / \tilde{\Gamma}_1$ . 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  $(\Gamma_1 + \tilde{\Gamma}_1)^\angle / (\ker \sigma|_{\Gamma_1 + \tilde{\Gamma}_1})$  can be identified with a subspace in  $\Gamma_1^\angle / \Gamma_1$  and at the same time with a subspace in  $\tilde{\Gamma}_1^\angle / \tilde{\Gamma}_1$ . Thus we see that a graph of the map  $\Phi_{02}$  is going to be a Lagrangian subspace in  $\Phi_{01}^{-1}(\Gamma_1^\angle / \Gamma_1) \times \Phi_{12}(\tilde{\Gamma}_1^\angle / \tilde{\Gamma}_1)$ , 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.  $\square$

### 3 Symplectic geometry and Morse type theorems

Under Morse type theorems we understand results that relate the behaviour of the Jacobi curve or its approximation with the index or the kernel of the corresponding Hessian. Before stating these results we need some definitions and theorems from symplectic geometry. We should note that the sign conventions in this work differ from the sign conventions in some of the previous articles like [9, 7]. The reason for this is that we would like to handle the normal and abnormal cases in a unified manner, and that we assume that the classical Jacobi equation for regular extremals produces a monotone increasing curve in the Lagrangian Grassmanian in the sense of Definition 3.1.

### 3.1 Linear symplectic geometry

Let  $(\Sigma, \sigma)$  be a symplectic space and denote by  $L(\Sigma)$  the corresponding Lagrangian Grassmanian. We will denote by  $\Pi^\natural$  the set of all Lagrangian planes transversal to a given Lagrangian plane  $\Pi \in L(\Sigma)$ .

Fix  $\Pi \in L(\Sigma)$  and choose a plane  $\Delta \in \Pi^\natural$  transversal to it. Then  $\Sigma = \Pi \oplus \Delta$  and any  $L \in \Pi^\natural$  can be identified with a graph  $(x, Sx)$  of a map  $S : \Delta \rightarrow \Pi$ . One can easily check that  $S$  defines a Lagrangian subspace if and only if  $S$  is symmetric. This construction gives a local chart, which allows to identify locally  $L(\Sigma)$  with the space of symmetric matrices.

A more intrinsic but essentially the same way to identify a neighbourhood of  $\Delta \in \Pi^\natural$  with symmetric quadratic forms goes as follows. Let  $P_\Delta$  be the projection operator from  $\Sigma$  to  $\Delta \in \Pi^\natural$  parallel to  $\Pi$ . Then we identify  $\Delta$  with

$$Q_\Delta(\lambda) = \sigma(P_\Delta \lambda, \lambda), \quad \lambda \in \Delta.$$

A simple equality

$$\sigma(P_\Delta \lambda, \mu) + \sigma(\lambda, P_\Delta \mu) = \sigma(\lambda, \mu), \quad \forall \lambda, \mu \in \Sigma.$$

implies that  $Q$  is symmetric. One can prove, that for any  $\Delta \in \Pi^\natural$  there exists a unique projection operator  $P_\Delta : \Sigma \rightarrow \Delta$ , s.t. the equality above holds and  $P_\Delta \Pi = \{0\}$ . For more details see [19].

Similarly one has an identification of the tangent space  $T_\Delta L(\Sigma)$  with the space  $\text{Sym}(\Delta)$  of all symmetric quadratic forms on  $\Delta$ . Indeed, given  $\Delta$  take any curve  $\Delta(\varepsilon) \in L(\Sigma)$ , s.t.  $\Delta(0) = \Delta$ , fix  $\lambda(0) \in \Delta(0)$  and take any curve  $\lambda(\varepsilon) \in \Delta(\varepsilon)$ . Then we identify  $\dot{\Delta}_0$  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 3.1.** We say that a  $C^1$ -curve  $\Delta(t) \in L(\Sigma)$  is *monotone increasing* if the corresponding matrix  $\dot{\Delta}(t) > 0$  as a quadratic form on  $\Delta(t)$  for every  $t$ . We say that it is *strictly monotone* if this inequality is strict.

We define similarly monotone decreasing curves, but it turns out that  $C^1$ -smooth Jacobi curves of minimum problems are always monotone increasing.

Now we define the Maslov index of a continuous curve. The Maslov train  $\mathcal{M}_\Pi$  is the set of all Lagrangian planes non-transversal to  $\Pi \in L(\Sigma)$ , i.e.  $\mathcal{M}_\Pi = L(\Sigma) \setminus \Pi^\natural$ . This is a stratified manifold where each strata  $\mathcal{M}_\Pi^k$  is the set of all  $\Delta \in L(\Sigma)$  s.t.  $\dim(\Delta \cap \Pi) = k$ . 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  $\mathcal{M}_\Pi^1$  has codimension one in  $L(\Sigma)$ . To define an intersection index we need to define a co-orientation on  $\mathcal{M}_\Pi^1$ . Suppose that



$\Lambda(\varepsilon) \in L(\Sigma)$  intersects  $\mathcal{M}_\Pi^1$  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 3.2.** *The Maslov index  $\text{Mi}_\Pi(\Lambda(t))$  of a curve  $\Lambda(t)$  is the intersection number of  $\Lambda(t)$  with  $\mathcal{M}_\Pi^1$ . For a curve in general position this is just a number of intersections  $\Lambda(t) \cap \Pi$  counted with signs.*

Since  $\mathcal{M}_\Pi^2$  has codimension three, the Maslov index is well defined and it is a homotopy invariant. The importance of the Maslov index comes from the following fact.

**Proposition 3.1.** *The Maslov index  $\text{Mi}_\Pi$  as a function on the loops in the Lagrangian Grassmanian  $L(\Sigma)$  induces an isomorphism  $\pi_1(L(\Sigma)) \rightarrow \mathbb{Z}$ . Moreover this isomorphism does not depend on the choice of  $\Pi$ , i.e.*

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

For the proof and more properties see [9]. Often we will just write  $\text{Mi}(\Lambda(t))$  for closed curves to emphasize, that the result does not depend on the choice of the Maslov train.

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  $\mathcal{M}_\Pi$ . To overcome this, one usually uses 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 [9].

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

$$q(\lambda) = \sigma(\lambda_1, \lambda_2), \quad \lambda = \lambda_1 + \lambda_2, \quad \lambda_i \in \Lambda_i. \quad (24)$$

**Definition 3.3.** *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 [9] for some others and the proves. We note again that in [9] a different sign convention was used and therefore the negative Maslov index played the central role.

**Lemma 3.1.** *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  $\Gamma \subset \Lambda_1 \cap \Lambda_2$ , we denote  $\Pi^{\Gamma} = (\Pi \cap \Gamma^{\perp}) + \Gamma$ , where  $\Gamma^{\perp}$  is the skew-orthogonal complement with respect to the symplectic form  $\sigma$ . Then*

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

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 3.4.** Let  $q(\lambda)$  be the quadratic form from (24), 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 3.2.** *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 [16] or [21]. The last one is proved in [9].

