



Bootstrapping globally optimal variational calculus solutions

Gregory S. Chirikjian¹

Received: 1 January 2022 / Accepted: 1 October 2022 / Published online: 24 December 2022
© The Author(s) 2022

Abstract

Whereas in a coordinate-dependent setting the Euler–Lagrange equations establish necessary conditions for solving variational problems in which both the integrands of functionals and the resulting paths are assumed to be sufficiently smooth, uniqueness and global optimality are generally hard to prove in the absence of convexity conditions, and often times they may not even exist. This is also true for variational problems on Lie groups, with the Euler–Poincaré equation establishing necessary conditions. The difficulties compound when integrands and/or the optimal paths are not sufficiently regular, since in this case the classical necessary conditions no longer apply. This article therefore reviews several nonstandard cases where unique globally optimal solutions can be guaranteed, and establishes a “bootstrapping” process to build new globally optimal variational solutions on larger spaces from existing ones on smaller spaces. Surprisingly, it is possible to prove global optimality in some nonconvex cases where even the regularity conditions required for classical necessary conditions do not hold. This general theory is then applied to several topics such as optimal framing of curves in three-dimensional Euclidean space, optimal motion interpolation, and optimal reparametrization of video sequences to compare salient actions.

Mathematics Subject Classification 49-XX · 22E70 · 53C22

1 Introduction

In classical variational calculus, the Euler–Lagrange (E–L) Equations provide necessary conditions for local optimality for functionals that have integrands that are sufficiently regular, under the assumption that the extremal solutions are also sufficiently regular. Stronger necessary conditions for local optimality such as those due to Jacobi and Legendre are also well known [1, 2]. But in many cases testable sufficient conditions for global optimality cannot be established unless the integrand is convex. Otherwise, global optimality may not exist, may not be unique, or may not be provable. A simple example of this is geodesics on the sphere. The Euler–Lagrange equations yield great arcs as the solution, but for most pairs

Communicated by A. Mondino.

✉ Gregory S. Chirikjian
mpegre@nus.edu.sg

¹ National University of Singapore, Singapore, Singapore

of points on the sphere there are two great arcs connecting them, one of which is globally minimal, and one of which takes the long way around. When the points are antipodal, all great arcs connecting them are globally optimal, hence there is a severe lack of uniqueness. In negatively curved (hyperbolic) spaces it is well known that geodesics generated by solving the Euler–Lagrange equations are globally optimal and unique [3, 4]. But this is a relatively rare situation, and uses specific geometric arguments. In contrast, the formulation presented here will rely minimally on geometry, yet will have implications for geometric problems.

A broad class of variational problems for which minimality of the solution can be guaranteed is when the functional is convex in both the generalized coordinates and their rates [5, 6]. But this is a tall order, since even in mechanics where the Lagrangian is the integrand of the action functional, joint convexity of the Lagrangian in coordinates and their rates is rare. Hence, it is well known that the “Principle of Least Action” is a misnomer. It is really the “Principal of Extremal Action” since the Euler–Lagrange equations are generated and solved but little attention is given as to whether the resulting solution is globally minimal.

1.1 The Euler–Lagrange equations

Perhaps the most trivial variational integral for which the E–L equations provide unique globally optimal solutions is that the arclength functional for curves in the Euclidean plane. The general single-degree-of-freedom variational problem is that of finding a function $y = y(x)$ such that the functional

$$J[y] = \int_{x_1}^{x_2} f(y(x), y'(x), x) dx \tag{1}$$

is extremized (where $y' = dy/dx$) for a given integrand $f(\cdot)$ subject to boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$. For example, when

$$f(y(x), y'(x), x) = \sqrt{1 + [y'(x)]^2}$$

the problem becomes that of finding the curve $y = y(x)$ that extremizes arclength.

The E–L equation corresponding to (1) is

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial y'} \right) = 0. \tag{2}$$

In the case of the arclength functional, this becomes $y'' = 0$, which when integrated twice gives $y(x) = ax + b$, the equation of a line. Even without invoking Euclid’s elements, within the class of problems with solutions of the form $y = y(x)$, the line is optimal because any perturbation added to it that observes the boundary conditions can only increase J . The details as to why this is so will be addressed shortly, and have far reaching implications for proving global optimality in other less trivial problems.

One might argue that the above is limiting because it does not allow for the possibility of curves in the plane connecting the points (x_1, y_1) and (x_2, y_2) that are not functions of the form $y = y(x)$. To consider the more general case, we would need to seek parameterized curves of the form $x = y_1(t)$ and $y = y_2(t)$, and solve a two-degree-of-freedom variational problem. The theory for such problems is also classical. Given

$$J[\mathbf{y}] = \int_{t_1}^{t_2} f(y_1, \dots, y_N, \dot{y}_1, \dots, \dot{y}_N, t) dt$$

where $\dot{y}_i = dy_i/dt$, the set of E–L equations for $i = 1, \dots, N$ are

$$\frac{\partial f}{\partial y_i} - \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{y}_i} \right) = 0. \tag{3}$$

This assumes that $f : \mathbb{R}^N \times \mathbb{R}^N \times [t_1, t_2] \rightarrow \mathbb{R}$ is sufficiently smooth for all derivatives to exist and the existence of extremal trajectories that are smooth enough for their evaluation in the E–L equations to make sense. Two special classes of problems will be considered later in which: (1) f is not even differentiable, and hence the E–L equations do not apply, and (2) the Lavrentiev phenomenon occurs in which f is regular but the resulting global minimizer is not, and therefore does not satisfy the E–L equations. Yet global optimality can be guaranteed in both cases.

Returning to the curve in the Euclidean plane connecting two points, the arclength functional can be written as

$$f(y_1, y_2, \dot{y}_1, \dot{y}_2, t) = \sqrt{[\dot{y}_1]^2 + [\dot{y}_2]^2},$$

and the resulting E–L equations are

$$\ddot{y}_i = 0 \text{ for } i = 1, 2,$$

thus resulting in the parameterized line $y_i^*(t) = a_i t + b_i$, or $\mathbf{y}^*(t) = \mathbf{a}t + \mathbf{b}$ where \mathbf{a} and \mathbf{b} are constant vectors that provide the freedom to match the boundary conditions. The strength of this argument relative to the previous one, again without invoking Euclid, is that it considers the possibility of curves in the plane that connect two points and are not necessarily plotted as a function $y = y(x)$.

More generally, given two points in a region of a manifold parameterized by coordinates $\mathbf{q} = [q_1, \dots, q_N]^T$, and metric tensor $M(\mathbf{q})$, the E–L equations (with q_i taking the place of y_i) and the functional

$$f(\mathbf{q}, \dot{\mathbf{q}}, t) = \sqrt{\dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}} \tag{4}$$

give necessary conditions for a geodesic. The square root is somewhat inconvenient when performing computations, and justification for eliminating it will be given shortly.

A metric tensor can arise in purely geometric contexts, e.g., for surfaces such as Minkowski sums of ellipsoids [7], or it can arise from the kinetic energy of a system of P particles with positions $\mathbf{x}_i \in \mathbb{R}^3$, masses m_i , and generalized coordinates $\mathbf{q} \in \mathbb{R}^N$ (with $N \leq 3P$) in which case it is well known from classical mechanics that the entries of the mass metric tensor become

$$m_{ij}(\mathbf{q}) = \sum_{k=1}^P m_k \frac{\partial \mathbf{x}_k}{\partial q_i} \cdot \frac{\partial \mathbf{x}_k}{\partial q_j}.$$

Hence the following well-known results from mechanics and a simple theorem will be helpful.

In Mechanics, the Lagrangian is defined to be the difference of kinetic and potential energy¹

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q})$$

¹ Explicit dependence of the functional on t is relatively rare in the mechanics of particles and rigid bodies and in geometry, and henceforth is dropped from the discussion at this point. It will be reintroduced in the discussion of the Lavrentiev phenomenon that follows later.

and the total energy is

$$E(\mathbf{q}, \dot{\mathbf{q}}) = T(\mathbf{q}, \dot{\mathbf{q}}) + V(\mathbf{q}).$$

The kinetic energy for a mechanical system always has the form

$$T(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}. \tag{5}$$

The mass matrix of a mechanical system, $M(\mathbf{q})$, satisfies the required properties of a metric tensor.

Lagrange’s equations of motion are the Euler–Lagrange equations (3) with L in place of f , and q_i in place of y_i , and they describe the motion of conservative systems. That is, if $\mathbf{q}^*(t)$ denotes the solution to Lagrange’s equations, it can be shown that energy is conserved:

$$E(\mathbf{q}^*, \dot{\mathbf{q}}^*) = \text{const}. \tag{6}$$

This classical fact will be used in the theorem presented below.

Lemma 1.1 *Suppose T is known a priori to be limited to the range $[x_1, x_2]$ where $0 < x_1 < x_2 < \infty$ and let $\Phi : [x_1, x_2] \rightarrow \mathbb{R}_{>0}$ be a twice differentiable monotonically increasing function. Then the solution to the variational problems with a functional of the form $f(\mathbf{q}, \dot{\mathbf{q}}) = \Phi(T(\mathbf{q}, \dot{\mathbf{q}}))$ will be the same as the solution to Lagrange’s equations in the absence of a potential energy term.²*

Proof From the chain rule

$$\frac{\partial f}{\partial q_i} = \Phi'(T) \frac{\partial T}{\partial q_i} \quad \text{and} \quad \frac{\partial f}{\partial \dot{q}_i} = \Phi'(T) \frac{\partial T}{\partial \dot{q}_i},$$

and

$$\frac{d}{dt} \left(\frac{\partial f}{\partial \dot{q}_i} \right) = \Phi''(T) \dot{T} \frac{\partial T}{\partial \dot{q}_i} + \Phi'(T) \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_i} \right).$$

But for a mechanical system with $V = 0$, $T = \text{const}$ from (6), and so $\dot{T} = 0$. Therefore

$$\Phi'(T) \left[\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_i} \right) - \frac{\partial T}{\partial q_i} \right] = \frac{d}{dt} \left(\frac{\partial f}{\partial \dot{q}_i} \right) - \frac{\partial f}{\partial q_i} = 0.$$

Since Φ is monotonically increasing on the interval over which it is defined and continuous (since it is twice differentiable), it is also invertible. Moreover, $\Phi'(T) > 0$. Therefore the variational problems for f and T are interchangeable.

In particular, restricting to when $T_0 > 0$ and letting $\Phi(x) = \sqrt{2x}$, we can solve the geodesic problem without the pesky square root, as if it were a mechanics problem.

The usefulness of this result is now illustrated with the line in the plane where it is easier to prove the global optimality for the integrand T than the integrand $f = \Phi(T)$ in the original arclength functional. Starting with the variational solution $\mathbf{y}^*(t) = \mathbf{a}t + \mathbf{b}$ and letting $\mathbf{y}(t) = \mathbf{y}^*(t) + \boldsymbol{\epsilon}(t)$, we see that the cost becomes

$$\int_{t_1}^{t_2} \|\dot{\mathbf{y}}\|^2 dt = \int_{t_1}^{t_2} \{ \|\mathbf{a}\|^2 + 2\mathbf{a} \cdot \dot{\boldsymbol{\epsilon}} + \|\dot{\boldsymbol{\epsilon}}\|^2 \} dt = (t_2 - t_1) \|\mathbf{a}\|^2 + 2\mathbf{a} \cdot \int_{t_1}^{t_2} \dot{\boldsymbol{\epsilon}} dt + \int_{t_1}^{t_2} \|\dot{\boldsymbol{\epsilon}}\|^2 dt. \tag{7}$$

² This result might be known, but the author could not find a clear statement of it other than in the case when $\Phi(x) = c\sqrt{x}$ which is mentioned in several works including [8–10]

But since the perturbed solution must satisfy the same boundary conditions,

$$\int_{t_1}^{t_2} \dot{\epsilon} dt = \epsilon(t_2) - \epsilon(t_1) = \mathbf{0} - \mathbf{0} = \mathbf{0}.$$

From this it is clear that adding ϵ to \mathbf{y}^* can only increase cost in (7). The same argument does not apply to the cost $\int_{t_1}^{t_2} \|\dot{\mathbf{y}}\| dt$ because the square root prevents the integral from acting directly on $\dot{\epsilon}$. That said, the Cauchy–Schwarz inequality provides the bound

$$\int_{t_1}^{t_2} \|\dot{\mathbf{y}}\| dt \leq \sqrt{t_2 - t_1} \left(\int_{t_1}^{t_2} \|\dot{\mathbf{y}}\|^2 dt \right)^{\frac{1}{2}}.$$

Substituting the minimal solution for the problem with integrand $\|\dot{\mathbf{y}}\|^2$ into the right hand side then provides an upper bound for the minimal solution to the variational problem with integrand $\|\dot{\mathbf{y}}\|$. And substituting the minimal solution for the problem with integrand $\|\dot{\mathbf{y}}\|^2$ into the cost with integrand $\|\dot{\mathbf{y}}\|$ provides an upper bound on the cost of the global minimizer for that problem. In Sect. 2, it will be shown that this second upper bound is in fact the global minimum.

