Project in Calculus of Variations

Holonomic and Non-Holonomic Constraints, with Lagrange Multipliers

Duarte Maia

May 2021

1 Introduction

This work originally started as a project for a class on the Calculus of Variations, at Instituto Superior Técnico. The topic of the project is classical optimization with holonomic and non-holonomic constraints. However, when doing research on the subject, I was shocked to find very few mathematical references with the Lagrange multiplier theorem for the non-holonomic case, and corresponding proof. Consequently, I have taken it upon myself to create such a reference.

This document is divided into two parts. The first part is a relatively simple exploration of holonomic constraints, that is, optimization on (possibly time-dependent) manifolds. In other words, we require that our paths satisfy equations of the form $\phi_i(t, u(t)) = 0$, $i = 1, \ldots, k$. The second part is trickier, in which the restrictions become differential equations of the form $\phi_i(t, u(t), \dot{u}(t))$, $i = 1, \ldots k$. This is where I had a problem finding references.

This work is the culmination of a couple of weeks spent interpreting 16 pages' worth of Bliss' *Lectures on the Calculus of Variations* [1, pp. 187-202]. I don't know whether to blame my unfamiliarity with the notational conventions, the great amount of generality in which the proof is done, or my own lack of experience with the topic, but it was with more effort than I am proud to admit that I could finally understand the proof which I reproduce in section 4, albeit in much less generality (though the essential ideas are there). Perhaps this work might serve as a primer for someone else's reading of Bliss' proof.

2 Notational Conventions

In this section we lay down the basic notational conventions that will be followed over the rest of this document.

The problem at hand is to find a curve $u\colon [t_{\text{start}}, t_{\text{end}}] \to \mathbb{R}^n$ that minimizes the value of a certain integral $I(u) = \int_{t_{\text{start}}}^{t_{\text{end}}} f(t, u(t), \dot{u}(t))\,\mathrm{d}t$. In order for this expression to make sense, it is assumed that the curves u being considered are at least C^1, and the function f is at least continuous, though the results to be obtained require in fact that f is at least C^2.

The function f is considered to be a function of 3 variables: time $t \in [t_{\text{start}}, t_{\text{end}}]$, position $u \in \mathbb{R}^n$ and velocity $\xi \in \mathbb{R}^n$. In truth, we could be considering f defined in more general (i.e. smaller) domains, but since our results are local in nature correctness is not harmed by omitting consideration of the domain.

We allow ourselves to consider partial derivatives in vector coordinates, for example $\frac{\partial f}{\partial \xi}(t, u, \xi)$. These derivatives are to be taken in the sense of multivariable calculus; for example, the expression we have just written is understood as a $1 \times n$ row-matrix.

Arguments of functions will often be omitted. For example, if u represents a curve, an expression like $f(t, u, \dot{u})$ is understood to mean $f(t, u(t), \dot{u}(t))$: the evaluations at time t are implied. Somewhat more extreme, if u is understood to be the curve in consideration, we might go as far as to write f without any arguments to mean $f(t, u(t), \dot{u}(t))$. So, for example, we might write $I(u) = \int_{\text{start}}^{\text{end}} f$. This applies especially to expressions involving partial derivatives. For example, the standard Euler-Lagrange equations might be written as

$$\frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi} = 0.$$

Note that that the total time derivative is evaluated *after* the implicit insertion of variables, unlike the partial derivatives. These conventions are common in physics books (see, e.g. [4]), so we hope that they are understandable to the reader.

In section 4.3, we introduce a notation for the 'Euler-Lagrange derivative', which we record here for completeness. We define the differential operator Θ as

$$\Theta(f) := \frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}.$$

3 Holonomic Constraints

3.1 Introduction

An optimization problem with holonomic constraints is one where the objective is to minimize $I(u)$ as u ranges over (say, C^1) paths satisfying the constraint $\phi(t, u) = 0$ for all t, where ϕ is a (nice enough) function $[t_{\text{start}}, t_{\text{end}}] \times \mathbb{R}^n \to \mathbb{R}^k$. In other words, u must obey a set of k cartesian equations. As in the statement for Lagrange multipliers in $\mathbb{R}^n$, it is necessary to assume that these equations are independent, in the sense that the k rows of $\frac{\partial \phi}{\partial u}$ are linearly independent everywhere in the domain of consideration; in other words, that 0 is a regular value of $\phi(t, u)$ as a function of u. Therefore, we will henceforth work under the following hypotheses:

Problem Statement. In this section, we endeavor to minimize the value of the functional $I(u)$ as u ranges over paths starting at u_{start} and ending at u_{end}, such that the k cartesian equations given by

$$\phi(t, u(t)) = 0$$

are satisfied for all $t \in [t_{\text{start}}, t_{\text{end}}]$.

It is assumed that ϕ is a C^2 function.

3.2 Local Coordinates (Time-Independent Conditions)

The most basic way to solve an optimization problem with holonomic constraints is to write it out in local coordinates.

The idea is easiest to state when $\phi \equiv \phi(u)$. In this case, the problem is reduced to optimization on a C^2-submanifold of $\mathbb{R}^n$. The usual proof of necessity of the Euler-Lagrange equations can be easily modified to work in coordinate charts. We state the result without proof (really, it's almost word-for-word the proof in euclidean space).

Theorem 1. *Let M be an m-dimensional C^2 manifold embedded in n-dimensional euclidean space. Let $f : [t_{\text{start}}, t_{\text{end}}] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, and u a C^2 path. Let $r : \mathbb{R}^m \to \mathbb{R}^n$ be a C^2 chart of M. Define $\tilde{f} : [t_{\text{start}}, t_{\text{end}}] \times \mathbb{R}^m \times \mathbb{R}^m$ (i.e. 'f in coordinates') as*

$$\tilde{f}(t, q, v) = f(t, r(q), r'(q)v).$$

Furthermore, define $\tilde{u}$ ('u in coordinates') as

$$\tilde{u}(t) = r^{-1}(u(t)), \text{ for applicable } t.$$

Then, if u_0 minimizes the value of $I(u) = \int f(t, u, \dot{u})$ as u ranges over paths starting at u_{start} and ending at u_{end}, u_0 must satisfy the Euler-Lagrange equations in every coordinate chart, i.e., for all C^2 charts,

$$\frac{\partial \tilde{f}}{\partial q} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \tilde{f}}{\partial v} = 0.$$

Remark 1. Even though the above theorem requires that the verification be made for every coordinate chart, it is enough to do the verification in a chart that covers u, as the Euler-Lagrange equations are invariant under change of coordinates. In physics parlance, we say that the Euler-Lagrange equations are invariant under point transformations. The proof is a trivial exercise in application of the chain rule.

Remark 2. Theorem 1 also holds for abstract C^2 manifolds, not embedded in $\mathbb{R}^n$, if f is defined on $[t_{\text{start}}, t_{\text{end}}] \times TM$, where TM is the tangent bundle of M.

Example 1. Suppose that we wish to find length-minimizing curves on the sphere of unit radius. Then, we begin by parametrizing the sphere using polar coordinates:

$$r(\theta, \varphi) = (\cos\theta\cos\varphi, \sin\theta\cos\varphi, \sin\varphi),$$

and now we calculate the function $f(t, u, \xi) = \|\xi\|$ in local coordinates:

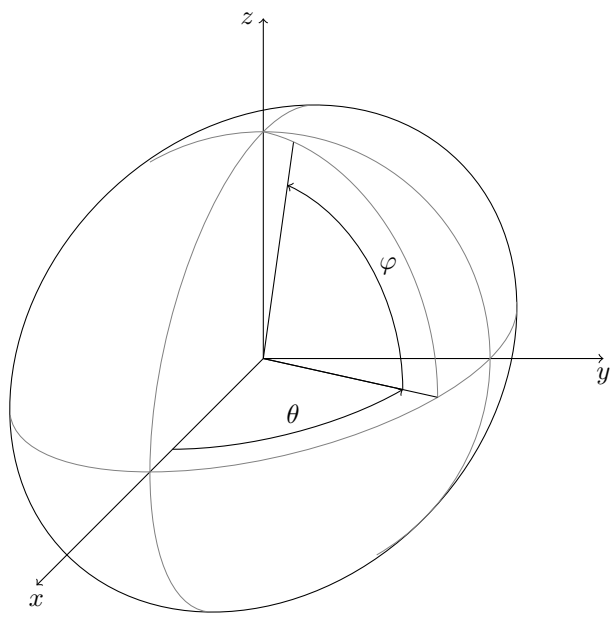

Figure 1: Polar coordinate system.

$$
\begin{aligned}
\tilde{f} &= \|r'(\theta,\varphi)(\dot{\theta},\dot{\varphi})\| \\
&= \left\| \begin{bmatrix} -\sin\theta\cos\varphi & -\cos\theta\sin\varphi \\ \cos\theta\cos\varphi & -\sin\theta\sin\varphi \\ 0 & \cos\varphi \end{bmatrix} \begin{bmatrix} \dot{\theta} \\ \dot{\varphi} \end{bmatrix} \right\| \\
&= \sqrt{\cos^2(\varphi)\,\dot{\theta}^2 + \dot{\varphi}^2}.
\end{aligned}
$$

The Euler-Lagrange equations then become, in local coordinates:

$$
\begin{aligned}
\frac{\partial \tilde{f}}{\partial \theta} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \tilde{f}}{\partial \dot{\theta}} = 0 &\Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\cos^2(\varphi)\dot{\theta}}{f}\right) = 0, \\
\frac{\partial \tilde{f}}{\partial \varphi} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \tilde{f}}{\partial \dot{\varphi}} = 0 &\Leftrightarrow \frac{\cos\varphi\sin\varphi\,\dot{\theta}^2}{f} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{\varphi}}{f}\right).
\end{aligned}
$$

To find a solution to these equations, we begin by pointing out that the length of a curve does not depend on the parametrization, as long as it does not double back. Therefore, we may, for simplicity, find a unit-speed parametrization. This simplification is equivalent to requiring that $f = 1$, so the equations become

$$
\dot{\theta} = \frac{C}{\cos^2(\varphi)}, \quad \ddot{\varphi} = \cos\varphi\sin\varphi\,\dot{\theta}^2,
$$

and the second equation can be simplified to $\ddot{\varphi} = C^2 \frac{\sin\varphi}{\cos^3\varphi}$.