Let us try to understand what are these indices geometrically in the simplest case, when  $\Sigma = \mathbb{R}^2$  (see picture 1). Fix some Darboux coordinates  $(p, q)$ , s.t.  $\Pi = \{(p, 0)\}$ . Then all the Lagrangian planes close to  $\Pi$  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  $\Pi$  in the clockwise direction, we add  $+1$  to the Maslov index, and  $-1$  if it crosses 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 depends 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. 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  $\Lambda_{-1}$  in the clockwise direction we meet  $\Lambda_1$  before  $\Pi$ , 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  $\Lambda_{-1}$  in the clockwise direction we meet  $\Pi$  before  $\Lambda_1$ , 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 3.2** ([16]). *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  $\text{Mi}$ ,  $\text{Ki}$  and  $\text{ind}$  we need the following definition

**Definition 3.5.** A curve  $\Lambda(t)$  is called *simple* if there exists  $\Delta \in L(\Sigma)$ , s.t.  $\Lambda(t) \in \Delta^{\text{h}}$ , i.e. it is entirely contained a single affine coordinate chart.

**Proposition 3.3** ([9]). *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*

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

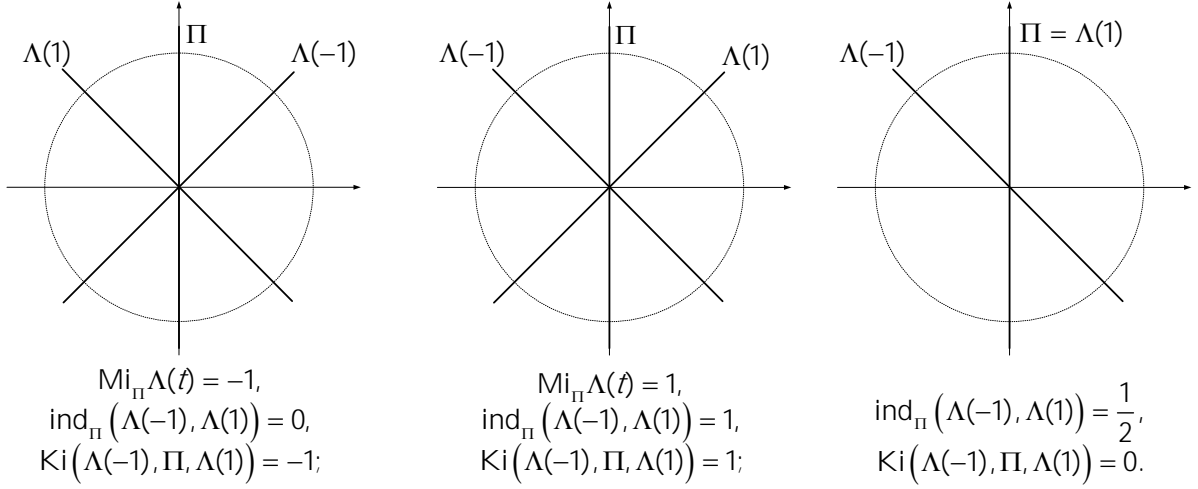


Figure 1: The Kashiwara and the positive Maslov indices in  $\mathbb{R}^2$

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 3.4** ([9]). *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 the Jacobi curves are discontinuous.

**Definition 3.6.** 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  $\Lambda_i$  are as in the Proposition 3.4. 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 curve*.

The Maslov index defined in such way inherits many useful properties of the usual Maslov index defined as an intersection number. For example, we have the following lemma.

**Lemma 3.3.** *Let  $\Lambda(t)$  be a closed monotone curve in the sense of Definition 3.6. 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 the Proposition 3.1

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

□

However we would also like to see that this definition is well defined, i.e. that it coincides with the previous one in the case of differentiable curves.

**Theorem 3.1.** *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.* Let  $\Lambda(t)$  be  $C^1$  on  $[0, T]$  and monotone in the sense of Definition 3.1. Let us assume by contradiction that it is not monotone in the sense of Definition 3.6. Then there exists a series of splittings  $D_n$  of the form  $0 = t_0^n < \dots < t_{k_n}^n$ , s.t.  $\text{ind}_\Pi^{D_n} \Lambda_t \rightarrow +\infty$  as  $n \rightarrow \infty$ . By refining  $D_n$  if necessary, we can always assume that each restriction  $\Lambda|_{[t_i^n, t_{i+1}^n]}$  is simple. Then by Proposition 3.4 we have  $\text{Mi}_\Pi \Lambda(t) = \infty$ .

We claim that this is impossible. Indeed in [7] it was shown, that one can introduce a Riemannian metric on  $L(\Sigma)$ , s.t. the length of the monotone curve  $\Lambda(t)$  can be bounded by the Maslov index as

$$\frac{\pi}{2\sqrt{n}} \text{Mi}_\Pi \Lambda(t) \leq \text{length}(\Lambda(t)) \leq \frac{\pi}{2} \text{Mi}_\Pi \Lambda(t).$$

Thus the length must be infinite and we arrive at a contradiction with the fact that the curve is  $C^1$ .

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 3.1. Let us prove the opposite statement that for a non-increasing curve  $\Lambda(t)$  the index  $\text{ind}_\Pi \Lambda(t)$  is infinite.

By Lemma 3.3 and the first property in Lemma 3.1, we can see that for any Lagrangian plane  $\Delta$

$$|\text{ind}_{\Pi} \Lambda(t) - \text{ind}_{\Delta} \Lambda(t)| \leq n.$$

So it is enough to prove that  $\text{ind}_{\Delta} \Lambda|_{[t_1, t_2]} = +\infty$  for any  $\Delta$  of our choice.

If  $\Lambda(t)$  is not monotone increasing in the sense of Definition 3.1, then there exists a moment of time  $t \in [0, T]$  and a vector  $\lambda$ , s.t.  $\dot{\Lambda}(t)(\lambda) < 0$ . We can choose a sufficiently small subinterval  $[t_1, t_2] \in [0, T]$  such that it contains the moment of time  $t$  and the restriction of  $\Lambda(t)$  is simple, i.e. lies entirely in  $\Delta^{\#}$ .