In other words, changing the original problem by an appropriate choice of Φ can preserve the E–L necessary conditions while making it easy to guarantee global optimality of the solution to the new problem. Though it is not obvious using elementary arguments that the global optimality of the new problem guarantees it for the original problem, such statements are made in differential geometric settings. This issue will be revisited without geometry in Sect. 2. But first a small detour will be taken to discuss variational problems for which the resulting optimal trajectories lack regularity (i.e., they are non-Lipschitz), and for which the E–L equations are not necessary conditions.

1.2 The Lavrentiev phenomenon

Whereas the classical variational calculus of Euler, Lagrange, Jacobi, etc. assumes functionals with sufficiently smooth integrand and well-behaved resulting extremal trajectories, over the past 100 years variational problems with globally minimal solution have been investigated in the absence of classical regularity conditions on the integrand and/or the globally minimal trajectories that result.

The *Lavrentiev phenomenon* has been of particular interest [11–13]. As an example, consider trajectories $x(t)$ that minimize the functional

$$J[x] = \int_0^1 (x^3 - t)^2 \dot{x}^6 dt \tag{8}$$

subject to boundary conditions $x(0) = 0$ and $x(1) = 1$. This is the Maniá example [12]. Though the integrand is smooth, it can be shown that the globally minimal solution is $x^*(t) = t^{1/3}$. As this is not Lipschitzian, but is absolutely continuous, this is an example where the infimum of the functional taken over the set of absolutely continuous trajectories is lower than that taken over Lipschitz trajectories. This is the essence of the Lavrentiev phenomenon.

Functionals such as (8) and others that demonstrate the Lavrentiev phenomenon, such as those in [14] are remarkable in that their integrand is regular, and yet the resulting trajectories are not [15]. Consequently, the E–L equations need not be necessary conditions [16]. This remains an active topic of research with more such examples continuing to emerge [17–19].

The following section examines an altogether different kind of variational problem with globally minimal solution. Then in Sect. 3 all of these pieces from previous sections are

assembled using a “bootstrapping” procedure in which globally minimal solutions on higher-dimensional spaces are constructed.

2 Nonstandard case studies in global optimality

Though the Euler–Lagrange equations provide necessary conditions for local optimality under the conditions that: (1) f is sufficiently smooth; and (2) a sufficiently smooth solution exists, they provide no guarantee that such a solution will be globally optimal. So-called second-order necessary conditions provide more confidence that the solution to the E–L equations might be minimal, but also provide no guarantee. In the absence of specific a priori knowledge of the properties of the functional, such as convexity, it is usually quite difficult to say anything other than that the solution generated by the E–L equations are extremals (rather than global minima or maxima), unless additional arguments can be employed as in the case of the line in the previous section.

That said, in certain special situations, the structure of the function $f(\cdot)$ will guarantee global optimality of the solution generated by the the Euler–Lagrange equations. Here special cases are presented in which global optimality of solutions of variational problems are guaranteed when satisfying the Euler–Lagrange necessary conditions. The emphasis here is on this sense of the word ‘global’ in contrast to global descriptions of geometry or dynamics which can be found elsewhere [3, 20].

2.1 Global optimality in 1D case

Consider the set of monotonically increasing bijective functions that map the closed interval $[0, 1]$ into itself and are twice differentiable on the open interval $(0, 1)$ and satisfy the boundary conditions $y(0) = 0$ and $y(1) = 1$. Such functions include but are not limited to $y(x) = x^p$ for any positive power $p \in \mathbb{R}_{>0}$. The set of all such functions are closed under function inversion and function composition $(y_1 \circ y_2)(x) = y_1(y_2(x))$, and have an identity element, $y_{id}(x) = x$. The result is an infinite dimensional group. In this section a variational problem is formulated that selects the one element of this group that globally minimizes a cost functional.

There exists a very general functional in the one-dimensional case for which the global minimality of the solution to the Euler–Lagrange equation is guaranteed. Specifically, let

$$f(y, y', x) = m(y)(y')^2$$

where $m : [0, 1] \rightarrow \mathbb{R}_{>0}$ is differentiable and let

$$J[y] = \int_0^1 m(y)(y')^2 dx. \tag{9}$$

Global minimization of this sort of functional has been examined informally in [21, 22]. A more rigorous treatment is given below.

2.1.1 Globally minimal reparametrization

Lemma 2.1 *Let m_1 and m_2 be real numbers with $0 < m_1 < m_2 < \infty$ and let $m : [0, 1] \rightarrow [m_1, m_2]$ be differentiable. Let $m^{\frac{1}{2}}(s)$ and $m^{-\frac{1}{2}}(s)$ respectively be shorthand for $\sqrt{m(s)}$ and $1/\sqrt{m(s)}$. The solution to the resulting Euler–Lagrange equations for (9) subject to the*

boundary conditions $y(0) = 0$ and $y(1) = 1$ satisfies the implicit equation

$$y^*(x) = \frac{\int_0^x m^{-\frac{1}{2}}(y^*(s))ds}{\int_0^1 m^{-\frac{1}{2}}(y^*(s))ds} \tag{10}$$

and $y^*(x)$ can be obtained explicitly by inverting the function $F : [0, 1] \rightarrow [0, \sqrt{m_2}]$ defined by

$$F(y) \doteq \int_0^y m^{\frac{1}{2}}(s)ds$$

in the expression

$$F(y^*(x)) = F(1)x. \tag{11}$$

Moreover

$$\int_0^1 m^{-\frac{1}{2}}(y^*(s)) ds = 1/F(1). \tag{12}$$

Proof Recognizing that $m' = (dm/dy)y'$, the E–L equation corresponding to (9) will be

$$my'' + \frac{1}{2}m'y' = 0.$$

Multiplying through by $m^{-\frac{1}{2}}(x) = 1/\sqrt{m(x)}$, the result then can be written as

$$\left(m^{\frac{1}{2}}y'\right)' = 0.$$

The solution is then of the form

$$m^{\frac{1}{2}}(y^*)\frac{dy^*}{dx} = c_1 \tag{13}$$

where c_1 is a constant of integration, and so

$$y^*(x) = c_2 + c_1 \int_0^x m^{-\frac{1}{2}}(y^*(s))ds.$$

Since $y(0) = 0$ and $y(1) = 1$ this becomes (10).

Alternatively,

$$m^{\frac{1}{2}}(y^*)dy^* = c_1dx$$

can be integrated to give

$$F(y^*) = c_1x.$$

The condition $F(y^*(1)) = c_1$ together with the boundary condition $y^*(1) = 1$ then gives (11). That is, $c_1 = F(1) \in [\sqrt{m_1}, \sqrt{m_2}]$. But since the same c_1 appears in the normalization of the implicit expression, (12) results.

The following theorem proves that the result of the above lemma is not only a solution to the E–L equations, but that it is globally optimal. Moreover, this global optimality generalizes to cases where the E–L equations no longer apply due to a lack of sufficient regularity, i.e., the case when the function m is not differentiable.

Theorem 2.2 *Let all of the conditions of Lemma 2.1 hold, except that the function $m : [0, 1] \rightarrow [m_1, m_2]$ need only be continuous rather than differentiable. Then the implicit solution (10) and explicit solution (11) globally minimize (9) among all differentiable functions subject to the boundary conditions $y(0) = 0$ and $y(1) = 1$.*

Proof (10) and (11) both satisfy (13) without invoking the E–L equations. Substituting $y^*(x)$ into (9) and observing (13) gives

$$J[y^*] = \int_0^1 \left(m^{\frac{1}{2}}(y^*) \frac{dy^*}{dx} \right) \left(m^{\frac{1}{2}}(y^*) \frac{dy^*}{dx} \right) dx = c_1^2$$

where $c_1 = F(1)$. The continuity and boundedness of m ensures that F is bounded, differentiable, and monotonically increasing. This together with the form of (10) or (11) ensures differentiability of y^* . Since $c_1^2 = \left(\int_0^1 c_1 dx \right)^2$,

$$J[y^*] = \left(\int_0^1 m^{\frac{1}{2}}(y^*(x)) \frac{dy^*}{dx} dx \right)^2 = \left(\int_0^1 m^{\frac{1}{2}}(y^*) dy^* \right)^2.$$

The above is simply a change of variables of integration and has nothing to do with the details of (10) other than the fact that $y^*(0) = 0$ and $y^*(1) = 1$ and $dy^*/dx > 0$, which is true even though m may not be monotonic, convex, or differentiable. Consequently, the value of the integral in the above expression for $J[y^*]$ is *independent* of the path $y^*(x)$. That is,

$$J[y^*] = \left(\int_0^1 m^{\frac{1}{2}}(y) dy \right)^2$$

can be computed without reference to $y^*(t)$ since the name of the variable of integration is irrelevant. Using the Cauchy–Schwarz inequality

$$\int_0^1 a(x)b(x)dx \leq \left(\int_0^1 |a(x)|^2 dx \right)^{\frac{1}{2}} \left(\int_0^1 |b(x)|^2 dx \right)^{\frac{1}{2}}$$

with $a(x) = m^{\frac{1}{2}}(y(x))dy/dx$ and $b(x) = 1$ then gives

$$J[y^*] \leq J[y] \tag{14}$$

for every possible differentiable $y(x)$ that satisfies the boundary conditions.

What is remarkable about this simple example is that none of the classical conditions stated in variational calculus regarding sufficient conditions for optimality apply. This is not a hyperbolic space, the functional is not convex, etc. And in addition to all of this, the regularity conditions on m needed for the E–L equations to hold can be relaxed. Moreover, as will be seen later, from this simple example it is possible to bootstrap up to cases in which $y(t)$ is replaced by a path defined in a multi-dimensional space and where global optimality is still guaranteed.

2.1.2 Reparameterizing an extremal solution only makes things worse

Given $T = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}$ and $V = 0$, with optimal solution $\mathbf{q}^*(t)$ that conserves kinetic energy $T(\mathbf{q}^*, \dot{\mathbf{q}}^*) = T_0$, it is clear that arclength along such a trajectory is

$$s(t) = \int_0^t \sqrt{\dot{\mathbf{q}}^*(t')^T M(\mathbf{q}^*(t')) \dot{\mathbf{q}}^*(t')} dt' = \int_0^t \sqrt{2T_0} dt' = \sqrt{2T_0} t.$$

If we were to attempt to reparameterize by substituting $\mathbf{q}^*(t) \rightarrow \mathbf{q}^*(\tau(t))$, with τ being the same sort of monotonically increasing function on $[0, 1]$ discussed earlier, the resulting cost

would be

$$\int_0^1 \sqrt{\dot{\mathbf{q}}^*(\tau(t))^T M(\mathbf{q}^*(\tau(t))) \dot{\mathbf{q}}^*(\tau(t))} \dot{\tau}^2(t) dt = \sqrt{2T_0} \int_0^1 \dot{\tau}^2(t) dt.$$

The optimal $\tau(t)$ in this context is $\tau^*(t) = t$ in which case $\int_0^1 (\dot{\tau}^*)^2(t) dt = 1$. For any other $\tau(t) = t + \epsilon(t)$, the result will be

$$\int_0^1 \dot{\tau}^2(t) dt = \int_0^1 (1 + \dot{\epsilon})^2 dt = 1 + \int_0^1 \dot{\epsilon}^2 dt,$$

which only can increase cost.