Finally, to simplify matters, let us suppose without loss of generality (by rotation) that the starting point of our curve lies in the equator ($\varphi = 0$), and its initial velocity is $\dot{\theta} = 1$, $\dot{\varphi} = 0$. Then, the Euler-Lagrange equations become easy to solve, as φ constant equal to 0 is a solution of $\ddot{\varphi} = C^2 \frac{\sin\varphi}{\cos^3\varphi}$, and $\dot{\theta} = C = 1$, and so our curve is just a constant speed rotation about the equator. This (and invariance by change of coordinates) proves that the (unit speed) solutions of the Euler-Lagrange equations are great circle rotations.

Of course, this does not solve the problem of length-minimizing curves in the circle, as we would need to ensure that length-minimizing curves are C^2, and solutions of the Euler-Lagrange equations do not necessarily minimize the value of the functional. However, this example has served the purpose of exemplifying the method.

Remark 3. This method also works for time-dependent conditions, but it is slightly more complicated, because the spaces that the conditions define are not just manifolds, but rather 'time-dependent' manifolds. Unfortunately, I do not know of any relevant bibliography on this matter.

3.3 Langrange Multipliers

When solving problems in calculus of optimization in manifolds, sometimes local parametrizations are very messy business. Indeed, even though the implicit function theorem guarantees that the solutions to independent cartesian equations are manifolds in euclidean space, to parametrize these manifolds explicitly is often messy at best, and impossible at worst. This motivates the study of parametrization-independent methods of optimization. The most well-known of these is the method of Lagrange multipliers, which can be extended to the Calculus of Variations in two ways: it can be applied to problems with integral restriction (e.g. curves of a fixed length), see [3, §12.1], or it can be applied to problems with pointwise (i.e. holonomic) restrictions, as we will see in this section. In section 4, we show a more general version of this principle, which can be applied to nonholonomic constraints, and see also remark 7 where we show how integral constraints are really just a particular case of nonholonomic constraints.

The main result of this section is the following:

Theorem 2. *Let u_0 be a curve that minimizes the functional $I(u) = \int f(t, u, \dot{u})$ as u ranges over C^1 curves starting at u_{start}, ending at u_{end}, and satisfying the k cartesian equations $\phi(t, u(t)) = 0$ for all t. Suppose that u_0 is C^2, that f and ϕ are C^2 in an open set containing all points of the form $(t, u_0, \dot{u}_0)$ and (t, u_0) respectively, and that $\frac{\partial\phi}{\partial u}$ has rank k on all points of the form (t, u_0). Then, there exists a continuous function $\lambda\colon\ ]t_{\text{start}}, t_{\text{end}}[\ \to \mathbb{R}^k$ such that*

$$\frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi} = \lambda(t)^{\intercal} \left.\frac{\partial \phi}{\partial u}\right|_{t, u_0(t)}. \tag{1}$$

Remark 4. The most common way to ensure the condition on the rank of $\frac{\partial \phi}{\partial u}$ is to request that 0 is a regular value of ϕ.

Proof. The proof of this theorem proceeds locally. We will show that, for all t, there exists a neighborhood of t and a function λ defined on this neighborhood, satisfying (1). The theorem follows immediately, as the functions thus defined are unique. Indeed, since $\frac{\partial \phi}{\partial u}$ has linearly independent rows, the equation in $\lambda(t)$ given by (1) has at most one solution for each t. The remarkable fact is that these solutions exist for all t, and form a continuous function.

To show the theorem locally, we use the inverse function theorem. Given t_0, since $\frac{\partial \phi}{\partial u}$ has linearly independent rows, we can without loss of generality (by swapping coordinates if necessary) suppose that the $n \times n$ matrix

$$A = \begin{bmatrix} \left.\frac{\partial \phi}{\partial u}\right|_{t_0, u_0(t_0)} \\ 0 \quad I \end{bmatrix}$$

is invertible. Now, let $\Phi(t, u) = (t, \phi(t, u), u_{k+1}, \ldots, u_n)$. It is obvious that

$$\Phi'(t_0, u_0(t_0)) = \begin{bmatrix} 1 & 0 \\ * & A \end{bmatrix},$$

which is invertible and so we are in the conditions of the inverse function theorem.

To simplify notation, let us define $u^2 = (u_{k+1}, \ldots, u_n)$, so that, for example, Φ can be written as $\Phi(t, u) = (t, \phi(t, u), u^2)$. Then, by the inverse function theorem, there exist open sets in $\mathbb{R}^{n+1}$, say U and V such that:

1. $(t_0, u_0(t_0)) \in U$, and
2. Φ is a C^2 diffeomorphism $U \to V$.

Since $(t_0, u_0(t_0)) \in U$ and $\Phi(t_0, u_0(t_0)) \in V$, by continuity there exists a neighborhood of t_0, say $B =]t_0 - \delta, t_0 + \delta[$, such that for all t in this neighborhood we have $(t, u_0(t)) \in U$ and $\Phi(t, u_0(t)) \in V$. Furthermore, suppose that δ is small enough such that, for all $t \in B$, for all $z \in \mathbb{R}^n$ of norm less than some d, $\Phi(t, u_0(t)) + (0, 0, z) \in V$. To show that such a δ exists, begin by picking a neighborhood of $\Phi(t_0, u_0(t_0))$ of the form $R =]t_0 - \delta_0, t_0 + \delta_0[\times B_{2d}(0, u_0^2(t_0))$ such that R is contained in V. Now pick $\delta < \delta_0$ such that $u_0(t)$ differs less than d from $u_0(t_0)$ for t within δ of t_0. This exists by continuity. Then, for $t \in]t_0 - \delta, t_0 + \delta[$ and z of norm less than d, we have that

$$\left\|(0, u_0^2(t_0)) - \big((0, u_0^2(t)) + (0, z)\big)\right\| \leq \|u_0^2(t_0) - u_0^2(t)\| + \|z\| \leq 2d.$$

This shows that $\Phi(t, u_0(t)) + (0, 0, z) \in R \subseteq V$, which allows us to define the variation of I at u_0 with respect to some variations contained in B. More specifically, let $v : [t_{\text{start}}, t_{\text{end}}] \to \mathbb{R}^{n-k}$ be a C^∞ function of support contained

in B. Then, for ε small enough (such that $\|\varepsilon v\| \leq d$), for all $t \in B$ we have $(t, \phi(t, u_0(t)), u_0^2(t) + \varepsilon v(t)) \in V$. Then, we may define u_ε as

$$u_\varepsilon(t) = \begin{cases} u_0(t), & t \notin \operatorname{supp} v, \\ \Psi(t, \phi(t, u_0(t)), u_0^2(t) + \varepsilon v(t)), & t \in B, \end{cases}$$

where Ψ is defined as follows. Since Φ is of the form $\Phi(t, u) = (t, \phi(t, u), u^2)$, its inverse must be of the form $\Phi^{-1}(t, w) = (t, \psi(t, w), w^2)$. We set $\Psi(t, w) = (\psi(t, w), w^2)$. In other words, $\Psi = \begin{bmatrix} 0 & I \end{bmatrix} \circ \Phi^{-1}$, i.e. Ψ is the inverse of Φ with the time component thrown out.

Since both branches of this definition are open, agree on the intersection, and each branch is C^1, we conclude that u_ε is also a C^1 path. It is also trivial to see that it satisfies the boundary conditions and that $\phi(t, u_\varepsilon(t)) = 0$. Therefore, we must have

$$I(u_\varepsilon) \geq I(u_0)$$

Now, the standard idea is to take the derivative in $\varepsilon = 0$ and apply Leibniz to get that a certain integral expression is null. However, for our purposes it is useful to first split the integral in two parts:

$$I(u_\varepsilon) = \int_{[t_{\text{start}}, t_{\text{end}}] \setminus B} f(t, u_0, \dot{u}_0) \, \mathrm{d}t + \int_B f(t, u_\varepsilon, \dot{u}_\varepsilon) \, \mathrm{d}t.$$

Now, the first expression is constant in ε, so differentiating and applying Leibniz we get

$$\frac{\mathrm{d}I(u_\varepsilon)}{\mathrm{d}\varepsilon} = \int_B \frac{\partial}{\partial \varepsilon} f(t, u_\varepsilon, \dot{u}_\varepsilon) \, \mathrm{d}t.$$

The usual chain rule and integration by parts of the calculus of variation yields[1]

$$\frac{\mathrm{d}I(u_\varepsilon)}{\mathrm{d}\varepsilon} = \int_B \frac{\partial f}{\partial u} \frac{\partial u_\varepsilon}{\partial \varepsilon} + \frac{\partial f}{\partial \xi} \frac{\partial \dot{u}}{\partial \varepsilon} \, \mathrm{d}t = \int_B \left(\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f}{\partial \xi} \right) \frac{\partial u}{\partial \varepsilon} \, \mathrm{d}t.$$