Let us consider the coordinate chart  $\Delta^{\#}$ . We can find  $[t_1, t_2]$  sufficiently small, so that  $S(t_2) - S(t_1) \not\geq 0$  (or else the curve would be monotone). Let us take any smooth simple monotone curve  $\alpha : [0, T] \rightarrow L(\Sigma)$  that joins  $\Lambda(t_1)$  with  $\Lambda(t_2)$ . The curve  $\alpha(t)$  can not lie entirely in  $\Delta^{\#}$ , because in this case we would have had by monotonicity  $S(t_2) - S(t_1) = S^{\alpha}(T) - S^{\alpha}(0) \geq 0$ . Therefore at some point  $\alpha$  must intersect the Maslov train  $\mathcal{M}_{\Delta}$  and  $\text{Mi}_{\Delta} \alpha(t) \geq 1$ . Then by Proposition 3.4 we have  $\text{ind}_{\Delta}(\Lambda(t_1), \Lambda(t_2)) \geq 1$  as well. But the same is true for any other subinterval of  $[t_1, t_2]$ . Thus by splitting it into smaller subintervals, we will find that  $\text{ind}_{\Delta}^D \Lambda(t) \rightarrow +\infty$ .  $\square$

Although these invariants were already successfully applied in [9, 7] 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 3.7.** 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 [16] or [19]. We are going to only list its main properties, that are going to be useful for the computations.

**Lemma 3.4.** *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  $\tilde{\Lambda}(t)$  as a lift 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 3.2.** *Let  $\tilde{\Lambda}_\alpha$  be a lift of an arbitrary Lagrangian plane  $\Lambda_\alpha$  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  $\Phi_\alpha$  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  $\Phi_\alpha$  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 [16].

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 3.4. So we just sketch the proofs.

**Lemma 3.5.** *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  $\Lambda^\natural$  is an open set in  $L(\Sigma)$ . One should just prove that those sets are dense, then the intersection

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

must be dense and therefore non-empty. □

**Lemma 3.6.** *Any two simple monotone curves connecting  $a, b \in L(\Sigma)$  are homotopic.*

This lemma is a direct consequence of Propositions 3.2 and 3.4. They show that Maslov index of monotone curves depends only on the relative position of its end-points.

**Remark 3.1.** 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 the 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.

### 3.2 Morse-type theory

Now we are ready to state and prove Morse-type theorems. The simplest one allows us to compute the dimension of the kernel of the Hessian. This is Lemma B.1 from Appendix B.

The next step is to extract the information about the index of the Hessian from the Jacobi curve. A simple theorem of such type is Theorem B.1 that is used in the proof of existence and uniqueness. But it is clear that in general one can not replace inequality in the statement by an equality. The right hand side of (46) 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 1.3) 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.

We will rely heavily on the following two lemmas from linear algebra

**Lemma 3.7.** *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

$$\text{ind}^+ Q = \text{ind}^+ Q|_V + \text{ind}^+ Q|_{V^\perp} + \dim(V \cap V^\perp) - \dim(V \cap \ker Q) \quad (25)$$

**Remark 3.2.** 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  $V$ .

**Lemma 3.8.** *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  $\xi$  in the annihilator  $N^\perp \subset (\mathbb{R}^n)^*$ , s.t.*

$$\langle \xi, Aw \rangle + Q(v, w) = 0, \quad \forall w \in V_1. \quad (26)$$

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

**Theorem 3.3.** *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 variations 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*

$$\text{ind}^- \text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]|_{V_D^0} = \sum_{i=-1}^N \text{ind}_\Pi(\mathcal{L}_i, \mathcal{L}_{i+1}) + \dim \left( \bigcap_{i=-1}^N \mathcal{L}_i \right) - n, \quad (27)$$



where for simplicity we wrote  $\mathcal{L}_i = \mathcal{L}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)]|_{\Omega_{N_0}^t \cap V_i}$ ,  $\mathcal{L}_{-1} = \mathcal{L}_{N+1} = \Pi$  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 (25). 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 subspaces

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

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

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

$$\int_0^t \sigma \left( \eta + \int_0^\tau X(\theta) v_2(\theta) d\theta, X(\tau) v_1(\tau) \right) + b(\tau)(v_2(\tau), v_1(\tau)) d\tau = 0, \quad \forall v_1(\tau) \in U_1 \quad (28)$$

This is a consequence of Lemma 3.8. Indeed, as we have discussed in Section 1.2 vector fields  $X(\tau)$  are lifts of  $g'(\tau)$ . Therefore from formula (42) it follows that we can characterize the kernel in terms of the Hamiltonian vector field  $X(\tau)$  as

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

We apply Lemma 3.8 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  $\sigma$  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  $\xi \in N^\perp \simeq N^\sphericalangle \simeq T_{q_0}^\perp N_0$  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 (43) 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

$$\int_0^{t_i} \sigma \left( \eta + \int_0^\tau X(\theta)v(\theta)d\theta, X(\tau)w(\tau) \right) + b(\tau)(v(\tau), w(\tau))d\tau = 0, \quad \forall w(\tau) \in V_i, \quad (29)$$

$$\eta + \int_0^{t_i} X(\tau)v(\tau)d\tau + \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \in \Pi. \quad (30)$$

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  $(\zeta, v(\tau), \alpha) \in W_i$  and  $\eta = (\mu, \zeta) \in T_{\bar{q}(0)}^\perp N_0 \times T_{\bar{q}(0)} N_0$ , 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 (30)

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

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

$$\lambda_1, \lambda_2 \in \mathcal{L}_i, \quad \lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau) d\tau \cdot \alpha \in \mathcal{L}_{i+1}, \quad \lambda_1 + \lambda_2 + \int_{t_i}^{t_{i+1}} X(\tau) d\tau \in \Pi. \quad (31)$$

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

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

If we denote

$$\tilde{S} = \{(\lambda_1 + \lambda_2, \alpha) \in \mathcal{L}_i \times \mathbb{R}^k : \lambda_j, \alpha \text{ satisfy (31)}\},$$

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.* We could have applied directly Theorem B.1 from the appendix, but in order to prove the other inequality, we need those specific expressions for  $Q|_{W_i}$  and  $q$ . We want to show, that all  $\lambda, \alpha$ , 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 1.3 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 to consider  $P$  on any complementary subspace  $L^\perp$  and  $E^\perp$ . But from Theorem 1.3 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

$$\begin{aligned} \text{ind}^+ Q_{i+1}|_{(V_i^0)^\perp} &= \text{ind}_\Pi(\mathcal{L}_i, \mathcal{L}_{i+1}) - \frac{1}{2} (\dim(\mathcal{L}_i \cap \Pi) + \dim(\mathcal{L}_{i+1} \cap \Pi)) \\ &\quad + \dim(\mathcal{L}_i \cap \mathcal{L}_{i+1} \cap \Pi) \end{aligned} \quad (32)$$

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

$$\text{ind}^+ Q_1 = \text{ind}_\Pi(\mathcal{L}_0, \mathcal{L}_1) - \frac{1}{2} (\dim(\mathcal{L}_0 \cap \Pi) + \dim(\mathcal{L}_1 \cap \Pi)) + \dim(\Pi \cap \mathcal{L}_0 \cap \mathcal{L}_1). \quad (33)$$

*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  $P_i$

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

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  $\lambda$ .

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  $(\zeta, v) \in V_i$ .