2.2 Implications for solving auxiliary variational problems

A result of combining Lemma 1.1 and Theorem 2.2 is

Corollary 2.3 *Let T be as in (5). Then the trajectory $\mathbf{q}^*(t)$ that minimizes³*

$$J_2[\mathbf{q}] \doteq \int_0^1 2T(\mathbf{q}, \dot{\mathbf{q}}) dt$$

globally subject to the end constraints $\mathbf{q}(0)$ and $\mathbf{q}(1)$ will also minimize

$$J_1[\mathbf{q}] \doteq \int_0^1 \sqrt{2T(\mathbf{q}, \dot{\mathbf{q}})} dt$$

globally subject to the same end constraints. That is, for any twice differentiable trajectory $\mathbf{q}(t)$ satisfying the boundary conditions

$$J_2[\mathbf{q}^*] \leq J_2[\mathbf{q}] \implies J_1[\mathbf{q}^*] \leq J_1[\mathbf{q}]. \tag{15}$$

Moreover

$$J_2[\mathbf{q}^*] = (J_1[\mathbf{q}^*])^2. \tag{16}$$

Proof If $\Phi(x) = \sqrt{x}$, then by Lemma 1.1 the E–L equations for both functionals are the same and since the boundary conditions are the same, the resulting solutions will both be $\mathbf{q}^*(t)$.

Let $\mathbf{q} = \mathbf{q}(t)$ be an arbitrary differentiable trajectory on $[0, 1]$ that satisfies the end constraints, and let $\tau(t)$ be a reparameterization of time. Then $(\mathbf{q} \circ \tau)(t) = \mathbf{q}(\tau(t))$ will also be a feasible trajectory. Suppressing the temporal dependence on τ , this will be referred to as $\mathbf{q}(\tau)$. Let

$$m(\tau) = \frac{d\mathbf{q}^T}{d\tau} M(\mathbf{q}) \frac{d\mathbf{q}}{d\tau}.$$

Then $2T(\mathbf{q}(\tau), \frac{d\mathbf{q}}{d\tau} \dot{\tau}) = m(\tau) \dot{\tau}^2$. If τ^* is the global minimizer of this one-dimensional functional, and if $\mathbf{q}^*(t)$ is the global minimizer of the functional J_2 , then

$$J_2[\mathbf{q}^*] \leq J_2[\mathbf{q}(\tau^*)] \leq J_2[\mathbf{q}]. \tag{17}$$

³ $J_1[\mathbf{q}]$ and $J_2[\mathbf{q}]$ are not functions of \mathbf{q} , but rather are functionals, the value of which is completely specified by the path $\mathbf{q} = \mathbf{q}(t)$.

Recall that the differentiable trajectory $\mathbf{q}(t)$ is arbitrary up to the same boundary conditions, and $\mathbf{q}(\tau^*(t))$ is the same path in configuration space parameterized differently by time from $\mathbf{q}(t) = \mathbf{q}(\tau(t) = t)$.

From Theorem 2.2, we know that $\mathbf{q}(\tau^*(t))$ will give

$$J_2[\mathbf{q}(\tau^*)] = (J_1[\mathbf{q}(\tau^*)])^2. \tag{18}$$

This is true for any differentiable $\mathbf{q}(\tau)$ satisfying the end constraints and evaluated at $\tau = \tau^*(t)$. The significance of (18) is that since both J_1 and J_2 are positive, then for any given $\mathbf{q}(\tau)$ they must both be minimized by the same $\mathbf{q}(\tau^*(t))$.

In the case when $\mathbf{q} = \mathbf{q}^*$ (the postulated global minimal solution for J_2), reparameterizing the temporal variable results in $\tau^*(t) = t$ since by definition there is no way to reparameterize to improve the global minimum. Therefore, $J_2[\mathbf{q}^*(\tau^*)] = J_2[\mathbf{q}^*]$.

On the other hand, the value of the integral

$$J_1[\mathbf{q}(\tau)] = \int_0^1 \sqrt{m(\tau)} \dot{\tau} dt = \int_0^1 \sqrt{m(\tau)} d\tau$$

is independent of the choice of the differentiable monotonically increasing bijective function $\tau : [0, 1] \rightarrow [0, 1]$ regardless of which path $\mathbf{q}(t)$ is taken. That is, for any valid $\tau(t)$,

$$J_1[\mathbf{q}(\tau)] = J_1[\mathbf{q}].$$

Therefore, $J_1[\mathbf{q}^*(\tau)] = J_1[\mathbf{q}^*]$ for every such τ , including τ^* and

$$J_1[\mathbf{q}^*] \leq J_1[\mathbf{q}(\tau^*)] = J_1[\mathbf{q}]. \tag{19}$$

Moreover, from (18), J_1 is globally minimized on the same global minimizer of J_2 , namely $\mathbf{q}^*(t)$ and the minimal values are related as (16).

2.3 The Poincaré half space model

As an example where the E–L equations provide a globally minimal solution, consider the Poincaré Half Space Model, which is a Riemannian space of constant negative curvature.

Given an $N \times N$ metric tensor $M(\mathbf{q}) = [m_{ij}(\mathbf{q})]$ and its inverse using the raised index notation $[M(\mathbf{q})]^{-1} = [m^{ij}(\mathbf{q})]$, the Christoffel symbols are computed as

$$\Gamma^i_{jk} \doteq \frac{1}{2} \sum_{l=1}^N m^{il} \left(\frac{\partial m_{lj}}{\partial q_k} + \frac{\partial m_{lk}}{\partial q_j} - \frac{\partial m_{jk}}{\partial q_l} \right), \tag{20}$$

and from these the Riemannian curvature tensor is computed as

$$R^i_{jkl} \doteq -\frac{\partial \Gamma^i_{jk}}{\partial q_l} + \frac{\partial \Gamma^i_{jl}}{\partial q_k} + \sum_{m=1}^N (-\Gamma^m_{jk} \Gamma^i_{ml} + \Gamma^m_{jl} \Gamma^i_{mk}). \tag{21}$$

The notation of raising and lowering indicies gives

$$R_{hijkl} = \sum_{i=1}^N m_{hi} R^i_{jkl}.$$

A space is said to have constant curvature if [23]

$$R_{hijkl} = K_0 (m_{hk} m_{jl} - m_{hl} m_{jk}). \tag{22}$$

Hyperbolic space defined by constant negative curvature at every point is an example where it is known that geodesics connecting any two points are unique and globally length minimizing. The Poincaré half space model of hyperbolic space is the open half space

$$\mathbb{H}^+ \doteq \{\mathbf{x} \in \mathbb{R}^N, \mathbf{x} \cdot \mathbf{e}_N > 0\}$$

endowed with the metric tensor

$$M(\mathbf{x}) = (\mathbf{x} \cdot \mathbf{e}_N)^{-2} \mathbb{I}.$$

For example, in the 2D case this consists of point $(x_1, x_2 > 0) \in \mathbb{R}^2$ with

$$\dot{\mathbf{x}}^T M(\mathbf{x}) \dot{\mathbf{x}} = \frac{\dot{x}_1^2 + \dot{x}_2^2}{x_2^2}.$$

In this case it can be shown that (22) holds with

$$K_0 = -1. \tag{23}$$

The family of geodesics in this case is parameterized by the group $PSL(2, \mathbb{R}) \doteq SL(2, \mathbb{R})/\{\mathbb{I}, -\mathbb{I}\}$ where

$$SL(2, \mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mid ad - bc = 1 \right\}.$$

Let $i = \sqrt{-1}$ and let a, b, c, d be real numbers satisfying $ad - bc = 1$. Define the curve in the complex plane

$$z^*(t) = \frac{b + iae^t}{d + ice^t} = x_1^*(t) + ix_2^*(t).$$

That is, $x_1^*(t) = Re(z^*(t))$ and $x_2^*(t) = Im(z^*(t))$. (Optimality properties of this curve will be demonstrated shortly, and $*$ denotes this, i.e., it is not Hermitian conjugation). Explicitly,

$$x_1^*(t) = (bd + ace^{2t})(d^2 + c^2e^{2t})^{-1} \text{ and } x_2^*(t) = e^t(d^2 + c^2e^{2t})^{-1}. \tag{24}$$

The signed curvature of these curves as viewed in the Euclidean plane, computed using the general formula

$$k(t) = \frac{\dot{x}_1\ddot{x}_2 - \dot{x}_2\ddot{x}_1}{(\dot{x}_1^2 + \dot{x}_2^2)^{3/2}} \tag{25}$$

is $k^*(t) = -2cd$, which is constant, indicating a clockwise bending circular arc. Since these are circles, they can be written as

$$(x_1^* - x_0)^2 + (x_2^*)^2 = r^2 \tag{26}$$

where $r = 1/|k^*| = 1/(2cd)$ and

$$x_0 = \frac{1}{2} \left(\frac{b}{d} + \frac{a}{c} \right).$$

It can be shown that on these circles

$$\frac{(\dot{x}_1^*)^2 + (\dot{x}_2^*)^2}{(x_2^*)^2} = 1, \tag{27}$$

and consequently t is arclength in the hyperbolic metric.

Moreover, the Euler–Lagrange equations for the functional

$$J = \int_{t_0}^{t_1} \frac{\dot{x}_1^2 + \dot{x}_2^2}{x_2^2} dt \tag{28}$$

can be written as

$$\frac{d}{dt} \left(\frac{\dot{x}_1}{x_2^2} \right) = 0 \tag{29}$$

$$\ddot{x}_2 = x_2^{-1} (\dot{x}_2^2 - \dot{x}_1^2). \tag{30}$$

$x_i^*(t)$ given above satisfy these equations. Consequently (24) is an extremal of (28). In particular, (29) gives

$$\frac{\dot{x}_1^*}{(x_2^*)^2} = 2cd.$$

From the fact that $\dot{x}_1^*/(x_2^*)^2$ is constant, it is clear that if $\epsilon_1(t)$ is a nonzero arbitrary differentiable function that vanishes at $t = t_0$ and $t = t_1$, it must be that

$$\int_{t_0}^{t_1} \frac{(\dot{x}_1^*)^2 + (\dot{x}_2^*)^2}{(x_2^*)^2} dt < \int_{t_0}^{t_1} \frac{(\dot{x}_1^* + \dot{\epsilon}_1)^2 + (\dot{x}_2^*)^2}{(x_2^*)^2} dt$$

based on the same argument as for the line. That is, when expanding the term in parenthesis, the resulting linear term in $\dot{\epsilon}$ integrates to zero.

Moreover, since these curves $\mathbf{x}^*(t)$ are already parameterized by arclength, replacing them with $\mathbf{y}(t) = \mathbf{x}^*(\tau(t))$ where $\tau(t) \neq t$ is monotonically increasing and $\tau(t_0) = t_0$ and $\tau(t_1) = t_1$ can only increase the value of J . Consequently,

$$\int_{t_0}^{t_1} \frac{(\dot{x}_1^*)^2 + (\dot{x}_2^*)^2}{(x_2^*)^2} dt < \int_{t_0}^{t_1} \frac{\dot{y}_1^2 + \dot{y}_2^2}{y_2^2} dt.$$

Writing $\mathbf{y}(t) = \mathbf{x}^*(t) + \mathbf{v}(t)$ (where there is one degree of freedom in $\mathbf{v}(t)$ following from the freedom of the scalar function $\tau(t)$) and combining the above two inequalities provides evidence for the following statement.

If $\epsilon_1(t)$ and $\epsilon_2(t)$ are two independent nonzero arbitrary differentiable functions that vanish at $t = t_0$ and $t = t_1$, then

$$\int_{t_0}^{t_1} \frac{(\dot{x}_1^*)^2 + (\dot{x}_2^*)^2}{(x_2^*)^2} dt < \int_{t_0}^{t_1} \frac{(\dot{x}_1^* + \dot{\epsilon}_1)^2 + (\dot{x}_2^* + \dot{\epsilon}_2)^2}{(x_2^* + \epsilon_2)^2} dt,$$

indicating that (24) are globally minimal geodesics.

2.4 Special higher dimensional vectorial cases

In the previous subsection, the Poincaré half-plane model was used. Another model of the hyperbolic plane is the Poincaré disk model in which

$$f(\mathbf{x}, \dot{\mathbf{x}}) = \dot{\mathbf{x}}^T M(\mathbf{x}) \dot{\mathbf{x}} \text{ where } M(\mathbf{x}) = \frac{4}{1 - \|\mathbf{x}\|^2} \mathbb{I}.$$

Extensions of the Poincaré half-plane and disk models to higher dimensions exist. For example, the above expression for $M(\mathbf{x})$ holds in n dimensions. The n -dimensional half-space

model has $M(\mathbf{x}) = x_n^{-2}\mathbb{I}$. From geometric considerations it is known that geodesics in these spaces are global minimizers of length.