Note that in the integration by parts we used the fact that partial derivatives commute, and so

$$\frac{\partial \dot{u}_\varepsilon}{\partial \varepsilon} = \frac{\partial^2 u_\varepsilon(t)}{\partial \varepsilon \, \partial t} = \frac{\partial}{\partial t} \frac{\partial u_\varepsilon}{\partial \varepsilon}.$$

The term $\frac{\partial u}{\partial \varepsilon}$ plays the part of the variation in the usual proof of the Euler-Lagrange equations. If $\frac{\partial u}{\partial \varepsilon}$ were arbitrary, we would in fact conclude Euler-Lagrange. However, they are not. Intuitively, they are arbitrary, though tangent to the manifold given by the cartesian equations $\phi = 0$. Therefore, one intuits that the term $\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f}{\partial \dot{u}}$ is always orthogonal to the manifold, which matches up with the expected conclusion, as we know from vector calculus that the

[1] Here is where we use the assumption that ϕ is C^2 and that Φ is a C^2 diffeomorphism, because we need that the function $\frac{\partial f}{\partial \xi}$ be C^1, and so that $\dot{u}_\varepsilon$ be C^1 in t and ε. To ensure that this is the case, we require the assumption that Ψ is C^2, and so that Φ is a C^2 diffeomorphism.

space orthogonal to a manifold defined by cartesian equations is generated by $\nabla_x\phi_1, \ldots, \nabla_x\phi_k$..

To proceed with the proof, let us investigate $\frac{\partial u_\varepsilon}{\partial \varepsilon}\big|_{\varepsilon=0}$. This can be computed using the chain rule:

$$\begin{aligned}
\left.\frac{\partial u_\varepsilon}{\partial \varepsilon}\right|_0 &= \left.\frac{\partial}{\partial \varepsilon}\Psi(t, \phi(t, u_0(t)), u_0^2(t) + \varepsilon v(t))\right|_0 \\
&= \begin{bmatrix} 0 & I \end{bmatrix} \left.\frac{\partial}{\partial \varepsilon}\Phi^{-1}(t, \phi(t, u_0(t)), u_0^2(t) + \varepsilon v(t))\right|_0 \\
&= \begin{bmatrix} 0 & I \end{bmatrix} (\Phi^{-1})'(t, \phi(t, u_0(t)), u_0^2(t))(0, 0, v(t)) \\
&= \begin{bmatrix} 0 & I \end{bmatrix} \left(\Phi'(\Phi^{-1}(t, \phi(t, u_0(t)), u_0^2(t)))\right)^{-1} (0, 0, v(t)) \\
&= \begin{bmatrix} 0 & I \end{bmatrix} (\Phi'(t, u_0(t)))^{-1} (0, 0, v(t)).
\end{aligned}$$

Therefore, the conclusion is that, for all smooth v defined in B,

$$\int_B \left(\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}\right) \begin{bmatrix} 0 & I \end{bmatrix} (\Phi'(t, u_0(t)))^{-1} (0, 0, v(t))\, \mathrm{d}t = 0.$$

Now, we may apply at once the Fundamental Lemma of the Calculus of Variations, to conclude that the last $n-k$ coordinates of the $(n+1)$-dimensional row-vector

$$\left(\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}\right) \begin{bmatrix} 0 & I \end{bmatrix} (\Phi'(t, u_0(t)))^{-1}$$

are null. In other words, we can write this vector in the form

$$\left(\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}\right) \begin{bmatrix} 0 & I \end{bmatrix} (\Phi'(t, u_0(t)))^{-1} = (\ell(t), \lambda(t), 0)^\intercal,$$

for some $\ell : B \to \mathbb{R}$ and $\lambda : B \to \mathbb{R}^k$. Note that these functions are continuous, as they are a product of continuous functions, and we cannot ensure more regularity under only C^2 assumptions because of the $\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}$ term.

To finalize, let us try to solve for $\frac{\partial f}{\partial u} + \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}$, which we will temporarily refer to as $T(t)$. Simply put, multiplying by Φ' and applying matrix block multiplication,

$$\begin{bmatrix} 0 & T(t) \end{bmatrix} = (\ell(t), \lambda(t), 0)^\intercal \Phi'(t, u_0(t)) = (\ell(t), \lambda(t), 0)^\intercal \begin{bmatrix} 1 & 0 & 0 \\ * & \frac{\partial \phi}{\partial u} & \\ * & 0 & I \end{bmatrix}.$$

Therefore, we conclude $T(t) = \lambda(t)^\intercal \frac{\partial \phi}{\partial u}$, and so the proof is complete. ■

Example 2. Let us reproduce example 1, i.e. to find paths of minimal length on the unit sphere. In this case,

$$\begin{aligned}
\phi &= x^2 + y^2 + z^2 - 1, \\
f &= \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}.
\end{aligned}$$

Our variables are $x(t)$, $y(t)$, $z(t)$ and $\lambda(t)$, and our equations are

$$\frac{\partial f}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \dot{x}} = \lambda(t)\frac{\partial \phi}{\partial x},$$
$$\text{(Same for } y \text{ and } z\text{)},$$
$$\phi = 0,$$

and these equations simplify to

$$-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\dot{x}}{f} = 2\lambda(t)x,$$
$$\text{(Same for } y \text{ and } z\text{)},$$
$$x^2 + y^2 + z^2 = 1,$$

As in example 1, let us make the simplifying assumption of unit speed parametrizations, i.e. $f = 1$. With this simplification, our equations become

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}(x, y, z) = -2\lambda(t)(x, y, z),$$
$$x^2 + y^2 + z^2 = 1.$$

In what follows, let $q = (x, y, z)$. The previous equations state that

$$\ddot{q} = -2\lambda q, \quad \|q\|^2 = 1.$$

If we can show that λ is constant equal to $1/2$, q obeys the Harmonic Oscillator equation, and it is easy to then write out the solutions explicitly as sines and cosines to obtain great circles. Therefore, it suffices to try and solve for λ. To this effect, we use the following trick. Since we are assuming that q has constant norm, we get that $\langle \dot{q}, q \rangle = 0$ by the product rule. Differentiating this expression a second time, we get

$$0 = \frac{\mathrm{d}\,\langle \dot{q}, q \rangle}{\mathrm{d}t} = \langle \ddot{q}, q \rangle + \langle \dot{q}, \dot{q} \rangle .$$

Since $\ddot{q} = -2\lambda q$ and we are assuming a unit speed parametrization, this gives

$$0 = -2\lambda + 1.$$

Consequently, λ is constant equal to $1/2$, and the example is complete.

4 Nonholonomic Constraints

4.1 Introduction

In the previous section, we went over problems with holonomic (pointwise) constraints, i.e. optimization problems where the path u to be optimized must

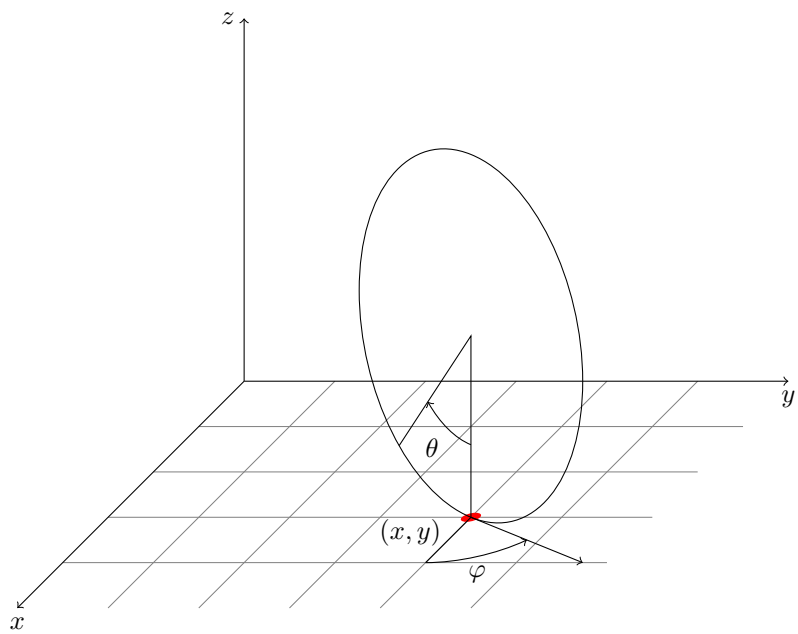

Figure 2: The coordinate system for the rolling upright disk.

satisfy k (independent) cartesian equations of the form $\phi(t, u) = 0$. For a *nonholonomic* problem, the constraints are now differential equations of the form $\phi(t, u, \dot{u}) = 0$.

Problems with nonholonomic constraints occur often when describing physical systems. The classical example is that of an upright disk rotating without slipping on the xy plane. It is described by four coordinates: the (x, y) coordinates of the disk's center, the angle of the disk about the vertical axis, which we call φ, and the disk's rotation about its normal axis, which we call θ. See figure 4.1.

Then, even though all four coordinates can attain any number, they are not independent. If the θ angle changes a little bit, that induces a change in x and y that depends on φ. In physics texts, this condition is often described using infinitesimals, as follows

$$\mathrm{d}x = \cos\varphi\,\mathrm{d}\theta, \quad \mathrm{d}y = \sin\varphi\,\mathrm{d}\theta.$$

For our mathematical purposes, we 'divide by $\mathrm{d}t$' to obtain the ordinary differential equation

$$\begin{cases} \dot{x} = \cos\varphi\,\dot{\theta}, \\ \dot{y} = \sin\varphi\,\dot{\theta}. \end{cases}$$

Therefore, if we intend to model the possible paths such a rolling disk might take, they ought to satisfy this differential equation.