Let  $L_m \in \mathcal{L}_m$ ,  $E_m \in \mathbb{R}^k$  be the subspace  $L$ ,  $E$  from the Theorem 1.3 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

$$\dim(V_i^0 \cap (V_i^0)^\perp) = \dim(\mathcal{L}_i \cap \Pi) + \sum_{m=0}^{i-1} \dim E_m - \dim \left( \bigcap_{m=-1}^i \mathcal{L}_m \right). \quad (34)$$

Now we compute  $\dim(\ker Q_{i+1} \cap V_i^0)$ . Using Lemma 3.8, 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.

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

$$\sigma \left( \eta + \int_0^{t_i} X(\tau)v(\tau)d\tau, \int_{t_i}^{t_{i+1}} X(\tau)d\tau \cdot \alpha \right) = 0, \quad \forall \alpha \in \mathbb{R}^k,$$

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

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

$$\dim(\ker Q_{i+1} \cap V_i^0) = \dim(\mathcal{L}_i \cap \mathcal{L}_{i+1} \cap \Pi) + \sum_{m=0}^{i-1} \dim E_m - \dim \left( \bigcap_{m=-1}^{i+1} \mathcal{L}_m \right). \quad (35)$$

So we sum over all  $i$  the formulas (32)-(35) 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 3.1.  $\square$

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 3.4.** *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  $\widetilde{L(\Sigma)}$ , 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 left-continuous 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*

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

**Remark 3.3.** The lifts  $\tilde{\Lambda}^D$  are constructed using simple monotone curves as described in Remark 3.1.

*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 3.3.

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 3.3 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 3.3. We also use notations analogous to Theorem 3.3 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 1.2 and Morse Theorem 3.3 for the piece-wise constant approximations, we obtain

$$\text{ind}^- \text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)] + n = \text{ind}_\Pi \Lambda_t^m(s) + \dim \left( \bigcap_{s=-1}^t \Lambda_t^m(s) \right). \quad (36)$$

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 1.3 and the definition of  $\Lambda_t^m(s)$  it follows that

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

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

$$\bigcap_{s=0}^t (\Pi \cap \mathcal{L}_s) = \bigcap_{s=0}^t (\Pi \cap \Lambda^m(s)) \tag{37}$$

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  $w \in V^m \cap \Omega_{N_0}^t$ , 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  $\mu \in \Lambda^m(s)$  by definition for  $m$  large enough and (37) 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  $\Delta_i^m$ , 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  $\Delta_i^m$ . Let  $\Delta$  be any Lagrangian plane. Then from the Maslov index formula in Proposition 3.4 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  $\Delta_i^m$ , we obtain by Lemma 3.4

$$\text{Li}(\tilde{\Delta}_i^m, \tilde{\Lambda}_t^m(t_i)) - \text{Li}(\tilde{\Delta}_i^m, \tilde{\Lambda}_t^m(t_{i+1})) = 0.$$

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  $\Delta'$ , s.t.  $\Pi \cap \Delta' = \{0\}$ . Then we obtain from the properties of Leray index

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

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  $\Delta$ , s.t. it does not intersect any  $\Lambda^m(t)$  or  $\mathcal{L}_t$  and moreover

$$\text{Li}(\tilde{\Lambda}_t^m(t+1), \tilde{\Delta}) = \text{Li}(\tilde{\Lambda}^m(t), \tilde{\Delta}). \quad (38)$$

Then by Lemma 3.3, formula (36) and steps 3 and 4 we have

$$\text{Li}(\tilde{\Lambda}^m(t), \tilde{\Delta}) = \text{Li}(\tilde{\Lambda}_{-1}, \tilde{\Delta}) - 2 \text{ind}^-(\text{Hess}(E_{N_0, t}, \nu J_t)[\tilde{\omega}, \lambda(t)]) - 2n + 2 \dim \left( \bigcap_{s=0}^t \mathcal{L}_s \cap \Pi \right) \quad (39)$$



for  $m$  sufficiently large. Therefore the limit on the left hand side exists. But we recall that by the Theorem 3.2 an open subset in  $\{\tilde{\Lambda} \in \widetilde{L(\Sigma)} : \Lambda \cap \Delta = \{0\}\}$  can be identified with  $\Delta^{\mathfrak{h}} \times \mathbb{Z}$ . 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  $\Delta$  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. The full argument can be found in the Appendix.

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

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

First of all, we note that directly from the definition, it follows, that variations from  $\ker Q \cap \ker A$  do not give any contribution to the  $\mathcal{L}$ -derivative. Next we refine our initial partition  $D_0$ . We assume that  $D_0$  is such that the space  $V_{D_0}$  is like the subspace  $V$  in Lemma B.2 with  $F = E_{N_0, t}$ . Then by Lemma B.1 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  $\Delta$  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  $\Delta$  that contains  $\mathcal{L}_t(V_{D_0}) \cap \Pi$ . Indeed as a result we will get monotone curves that have constant intersection with  $\Delta$ , so it is going to be enough to replace it with  $\Delta^\Gamma$ , where  $\Gamma$  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 B.1. Therefore

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

We note that from the definition of the positive Maslov index 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  $\Lambda^m(t) \in \Delta^{\mathfrak{h}}$ . 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 from property 1 of Lemma 3.2 we obtain

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

This establishes the existence of  $\Delta$ . 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  $\Pi$  by a monotone curve that does not intersect  $\Delta$ . By properties of the Leray index it now follows that (38) holds. □

## Appendices

### A Chronological calculus and second variation

In this Appendix we give some basic facts from the chronological calculus and derive a formula for the second variation in terms of symplectic geometry. A more detailed exposition can be found in [8].

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

$$\dot{P}^t = P^t \circ V(t). \quad (40)$$

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 (40) 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

$$P^t = \text{id} + \int_0^t V(\tau) d\tau + \int_0^t \int_0^\tau V(\theta) \circ V(\tau) d\theta d\tau + \dots \quad (41)$$

The last thing that we need is the variation formulae for the right chronological exponent. Suppose that  $V(t), W(t)$  are non-autonomous vector fields and  $P^t$  satisfies (40). Then the following formulae is true

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

Here  $P_*^t$  should be understood as a normal 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].

Now we would like to obtain an explicit expression for the first and second differential of the extended end-point map  $\widehat{E}_{N_0, t} = (E_{N_0, t}, J_t)$ .

Using the notations of Section 1.2 by the variation formulae we then find

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

Note that  $\hat{g}_{\tau, \tilde{u}(\tau)} \equiv 0$ . Using the Volterra expansion (41) 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

$$d\widehat{E}_{N_0, t}[\tilde{\omega}, \hat{q}(0)](v, \hat{\zeta}) = \widehat{P_*^t} \hat{\zeta} + \left( \widehat{P_*^t} \int_0^t \left( \frac{\partial}{\partial u} \Big|_{u=\tilde{u}(\tau)} \hat{g}_{\tau, u} \right) v(\tau) d\tau \right) (\hat{q}(t)). \quad (42)$$

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

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

where  $h$  is the Hamiltonian of PMP defined in Subsection 1.2. If we restrict the equation (11) to  $w$  in  $L_k^\infty[0, t]$ , we obtain

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

Since this equality holds for any  $v(t) \in L_k^\infty[0, t]$ , we obtain this way the extremality condition 12.