All of the above are of the form

$$f(\mathbf{x}, \dot{\mathbf{x}}) = m \left(\sum_{i=1}^N a_i x_i^2 \right) \sum_{j=1}^N b_j \dot{x}_j^2$$

where $b_j = 1$ and $a_i \in \{0, 1\}$. That is, the mass matrix consists of a scalar mass function, $m(\cdot)$, which depends on coordinates in a particular way, which multiplies the identity matrix. Therefore, functions of the form above are a reasonable candidate for exploring wider conditions on the parameters $\{a_i\}$ and $\{b_j\}$ such that global minimality of the variational problem is guaranteed.

Another very special vectorial case is when there are coordinates $\{q_i'\}$ such that

$$f(\mathbf{q}', \dot{\mathbf{q}}', t) = \sum_{i=1}^N \lambda_i(q_i') (\dot{q}_i')^2$$

where each λ_i is a continuous function on the interval $[0, 1]$ that takes finite positive values. Then global optimization in each coordinate can be performed independently as per Theorem 2.2. Likewise, if a coordinate transformation can reduce a coupled vectorial problem to this form, then the global optimum can be found in each new coordinate, and the result can be transformed back. Such cases are not the norm. That is, if the matrix

$$M(\mathbf{q}) = S^T \Lambda(S\mathbf{q}) S$$

for some invertible constant matrix S , then the variational problem for $f(\mathbf{q}, \dot{\mathbf{q}}, t) \doteq \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}$ will have a globally minimal solution if it is defined on the region $S \cdot [0, 1]^N$ because then in the coordinates $\mathbf{q}' = S\mathbf{q}$, everything will reduce to $f(\mathbf{q}', \dot{\mathbf{q}}', t)$.

2.5 A suboptimal ansatz

For more complicated variational problems, employing numerical shooting methods to meet boundary conditions can be computationally costly. Moreover, since globally optimal solutions might not even exist, it can sometimes be useful to obtain a fast suboptimal solution to a variational problem rather than an exact extremal.

Motivated by the splitting in Theorem 2.2, we can define a trajectory $\mathbf{q}^\circ(t)$ such that for symmetric positive definite $M(\mathbf{q}) \in \mathbb{R}^{n \times n}$ (as in kinetic energy or metric tensor) and $R(\mathbf{q}) \in SO(n)$,

$$R^T(\mathbf{q}^\circ) M^{\frac{1}{2}}(\mathbf{q}^\circ) \dot{\mathbf{q}}^\circ = \mathbf{c}$$

where \mathbf{c} is a constant vector. Then the kinetic energy $T = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}$ along this path will remain constant and

$$\mathbf{q}^\circ(t) = \mathbf{q}(0) + \int_0^t M^{-\frac{1}{2}}(\mathbf{q}^\circ(\tau)) R(\mathbf{q}^\circ(\tau), \tau) \mathbf{c} d\tau. \tag{31}$$

Here $M^{\frac{1}{2}}(\mathbf{q})$ is the matrix square root of $M(\mathbf{q})$ such that $M^{\frac{1}{2}}(\mathbf{q}) M^{\frac{1}{2}}(\mathbf{q}) = M(\mathbf{q})$ and $M^{-\frac{1}{2}}(\mathbf{q}) = [M^{\frac{1}{2}}(\mathbf{q})]^{-1}$ and $R(\mathbf{q}, t) \in SO(n)$ is differentiable trajectory with $R(\mathbf{q}(0), 0) = \mathbb{I}$ that can be used to steer the behavior of $\mathbf{q}^\circ(t)$ without adversely affecting the cost.

The constant vector \mathbf{c} provides sufficient freedom to match arbitrary $\dot{\mathbf{q}}(0)$ or $\mathbf{q}(1)$, but not both.⁴ For example, when specifying initial conditions,

$$\mathbf{q}^\circ(t) = \mathbf{q}(0) + \left(\int_0^t M^{-\frac{1}{2}}(\mathbf{q}^\circ(\tau))R(\mathbf{q}^\circ(\tau), \tau) d\tau \right) M^{\frac{1}{2}}(\mathbf{q}(0)) \dot{\mathbf{q}}(0). \tag{32}$$

from this $\dot{\mathbf{q}}(0)$ can be selected so as to match $\mathbf{q}(1)$ when the matrix in parenthesis is invertible. Of course, it is not possible to specify both $\dot{\mathbf{q}}(0)$ and $\mathbf{q}(1)$ independently.

In general this will not be a solution to the E–L equations for the variational problem with integrand (5), unless $M(\mathbf{q}) = M_0$ is constant. And as such the solution will not be optimal in general. But since the E–L solution has $T = T_0 = const$), and

$$\frac{1}{2}(\dot{\mathbf{q}}^\circ)^T M(\mathbf{q}^\circ)\dot{\mathbf{q}}^\circ = \frac{1}{2}\dot{\mathbf{q}}(0)^T M(\mathbf{q}(0))\dot{\mathbf{q}}(0) = T_0,$$

the cost of this ansatz appears not to be so bad. Moreover its structure lends itself to rapid numerical shooting to match the distal boundary condition.

For example, consider the case of geodesics in the Poincaré half plane model. Suppose that the solution were not known and we seek a reasonable approximation. Using this ansatz with $R = \mathbb{I}$, the result would be

$$\dot{x}_1^\circ = c_1 x_2^\circ \text{ and } \dot{x}_2^\circ = c_2 x_2^\circ.$$

This means that

$$x_2^\circ(t) = x_2(0)e^{c_2 t}$$

which can be back substituted to give

$$x_1^\circ(t) = \alpha + \beta e^{c_2 t}.$$

Now suppose we want to compare to the exact solution presented in the previous section. One way to compare would be to examine how different the trajectory with the same initial conditions is. However, a more telling comparison is to examine the true solution and ansatz approximation for given boundary conditions. The above solutions are written in terms of boundary conditions as

$$x_2(t) = x_2(0)e^{t \ln(x_2(1)/x_2(0))} \tag{33}$$

which can be back substituted to give

$$x_1^\circ(t) = \left(x_1(0) - x_2(0) \frac{x_1(1) - x_1(0)}{x_2(1) - x_2(0)} \right) + x_2(0) \frac{x_1(1) - x_1(0)}{x_2(1) - x_2(0)} e^{t \ln(x_2(1)/x_2(0))}. \tag{34}$$

Consider the boundary conditions $\mathbf{x}(0)$ and $\mathbf{x}(1)$ from the solutions in the previous section:

$$\mathbf{x}^*(0) = \begin{pmatrix} (bd + ac)(d^2 + c^2)^{-1} \\ (d^2 + c^2)^{-1} \end{pmatrix} \text{ and } \mathbf{x}^*(1) = \begin{pmatrix} (bd + ace^2)(d^2 + c^2e^2)^{-1} \\ e(d^2 + c^2e^2)^{-1} \end{pmatrix}.$$

Though it meets the boundary condtions, the energy for the ansatz is a larger constant than for the geodesics. Introducing the rotation $R(x_1, x_2, t)$ provides freedom to reduce this

⁴ As is the case in solving the true variational problem, here too in the case of the ansatz it is not possible to specify both $\dot{\mathbf{q}}(0)$ and $\mathbf{q}(1)$ independently. But if a family of solutions parameterized by varying $\dot{\mathbf{q}}(0)$ is generated, then a solution within this family that matches the condition $\mathbf{q}(1)$ can be chosen.

energy while still meeting the same boundary conditions. For example, suppose $R = R(t)$ defined by counterclockwise measured angle $\theta(t)$. Then

$$\frac{\dot{x}_2^\circ}{x_2^\circ} = c_1 \sin \theta(t) + c_2 \cos \theta(t)$$

or

$$x_2^\circ(t) = x_2^\circ(0) \exp\left(c_1 \int_0^t \sin \theta(\tau) d\tau + c_2 \int_0^t \cos \theta(\tau) d\tau\right).$$

In the simplest case the rotation has one variable to use to steer the solution: $\theta(t) = \omega_0 t$, and $x_1^\circ(t)$ can be obtained numerically.

3 Global optimality and bootstrapping

In the general vectorial case there are many possible paths connecting the initial and final values $\mathbf{q}(0)$ and $\mathbf{q}(1)$. For this reason the argument used in the 1D case earlier cannot generalize and the Euler–Lagrange equation does not in general provide a globally optimal solution. The vectorial extension of the integrand in (9) is of the form $\dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}$. Here $M(\mathbf{q})$ might be the metric tensor for a Riemannian manifold, in which case the solutions of the Euler–Lagrange equations will be equivalent to the equations giving geodesic paths. As is clear even in the simple cases such as the spheres S^1 and S^2 , the geodesics connecting any two points are not unique. Moreover, both the shorter or the longer of the two great arcs connecting two points are solutions to the variational problem when the points are not antipodal. And when they are, an infinite number of great arcs connect them, illustrating that shortest paths need not be unique.

From these examples it is clear that global optimality is closely related to the statement of the problem as a boundary value problem. If instead the problem is restated as that of finding the path from $\mathbf{q}(0)$ with specified value of $\dot{\mathbf{q}}(0)$ that minimizes the cost functional $\int_0^1 f(\mathbf{q}, \dot{\mathbf{q}}, t) dt$, then the nonuniqueness of solutions of the Euler–Lagrange equations disappears. In the case of the sphere, even the longer great arc connecting two points then will be globally optimal. In the one-dimensional case with cost of the form $m(y)(y')^2$ on a domain that is an interval, the paths had only one possible direction to go and so the distinction between the initial-value and boundary-value problems was not important. While this is not true in general for functionals of multi-dimensional vectorial paths $\mathbf{q}(t) \in \mathbb{R}^N$, special cases still do exist in which global optimality can be proved. For example, when $M(\mathbf{q}) = M_0$ is constant, or when $m_{ij}(\mathbf{q}) = m_i(q_i)\delta_{ij}$, then when the domain is a product of intervals so that $\mathbf{q} \in [0, 1]^n$, then the resulting boundary value problem will be globally minimized by the solution of the Euler–Lagrange equations. Global optimality is verified by evaluating the cost of $\mathbf{q}(t) = \mathbf{q}^*(t) + \epsilon(t)$ where $\epsilon(0) = \epsilon(1) = \mathbf{0}$. As with the case of the line, the effect of adding $\epsilon(t)$ is that the cost will only increase relative to that for $\mathbf{q}^*(t)$.

3.1 Bootstrapping globally optimal Euler–Lagrange solutions

Let

$$J_0[\theta, \mathbf{q}] = \int_0^1 f_0(\theta, \dot{\theta}, \mathbf{q}, \dot{\mathbf{q}}, t) dt$$

where

$$f_0(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}, \mathbf{q}, \dot{\mathbf{q}}, t) \doteq \frac{1}{2} \|\dot{\boldsymbol{\theta}} - A(\mathbf{q}, t)\dot{\mathbf{q}}\|_W^2, \tag{35}$$

$\mathbf{q} \in \mathbb{R}^N, \boldsymbol{\theta} \in \mathbb{R}^P$, and $\|B\|_W^2 = \text{tr}(B^T W B)$ is the square of the weighted Frobenius norm with the weighting defined by a given constant positive definite symmetric matrix $W = W^T > 0$.

If $\mathbf{q}(t)$ is fixed in advance, then the globally minimal solution for $\boldsymbol{\theta}(t)$ is obviously defined by $\dot{\boldsymbol{\theta}} = A(\mathbf{q}, t)\dot{\mathbf{q}}$. This is true regardless of whether E–L regularity conditions are imposed on A or not.