In dealing with holonomic constraints, it was essential to assume that the equations of constraint were independent, in the sense that the space-coordinate derivative had full rank. Such an assumption is also necessary here, except that as we now wish to ensure that a collection of *differential* equations is independent, we consider the derivative in the velocity-coordinate.

Problem Statement. A problem of optimization with nonholonomic constraints is that in which we intend to minimize the value of the functional $I(u) = \int f(t, u, \dot{u})\,\mathrm{d}t$ under the constraints:

1. u has prescribed initial and final conditions, i.e. $u(t_{\text{start}}) = u_{\text{start}}$ and $u(t_{\text{end}}) = u_{\text{end}}$, and
2. u satisfies the equation $\phi(t, u, \dot{u}) = 0$, where

$$\phi \equiv \phi(t, u, \xi) \colon \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^k.$$

The k differential equations $\phi = 0$ are said to be *independent* if the $k \times n$ matrix given by $\frac{\partial \phi}{\partial \xi}$ has rank k, i.e. has linearly independent rows.

Remark 5. In most applications, ϕ is given in the so-called Pfaffian form, i.e.

$$\phi(t, u, \xi) = A(t, u)\xi + b(t, u),$$

for some matrix A and vector b. In this case, the requirement of independence of the equations corresponds to requiring that the rows of A are linearly independent.

Remark 6. The definition we have just given is not universal. Indeed, in physics, the term nonholonomic is usually reserved for constraints which *cannot* be made holonomic. So, for example, a constraint of the form $\mathrm{d}x = \mathrm{d}y + \mathrm{d}z$ would not be nonholonomic, because it is equivalent to $x = y + z + C$ for some fixed C.

However, in this document the distinction is not useful, and we will use the term nonholonomic for any problem whose constraints are differential equations. Indeed, we consider holonomic problems a subset of nonholonomic problems: see remark 7 below.

Remark 7. A wide class of problems can be seen as nonholonomic. For example, any holonomic problem can be seen in this framework.

Consider a holonomic problem with restriction $\phi(t, u) = 0$. The obvious way to transform it into a nonholonomic problem would be to simply write $\widetilde{\phi}(t, u, \xi) = \phi(t, u)$. However, $\widetilde{\phi}$ does not satisfy the independence requirement for nonholonomic problems, which means that our methods will not be applicable. Instead, a better way to turn a holonomic problem into a nonholonomic one would be to note that $\phi(t, u)$ is constant, and so, differentiating in time, $\frac{\partial \phi}{\partial u}\dot{u} + \frac{\partial \phi}{\partial t} = 0$. This can be seen as a nonholonomic constraint in Pfaffian form, and it is easy to see that independence of the holonomic equations $\phi = 0$

is equivalent to independence of the nonholonomic equations $\psi = 0$, where $\psi(t, u, \xi) = \frac{\partial \phi}{\partial u}\Big|_{t,u} \xi + \frac{\partial \phi}{\partial t}\Big|_{t,u}$ (see remark 5).

Another class of problems that can be seen in nonholonomic form is problems with functional constraints. That is, suppose we wish to consider paths u satisfying $\int g(t, u, \dot{u})\mathrm{d}t = c$. Such a constraint can be written as a differential equation by adding an auxiliary quantity, which we call $j(t)$, which represents $\int_{\text{start}}^{t} g$. Then, we add the constraints

$$j(t_{\text{start}}) = 0, \quad j(t_{\text{end}}) = c, \quad \frac{\mathrm{d}j}{\mathrm{d}t} = g(t, u, \dot{u}).$$

4.2 Valid Displacements

The strategy used to prove the basic form of the Euler-Lagrange equations, which consists of adding a perturbation to a path u_0 which is assumed to minimize $I(u)$, requires vector space structure, which is lost when we restrict our paths to those satisfying nonlinear conditions. In the holonomic case, this problem was recoverable by using the inverse/implicit function theorem to locally 'distort the space into $\mathbb{R}^n$'.

Unfortunately, such strategies fall completely flat when passing to the nonholonomic case. Because of the differential nature of the restrictions, small perturbations will propagate from the place of perturbation to the end of the path. The easiest way to see this is to imagine a problem in which the constraint is given by $\|\dot{u}\| = 1$. Then, a straight line cannot possibly be perturbed without changing the initial or final location. This means that we necessarily have to give up on the strategy of considering small admissible perturbations of a minimal path.

Therefore, our strategy will be to consider perturbations that keep only the start of the path fixed. We will concern ourselves with the end later on. But even lifting this restriction, it is not obvious that *any* nontrivial perturbation exists, or how to construct it. It is here that the hypothesis of independence of the differential equations comes into play, because it allows is to apply a kind of implicit/inverse function theorem, which we will now state. The proof of this theorem is left to appendix A. For an alternative proof, we direct the reader to [2, §3].

Theorem 3. *Let $u\colon [t_{\text{start}}, t_{\text{end}}] \to \mathbb{R}^n$ be a C^p curve, for $p \in \mathbb{N}_0$, Ω an open subset of $X = \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n$ such that $(t, u(t), \dot{u}(t)) \in \Omega$ for all t, and $\phi \equiv \phi(t, u, \xi)$ a C^{p+1} function $\phi\colon \Omega \to \mathbb{R}^k$.*

Suppose that for all t the matrix $\frac{\partial \phi}{\partial \xi}$ has rank k. Then there exist open subsets of X, say U and V, and a function $\widetilde{\phi}\colon U \to \mathbb{R}^n$, such that

- *U contains all points of the form $(t, u(t), \dot{u}(t))$,*
- *U is contained in Ω,*
- *The first k coordinates of $\widetilde{\phi}$ agree with ϕ*

- *The function given by*

$$\begin{aligned}\Phi\colon U &\to V\\ (t,u,\xi) &\mapsto (t,u,\widetilde{\phi}(t,u,\xi))\end{aligned}$$

is a C^p diffeomorphism.

We will now use this theorem to build our perturbations of a path. Indeed, given a path u_0, construct the U, V, and Φ given above. Let the inverse of Φ be denoted $(t,u,\xi)\mapsto(t,u,\psi(t,u,\xi))$.

We can use ψ to build a valid displacement of u as follows. We wish to find a perturbation of u, let's say $\tilde{u}$, such that $\phi(t,u,\dot{u})=0$. Therefore, $\widetilde{\phi}(t,u,\dot{u})$ is of the form $(0,w(t))$ for some v. This can be solved for $\dot{u}$, and we get the differential equation

$$\dot{u}(t)=\psi(t,u(t),w(t)).$$

We can now use ODE theory to guarantee that this has a solution for w 'close enough to w_0', for 'w_0 corresponding to the path u_0'.

Theorem 4. *Let $p\in\mathbb{N}^+$.*[2] *Let u be a C^{p+1} path satisfying $\phi(t,u,\dot{u})=0$ for ϕ a C^p function such that $\frac{\partial\phi}{\partial\xi}$ has linearly independent rows. Extend ϕ to $\widetilde{\phi}$ (in a C^p way) using the methods of theorem 3. Let $\widetilde{\phi}(t,u,\dot{u})=(0,w(t))$. Then, given a curve $v:[t_{\text{start}},t_{\text{end}}]\to\mathbb{R}^{n-k}$ of class C^p in this interval, for real numbers b close enough to zero the following ODE has a unique solution with $(t,u,\dot{u})\in U$:*

$$\begin{cases}\widetilde{\phi}(t,u,\dot{u})=(0,w(t)+b\,v(t)),\\ u(t_{\text{start}})=u_{\text{start}}.\end{cases}$$

For b small enough that this exists, we refer to this solution as $u_b^v(t)$, or u_b if v is implied.

The resulting function of two parameters (b and t) is C^p.

Proof. We begin by applying theorem 3 to ϕ, to obtain the U, V and Φ therein described. Let

$$\Phi^{-1}(t,u,\xi)=(t,u,\psi(t,u,\xi)).$$

Note that all points of the form $(t,u(t),0,w(t))$ are in V, for $t\in[t_{\text{start}},t_{\text{end}}]$. The set of these points form a compact set, and it is known from metric space theory that there exists a $\delta>0$ such that for all z with $\|z\|<\delta$ we have $(t,u,0,w)+z\in V$ for all t. Consequently, if $b<\delta/\max\|v\|$, $(t,u,0,w+bv)\in V$, and so[3]

$$\widetilde{\phi}(t,u,\dot{u})=(0,w(t)+b\,v(t))\text{ iff }\dot{u}(t)=\psi(t,u(t),w(t)+b\,v(t)).$$

Therefore, (for small enough b) our ODE is of the form

$$\begin{cases}\dot{u}(t)=F(t,u(t),b),\\ u(t_{\text{start}})=u_{\text{start}},\end{cases}$$

[2] We will only use the case $p=1$.

[3] It is subtle, but this 'if and only if' is where we use the hypothesis that $(t,u,\dot{u})\in U$.

to which elementary ODE theory is applicable. Note that since ψ, w, and v are C^p, F is also C^p. In particular, it is C^1, and so (see [5, p.74]) the solution exists and is unique for small enough t. Moreover, it is known to exist in $[t_{\text{start}}, t_{\text{end}}]$ for $b = 0$ (it is u_0 in this case), and consequently (see [5, thm.2, p.84]) the solution exists and is unique for $t \in [t_{\text{start}}, t_{\text{end}}]$, for b close enough to 0.