Note that since we do not vary the initial value of the functional (i.e.  $J_0 = 0$ ), we have  $\hat{\zeta} = (\zeta, 0) \in T_{\hat{q}(0)}M \times \mathbb{R}$ . Thus if we restrict the equation (11) to  $w \in T_{\hat{q}(0)}N_0$ , 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 13.

Now we can find an explicit formula for the Hessian, that we use in Subsection 3.2.

**Proposition A.1.** *The Hessian  $\text{Hess}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)]$  has the following form*

$$\begin{aligned} &\text{Hess}(E_{N_0,t}, \nu J_t)[\tilde{\omega}, \lambda(t)]((\zeta_1, v_1), (\zeta_2, v_2)) = \\ &= - \int_0^t \sigma \left( \zeta_1 + \int_0^\tau X(\theta) v_1(\theta) d\theta, X(\tau) v_2(\tau) \right) + b(\tau)(v_1(\tau), v_2(\tau)) d\tau. \end{aligned} \quad (43)$$

Note that there is no  $\zeta_2$  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

$$g'(\tau) = \frac{\partial}{\partial u} \Big|_{u=\hat{u}(\tau)} g_{\tau,u},$$

and similarly the ‘‘hatted’’  $\hat{g}'_\tau$  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  $\sigma$  is the standard symplectic form on  $T^*M$ .

Using the Volterra expansion once more we obtain an explicit formula

$$\begin{aligned}
& d^2\hat{G}_t[\tilde{\omega}]((\hat{\zeta}_1, v_1), (\hat{\zeta}_2, v_2)) = \\
& = \int_0^t d_{\hat{q}}\hat{g}'(\tau)(\hat{\zeta}_1 v_2(\tau) + \hat{\zeta}_2 v_1(\tau))d\tau + \int_0^t \hat{g}''(\tau)(v_1(\tau), v_2(\tau))d\tau + \\
& + \int_0^t \left( \int_0^\tau \hat{g}'(\theta)v_1(\theta)d\theta \circ \hat{g}'(\tau)v_2(\tau) + \int_0^\tau \hat{g}'(\theta)v_2(\theta)d\theta \circ \hat{g}'(\tau)v_1(\tau) \right) d\tau.
\end{aligned} \tag{44}$$

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

$$\int_0^t d_{\hat{q}}\hat{g}'(\tau)\hat{\zeta}_1v_2(\tau)d\tau = \int_0^t g'(\tau)v_2(\tau)d\tau \circ \int_0^t \hat{g}'(\tau)v_1(\tau)d\tau$$

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

$$\begin{aligned} & \int_0^t d_{\hat{q}}\hat{g}'(\tau)\hat{\zeta}_2v_1(\tau)d\tau + \int_0^t \hat{g}'(\tau)v_2(\tau)d\tau \circ \int_0^t \hat{g}'(\tau)v_1(\tau)d\tau = \\ & = \int_0^t (g^0)'(\tau)v_2(\tau)d\tau \partial_y \circ \int_0^t \hat{g}'(\tau)v_1(\tau)d\tau, \end{aligned}$$

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  $\nu$  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. □

## B Existence and uniqueness of $\mathcal{L}$ -derivatives

In this appendix we prove some lemmas needed in Subsection 3.2 and repeat in more detail the proof of Theorem 1.2 compared to the original article [1].

We begin with proof of Lemma 3.8, that we have used several times in the text.

*Proof of Lemma 3.8.* We denote by  $W$  the subspace consisting of  $v \in V_1^N$  as defined in the statement. By restricting (26) 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 compliment of  $N$  in  $\mathbb{R}^n$ . Let us then take a compliment of  $A^{-1}(N)$  in  $V_2$  which will be

isomorphic to  $\text{im } A \cap N^\perp$  under  $A$  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  $v \in (V_1^N)_Q^\perp$ , by identifying  $(\mathbb{R}^n)^*$  with  $\mathbb{R}^n$  using the Euclidean inner product, we find that (26) 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  $W \supset (V_1^N)_Q^\perp$ .

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

**Lemma B.1.** *Let  $F : \mathcal{U} \rightarrow M$  be a smooth map from a finite dimensional manifold  $\mathcal{U}$  to a finite dimensional manifold  $M$ ,  $J : \mathcal{U} \rightarrow \mathbb{R}$  be a smooth functional, and let  $(u, \lambda)$  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

$$\langle \xi, dF[u](w) \rangle + Q(v, w) = 0, \quad \forall w \in T_u \mathcal{U}. \quad (45)$$

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  $\xi$ , s.t. (45) holds.  $\square$

**Lemma B.2.** *Let  $(u, \lambda)$  be a Lagrangian point of  $(F, J) : \mathcal{U} \rightarrow M \times \mathbb{R}$ . If  $\text{ind}^+ Q|_{\ker dF[u]} < \infty$ , then there exists a finite-dimensional subspace  $V \subset T_u \mathcal{U}$ , s.t. for any  $W \supset V$  one has:*

1.  $\text{rank } dF[u]|_W = \text{rank } dF[u]|_V$ ,
2.  $\text{ind}^+ Q|_W = \text{ind}^+ Q|_V$ ,
3.  $v \in \ker Q|_{W_0}$  and  $v \notin V$  iff  $v \in \ker Q \cap \ker dF[u]$ .

The proof is obvious. We just need to take a direct sum of a subspace isomorphic to  $\text{im } dF[u]$ , a maximal positive subspace and a subspace complementary to  $\ker Q \cap \ker dF[u]$  complementary to  $\ker Q|_{\ker dF[u]}$ .

Using these lemmas we can now give the proof of the following Morse-type result that appeared first in [1].