Alternatively, if boundary conditions on $\boldsymbol{\theta}$ are imposed and if A does satisfy E–L regularity conditions (e.g., $A : \mathbb{R}^N \times [0, 1] \rightarrow \mathbb{R}^{P \times N}$ is twice continuously differentiable), then the E–L equations in $\boldsymbol{\theta}$ give

$$\frac{d}{dt} \{W [\dot{\boldsymbol{\theta}} - A(\mathbf{q}, t)\dot{\mathbf{q}}]\} = \mathbf{0}$$

or

$$\dot{\boldsymbol{\theta}} - A(\mathbf{q}, t)\dot{\mathbf{q}} = \mathbf{a}, \tag{36}$$

where the vector \mathbf{a} is constant. Integrating again gives another constant \mathbf{b} and the solution is $\boldsymbol{\theta}(t) = \boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{a}, \mathbf{b})$ where

$$\boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{a}, \mathbf{b}) \doteq \mathbf{a}t + \mathbf{b} + \int_0^t A(\mathbf{q}(\tau), \tau)\dot{\mathbf{q}}(\tau)d\tau. \tag{37}$$

Here $*$ denotes that this is the extremal solution the functional $J_0[\boldsymbol{\theta}, \mathbf{q}]$, the subscript denotes that this solution is a functional of \mathbf{q} , and the solution is conditioned on boundary conditions specified by \mathbf{a} and \mathbf{b} . When no boundary conditions are imposed, $\boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{0}, \mathbf{b})$ makes $f_0 = 0$, which is the global minimum value and is independent of how the differentiable differentiable $\mathbf{q}(t)$ is chosen. When $\mathbf{a} \neq \mathbf{0}$, $\boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{a}, \mathbf{b})$ has sufficient freedom to match boundary conditions and for any fixed given $\mathbf{q}(t)$ this choice of $\boldsymbol{\theta}(t)$ is again the global minimum since a perturbation of the form $\boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{a}, \mathbf{b}) \rightarrow \boldsymbol{\theta}_{[\mathbf{q}]}^*(t | \mathbf{a}, \mathbf{b}) + \boldsymbol{\epsilon}(t)$ with $\boldsymbol{\epsilon}(0) = \boldsymbol{\epsilon}(1) = \mathbf{0}$ will only increase the value of the quadratic functional. But when $\mathbf{a} \neq \mathbf{0}$, the choice of $\mathbf{q}(t)$ is not completely free for reasons explained below.

Assuming sufficient regularity,

$$\begin{aligned} & \frac{d}{dt} \left(\frac{\partial f_0}{\partial \dot{q}_i} \right) - \frac{\partial f_0}{\partial q_i} \\ &= - \left(\frac{dA}{dt} \mathbf{e}_i \right)^T W (\dot{\boldsymbol{\theta}} - A\dot{\mathbf{q}}) - (A\mathbf{e}_i)^T W \frac{d}{dt} (\dot{\boldsymbol{\theta}} - A\dot{\mathbf{q}}) + \left(\frac{\partial A}{\partial q_i} \dot{\mathbf{q}} \right)^T W (\dot{\boldsymbol{\theta}} - A\dot{\mathbf{q}}). \end{aligned} \tag{38}$$

Substituting in (36) gives

$$\begin{aligned} & \frac{d}{dt} \left(\frac{\partial f_0}{\partial \dot{q}_i} \right) - \frac{\partial f_0}{\partial q_i} \\ &= - \left(\frac{dA}{dt} \mathbf{e}_i \right)^T W \mathbf{a} + \left(\frac{\partial A}{\partial q_i} \dot{\mathbf{q}} \right)^T W \mathbf{a}. \end{aligned} \tag{39}$$

Consequently, if $\mathbf{a} = \mathbf{0}$, these equations vanish. Otherwise, equating the above expression to zero provides a constraint on $\mathbf{q}(t)$.

That said, the goal in these computations is not to derive E–L Equations in \mathbf{q} since $J_0[\boldsymbol{\theta}, \mathbf{q}]$ is a degenerate variational problem (in the sense that the mass matrix is singular). Rather, the above will be used in the ‘‘bootstrapping’’ procedure described below.

Now suppose that

$$f_1(\mathbf{q}, \dot{\mathbf{q}}, t) = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}} \tag{40}$$

is given for which the corresponding Euler–Lagrange equations globally solve a variational problem with $\mathbf{q}(0)$ and $\mathbf{q}(1)$ specified. Here it is shown that this solution then can be used as a seed to generate a globally optimal solution to a larger variational problem. This is referred to here as “bootstrapping”. Define

$$f_2(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}, \mathbf{q}, \dot{\mathbf{q}}, t) = f_1(\mathbf{q}, \dot{\mathbf{q}}, t) + f_0(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}, \mathbf{q}, \dot{\mathbf{q}}, t) \tag{41}$$

For example, f_1 could correspond to the situation in Sect. 2.1. This scenario is akin to the “warped product” defined in [24] and “Kaluza–Klein Lagrangian” described in [25], though is more general and does not rely on underlying geometric properties, though the result has geometric implications.

In the following let

$$J_1[\mathbf{q}] \doteq \int_0^1 f_1(\mathbf{q}, \dot{\mathbf{q}}, t) dt \text{ and } J_2[\boldsymbol{\theta}, \mathbf{q}] = \int_0^1 f_2(\boldsymbol{\theta}, \dot{\boldsymbol{\theta}}, \mathbf{q}, \dot{\mathbf{q}}, t) dt. \tag{42}$$

Theorem 3.1 *If f_1 is sufficiently regular such that the E–L equations provide extremals, and if \mathbf{q}^* is the global minimizing extremal of J_1 , then when boundary conditions are not imposed on $\boldsymbol{\theta}$, a family of global minimizers of J_2 exists and is of the form $(\mathbf{q}^*, \boldsymbol{\theta}_{[\mathbf{q}^*]}^*(t | \mathbf{0}, \mathbf{b}))$ where $\mathbf{q}^*(t)$ is the global minimizer of J_1 subject to specified $\mathbf{q}(0)$ and $\mathbf{q}(1)$ and $\mathbf{b} \in \mathbb{R}^P$ parameterizes the family.*

Proof By definition, in component form

$$\frac{d}{dt} \left(\frac{\partial f_1}{\partial \dot{q}_i} \right) - \frac{\partial f_1}{\partial q_i} = 0, \tag{43}$$

or explicitly when $f_1(\mathbf{q}, \dot{\mathbf{q}}, t) = \frac{1}{2} \dot{\mathbf{q}}^T M(\mathbf{q}) \dot{\mathbf{q}}$,

$$\mathbf{e}_i^T M(\mathbf{q}) \ddot{\mathbf{q}} + \sum_k \dot{q}_k \mathbf{e}_i^T \frac{\partial M}{\partial q_k} \dot{\mathbf{q}} - \frac{1}{2} \dot{\mathbf{q}}^T \frac{\partial M}{\partial q_i} \dot{\mathbf{q}} = 0,$$

(though restricting to this form is not required in the proof).

Let the solution to the system of equations (43) subject to boundary conditions $\mathbf{q}(0) = \mathbf{q}_0$ and $\mathbf{q}(1) = \mathbf{q}_1$ be denoted as $\mathbf{q}^*(t)$.

On the other hand, evaluating

$$\frac{d}{dt} \left(\frac{\partial f_2}{\partial \dot{\theta}_i} \right) - \frac{\partial f_2}{\partial \theta_i} = 0$$

for all values of i simply results in (37).

The Euler–Lagrange equations for the new system corresponding to coordinates $\{q_i\}$, written in component form, will be

$$\frac{d}{dt} \left(\frac{\partial f_2}{\partial \dot{q}_i} \right) - \frac{\partial f_2}{\partial q_i} = 0.$$

From this it is clear that all of the f_1 terms disappear when evaluated at $\mathbf{q} = \mathbf{q}^*$, and if $\mathbf{a} = \mathbf{0}$, then (39) gives that $\boldsymbol{\theta}_{[\mathbf{q}^*]}^*(t | \mathbf{0}, \mathbf{b})$ causes all other terms to vanish. This means that $(\mathbf{q}^*(t), \boldsymbol{\theta}_{[\mathbf{q}^*]}^*(t | \mathbf{0}, \mathbf{b}))$ is an E–L extremizer of J_2 .

The global minimality of the solutions $(\mathbf{q}^*(t), \theta_{[\mathbf{q}^*]}^*(t | \mathbf{0}, \mathbf{b}))$ is guaranteed by the postulated condition that $\mathbf{q}^*(t)$ is obtained a priori, and the global minimality of $\theta_{[\mathbf{q}^*]}^*(t | \mathbf{0}, \mathbf{b})$ in (37) as a solution to a quadratic cost problem. That is, if $\mathbf{q}^*(t) \rightarrow \mathbf{q}^*(t) + \epsilon_1(t)$ it will only increase the cost of J_1 without affecting the cost of J_0 , since $\theta^*(t)$ is keyed to changes in $\mathbf{q}(t)$. Therefore, if $\theta(t) = \theta_{[\mathbf{q}^* + \epsilon_1]}^*(t | \mathbf{0}, \mathbf{b}) + \epsilon_2(t)$ is substituted into the cost function, $\epsilon_1(t)$ will have no effect on J_0 , which will remain zero, and the addition of $\epsilon_2(t)$ only increases the cost from zero to $\int_0^1 \|\dot{\epsilon}\|_W^2 dt$. This does not depend on $\|\dot{\epsilon}\|_W$ being small. (This is the same as the argument in (7) but with $\mathbf{a} = \mathbf{0}$.) That is, the f_0 term in f_2 is uniquely zero when $\mathbf{a} = \mathbf{0}$, and there is no way to do better. \square

Note that when $\mathbf{a} = \mathbf{0}$ it is only possibly to specify boundary conditions on θ^* at one end. If $\mathbf{a} \neq \mathbf{0}$ more freedom is allowed, but in general \mathbf{q}^* will no longer satisfy the E–L equations for J_2 due to the nonzero terms contributed from J_0 . However, in some special cases those J_0 terms vanish, as described below.

Theorem 3.2 *Two special cases for which global minimality is guaranteed without making the restriction $\mathbf{a} = \mathbf{0}$ are: (1) When $A(\mathbf{q}, t) = A_0$ is constant; (2) When $N = 1$ and $A(q, t) = \mathbf{a}_1(q) \in \mathbb{R}^{P \times 1}$ with q being scalar. The globally minimal solutions in these two cases are respectively $(\mathbf{q}^*, \theta_{\mathbf{q}^*}^*(t | \mathbf{a}, \mathbf{b}))$ and $(q^*, \theta_{[q^*]}^*(t | \mathbf{a}, \mathbf{b}))$ where⁵*

$$\theta_{\mathbf{q}^*}^*(t | \mathbf{a}, \mathbf{b}) \doteq \mathbf{a}t + \mathbf{b} + A_0 (\mathbf{q}^*(t) - \mathbf{q}^*(0)) \tag{44}$$

and

$$\theta_{[q^*]}^*(t | \mathbf{a}, \mathbf{b}) \doteq \mathbf{a}t + \mathbf{b} + \int_0^t \mathbf{a}_1(q^*(s))\dot{q}^*(s)ds \tag{45}$$

satisfies the expanded E–L equations corresponding to (41) and can satisfy all boundary conditions $\theta(0)$ and $\theta(1)$ in addition to those in \mathbf{q} or q .

Proof If $\mathbf{a} \neq \mathbf{0}$ in (36) then the necessary conditions established by the E–L equations become (39). If $A(\mathbf{q}) = A_0$ is constant, then the terms in (39) vanish and the case 1 solution results. This condition also can be satisfied if the dimension of \mathbf{q} is 1, wherein both terms in parenthesis in (39) become the same and hence cancel. In this second case, denote $A = \mathbf{a}_1$, since it consists of one column vector, gives (45).

Global optimality is ensured as before, by introducing a perturbation $\epsilon(t)$ of any size satisfying the boundary conditions $\epsilon(0) = \epsilon(1) = \mathbf{0}$.

The class of problems addressed in this section can be viewed in a slightly different way when (40) holds by rewriting (41) as

$$f(\mathbf{q}, \theta, \dot{\mathbf{q}}, \dot{\theta}, t) = \frac{1}{2} \begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\theta} \end{bmatrix}^T \begin{pmatrix} M(\mathbf{q}) + A^T(\mathbf{q}, t)WA(\mathbf{q}, t) & A^T(\mathbf{q}, t)W \\ WA(\mathbf{q}, t) & W \end{pmatrix} \begin{bmatrix} \dot{\mathbf{q}} \\ \dot{\theta} \end{bmatrix}. \tag{46}$$

Consequently, a quadratic cost that can be decomposed in this way will have solutions that inherit the global optimality from the original smaller problem under appropriate boundary conditions. It also means that the above reasoning can be used recursively. If the matrix in (46) is called $M'(\mathbf{q})$ and we define $\mathbf{q}' \doteq [\mathbf{q}^T, \theta^T]^T$, then the problem can be expanded further by adding an additional terms such as $\|\dot{\phi} - A'(\mathbf{q}', t)\dot{\mathbf{q}}'\|_W^2$.