To show differentiability in these two variables, the cited reference proves only the case $p = 1$. However, the methods therein found can be used to prove inductively for all $p \in \mathbb{N}^+$ that if F is C^p then u is also C^p in time and the parameters. ■

Remark 8. Though the above proof was stated for only one parameter b, it could easily be stated and extended for several parameters and v's. In other words, we could have considered the ODE

$$\begin{cases} \widetilde{\phi}(t, u, \dot{u}) = (0, w(t) + b_1\, v_1(t) + \cdots + b_q v_q(t)), \\ u(t_{\text{start}}) = u_{\text{start}}. \end{cases}$$

The same result and notation applies, except that we collect the functions v_i in a $(n-k) \times q$ matrix we call $V(t)$, and the scalars b_i form a vector we call $\bar{b}$. This will be crucial in section 4.4, which contains the main result of this document: the multiplier rule.

4.3 First Variation

Consider a one-parameter family of arcs of the form $u_b^v(t)$, constructed as described above. We will abbreviate this to $u_b(t)$. We will now investigate the behavior of $I(u_b)$ around zero: more precisely, we will study $\left.\frac{\mathrm{d}I(u_b)}{\mathrm{d}b}\right|_0$, i.e. the first variation of I along this family.

Recall that $I(u_b)$ is defined as $\int f(t, u_b(t), \dot{u}_b(t))\mathrm{d}t$, and since this is integration over a compact interval of a C^1 function in both its variables (assuming f is C^1 and $u_b(t)$ is C^2 jointly in both variables) Leibniz' rule is applicable and so

$$\frac{\mathrm{d}I(u_b)}{\mathrm{d}b} = \int_{\text{start}}^{\text{end}} \frac{\partial}{\partial b} f(t, u_b, \dot{u}_b)\mathrm{d}t = \int_{\text{start}}^{\text{end}} \frac{\partial f}{\partial u}\frac{\partial u_b}{\partial b} + \frac{\partial f}{\partial \xi}\frac{\partial \dot{u}_b}{\partial b}\,\mathrm{d}t.$$

If we now assume that f is C^2, the usual integration by parts of the calculus of variations can be performed. In doing so, we are implicitly using the fact that, since u is C^2 in b and t, the order of derivatives is irrelevant, and thus $\frac{\partial \dot{u}_b}{\partial b} = \frac{\partial^2 u_b(t)}{\partial t \partial b}$. We conclude

$$\frac{\mathrm{d}I}{\mathrm{d}b} = \int_{\text{start}}^{\text{end}} \left(\frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}\right)\frac{\partial u_b}{\partial b}\,\mathrm{d}t + \left[\frac{\partial f}{\partial \xi}\frac{\partial u_b}{\partial b}\right]_{\text{start}}^{\text{end}}.$$

Finally, note that we are keeping the first end of u fixed, so that $u_b(t_{\text{start}})$ is constant and so $\left.\frac{\partial u_b}{\partial b}\right|_{\text{start}} = 0$. Therefore, the bracket term can be simplified, yielding

$$\frac{\mathrm{d}I}{\mathrm{d}b} = \int_{\text{start}}^{\text{end}} \left(\frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}\right)\frac{\partial u_b}{\partial b}\,\mathrm{d}t + \left.\frac{\partial f}{\partial \xi}\right|_{\text{end}}\left.\frac{\partial u_b}{\partial b}\right|_{\text{end}}. \tag{2}$$

In order to avoid writing a burdensome amount of terms, we now introduce a notation for the 'Euler-Lagrange derivative'. Let Θ be the differential operator given by

$$\Theta(f) := \frac{\partial f}{\partial u} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial f}{\partial \xi}.$$

With this notation, (2) can be written

$$\frac{\mathrm{d}I}{\mathrm{d}b} = \int_{\text{start}}^{\text{end}} \Theta(f)\frac{\partial u_b}{\partial b}\,\mathrm{d}t + \left.\frac{\partial f}{\partial \xi}\right|_{\text{end}} \left.\frac{\partial u_b}{\partial b}\right|_{\text{end}}. \tag{3}$$

There exists a trick that will later be used to get rid of the non-integral terms, which we now describe. Recall that u_b satisfies the differential equation

$$\widetilde{\phi}(t, u_b(t), \dot{u}_b(t)) = (0, w(t) + b\,v(t)),$$

where $\widetilde{\phi}(t, u_0, \dot{u}_0) = (0, w(t))$. This actually consists of n differential equations, which can be added together. Indeed, if $\mu(t) = (\mu_1(t), \mu_2(t))$ is a regular enough time-dependent n-vector,

$$\mu \cdot \widetilde{\phi}(t, u, \dot{u}) = \mu_2 \cdot w + b\,\mu_2 \cdot v.$$

Taking the partial derivative in b, we obtain

$$\mu \cdot \frac{\partial \widetilde{\phi}}{\partial u}\frac{\partial u}{\partial b} + \mu \cdot \frac{\partial \widetilde{\phi}}{\partial \xi}\frac{\partial \dot{u}}{\partial b} = \mu_2 \cdot v,$$

which can be rewritten in terms of Θ as

$$\Theta(\mu \cdot \widetilde{\phi})\frac{\partial u}{\partial b} = \mu_2 \cdot v - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial(\mu \cdot \widetilde{\phi})}{\partial \xi}\frac{\partial u}{\partial b}\right).$$

Integrating, this yields

$$\int_{\text{start}}^{\text{end}} \Theta(\mu \cdot \widetilde{\phi})\frac{\partial u}{\partial b}\,\mathrm{d}t = \int_{\text{start}}^{\text{end}} \mu_2 \cdot v\,\mathrm{d}t - \left[\frac{\partial(\mu \cdot \widetilde{\phi})}{\partial \xi}\frac{\partial u}{\partial b}\right]_{\text{start}}^{\text{end}}.$$

Using the same trick as before, the bracket can be simplified to get

$$\int_{\text{start}}^{\text{end}} \Theta(\mu \cdot \widetilde{\phi})\frac{\partial u}{\partial b}\,\mathrm{d}t = \int_{\text{start}}^{\text{end}} \mu_2 \cdot v\,\mathrm{d}t - \mu(t_{\text{end}}) \cdot \left.\frac{\partial \widetilde{\phi}}{\partial \xi}\right|_{\text{end}} \left.\frac{\partial u}{\partial b}\right|_{\text{end}}.$$

Adding and subtracting this expression to a multiple of $\frac{\mathrm{d}I}{\mathrm{d}b}$, using (3) we get

$$\begin{aligned}\ell_0\frac{\mathrm{d}I}{\mathrm{d}b} = \int_{\text{start}}^{\text{end}} \Theta(\ell_0 f + \mu \cdot \widetilde{\phi})\frac{\partial u_b}{\partial b}\,\mathrm{d}t - \int_{\text{start}}^{\text{end}} \mu_2 \cdot v\,\mathrm{d}t \\ + \left(\ell_0 \left.\frac{\partial f}{\partial \xi}\right|_{\text{end}} + \mu(t_{\text{end}})^{\mathsf{T}} \left.\frac{\partial \widetilde{\phi}}{\partial \xi}\right|_{\text{end}}\right) \left.\frac{\partial u_b}{\partial b}\right|_{\text{end}}.\end{aligned} \tag{4}$$

Even though this expression looks more complex than (3), it has a large degree of freedom to exploit: μ has been left arbitrary (as long as it is C^1 so that the Θ term exists). For example, we could cross out the non-integral term by choosing an appropriate value for $\mu(t_{\text{end}})$. We will not do so yet, however, because an even more appropriate choice will present itself in due time. The mysterious ℓ_0 coefficient appears for technical reasons, and is necessary in a proof. If the reader is mystified by its presence, he or she may ignore it for now.

4.4 The Multiplier Rule

In this section, we will finally prove the necessary condition for a minimum that we have been building up to: the Lagrange Multipler Rule.

Suppose that that u_0 is a minimum of the functional I over the class of paths such that $u(t_{\text{start}}) = u_{\text{start}}$, $u(t_{\text{end}}) = u_{\text{end}}$ and $\phi(t, u, \dot{u}) = 0$, under the hypothesis that f is C^2, u_0 is C^2 and ϕ is C^3 and that $\frac{\partial \phi}{\partial \xi}$ has full rank over the path u_0.

We might be tempted to apply the formula (3) or (4) for $\frac{\mathrm{d}I}{\mathrm{d}b}$ and claim that $\frac{\mathrm{d}I}{\mathrm{d}b} = 0$ for all v, since u_0 minimizes the value of I. However, that is neglecting the requirement that $u(t_{\text{end}}) = u_{\text{end}}$: since the variations we have created break this requirement, $I(u_b)$ has no obligation to be minimal at $b = 0$. One could consider restricting themselves to variations v that guarantee the boundary conditions, but such v might not even exist: consider the example where u_0 is a straight line and $\phi(t, u, \xi) = \|\xi\|$.