**Theorem B.1.** *Let  $F : \mathcal{U} \rightarrow M$  be a map from a possibly infinite dimensional smooth manifold  $\mathcal{U}$  to a finite dimensional manifold  $M$ , whose restrictions to finite-dimensional subspaces are smooth, and let  $J : \mathcal{U} \rightarrow \mathbb{R}$ ,  $(u, \lambda)$  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*

$$\begin{aligned} & \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{V_2^0} - \text{ind}^- \text{Hess}(F, \nu J)[u, \lambda]|_{V_1^0} \geq \\ & \geq \text{ind}_\Pi(\mathcal{L}(F, \nu J)[u, \lambda](V_1), \mathcal{L}(F, \nu J)[u, \lambda](V_2)). \end{aligned} \quad (46)$$

*Proof.* From the assumption on the kernels we have by Lemma B.1, that

$$\mathcal{L}(F, \nu J)[u, \lambda](V_1) \cap \Pi = \mathcal{L}(F, \nu J)[u, \lambda](V_2) \cap \Pi,$$

and by Lemma 3.7 that

$$\text{ind}^+ Q|_{V_2^0} - \text{ind}^+ Q|_{V_1^0} = \text{ind}^+ Q_{(V_1^0)^\perp},$$

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

Let us write down explicitly the quadratic form  $q$  from the definition of the positive index. Suppose that  $(\xi_i, dF[u](v_i)) \in \mathcal{L}(F, \nu J)[u, \lambda](V_i)$ . The quadratic form  $q$  is defined on the intersection of  $\Pi$  and the sum of  $\mathcal{L}(F, \nu J)[u, \lambda](V_i)$ . Thus we define  $\xi = \xi_1 + \xi_2$  and assume that  $dF[u](v_1 + v_2) = 0$ . Then we get

$$q(\xi) = \langle \xi_1, dF[u](v_2) \rangle - \langle \xi_2, dF[u](v_1) \rangle = -\langle \xi_2, dF[u](v_1) \rangle - \langle \xi_1, dF[u](v_1) \rangle = -\langle \xi, dF[u](v_1) \rangle.$$

On the other side, from Lemma 3.8 it follows that in this case  $v = v_1 + v_2$  lies in  $(V_1^0)^\perp$ . So it only remains to check that  $Q(v, v)$  is the same as  $q(\xi)$ . And indeed, we have

$$Q(v, v) = Q(v_2, v_1 + v_2) + Q(v_1 + v_2, v_1) = -\langle \xi_2, dF[u](v_1 + v_2) \rangle - \langle \xi_1, dF[u](v_1) \rangle = -\langle \xi, dF[u](v_1) \rangle.$$

□

Now we are ready to prove Theorem 1.2. First we show the existence part, since we already have all the necessary results.

*Proof of the first part of Theorem 1.2.* For this prove we shorten the notation  $\mathcal{L}(F, \nu J)[u, \lambda](W)$  to  $\mathcal{L}(W)$ . Let us take three finite-dimensional subspaces  $V \subset V_1 \subset V_2$ , where  $V$  is like in Lemma B.2. Then as a consequence we have

$$\mathcal{L}(V_i) \cap \Pi = \mathcal{L}(V) \cap \Pi.$$



Therefore we can assume that  $\mathcal{L}(V_i) \cap \Pi = \{0\}$  by considering  $\Sigma = (\mathcal{L}(V) \cap \Pi)^\perp / (\mathcal{L}(V) \cap \Pi)$ . By the assumptions on  $V$  and Theorem B.1 we also find that

$$\text{ind}_\Pi(\mathcal{L}(V_1), \mathcal{L}(V_2)) = 0.$$

Take any  $\Delta \in L(\Sigma)$ , s.t.  $\text{ind}_\Pi(\mathcal{L}(V), \Delta) = \dim \Sigma / 2$ . Then by the triangle inequality we have

$$\text{ind}_\Pi(\mathcal{L}(V_i), \Delta) \geq \text{ind}_\Pi(\mathcal{L}(V), \Delta) - \text{ind}_\Pi(\mathcal{L}(V), \mathcal{L}(V_i)) = \frac{\dim \Sigma}{2}.$$

Therefore  $\text{ind}_\Pi(\mathcal{L}(V_i), \Delta) = \dim \Sigma / 2$  and  $\mathcal{L}(V_i) \pitchfork \Delta$ .

So we consider a coordinate chart  $\Delta^\flat$  centered at  $\Pi$ . In this chart  $\mathcal{L}(V_i)$  are identified with symmetric operators  $S_i : \Pi \rightarrow \Delta$ . Then for the quadratic form in the definition of  $\text{ind}_\Pi(\mathcal{L}(V_i), \Delta)$  we find that

$$q(p) = \sigma((p, S_i p), (0, -S_i p)) = -p^T S_i p.$$

Therefore  $S_i$  must be negative-definite symmetric matrices.

We want to show that  $\text{ind}_\Delta(\mathcal{L}(V_1), \mathcal{L}(V_2)) = 0$ . Then using the same reasoning we find that  $S_2 \geq S_1$ , whenever  $V_2 \supset V_1$ . Thus we obtain a monotone and bounded generalized sequence of symmetric finite-dimensional matrices that must have a limit. In order to compute  $\text{ind}_\Delta(\mathcal{L}(V_1), \mathcal{L}(V_2))$  we use property 4 from Lemma 3.2 and the cocycle identity. We have that

$$\text{Ki}(\mathcal{L}(V_i), \Pi, \Delta) = 2 \text{ind}_\Pi(\mathcal{L}(V_i), \Delta) + \dim(\mathcal{L}(V_i) \cap \Delta) - \frac{\dim \Sigma}{2} = \frac{\dim \Sigma}{2}.$$

Then by the cocycle identity we have

$$\begin{aligned} \text{Ki}(\mathcal{L}(V_1), \Delta, \mathcal{L}(V_2)) &= \text{Ki}(\mathcal{L}(V_1), \Pi, \mathcal{L}(V_2)) + \text{Ki}(\Delta, \mathcal{L}(V_2), \Pi) + \text{Ki}(\mathcal{L}(V_1), \Delta, \Pi) = \\ &= \text{Ki}(\mathcal{L}(V_1), \Pi, \mathcal{L}(V_2)). \end{aligned}$$

Therefore we find that

$$\begin{aligned} 2 \text{ind}_\Delta(\mathcal{L}(V_1), \mathcal{L}(V_2)) &= \frac{\dim \Sigma}{2} - \dim(\mathcal{L}(V_1) \cap \mathcal{L}(V_2)) + \text{Ki}(\mathcal{L}(V_1), \Delta, \mathcal{L}(V_2)) = \\ &= \frac{\dim \Sigma}{2} - \dim(\mathcal{L}(V_1) \cap \mathcal{L}(V_2)) + \text{Ki}(\mathcal{L}(V_1), \Pi, \mathcal{L}(V_2)) = \\ &= 2 \text{ind}_\Pi(\mathcal{L}(V_1), \mathcal{L}(V_2)) = 0. \end{aligned}$$