⁵ In (44) the subscript \mathbf{q}^* is used instead of $[\mathbf{q}^*]$ because in this case $\theta_{\mathbf{q}^*}^*$ depends on \mathbf{q}^* as a function rather than as a functional.

3.2 Bootstrapping globally optimal solutions that lack regularity

If the functional f_1 or the resulting global optimizers lack sufficient regularity for the E–L equations to be applicable, the bootstrapping process can nevertheless be used in some cases to generate globally optimal solutions on higher dimensional spaces.

For example, if $f_1(q, \dot{q}) = m(q)\dot{q}^2$ and m is as in Theorem 2.2, then the globally minimal solution of the variational problem with functional

$$f_2(q, \dot{q}, \theta, \dot{\theta}) = m(q)\dot{q}^2 + [\dot{\theta} - a(q, t)\dot{q}]^2$$

with $t \in [0, 1]$ and $a : \mathbb{R} \times [0, 1] \rightarrow \mathbb{R}$ continuous and bounded will be $q^*(t)$ from Theorem 2.2, and $\theta^*(t) = a_0t + b_0 + \int_0^t a(q^*(s), s)\dot{q}^*(s)ds$. This is true even if m is not differentiable, and hence lacks the regularity to use the E–L equations. Moreover, in cases such as this where q and θ are both one-dimensional, when $a = a(q)$ is autonomous, the integral defining θ^* can be rewritten as

$$\theta^*(t) = a_0t + b_0 + \int_{q^*(0)}^{q^*(t)} a(q)dq, \tag{47}$$

making it an explicit function of $q^*(t)$.

As a second example, we can bootstrap off of the Maniá example of the Lavrentiev phenomenon in (8) and seek a global minimum of

$$J[x, \theta] = \int_0^1 \{(x^3 - t)^2\dot{x}^6 + (\dot{\theta} - a(x, t)\dot{x})^2\} dt \tag{48}$$

subject to boundary conditions $x(0) = 0$ and $x(1) = 1$. We can simply write $x^*(t) = t^{1/3}$ and $\theta^*(t) = a_0t + b_0 + \int_0^t a(x^*(s), s)\dot{x}^*(s)ds$ for a wide variety of functions a , as long as the integral defining $\theta^*(t)$ exists and is finite.

That said, most of the applications that follow involve bootstrapping in the case when the paths and functional are regular.

4 Examples from the theory of framed space curves

In the classical theory of curves in three-dimensional space, the Frenet–Serret Apparatus is used to frame arclength-parametrized curves. In the first subsection of this section, a brief review of that classical theory is given, adapted from Chapter 5 of the author’s book [22]. This material is classical (dating back 150 years) and can be found stated using other notation in [9, 26]. The material is presented here primarily because subtle differences in the notation used here relative to the standard presentation will be convenient later.

4.1 The Frenet–Serret apparatus

Without loss of generality, every curve can be parameterized by its arclength, s . Associated with every three-times-differentiable arclength-parameterized curve, $\mathbf{x}(s) \in \mathbb{R}^3$, are two functions: *curvature*, $\kappa(s)$, and *torsion*, $\tau(s)$. These functions completely describe the shape of the curve up to arbitrary rigid-body motion.

As long as $\kappa(s) \neq 0$, a unique Frenet frame can be attached at every point on the curve. The orientation of this frame relative to the world frame will be

$$R_{FS}(s) \doteq [\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)] \in SO(3)$$

where $\mathbf{t}(s)$, $\mathbf{n}(s)$, and $\mathbf{b}(s)$ are respectively the unit tangent, normal, and binormal. The Frenet frames, the curvature, and the torsion, together are known as the Frenet–Serret apparatus.

The way that the orientation of a Frenet frame evolves along the curve is tied to the curvature and torsion through the system of first order differential equations⁶

$$\frac{dR_{FS}}{ds} = -R_{FS}\Omega_{FS} \tag{49}$$

where Ω_{FS} is a special skew-symmetric matrix with only two independent nonzero entries

$$\Omega_{FS}(s) = \begin{pmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{pmatrix}.$$

The Frenet frame at each value of arclength is then the rotation-translation pair $(R_{FS}(s), \mathbf{x}(s)) \in SO(3) \times \mathbb{R}^3$. No group law is defined. Later in this article several possibilities are explored.

4.2 Minimally varying frames

The classical Frenet frames are not the only way to attach reference frames to space curves, as pointed out by Bishop [27]. Moreover, the Frenet frames are not optimal in the sense of twisting as little as possible around the tangent as the curve is traversed. The word ‘twist’ is somewhat problematic, as it can be confused with torsion. Whereas torsion is a twisting of the curve, the twisting of the frame attached to the curve does not affect the curve shape. For this reason, the word “roll” can be used in place of “twist”, which is consistent with the terminology used to describe motions of aircraft and ships.

Two ways to modify Frenet frames are described here, which demonstrate the variational formulation from earlier in this paper. First the optimal amount of roll along the tangent relative to the Frenet frames is established. Next, alternative parameterizations in place of arclength of the curve that have optimality properties are examined. Finally, the optimal set of frames and reparameterization are combined, which represents an example of the bootstrapping described previously.

Prior to conceiving the general concept of bootstrapping globally optimal variational problems, the author considered special cases that arose from the field of robotics [28, 29], as reviewed in [21, 22], from which some of this presentation is adapted.

4.2.1 Minimally rolling frames

Consider the Frenet frames as a start, and allow freedom to roll around the tangent so that the new frames have orientation

$$R = R_{FS} R_1(\theta) \text{ where } R_1(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{pmatrix}.$$

The optimal $\theta(s)$ will minimize the functional

$$J = \frac{1}{2} \int_0^1 \text{tr} \left(\frac{dR}{ds} \frac{dR^T}{ds} \right) ds. \tag{50}$$

⁶ In most presentations the vectors \mathbf{t} , \mathbf{n} , \mathbf{b} are stacked rather than being written side-by-side, resulting in an equation deceptively similar in appearance to (49) but without introducing R_{FS} .

Computing the derivatives and expanding out gives the explicit expression

$$\frac{1}{2} \operatorname{tr} \left(\frac{dR}{ds} \frac{dR^T}{ds} \right) = \kappa^2 + \tau^2 + 2 \frac{d\theta}{ds} \mathbf{e}_1 \cdot \boldsymbol{\omega}_{FS} + \left(\frac{d\theta}{ds} \right)^2 \tag{51}$$

where

$$\boldsymbol{\omega}_{FS} \doteq \begin{pmatrix} \tau \\ 0 \\ \kappa \end{pmatrix}.$$

This is obtained by using the properties of the trace. For example, $\operatorname{tr} \left(\frac{dR}{ds} \frac{dR^T}{ds} \right) = \operatorname{tr} \left(R^T \frac{dR}{ds} \frac{dR^T}{ds} R \right)$ which is the trace of the product of two skew-symmetric matrices. And for skew-symmetric matrices Ω_1 and Ω_2

$$\frac{1}{2} \operatorname{tr} (\Omega_1 \Omega_2) = \boldsymbol{\omega}_1 \cdot \boldsymbol{\omega}_2$$

where $\boldsymbol{\omega}_i$ is the unique vector such that

$$\boldsymbol{\omega}_i \times \mathbf{v} = \Omega_i \mathbf{v}$$

for every $\mathbf{v} \in \mathbb{R}^3$. This relationship between $\boldsymbol{\omega}_i$ and Ω_i is denoted as

$$\boldsymbol{\omega}_i = (\Omega_i)^\vee \text{ and } \Omega_i = \hat{\boldsymbol{\omega}}_i. \tag{52}$$

From (51) and the expression for $\boldsymbol{\omega}_{FS}$ it is clear that

$$\frac{1}{2} \operatorname{tr} \left(\frac{dR}{ds} \frac{dR^T}{ds} \right) = \kappa^2 + \left(\tau + \frac{d\theta}{ds} \right)^2, \tag{53}$$

and the cost functional is minimized when

$$\frac{d\theta}{ds} = -\tau.$$

Though this had been known for many decades as explained by Bishop [27], it became known to the author as part of studies on snakelike robot arms [28, 29]. In that work the goal was to use a “backbone curve” to capture the overall shape of a physical robot made from joints and links, and since each section of the robot was commanded to adhere to a section of the curve, by limiting the amount of change in position and orientation between reference frames at the end of each section of curve it was easier for the robot to conform.

The above defines frames that have globally minimal roll. In some applications it is required to satisfy conditions at the end points. Such solutions will be of the form

$$\theta^*(s) = c_1 + c_2 s - \int_0^s \tau(\sigma) d\sigma$$

where c_1 and c_2 are determined by the end conditions $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$.

The fact that this is globally optimal follows from the same arguments as in the example of the line (i.e., let $\theta(s) = \theta^*(s) + \epsilon(s)$ and show that this can only make things worse).

4.2.2 Optimal reparameterization for least variation

Within the same motivating snakelike robot application, it was reasoned that if the spacing along the curve was relaxed from being arclength, then this would provide additional freedom for the physical robot to conform to the backbone curve. In analogy with how the Frenet frames are not the only (or perhaps not even the most desirable) orientations to attach to a curve, arclength may not be the best parameter to use. In optimal reparameterization, the idea is to no longer use the arclength, s , but instead use a new curve parameter t such that $s(t)$ satisfies $s(0) = 0$ and $s(1) = 1$ such that a cost functional of the form

$$J = \int_0^1 \left\{ \frac{1}{2} r^2 \operatorname{tr} \left(\frac{dR(s(t))}{dt} \frac{dR^T(s(t))}{dt} \right) + \left(\frac{ds}{dt} \right)^2 \right\} dt$$

is minimized. The constant r is used to introduce units of length because the units to evaluate bending are measured in angular units whereas extension is measured in units of length. For example, if a ball of radius r and unit mass is traversing this trajectory, J is a measure of the total rotational and translational kinetic energy of the ball integrated over the trajectory. If it speeds up or slows down more quickly than the optimal solution, this cost will increase. This idea will be generalized later in the paper.

The integrand in the above functional is of the form $m(s)(s')^2$ where $s' = ds/dt$ and

$$m(s) = \frac{1}{2} r^2 \operatorname{tr} \left(\frac{dR}{ds} \frac{dR^T}{ds} \right) + 1,$$

and therefore the 1D globally optimal solution discussed previously in Theorem 2.2 applies. Call this solution $s^*(t)$.

4.3 Simultaneous optimal reparameterization and roll modification

In the physical robotics problem discussed earlier, it was desirable to combine the optimal roll and curve reparameterization results. This can be done by first solving for the optimal roll and setting $R(s)$, followed by performing the optimal curve reparameterization, resulting in the frame $(R_{FS}(s^*(t))R_1(\theta^*(s^*(t))), \mathbf{x}(s^*(t)))$. A natural question to ask is if this composition of optimal solutions to subproblems optimally solves the larger composite problem. In general one would expect composition of optimal solutions to smaller problems to be suboptimal on the larger space. But in the case of optimal curve reparameterization and rolling, the directions are in a sense orthogonal and commuting (local rolling and extension about/along the same vector).

Moreover, this problem of simultaneous curve reparameterization and optimal roll distribution serves as a concrete application of the class of bootstrapping problems discussed earlier in this paper. Start with a curve initially parameterized by arclength, $\mathbf{x}(s)$ for $s \in [0, 1]$, and with the Frenet frames $[\mathbf{x}(s), R_{FS}(s)]$. We seek a new set of smoothly evolving reference frames of the form $[\mathbf{x}(t), R(t)] = [\mathbf{x}(s(t)), R_{FS}(s(t))R_1(\theta(s(t)))]$, where $R_1(\theta)$ is an added roll, of the Frenet frames about the tangent together with a reparameterization $s = s(t)$.

A cost function in the two variables $(s(t), \theta(t))$ can be constructed of the form

$$C \doteq \frac{1}{2} \int_0^1 \left\{ \frac{1}{2} r^2 \operatorname{tr}(\dot{R}\dot{R}^T) + \dot{\mathbf{x}} \cdot \dot{\mathbf{x}} \right\} dt \tag{54}$$

$$= \frac{1}{2} \int_0^1 \left\{ (r^2 \kappa^2(s) + 1) \dot{s}^2 + r^2 (\tau(s) \dot{s} + \dot{\theta})^2 \right\} dt. \tag{55}$$

This functional is the same form as in (41) with θ taking the place of $\boldsymbol{\theta}$ and s taking the place of \mathbf{q} . If the second term in the integral is zero, then the remaining term is a one-dimensional variational problem with $f(s, \dot{s}, t) = \frac{1}{2}m(s)(\dot{s})^2$ for which a globally minimal solution can be obtained. From the previous discussion about bootstrapping, Theorem 3.2 guarantees the global optimality of the composite problem.