To get around this, consider an arbitrary collection $v_1, \ldots, v_{n+1}$ of smooth functions $[t_{\text{start}}, t_{\text{end}}] \to \mathbb{R}^{n-k}$, which we collect into the matrix $V(t)$. We can see I as function in $n+1$-variables, given by the vector $\bar{b}$ (see remark 8). Furthermore, we can see $u(t_{\text{end}})$ as a function of $\bar{b}$ as well, and so we can consider the function (with $n+1$ variables and values in $\mathbb{R}^{n+1}$) given by

$$\bar{b} \mapsto (I(u_{\bar{b}}^V), u_{\bar{b}}(t_{\text{end}})). \tag{5}$$

This function is C^1 in a neighborhood of the origin. To see why, first note that theorem 4 guarantees that $u_{\bar{b}}(t)$ is C^2 as a function of both parameters, and so in particular is C^1 if we fix $t = t_{\text{end}}$. This shows that the last n coordinates of (5) are C^1. It now suffices to show that $I(u_{\bar{b}}^V)$ is C^1. To do so, note that f is C^1 in its three arguments: t, $u_{\bar{b}}^V(t)$ and $\dot{u}_{\bar{b}}^V(t)$. Furthermore, since all its arguments are C^1 in t and $\bar{b}$ (here we use that ϕ is C^3 to show that u is C^2 in t and $\bar{b}$ and so $\dot{u}$ is C^1 in $\bar{b}$), we get that so is the composite, and Leibniz' theorem guarantees that an integral in t of a C^1 function in t and $\bar{b}$ is C^1 in $\bar{b}$.

Since (5) is C^1 from $n+1$-dimensional space to $n+1$-dimensional space, we are in conditions of applying the inverse function theorem. We conclude that if the jacobian of (5) at $\bar{b} = 0$ were invertible, (5) would be a diffeomorphism in a neighborhood of the origin. But this contradicts the assumption that u_0 is a minimum of I, for in this case we could find $\bar{b}$ such that $u_{\bar{b}}(t_{\text{end}}) = u_{\text{end}}$, and $I(u_{\bar{b}}^V)$ is slightly smaller than $I(u_0)$. In other words, under the assumption that u is a minimum of I, the following matrix has linearly dependent columns: (evaluated at $\bar{b} = 0$)

$$\begin{bmatrix} \dfrac{\partial}{\partial b_1} I(u_{\bar{b}}^V) & \cdots & \dfrac{\partial}{\partial b_{n+1}} I(u_{\bar{b}}^V) \\ \dfrac{\partial}{\partial b_1} u_{\bar{b}}^V(t_{\text{end}}) & \cdots & \dfrac{\partial}{\partial b_{n+1}} u_{\bar{b}}^V(t_{\text{end}}) \end{bmatrix}.$$

Note that, at $\bar{b} = 0$, these columns can be simplified:

$$\begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}b} I(u_b^{v_1}) & \dots & \frac{\mathrm{d}}{\mathrm{d}b} I(u_b^{v_{n+1}}) \\ \frac{\mathrm{d}}{\mathrm{d}b} u_b^{v_1}(t_{\mathrm{end}}) & \dots & \frac{\mathrm{d}}{\mathrm{d}b} u_b^{v_{n+1}}(t_{\mathrm{end}}) \end{bmatrix}.$$

Since this holds for arbitrary collections $v_1, \dots, v_{n+1}$, we can actually conclude a surprisingly strong result. Let q be the maximal possible number of $v_1, \dots, v_q$ such that the following matrix has linearly independent columns

$$A = \begin{bmatrix} \frac{\mathrm{d}}{\mathrm{d}b} I(u_b^{v_1}) & \dots & \frac{\mathrm{d}}{\mathrm{d}b} I(u_b^{v_q}) \\ \frac{\mathrm{d}}{\mathrm{d}b} u_b^{v_1}(t_{\mathrm{end}}) & \dots & \frac{\mathrm{d}}{\mathrm{d}b} u_b^{v_q}(t_{\mathrm{end}}) \end{bmatrix}.$$

By the above, $q < n+1$. Therefore, for any v, the vector $(\frac{\mathrm{d}}{\mathrm{d}b} I(u_b^v), \frac{\mathrm{d}}{\mathrm{d}b} u_b^v(t_{\mathrm{end}}))$ is a linear combination of the columns of A, i.e. of the form Aw for some $w \in \mathbb{R}^{n+1}$.

Since the matrix A has more rows than columns, there exists a row-matrix, say $E = \begin{bmatrix} \ell_0 & e \end{bmatrix} = \begin{bmatrix} \ell_0 & e_1 & \dots & e_n \end{bmatrix}$, such that $EA = 0$. By the previous paragraph, we conclude that *for this specific collection of scalars*(!), for any smooth v we have

$$\ell_0 \frac{\mathrm{d}}{\mathrm{d}b} I(u_b^v) + e \frac{\mathrm{d}}{\mathrm{d}b} u_b^v(t_{\mathrm{end}}) = 0. \tag{6}$$

We can now rewrite the variation in I using (4) to get

$$\ell_0 \frac{\mathrm{d}I}{\mathrm{d}b} = \int_{\mathrm{start}}^{\mathrm{end}} \Theta(\ell_0 f + \mu \cdot \tilde{\phi}) \frac{\partial u_b}{\partial b} \mathrm{d}t - \int_{\mathrm{start}}^{\mathrm{end}} \mu_2 \cdot v \, \mathrm{d}t + \left(\ell_0 \left. \frac{\partial f}{\partial \xi} \right|_{\mathrm{end}} + \mu(t_{\mathrm{end}})^{\intercal} \left. \frac{\partial \tilde{\phi}}{\partial \xi} \right|_{\mathrm{end}} \right) \left. \frac{\partial u_b}{\partial b} \right|_{\mathrm{end}},$$

where, as the reader will recall, μ has been left as an arbitrary C^1 function. Substituting in (6), we obtain

$$\int_{\mathrm{start}}^{\mathrm{end}} \Theta(\ell_0 f + \mu \cdot \tilde{\phi}) \frac{\partial u_b}{\partial b} \mathrm{d}t - \int_{\mathrm{start}}^{\mathrm{end}} \mu_2 \cdot v \, \mathrm{d}t + \left(\ell_0 \left. \frac{\partial f}{\partial \xi} \right|_{\mathrm{end}} + \mu(t_{\mathrm{end}})^{\intercal} \left. \frac{\partial \tilde{\phi}}{\partial \xi} \right|_{\mathrm{end}} + e \right) \left. \frac{\partial u_b}{\partial b} \right|_{\mathrm{end}} = 0, \text{ for all regular } v. \tag{7}$$

Even though this looks like a very complicated expression, this is where the arbitrariness of μ comes into play. For instance, we can now pick $\mu(t_{\mathrm{end}})$ to make

$$\ell_0 \left. \frac{\partial f}{\partial \xi} \right|_{\mathrm{end}} + \mu(t_{\mathrm{end}})^{\intercal} \left. \frac{\partial \tilde{\phi}}{\partial \xi} \right|_{\mathrm{end}} + e = 0.$$

This is possible because $\frac{\partial \tilde{\phi}}{\partial \xi}$ is invertible. Note that this choice is independent of v.

As a further simplification, we may wish to ensure that $\Theta(\ell_0 f + \mu \cdot \widetilde{\phi})$. Expanding the expression, we get the following linear ODE:

$$\ell_0 \frac{\partial f}{\partial u} + \mu^\intercal \frac{\partial \widetilde{\phi}}{\partial u} = \ell_0 \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial f}{\partial \xi} + \mu^\intercal \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial \widetilde{\phi}}{\partial \xi} + \dot{\mu}^\intercal \frac{\partial \widetilde{\phi}}{\partial \xi}, \tag{8}$$

where all partial derivatives are evaluated on the vector $(t, u_0(t), \dot{u}_0(t))$. Since $\frac{\partial \widetilde{\phi}}{\partial \xi}$ is invertible, this can be rewritten as a linear ODE of the form

$$\dot{\mu}(t) = A(t)\mu(t) + B(t). \tag{9}$$

Classical ODE theory guarantees that (9) has unique solution for all time, given initial conditions (see [5, pp. 60,61]). Therefore, we may construct a concrete μ as the unique solution to

$$\begin{cases} \Theta(\ell_0 f|_{t,u_0(t),\dot{u}_0(t)} + \mu(t) \cdot \widetilde{\phi}(t, u_0(t), \dot{u}_0(t)), \\ \ell_0 \left.\frac{\partial f}{\partial \xi}\right|_{\text{end}} + \mu(t_{\text{end}})^\intercal \left.\frac{\partial \widetilde{\phi}}{\partial \xi}\right|_{\text{end}} + e = 0. \end{cases} \tag{10}$$

For this particular μ, (7) simplifies to

$$\int_{\text{start}}^{\text{end}} \mu_2 \cdot v \, \mathrm{d}t = 0 \text{ for all regular } v.$$

We may now apply the FLCV to get that the last k coordinates of μ are null, and so

$$\Theta(\ell_0 f + \mu_1 \cdot \phi) = 0$$

This is good, because this result does not depend on any arbitrary choices, such as the extension of ϕ to $\widetilde{\phi}$. Therefore, we have the first version of the Lagrange multiplier rule:

Theorem 5. *Let u_0 be a C^2 minimum of the optimization problem*

$$\min I(u) = \int f,$$

where f is C^2 and u varies over C^1 paths satisfying $u(t_{\text{start}}) = u_{\text{start}}$, $u(t_{\text{end}}) = u_{\text{end}}$, and the k differential equations given by

$$\phi(t, u(t), \dot{u}(t)) = 0,$$

where ϕ is a C^3 function with k coordinates. Suppose that these k equations are independent differential equations, in the sense that $\frac{\partial \phi}{\partial \xi}$ has rank k over the whole path. Then, there exists a constant ℓ_0 and k C^1 functions, say $\lambda_1, \ldots, \lambda_k$, such that the following differential equation is satisfied in $[t_{\text{start}}, t_{\text{end}}]$.

$$\ell_0 \frac{\partial f}{\partial u} + \lambda \cdot \frac{\partial \phi}{\partial u} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\ell_0 \frac{\partial f}{\partial \xi} + \lambda^\intercal \frac{\partial \phi}{\partial \xi} \right). \tag{11}$$

Furthermore, it is not the case that ℓ_0, λ_1, …, λ_k are all null.