□

In order to prove the second part of Theorem 1.2, one can try to show that if  $U_0 \subset T_u \mathcal{U}$  is a dense subset, then

$$\text{ind}_\Delta(\mathcal{L}(F, \nu J)[u, \lambda](U_0), \mathcal{L}(F, \nu J)[u, \lambda]) = 0,$$

for all  $\Delta$  in a dense subset of  $L(\Sigma)$ . In order to do this we must define  $\mathcal{L}$ -derivatives of maps similar to the  $\mathcal{L}$ -derivatives of constrained variational problems. Namely if  $F : \mathcal{U} \rightarrow M$  is a smooth map, then  $(u, \lambda)$  is a Lagrangian point if

$$\langle \lambda, dF[u](w) \rangle = 0, \quad \forall w \in T_u \mathcal{U}$$

and  $\mathcal{L}$ -derivative of this map at  $(u, \lambda)$  that we denote by  $\mathcal{L}F[u, \lambda](V)$  consists of vectors  $(\xi, dF[u](v))$  which satisfy

$$\langle \xi, dF[w] \rangle + \langle \lambda, d^2 F[u](v, w) \rangle = 0, \quad \forall w \in T_u \mathcal{U}$$

just like in Definition 1.1. One can easily check that  $(u, \lambda)$  is a Lagrange point of the constrained variational problem  $(F, J)$  iff  $(u, (\lambda, -\nu))$  is a Lagrange point of the extended map  $\hat{F} = (F, J)$ . Moreover one can easily reconstruct  $\mathcal{L}(F, \nu J)[u, \lambda](V)$  from  $\mathcal{L}\hat{F}[u, (\lambda, -\nu)](V)$ . So we just prove uniqueness of the latter one. To do this we need a lemma similar to Theorem B.1.

**Lemma B.3.** *Let  $F : \mathcal{U} \rightarrow M$  be a smooth map,  $a : M \rightarrow \mathbb{R}$  be a smooth function and  $(u, \lambda)$  a Lagrange point of  $F$ , s.t.  $\lambda = da$ . We define a Lagrangian subspace  $\Pi_a(\lambda) = T_\lambda(da)$  and the Hessian of  $F$  at  $u$  to be  $\text{Hess}(a \circ F)[u] = -d^2(a \circ F)[u]$ . Let  $V_1 \subset V_2$  be two finite-dimensional subspaces, s.t.  $\text{rank } dF[u]|_{V_1} = \text{rank } dF[u]|_{V_2}$  and  $\ker \text{Hess}(a \circ F)[u]|_{V_1} = \ker \text{Hess}(a \circ F)[u]|_{V_2}$ . Then*

$$\text{ind}^- \text{Hess}(a \circ F)[u]|_{V_2} - \text{ind}^- \text{Hess}(a \circ F)[u]|_{V_1} \geq \text{ind}_{\Pi_a(\lambda)}(\mathcal{L}F[u, \lambda](V_1), \mathcal{L}F[u, \lambda](V_2))$$

*Proof.* Although similar to Theorem B.1, the proof of this result is somewhat easier. First of all, like in Lemma B.1 variations  $v \in \ker \text{Hess}(a \circ F)[u]|_V$  are in one to one correspondence with  $\mathcal{L}(F)[u, \lambda](V) \cap \Pi_a(\lambda)$ . Indeed, if  $v \in \ker \text{Hess}(a \circ F)[u]|_V$ , then

$$0 = d^2(a \circ F)(v, w) = d^2 a[F(u)](dF[u](v), dF[u](w)) + da[F(u)] \circ d^2 F[u](v, w), \quad \forall w \in V.$$

On the other hand if a vector lies in  $\mathcal{L}(F)[u, \lambda](V) \cap \Pi_a(\lambda)$ , then it must be of the form  $(d^2 a[F(u)] \circ dF[u](v), dF[u](v))$ , where  $v$  satisfies the same equation as above. Thus from the assumptions on the ranks and kernels we have that

$$\text{ind}^- \text{Hess}(a \circ F)[u]|_{V_2} - \text{ind}^- \text{Hess}(a \circ F)[u]|_{V_1} = \text{ind}^- \text{Hess}(a \circ F)[u]|_{(V_1)^\perp}$$

and  $\mathcal{L}(F)[u, \lambda](V_1) \cap \Pi_a(\lambda) = \mathcal{L}(F)[u, \lambda](V_2) \cap \Pi_a(\lambda)$ .

Like in the proof of Theorem B.1, it is enough to show that the quadratic form  $q$  from the definition of the positive index can be identified with the restriction of minus  $\text{Hess}(a \circ F)[u]|_{(V_1)^\perp}$  to a smaller subspace. And indeed, let  $(\xi_i, dF[u](v_i)) \in \mathcal{L}F[u, \lambda](V_i)$  be such that  $\xi = \xi_1 + \xi_2 = d^2 a[F(u)] \circ dF[u](v_1 + v_2)$ . From the definition of  $\mathcal{L}F[u, \lambda](V_i)$  it is clear that in this case  $v_1 + v_2 \in (V_1)^\perp$  and for the form  $q$ , using the fact that  $\lambda = da$ , we find that for  $\mu \in \Pi_a(\lambda) \cap (\mathcal{L}F[u, \lambda](V_1) + \mathcal{L}F[u, \lambda](V_2))$

$$\begin{aligned} q(\mu) &= \langle \xi_1, dF[u](v_2) \rangle - \langle \xi_2, dF[u](v_1) \rangle = \langle \xi, dF[u](v_2) \rangle - \langle \xi_2, dF[u](v_1 + v_2) \rangle = \\ &= d^2 a[F(u)](dF[u](v_1 + v_2), dF[u](v_2) + da[F(u)] \circ d^2 F[u](v_1 + v_2, v_2) = \\ &= -\text{Hess}(a \circ F)[u](v_1 + v_2, v_2) = -\text{Hess}(a \circ F)[u](v_1 + v_2, v_1 + v_2), \end{aligned}$$

which ends the proof. □

*Proof of the second part of Theorem 1.2.* We again employ the short notation  $\mathcal{L}F[u, \lambda](W) = \mathcal{L}(W)$ .

Let us assume first that  $\mathcal{L}(U_0) \cap \mathcal{L} = \{0\}$ . Then we can find  $\Delta \in \mathcal{L}(U_0)^\natural \cap \mathcal{L}^\natural \cap \Pi^\natural$ , s.t.  $\text{ind}_\Delta(\mathcal{L}(U_0), \mathcal{L}) \neq 0$ . But then for any two neighborhoods  $O_{\mathcal{L}(U_0)}$  and  $O_{\mathcal{L}}$  there exist subspaces of variations  $V_0 \subset U_0$  and  $V$ , s.t.  $\mathcal{L}(V_0) \in O_{\mathcal{L}(U_0)}$  and  $\mathcal{L}(V_0 + V) \in O_{\mathcal{L}}$ . By Proposition 3.2 we have  $\text{ind}_\Delta(\mathcal{L}(V_0), \mathcal{L}(V_0 + V)) = \text{ind}_\Delta(\mathcal{L}(U_0), \mathcal{L}) \neq 0$  for sufficiently small  $O_{\mathcal{L}(U_0)}$  and  $O_{\mathcal{L}}$ .

Since  $\Delta \in \Pi^\natural$ , there exists a smooth function  $a : M \rightarrow R$ , s.t.  $\Delta = T_\lambda(da)$ . Therefore if we take  $V_0$  to be such that  $\text{ind}^- \text{Hess}(a \circ F)|_{V_0}$  is maximal (we can always do this by the definition of a generalized sequence), then from the previous lemma we obtain that  $\text{ind}_\Delta(\mathcal{L}(V_0), \mathcal{L}(V_0 + V)) = 0$ , which gives us a contradiction. Thus  $\mathcal{L}(U_0) \cap \mathcal{L} \neq \{0\}$ .