What is also interesting here is that the two problems (optimal roll distribution for an arclength-parameterized curve and optimal curve reparametrization) can be considered sequentially and then merged, and this merged solution will be the same as the joint globally optimal one. Moreover, this example serves as a model for far more general problems with these characteristics.

5 Global optimality in the Lie group setting

Let G be a real matrix Lie group. The dimension of G as a manifold is denoted as N , and the constituent matrices, $g \in G$, are elements of $\mathbb{R}^{n \times n}$.

The variational calculus problem on matrix Lie groups can be formulated as that of finding paths $g(t) \in G$ that extremize a functional

$$J = \int_{t_1}^{t_2} f(g; \boldsymbol{\xi}; t) dt \tag{56}$$

where $\boldsymbol{\xi} = (g^{-1}\dot{g})^\vee$ is a kind of velocity, which is computed simply as the matrix product of g^{-1} and \dot{g} , the latter of which is not an element of G . It can be shown that $g^{-1}\dot{g} \in \mathcal{G}$, the Lie algebra corresponding to G , which has N basis elements $\{E_i\}$, each of which are in $\mathbb{R}^{n \times n}$. This means that

$$g^{-1}\dot{g} = \sum_{i=1}^N \xi_i E_i$$

and a vector $\boldsymbol{\xi} = [\xi_1, \dots, \xi_N]^T \in \mathbb{R}^N$ can be assigned to each element of \mathcal{G} . The notation \vee takes elements of \mathcal{G} and returns vectors in \mathbb{R}^N , and the opposite is the $\hat{\cdot}$ which converts $\boldsymbol{\xi} \in \mathbb{R}^N$ into $\sum_{i=1}^N \xi_i E_i \in \mathcal{G}$. An example of this was demonstrated for $\mathcal{G} = so(3)$ in (43). The choice of $\{E_i\}$ with the definition $E_i^\vee = \mathbf{e}_i$ induces an inner product on \mathcal{G}

$$(E_i, E_j) = \mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$$

and fixes the metric for G .

Each $g_i(t) \doteq \exp(tE_i)$ is a curve in G defining one parameter subgroups, where $\exp(\cdot)$ is the matrix exponential. These curves are geodesics relative to the metric for which $(E_i, E_j) = \delta_{ij}$.

The necessary conditions for optimality in this setting are the *Euler–Poincaré equations*, which can be written as [21, 30–32]

$$\frac{d}{dt} \left(\frac{\partial f}{\partial \xi_i} \right) + \sum_{j,k=1}^N \frac{\partial f}{\partial \xi_k} C_{ij}^k \xi_j = \tilde{E}_i f. \tag{57}$$

where C_{ij}^k are the structure constants for the Lie algebra \mathcal{G} relating the Lie bracket of basis elements as a linear combination of basis elements

$$[E_i, E_j] = \sum_{k=1}^N C_{ij}^k E_k.$$

For a function $f : G \rightarrow \mathbb{R}$,

$$(\tilde{E}_i f)(g) \doteq \left. \frac{d}{ds} f(g \circ \exp(sE_i)) \right|_{s=0}$$

is a directional derivative in the direction of the i^{th} one-dimensional subgroup. Throughout this section, \exp denotes the matrix exponential, which when applied to Lie algebra elements produces Lie group elements. (In (57) $f : G \times \mathcal{G} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}$, but the \tilde{E}_i is essentially a partial derivative, hence only the dependence on G is pertinent).

A version of the above Euler–Poincaré equation with additional end constraints imposed with Lagrange multipliers can be found in [33, 34] in the context of modeling DNA loops and thin steerable elastic tubes proposed for medical procedures.

5.1 Globally minimal solutions of the Euler–Poincaré equation

Consider the integrand of a cost functional

$$f(g, \xi, t) = \frac{1}{2} (\xi - \mathbf{c})^T K (\xi - \mathbf{c}) = \frac{1}{2} \sum_{i,j=1}^N K_{ij} (\xi_i - c_i) (\xi_j - c_j)$$

with constant $K = K^T > 0$, the absolute minimal value is obtained when $\xi = \mathbf{c}$. The corresponding globally minimal trajectory on the group is then defined by

$$g^{-1} \dot{g} = \sum_{i=1}^N c_i E_i.$$

If \mathbf{c} is a constant vector, then

$$g(t) = g(0) \exp \left(t \sum_{i=1}^N c_i E_i \right).$$

However, this solution is not able to satisfy arbitrary distal end conditions, such as specifying $g(1)$. If the goal is to do that, then the Euler–Poincaré equations of the form

$$\frac{d}{dt} \left(\sum_{j=1}^N K_{ij} (\xi_j - c_j) \right) + \sum_{j,k=1}^N \left(\sum_{l=1}^N K_{kl} (\xi_l - c_l) \right) C_{ij}^k \xi_j = 0 \tag{58}$$

need to be satisfied. In general, this requires numerical solution, but in some cases things simplify. In particular, consider the case when $\mathbf{c} = \mathbf{0}$ and let

$$S_{lj}^i \doteq \sum_{k=1}^N K_{kl} C_{ij}^k. \tag{59}$$

If $S_{lj}^i = -S_{jl}^i$ then $\sum_{j,l=1}^N S_{lj}^i \xi_l \xi_j = 0$ and (58) reduces to

$$K \dot{\xi} = \mathbf{0} \implies \xi(t) = \xi(0) \implies g(t) = g(0) \circ e^{t \dot{\xi}(0)}. \tag{60}$$

The shortest path connecting $g(0) = g_0$ and $g(1) = g_1$ is then

$$g^*(t) = g_0 \circ \exp(t \cdot \log(g_0^{-1} \circ g_1)) \tag{61}$$

where \log is the matrix logarithm. However, if $S_{lj}^i \neq -S_{jl}^i$ then the path generated by solving the E–P variational equations is generally not this geometric one. For example, when $G = SO(3)$ and when the moment of inertia is not isotropic, the optimal path between rotations can become very complicated. This is relevant to satellite attitude reorientation problems, as discussed in [35, 36].

5.2 The rotation group $SO(3)$ and the sphere \mathbb{S}^2

5.2.1 Euler–Poincaré Equations for $SO(3)$

For the case when $G = SO(3)$ group elements are 3×3 special orthogonal (rotation) matrices, denoted as R , and the Lie algebra elements are skew symmetric matrices. In particular $R^T \dot{R} = \Omega$ corresponds to the body-fixed description of angular velocity, ω , and the notation in (52) is used, i.e., $\omega = \Omega^\vee$ and $\Omega = \hat{\omega}$.

Minimizing a functional of the form

$$J = \frac{1}{2} \int_{t_1}^{t_2} (\omega - \omega_0)^T \mathcal{I} (\omega - \omega_0) dt \tag{62}$$

arises in several applications. For example, when $\omega_0 = \mathbf{0}$ this corresponds to minimizing kinetic energy due to rotation where \mathcal{I} is the moment of inertia matrix. The Euler–Poincaré equations corresponding to the above functional are

$$\mathcal{I} \dot{\omega} + \omega \times \mathcal{I} (\omega - \omega_0) = \mathbf{0}, \tag{63}$$

and these become the classical Euler equations of motion without applied moments when $\omega_0 = \mathbf{0}$.

Trajectory planning on $SO(3)$ is both of fundamental importance as an example of more general geometric methods as in [37–43], as well as practical and elegant engineering and computer graphics applications such as motion interpolation [44, 45] and spacecraft attitude reorientation [35, 36]. In the subsections that follow, specific variational solutions are examined, and their minimality properties are queried.

5.2.2 Geodesics on $SO(3)$

The situation in (60) occurs for the functional in (62) in the special case when $\mathcal{I} = \mathbb{I}$ and $\omega = \mathbf{0}$ and (61) applies. That is, the geodesic path connecting $R_0, R_1 \in SO(3)$ with $\text{tr}(R_1^T R_0) > -1$ (so that they are not rotations related by a rotation of angle π) is

$$R^*(t) = R_0 \circ \exp(t \cdot \log(R_0^T R_1)). \tag{64}$$

Furthermore, this path can be shown to be globally optimal in this case because of the structure of the cost function. To verify this statement, suppose that $R^*(t)$ is replaced with

$R(t) = R^*(t)Q(t)$ where $Q(t)$ is an arbitrary differentiable path in G with end points $Q(0) = Q(1) = e$. Then

$$R^T \dot{R} = (R^* Q)^T \{ \dot{R}^* Q + R^* \dot{Q} \} = Q^T \left((R^*)^T \dot{R}^* \right) Q + Q^T \dot{Q}.$$

With $\omega \doteq (R^T \dot{R})^\vee$, $\omega^* = ((R^*)^T \dot{R}^*)^\vee$, and $\omega_Q \doteq (Q^T \dot{Q})^\vee$ we have

$$\omega = Q^T \omega^* + \omega_Q.$$

Then

$$\|\omega\|^2 = \|\omega^*\|^2 + 2(\omega^*)^T Q \omega_Q + \|\omega_Q\|^2$$

Note that ω^* is constant and $Q \omega_Q = (\dot{Q} Q^T)^\vee = \omega_Q^s$ (space-fixed description of angular velocity corresponding to $Q(t)$). Therefore

$$\int_0^1 (\omega^*)^T Q \omega_Q dt = (\omega^*)^T \int_0^1 \omega_Q^s dt.$$

For small rotations $Q(t)$ that do not move far from the identity for $t \in [0, 1]$, the boundary conditions $Q(0) = Q(1) = \mathbb{I}$ force

$$\int_0^1 \omega_Q^s dt = \mathbf{0}.$$

Moreover, both ω^* and any perturbation Q (whether small or large) must obey the invariance of the problem. For example, if the start and end points are reversed, we get $\omega^* \rightarrow -\omega^*$. And since $Q(0) = Q(1) = \mathbb{I}$, and since $SO(3)$ is a symmetric space, then unless $Q(t)$ executes a full 2π rotation (which cannot help since the total angle traversed from R_0 to R_1 is less than π), symmetry under swapping the roles of the end points indicate that the angular velocity ω_Q^s must integrate to zero. That is, there is no way to do better than $R^*(t)$, and the argument is essentially the same as for the line in the plane.

5.2.3 The case when $\omega_0 \neq 0$

The E-P equation in (63) when $\mathcal{I} = \mathbb{I}$ and $\omega_0 \neq 0$ results in

$$\dot{\omega} - \omega \times \omega_0 = \mathbf{0}. \tag{65}$$

This is a linear system of equations with solution

$$\omega(t) = \exp(-t\hat{\omega}_0)\omega(0) = \left(\mathbb{I} - \sin(\|\omega_0\|t) \frac{\hat{\omega}_0}{\|\omega_0\|} + (1 - \cos(\|\omega_0\|t)) \frac{\hat{\omega}_0^2}{\|\omega_0\|^2} \right) \omega(0). \tag{66}$$

Since $\exp(t\hat{\omega}_0)\omega_0 = \omega_0$ it follows that

$$\|\omega(t) - \omega_0\|^2 = \|\omega(0) - \omega_0\|^2.$$

With initial conditions $R(0) = \mathbb{I}$, the small-time solution for $R(t)$ is then

$$R(t \ll 1) = \mathbb{I} + \int_0^t \hat{\omega}(\tau) d\tau = \mathbb{I} + \hat{\theta}$$

where

$$\theta(t) = \int_0^t \omega(\tau) d\tau = \left(t\mathbb{I} + [\cos(\|\omega_0\|t) - 1] \frac{\hat{\omega}_0}{\|\omega_0\|^2} + (t\|\omega_0\| - \sin(\|\omega_0\|t)) \frac{\hat{\omega}_0^2}{\|\omega_0\|^3} \right) \omega(0).$$

A small perturbation of this will be $R(t) \exp(\hat{\epsilon})$ with $\epsilon(0) = \epsilon(1) = \mathbf{0}$, or $\theta \rightarrow \theta + \epsilon$, leading to the cost $\|\omega(t) - \omega_0 + \dot{\epsilon}\|^2$. As with the case of the line in the plane, this reduces to $\|\omega(t) - \omega_0\|^2 + \|\dot{\epsilon}\|^2$ because the cross term integrates to zero.