Remark 9. The condition that not all be null is essential to stop the theorem from becoming a triviality: otherwise, the trivial solution would always be possible.

To show that not all are null, suppose that ℓ_0 is null. Then, by (10) λ satisfies

$$\lambda(t_{\text{end}})^{\intercal} \left.\frac{\partial \widetilde{\phi}}{\partial \xi}\right|_{\text{end}} + e = 0.$$

Since by hypothesis the row-matrix $E = \begin{bmatrix} \ell_0 & e \end{bmatrix}$ is non-null, and we are assuming that ℓ_0 is null, and so e cannot be, and therefore $\lambda(t_{\text{end}})$ cannot be as well. By continuity, we conclude that λ cannot be the null function.

Remark 10. As it ought to, the problem posed by (11) has as many equations as it has variables, if we consider the solution of the problem to be the process of finding the u_0. Indeed, we have $n + k$ functions to find:

$$u_1, \ldots, u_n, \lambda_1, \ldots, \lambda_k,$$

and $k+n$ equations: k are given by the condition $\phi(t, u, \dot{u}) = 0$, and n are given by (11). It is by solving this problem in $n + k$ variables and equations that one usually solves this kind of variational problems.

4.5 Normality and Abnormality

Consider the parameter ℓ_0 in the Lagrange multiplier equations, (**??**). There exist two possible cases:

- If $\ell_0 \neq 0$, the whole equation can be divided by ℓ_0, yielding solutions to the equation
$$\frac{\partial f}{\partial u} + \lambda \cdot \frac{\partial \phi}{\partial u} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial f}{\partial \xi} + \lambda \cdot \frac{\partial \phi}{\partial \xi}\right). \tag{12}$$
This is the so-called *normal case*;

- If $\ell_0 = 0$, we obtain a *nontrivial* solution to the equation
$$\lambda \cdot \frac{\partial \phi}{\partial u} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\lambda \cdot \frac{\partial \phi}{\partial \xi}\right). \tag{13}$$
This is the so-called *abnormal* case.

The solutions to the Lagrange multiplier equation can be divided into two types: normal solutions, wherein $\ell_0 \neq 0$, and abnormal solutions, where $\ell_0 = 0$.

An analogy can be made between the distinction normal/abnormal solutions and the distinction interior/border of a manifold. In some sense, as the name would indicate, abnormal solutions are 'undesired', or uncommon. In fact, note that the existence (or not) of abnormal solutions *does not depend on the functional being considered*, but only on the constraints placed on the paths! This is by analogy to optimization in calculus, where, besides checking critical points of

the function to be optimized, one must usually also check the border of the space for extrema. According to Brunt [6, p. 129], the usual approach to variational problems with nonholonomic constraints is to work under the assumption that the problem is normal, i.e. that there are no abnormal extrema, and only afterwards study the abnormal solutions and compare them to the normal critical points found.

Geometrically, abnormal solutions correspond to solutions that cannot be perturbed. For example, in a constant-speed nonholonomic problem, a straight line cannot be perturbed while preserving the boundary conditions, while a non-straight line can be slightly perturbed. This particular case can easily be verified:

Theorem 6. *Let $\phi(t, u, \xi) = \xi^2 - 1$. Then, u is an abnormal solution to the Euler-Lagrange equations if and only if it is a straight line.*

Proof. The equation to be solved is given by (13), which in this case reduces to

$$0 = \frac{\mathrm{d}(2\lambda \dot{u})}{\mathrm{d}t},$$

i.e. the vector $\lambda \dot{u}$ is constant. Furthermore, we are under the constraint $\|\dot{u}\| = 1$, so taking the norm of $\lambda \dot{u}$ we obtain that λ is constant, and by nontriviality of the solution must be a constant different from zero. Therefore, $\dot{u}$ must be constant, and so u is a (speed 1) straight line. ■

4.6 Example of application

Let us apply the method of Lagrange multipliers to Dido's problem, of finding a curve of fixed length $L > 2$, starting at $(-1, 0)$ and ending at $(1, 0)$, that encompasses the most area between it and the x axis. To do so, we encode the problem as a variational problem with nonholonomic constraints.

In order to consider curves of fixed length L, instead of considering an integral constraint on the paths, we will consider paths of unit speed defined on $[0, L]$. That is, if we let $u = (x, y)$, our restrictions are given by

$$x(0) = -1,\ x(L) = 1, \quad y(0) = y(L) = 0, \quad \dot{x}^2 + \dot{y}^2 - 1 = 0. \tag{14}$$

To measure the area encompassed by the curve[4] $u = (x, y)$, the usual trick is to apply Green's formula on the integral $\int_\Omega 1 \mathrm{d}x\mathrm{d}y$ to get a formula for the area such as

$$I(u) = -\int_0^L x\dot{y}\,\mathrm{d}t.$$

Let us now find the Euler-Lagrange equations. By theorem 6, the abnormal solutions to the Euler-Lagrange equations are straight lines, and straight lines cannot satisfy the boundary conditions as the start and end points are too

[4]The usual remark about the equivalence of maximization and minimization problems applies.

close together, at a distance of $2 < L$. Therefore, we need only consider the Euler-Lagrange equations for the normal case, which are

$$\begin{cases} \dfrac{\partial f}{\partial x} + \lambda \dfrac{\partial \phi}{\partial x} = \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\partial f}{\partial \dot{x}} + \lambda \dfrac{\partial \phi}{\partial \dot{x}}\right), \\ \dfrac{\partial f}{\partial y} + \lambda \dfrac{\partial \phi}{\partial y} = \dfrac{\mathrm{d}}{\mathrm{d}t}\left(\dfrac{\partial f}{\partial \dot{y}} + \lambda \dfrac{\partial \phi}{\partial \dot{y}}\right), \\ \dot{x}^2 + \dot{y}^2 = 1, \end{cases}$$

where λ is a C^1 real function, and u satisfies (14). Expanding the partial derivatives,

$$\begin{cases} -\dot{y} = 2\dot{\lambda}\dot{x} + 2\lambda\ddot{x}, \\ 0 = -\dot{x} + 2\dot{\lambda}\dot{y} + 2\lambda\ddot{y}, \\ \dot{x}^2 + \dot{y}^2 = 1. \end{cases} \tag{15}$$

To solve these equations, note that from the restriction $\dot{x}^2 + \dot{y}^2 = 1$ we conclude

$$2\lambda\ddot{x}\dot{x} + 2\lambda\ddot{y}\dot{y} = 0.$$

Therefore, adding appropriate multiples of the other two equations, we obtain

$$\cancel{-\dot{y}\dot{x}} = 2\dot{\lambda}\dot{x}^2 + \cancel{-\dot{x}\dot{y}} + 2\dot{\lambda}\dot{y}^2,$$

and using $\dot{x}^2 + \dot{y}^2 = 1$ we get

$$\dot{\lambda} = 0.$$

Consequently, λ is a constant, and (15) reduces to

$$\begin{cases} \dot{y} = -2\lambda\ddot{x}, \\ \dot{x} = 2\lambda\ddot{y}, \\ \dot{x}^2 + \dot{y}^2 = 1. \end{cases} \tag{16}$$

Consequently, $\dot{x} = 2\lambda\ddot{y} = -4\lambda^2\ddot{\dot{x}}$, and so $\dot{x}$ is of the form

$$\dot{x} = A\sin(2\lambda t) + B\cos(2\lambda t).$$

If $\lambda = 0$, x would be a constant and so $\ddot{x} = 0$, implying $\dot{y} = 0$ and so (x, y) is constant, contradicting the hypothesis of unit speed. Therefore, $\lambda \neq 0$ and so we can integrate $\dot{x}$ to get

$$x = A'\sin(2\lambda t) + B'\cos(2\lambda t) + C,$$

for $A' = \frac{1}{2\lambda}B$ and $B' = -\frac{1}{2\lambda}A$. Integrating $\dot{y} = 2\lambda\ddot{x}$, we get

$$y = A'\cos(2\lambda t) - B'\sin(2\lambda t) + C'.$$

We conclude that the curve (x, y) describes a circle around the point (C, C'), and so the (multiple) solutions to the Euler-Lagrange equations are unit speed circle sections starting at $(-1, 0)$ and ending at $(1, 0)$. Further investigation would be necessary to find expressions for these circles and to discover which of them is the best candidate for a maximum.

4.7 Further reading

As mentioned in the introduction, this proof is, for the most part, a particularization of the proof found in Bliss [1, pp. 187-202]. We direct the interested reader there, as the theorem is stated and proved in much, much greater generality: the paths considered are sectionally C^1, the endpoints of the curves are not fixed but rather obey some set of equations (imagine a chain with endpoints on a rail), t_{start} and t_{end} are likewise not fixed, and the functional to be optimized might have a term depending on the endpoints. It is reasonably well-written, though in a terse style which, when combined with unusual notational conventions (e.g. the Euler-Lagrange equations are written in integral form) might leave a first-time reader perplexed. I certainly was.

The section on normality and abnormality was loosely based on Brunt [6, pp. 128,129], which also has a discussion of the (non)applicability of the Lagrange multiplier method to nonholonomic problems in mechanics, as well as further bibliographic references.