So let us assume that  $\mathcal{L}(U_0) \cap \mathcal{L} \neq \{0\}$ , but  $\mathcal{L}(U_0) \neq \mathcal{L}$ . We can reduce this case to the one we have just considered. Indeed, let us denote  $\Gamma = \mathcal{L}(U_0) \cap \mathcal{L}$ . Then we can consider the symplectic space  $\Gamma^\perp/\Gamma$  and  $\mathcal{L}(U_0)/\Gamma, \mathcal{L}/\Gamma$  can be identified with two Lagrangian subspaces in this symplectic space. Similarly we can identify  $\Delta \cap \Gamma^\perp, (\mathcal{L}(V_0) \cap \Gamma^\perp)/(\mathcal{L}(V_0) \cap \Gamma)$  and  $(\mathcal{L}(V_0 + V) \cap \Gamma^\perp)/(\mathcal{L}(V_0 + V) \cap \Gamma)$  with Lagrangian subspaces in  $\Gamma^\perp/\Gamma$ . But then we have

$$\begin{aligned} & \text{ind}_\Delta(\mathcal{L}(V_0), \mathcal{L}(V_0 + V)) \geq \\ & \geq \text{ind}_{\Delta \cap \Gamma^\perp}((\mathcal{L}(V_0) \cap \Gamma^\perp)/(\mathcal{L}(V_0) \cap \Gamma), (\mathcal{L}(V_0 + V) \cap \Gamma^\perp)/(\mathcal{L}(V_0 + V) \cap \Gamma)), \end{aligned}$$

because it corresponds to the restriction of the form  $q$  from the definition of the positive Maslov index to a smaller subspace. Thus again by Lemma B.3 the expression on the right must be zero as well, and we arrive to a contradiction. Therefore  $\mathcal{L}(U_0) = \mathcal{L}$ .  $\square$

## References

- [1] A. Agrachev. Feedback-invariant optimal control theory and differential geometry, ii. jacobi curves for singular extremals. *J. Dynamical and Control Systems*, pages 583–604, 1998.
- [2] A. Agrachev. *Nonlinear and Optimal Control Theory*, chapter Geometry of Optimal Control Problems and Hamiltonian Systems, pages 1–59. Springer, 2004.
- [3] A. Agrachev, D. Barilari, and U. Boscain. Introduction to riemannian and subriemannian geometry. <https://webusers.imj-prg.fr/~davide.barilari/Notes.php>, Ver. Nov 17, 2017.
- [4] A. Agrachev and I. Beschastnyi. Symplectic geometry of constrained optimization. *Regular and Chaotic Dynamics*, 22:750–770, 2017.
- [5] A. Agrachev and R. Gamkrelidze. *Nonlinear Controllability and Optimal Control*, chapter Symplectic geometry for optimal control, pages 263–277. CRC Press Book, 1990.

- [6] A. Agrachev and R. Gamkrelidze. Feedback-invariant optimal control theory and differential geometry, i. regular extremals. *J. Dynamical and Control Systems*, pages 343–389, 1997.
- [7] A. Agrachev and R. Gamkrelidze. *Geometry of Feedback and Optimal Control*, chapter Symplectic methods in optimization and control, pages 19–77. CRC Press Book, 1998.
- [8] A. Agrachev and Yu. Sachkov. *Control theory from a geometric point of view*. Springer, 2004.
- [9] A. Agrachev, G. Steffani, and P. Zezza. Quadratic mappings in the geometric control theory. *Itogi nauki. VINITI. Problemy geometrii*, 20:111–205, 1988.
- [10] A. Agrachev, G. Steffani, and P. Zezza. Strong optimality of a bang-bang trajectory. *SIAM J. on Control and Optimization*, 41:981–1014, 2002.
- [11] V Arnold and A Givental. Dynamical systems iv: Symplectic geometry and its applications. chapter Symplectic geometry, pages 1–138. Springer, 2001.
- [12] A. Aronna, F. Bonnans, A. Dmitruk, and P. Lolito. Quadratic order conditions of a local minimum for abnormal extremals. *Nonlinear Analysis: Theory, Methods & Applications*, 30:2439–2448, 1997.
- [13] A. Aronna, F. Bonnans, A. Dmitruk, and P. Lolito. Quadratic order conditions for bang-singular extremals. *Numerical Algebra, Control and Optimization*, 2:511–546, 2012.
- [14] D. Barilari and L. Rizzi. Comparison theorems for conjugate points in sub-riemannian geometry. *ESAIM: COCV*, 22:439–472, 2016.
- [15] B. Bonnard and E. Trelat. On the role of abnormal minimizers in sub-riemannian geometry. *Ann. Fac. Sci. Toulouse*, 6:405–491, 2001.
- [16] M. de Gosson. *Symplectic Geometry and Quantum Mechanics*. Birkhauser, 2000.
- [17] A. Dmitruk and N. Osmolovskii. Variations of the  $v$ -change of time in problems with state constraints. *Trudy Inst. Mat. i Mekh. UrO RAN*, 24:76–92, 2018.
- [18] J. Duistermaat. On the morse index in variational calculus. *Adv. in Math.*, 21:173–195, 1976.
- [19] V. Guillemin and S. Sternberg. *Geometric Asymptotics*. American Mathematical Society, 1977.
- [20] K. Hoffman and R. Manning. An extended conjugate point theory with application to the stability of planar buckling of an elastic rod subject to a repulsive self-potential. *SIAM J. Math. Anal.*, 41:465–494, 1998.

- [21] G. Lion and M. Vergne. *The Weil representation, Maslov index and Theta series*. Birkhauser, 1980.
- [22] J. Reeds and L. Shepp. Optimal paths for a car that goes both forwards and backwards. *Pacific J. Math.*, (2):367–393, 1990.
- [23] Yu. Sachkov. Conjugate points in euler’s elastic problem. *Journal of Dynamical and Control Systems*, 14:409–439, 2008.
- [24] H. Shättler and U. Ledzewicz. *Geometric Optimal Control*. Springer, 2012.
- [25] H. Sussmann and G. Tang. Shortest paths for the reeds-shepp car: A worked out example of the use of geometric techniques in nonlinear optimal control. Technical report, Rutgers University, 1991.
- [26] M. Zelikin and V. Borisov. *Theory of Chattering Control*. Birkhäuser, 1994.
- [27] M. Zelikin, L. Lokutsievskii, and R. Hildebrand. Typicality of chaotic fractal behavior of integral vortices in hamiltonian systems with discontinuous right hand side. *Journal of Mathematical Sciences*, 221:1–136, 2017.