Appendix A An implicit function-like theorem

In this appendix, we prove the following auxiliary result, which was used in section 4.2

Theorem 3. *Let $u\colon [t_{\text{start}}, t_{\text{end}}] \to \mathbb{R}^n$ be a C^p curve, for $p \in \mathbb{N}_0$, Ω an open subset of $X = \mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n$ such that $(t, u(t), \dot{u}(t)) \in \Omega$ for all t, and $\phi \equiv \phi(t, u, \xi)$ a C^{p+1} function $\phi\colon \Omega \to \mathbb{R}^k$.*

Suppose that for all t the matrix $\frac{\partial\phi}{\partial\xi}$ has rank k. Then there exist open subsets of X, say U and V, and a function $\widetilde{\phi}\colon U \to \mathbb{R}^n$, such that

- *U contains all points of the form $(t, u(t), \dot{u}(t))$,*
- *U is contained in Ω,*
- *The first k coordinates of $\widetilde{\phi}$ agree with ϕ*
- *The function given by*

$$\begin{aligned} \Phi\colon U &\to V \\ (t, u, \xi) &\mapsto (t, u, \widetilde{\phi}(t, u, \xi)) \end{aligned}$$

  is a C^p diffeomorphism.

The kernel of the proof is the following lemma:

Lemma 1. *Let $A(t)$ be a variable $k \times n$ matrix, for $t \in [t_{\text{start}}, t_{\text{end}}]$. Suppose that A is a C^p function for some $p \in \mathbb{N}_0$. Then, it is possible to add $n - k$ (variable) rows to A, yielding a function $\tilde{A}(t)$, such that $\tilde{A}$ is C^p and invertible for all t.*

Proof. To prove this lemma, we will actually prove something stronger yet. We will show that if $k < n$ then there exists $v(t)$, different from zero, defined on $[t_{\text{start}}, t_{\text{end}}]$ such that v is orthogonal to all rows of A. Such a v can clearly be appended as a row to A, creating a $(k+1) \times n$ matrix of rank $k+1$, and repeating the process $n-k$ times we obtain the desired $\tilde{A}$.

The main trick in this proof is that, locally, such a v is determined uniquely by $n-k$ well-chosen coordinates. Indeed, we wish to solve the equation $Av = 0$. We know from linear algebra that there exists a $k \times k$ invertible minor of A (which corresponds to the pivots in gaussian elimination). Without loss of generality we can suppose that, for some fixed t_1, this minor is the leftmost, i.e. given t_1 we can by reordering of coordinates assume that $A(t) = \begin{bmatrix} A_1(t) & A_2(t) \end{bmatrix}$, with $A_1(t_1)$ invertible. Of course, by continuity of the determinant, A_1 is actually invertible for t in a neighborhood of t_1.

Consequently, the problem can be solved easily in a neighborhood of t_{start}. Indeed, choose $v_{k+1}(t), \ldots, v_n(t)$ arbitrarily, as long as they are not all zero, and $v_1(t), \ldots, v_k(t)$ is uniquely determined as

$$(v_1, \ldots, v_k) = -A_1(t)^{-1} A_2(t)(v_{k+1}, \ldots, v_n).$$

Note that if the $v_{k+1}, \ldots, v_n$ are chosen to be C^p functions, then v is also C^p. This uses the well-known fact that if an invertible matrix is C^p then so is its inverse, by the cofactor formula.

To solve the problem globally, let us consider the following number:

$$t_0 = \sup\{\, t \in [t_{\text{start}}, t_{\text{end}}] \mid \begin{array}{c} \text{There exists } v \text{ of class } C^p \text{ defined on } [t_{\text{start}}, t] \\ \text{satisfying the hypotheses of the theorem} \end{array} \}.$$

We have just shown that $t_0 > t_{\text{start}}$. Now, we will show that $t_0 = t_{\text{end}}$.

Suppose that $t_0 < t_{\text{end}}$. Then, consider an invertible $k \times k$ minor of $A(t_0)$, which we will again suppose without loss of generality is A_1 above. Suppose A_1 is invertible for t in a neighborhood of t_0, say $V =]t_0 - \delta, t_0 + \delta[$. Pick some number τ strictly between t_0 and $t_0 - \delta$. Then, by definition of supremum, we know that there exists a v satisfying the hypotheses of the theorem defined up until τ. However, by the paragraph at the start of the proof, such a v is uniquely determined on V by its last $n-k$ coordinates. Therefore, all one has to do is to extend $v_{k+1}, \ldots, v_n$ from $[t_{\text{start}}, \tau]$ to $[t_{\text{start}}, t + \delta/2]$, say, in a C^p way, taking care to avoid zero. The C^p extension can be done by considering the Taylor expansion of each v_i at τ and extending it with that formula. To avoid zero, pick a coordinate v_i such that $v_i(\tau) \neq 0$, and add a large enough multiple of $\pm(x-\tau)^{p+1}$. Then, we have just defined a nonzero $v(t)$ such that $Av = 0$ on the interval $[t_{\text{start}}, t_0 + \delta/2]$, contradicting the hypothesis that t_0 was the supremum of that set.

This concludes the proof that $t_0 = t_{\text{end}}$. Clearly this proof can be easily adjusted to show that there exists $v(t)$ defined over $[t_{\text{start}}, t_{\text{end}}]$, C^p and never null, such that $Av = 0$. Therefore, the proof is complete. ■

Proof of theorem 3. Define $A(t) = \left.\frac{\partial \phi}{\partial \xi}\right|_{t,u(t)}$. By hypothesis, $A(t)$ is a $k \times n$ matrix of rank k, and its coordinates are C^p in time. Therefore, we are in the conditions of lemma 1, and can find a $(n-k) \times n$ matrix, let's say $B(t)$, such that $\tilde{A} = \begin{bmatrix} A \\ B \end{bmatrix}$ is invertible for all t. Furthermore, we may suppose that B is C^p. Then, we define

$$\widetilde{\phi}(t, u, \xi) = (\phi(t, u, \xi), B(t)\xi).$$

This function is clearly C^p in all its variables, and well-defined whenever ϕ is. Now, put

$$\Phi(t, u, \xi) = (t, u, \widetilde{\phi}(t, u, \xi).$$

This is a function from $\Omega \subseteq \mathbb{R}^{2n+1}$ to $\mathbb{R}^{2n+1}$, and computing the derivative one easily finds that $\Phi'(t, u(t), \dot{u}(t))$ is always invertible. Therefore, by the inverse function theorem, for all t there exists a neighborhood $U(t)$ of $(t, u(t), \dot{u}(t))$ where Φ is a diffeomorphism. By compactness, a finite number of these is needed to cover all points of the form $(t, u(t), \dot{u}(t))$. If we suppose, without loss of generality, that these neighborhoods are tubular in ξ in $\mathbb{R}^{2n+1}$, then we can ensure that Φ is a bijection (and thus a C^p diffeomorphism by the IFT) from the union of these balls (let's call it $U = U(t_1) \cup \cdots \cup U(t_N)$) to $\Phi(U)$ (which we call V).

To be more specific, suppose that $U(t)$ is of the following form: there exists an interval $]a, b[$ and two radii r_1, r_2, such that

$$(t, u, \xi) \in U(t) \text{ iff } t \in \,]a, b[\, , \|u - u(t)\| < r_1 \text{ and } \|\xi - \dot{u}(t)\| < r_2,$$

where we have extended u in a C^p to an open interval containing $[t_{\text{start}}, t_{\text{end}}]$. This can be done by using the Taylor expansion of u at the ends of the interval.

The tubularity assumption on the neighborhoods can be made without loss of generality because given an open containing $(t_i, u(t_i), \dot{u}(t_i))$, one may assume that it is of the form $]a, b[\times W$ for some open W, decrease the interval to ensure that $(u(t), \dot{u}(t))$ is always contained in W, and then set $r_1 = r_2$ as the minimum distance from $(u(t), \dot{u}(t))$ to the outside of W as t ranges over $[a, b]$.

Now, under the assumption that the $U(t_i)$ are tubular, suppose that (t, u, ξ) and (t', u', ξ') are two points of U such that $\Phi(t, u, \xi) = \Phi(t', u', \xi')$. If we show that there exists some $U(t_i)$ which contains (t, u, ξ) and (t', u', ξ') then the result will be proven, since Φ is injective when restricted to $U(t_i)$.

Since $\Phi(t, u, \xi) = \Phi(t', u', \xi')$ we have $t = t'$ and $u = u'$. Now, consider the distance from ξ to $\dot{u}(t)$ and ξ' to $\dot{u}(t)$. If, say, ξ is closer to $\dot{u}(t)$ than ξ' is, then the tubular neighborhood that contains (t, u, ξ') also contains (t, u, ξ). If ξ is not closer than ξ', the same argument can be made with the neighborhoods swapped. Therefore, these two points are in a neighborhood in which Φ is known to be injective, and the proof is complete. ■

References

[1] Gilbert A. Bliss. *Lectures on the Calculus of Variations.* University of Chicago Press, Chicago, Ill., 1946.

[2] Gilbert Ames Bliss. The problem of Mayer with variable end points. *Trans. Amer. Math. Soc.*, 19(3):305–314, 1918.

[3] I. M. Gelfand and S. V. Fomin. *Calculus of variations*. Prentice-Hall, Inc., Englewood Cliffs, N.J., 1963. Revised English edition translated and edited by Richard A. Silverman.

[4] Herbert Goldstein. *Classical Mechanics*. Addison-Wesley, 1980.

[5] Lawrence Perko. *Differential equations and dynamical systems*, volume 7 of *Texts in Applied Mathematics*. Springer-Verlag, New York, third edition, 2001.

[6] Bruce van Brunt. *The calculus of variations*. Universitext. Springer-Verlag, New York, 2004.