5.2.4 Geodesics on the sphere

A convenient way to describe geodesics connecting two non-colinear points $\mathbf{a}, \mathbf{b} \in \mathbb{S}^2$ is via the rotation matrix $R(\mathbf{a}, \mathbf{b})$, which most directly transforms a unit vector \mathbf{a} into the unit vector \mathbf{b} :

$$\mathbf{b} = R(\mathbf{a}, \mathbf{b})\mathbf{a}.$$

If $\theta_{ab} \in [0, \pi]$ denotes the angle of rotation measured counterclockwise from \mathbf{a} to \mathbf{b} around the axis defined by $\mathbf{a} \times \mathbf{b}$. Then again using the notation in (52) it can be shown that [22]

$$R(\mathbf{a}, \mathbf{b}) = \exp\left(\frac{\theta_{ab}}{\sin \theta_{ab}} \widehat{\mathbf{a} \times \mathbf{b}}\right) = \mathbb{I} + \widehat{\mathbf{a} \times \mathbf{b}} + \frac{(1 - \mathbf{a} \cdot \mathbf{b})}{\|\mathbf{a} \times \mathbf{b}\|^2} \left(\widehat{\mathbf{a} \times \mathbf{b}}\right)^2. \tag{67}$$

Then the minimal length geodesic arc connecting \mathbf{a} and \mathbf{b} is

$$\mathbf{u}(t) = \exp(t \log R(\mathbf{a}, \mathbf{b}))\mathbf{a} = \mathbf{a} + t \left(\widehat{\mathbf{a} \times \mathbf{b}}\right) \mathbf{a} + t^2 \frac{(1 - \mathbf{a} \cdot \mathbf{b})}{\|\mathbf{a} \times \mathbf{b}\|^2} \left(\widehat{\mathbf{a} \times \mathbf{b}}\right)^2 \mathbf{a}$$

for $t \in [0, 1]$. In the above, the hat notation in (52) is used.

5.3 Globally optimal solutions on direct and semi-direct products

Here the concept of (θ^*, s^*) from optimal curve rolling and reparameterization is generalized where each variational subproblem is in a subgroup of a larger group, which then are bootstrapped to obtain solutions on the larger space.

5.3.1 Direct products

Let G and H be a matrix Lie groups with respective Lie algebra basis elements $\{E_i\}$ and let $\{\mathcal{E}_j\}$. Given a direct product with elements $(g, h) \in G \times H$ and the direct product group law

$$(g_1, h_1)(g_2, h_2) = (g_1g_2, h_1h_2).$$

Let $\omega = (g^{-1}\dot{g})^\vee$ and $\mathbf{v} = (h^{-1}\dot{h})^\vee$ denote the velocities. (The reason for this choice of names will become clear soon.) As an example, the so-called *pose change group* [46] is the direct product $SO(3) \times \mathbb{R}^3$. This can be described as a matrix Lie group with elements of the form of a direct sum

$$(R, t) = R \oplus \begin{pmatrix} \mathbb{I} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \in \mathbb{R}^{7 \times 7}.$$

In this context

$$h^{-1}\dot{h} = \begin{pmatrix} \mathbb{I} & -\mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \begin{pmatrix} \mathbb{O} & \dot{\mathbf{t}} \\ \mathbf{0}^T & 0 \end{pmatrix} = \begin{pmatrix} \mathbb{O} & \dot{\mathbf{t}} \\ \mathbf{0}^T & 0 \end{pmatrix}$$

and $\mathbf{v} = (h^{-1}\dot{h})^\vee = \dot{\mathbf{t}}$.

$SO(3) \times \mathbb{R}^3$ is **not** the group of handedness preserving Euclidean motions, but it still is relevant in describing changes in pose. For example, if an aerial vehicle is moving, then it can be convenient to track its relative orientation and absolute position as described in [47]. Similarly, for protein-protein docking this direct product approach is often taken [48–50].

Moreover, if $(R, \mathbf{t}) \in SO(3) \times \mathbb{R}^3$, it is convenient to define body-fixed angular velocity $\boldsymbol{\omega} = (R^T \dot{R})^\vee$ and the space fixed view of translational velocity $\mathbf{v} = \dot{\mathbf{t}}$. Indeed, in the classical expression for kinetic energy

$$T = \frac{1}{2} \boldsymbol{\omega}^T \mathcal{I} \boldsymbol{\omega} + \frac{1}{2} m \mathbf{v} \cdot \mathbf{v}$$

this is exactly what is done.

It is clear that if the globally minimal solutions to

$$J_1 = \int_{t_1}^{t_2} f_1(g, \boldsymbol{\omega}, t) dt \quad \text{and} \quad J_2 = \int_{t_1}^{t_2} f_2(h, \mathbf{v}, t) dt$$

are obtained separately, then the globally minimal solution for $J_1 + J_2$ on the product group will be obtained by pairing the solutions for each independently as $(g^*(t), h^*(t))$. In the case of the direct product $SO(3) \times \mathbb{R}^3$

$$(R^*(\tau), \mathbf{t}^*(\tau)) = \left(R_0 \circ \exp(\tau \cdot \log(R_0^T R_1)), \mathbf{t}_0 + (\mathbf{t}_1 - \mathbf{t}_0)\tau \right). \tag{68}$$

Slightly more interesting than that is when the cost is of the form⁷

$$f(g, h, \boldsymbol{\omega}, \mathbf{v}) = f_1(g, \boldsymbol{\omega}) + \frac{1}{2} \|\mathbf{v} - A(g)\boldsymbol{\omega}\|_W^2 \tag{69}$$

in analogy with the coordinate-dependent discussion of bootstrapping earlier in the paper. Assuming that f_1 satisfies the E–P equation, then the E–P equation for g will reduce to

$$\frac{d}{dt} \mathbf{e}^T A^T W (A\boldsymbol{\omega} - \mathbf{v}) - \boldsymbol{\omega} \times (A\boldsymbol{\omega} - \mathbf{v}) = (A\boldsymbol{\omega} - \mathbf{v}) \tilde{E}_i A \boldsymbol{\omega}.$$

and the E–P equation for h will reduce to

$$\frac{d}{dt} (A\boldsymbol{\omega} - \mathbf{v}) = \mathbf{0}.$$

The condition for global minimality

$$\mathbf{v} = A(g)\boldsymbol{\omega}$$

is obtained when $\boldsymbol{\omega}(0) = \mathbf{0}$.

5.3.2 Semidirect products

Sometimes cost functions such as those in (69) arise naturally in the context of semi-direct products rather than direct. The group of special Euclidean motions

$$SE(3) = \mathbb{R}^3 \rtimes SO(3)$$

⁷ The case $f(g, h, \boldsymbol{\omega}, \mathbf{v}) = f_1(h, \mathbf{v}) + \|\boldsymbol{\omega} - A(h)\mathbf{v}\|_W^2$ can be handled analogously.

is such semi-direct product. Its elements reside in the same underlying space as the pose change group, $g_i = (R_i, \mathbf{t}_i)$, but the group law is different:

$$g_1 g_2 = (R_1 R_2, R_1 \mathbf{t}_2 + \mathbf{t}_1).$$

The standard way to describe $SE(3)$ as a matrix Lie group is to write its elements as

$$g = \begin{pmatrix} R & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \in \mathbb{R}^{4 \times 4}.$$

The corresponding velocities can be written as

$$\xi = (g^{-1} \dot{g})^\vee = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} (R^T \dot{R})^\vee \\ R^T \mathbf{v} \end{pmatrix}$$

For example, in modeling the minimal energy shapes of DNA molecules treated as elastic filaments, a common model to describe the strain energy at the point s along a backbone curve is of the form

$$f(g, \xi) = \frac{1}{2} \|\xi - \xi_0\|_K.$$

For special cases of K , the theory in (5.1) can be applied. The resulting geodesics will be different than in the direct product case, with the translational parts taking helical paths. In principle, similar expressions can be used when modeling large elastic deformations in sheets and solids comprised of elastic filaments in multiple directions. Then such a formulation may be applied to robots powered by transparent elastic actuators, as in [51].

For special cases of $A(g)$ it is also possible to capture a cost function such as (69) in the form of $\frac{1}{2} \|\xi\|_K$. In particular, writing

$$\|\xi\|_K = \xi_1^T K_{11} \xi_1 + \|\xi_2 - A_0 \xi_1\|_W^2$$

we recognize that $\xi_2 = R^T \mathbf{v}$, and when $W = c\mathbb{I}$, then

$$\|\xi_1 - A_0 \xi_2\|_W^2 = c^2 \|\xi_2 - RA_0^{-1} \xi_1\|^2.$$

Then (69) can be written as a homogenous problem on $SE(3)$ rather than as a bootstrapped problem from $SO(3)$ to $SO(3) \times \mathbb{R}^3$. In other words, we can minimize the bootstrapped problem on $SO(3) \times \mathbb{R}^3$ as a homogeneous one on $SE(3)$ with $f(\xi) = \frac{1}{2} \|\xi\|_K$ and where

$$K = \begin{pmatrix} K_{11} & cA_0 \\ cA_0^T & c\mathbb{I} \end{pmatrix}.$$

5.4 Application: extraction of salient actions from a video sequence

An $M \times N$ black-and-white image consists of $M \cdot N$ pixels, each with fixed location and pixel intensity (grayscale) values. When there is color, instead of a scalar, each pixel has a vector of RGB (red, green, blue) values. A black-and-white video can be viewed as a curve in $\mathbb{R}^{M \times N}$ parameterized by time, and an individual image in the video is a point in this space. Similarly, a color video can be thought of as a curve in $\mathbb{R}^{M \times N \times 3}$.

Given a video sequence, a problem of importance is activity recognition. For example, if a humanoid robot is charged with taking care of grandma, the robot should decode grandma’s actions and gestures such as waving, walking, and summoning. As such, salient actions need to be extracted from video sequences and matched with stored representatives of such action

classes in a database. A first step in comparing two video sequences $X_1(\tau)$ and $X_2(\tau)$ for $\tau \in [0, 1]$ is to put the salient actions on a common timescale. Suppose that each video illustrates a human hand waving from left to right and back. One person may wave faster at the beginning and slower at the end, and the other person may wave slower at the beginning and faster at the end. This means that computing a quantity such as $\int_0^1 \|X_1(\tau) - X_2(\tau)\|^2 d\tau$ to measure the differences between the videos will give a large value, and miss the fact that both videos describe essentially the same waving simply because they are out of sync. A way to reconcile the differences in the trajectories described by the videos is to assign a different timescale for one relative to the other. This is one of the goals of dynamic time warping. The literature on this topic is immense, and some well known articles include [52–54].

An alternative to dynamic time warping is to optimally reparameterize both of the videos as $X_i(\tau_i(t))$. Since $dX_i/dt = X'_i(\tau)\dot{\tau}$, a measure of the amount of change in video i at any instant is $\|\dot{X}_i(\tau)\|^2 = \|X'_i(\tau)\|^2 \dot{\tau}^2$. An optimal timescale $\tau_i^*(t)$ for each video can be defined as one that expands time when there is a lot of change in the video (i.e., large value of derivative), and compresses time when nothing is happening (i.e., small value of derivative). In other words, we seek $X_i(\tau_i^*(t))$ where τ_i^* minimizes the functional

$$J = \int_0^1 \|X'_i(\tau)\|^2 \dot{\tau}^2 dt.$$

This is analogous to the curve reparametrization problem reviewed earlier in this paper, and normalizes the timescale of both videos according to how much is changing in each. This is the approach taken in [55, 56].

The curvature of such an arclength parameterized curve in pixel space can be described using a higher-dimensional version of the Frenet–Serret apparatus, and the analogs of curvature and torsion can be used to identify motion features as explained in [57].

Moreover, rather than considering only the raw images in the video sequence, it is common to use image processing techniques to detect edges or extract other features that are tracked from frame to frame. These features undergo motions within the image plane, which can be thought of as being orthogonal to the temporal direction. Motions can be intrinsic to the the action being observed by the camera, or they can be a result of nuisances, as would be the case if the camera is shaking when the video is taken, or one person is waving from a stationary location and another is waving while riding a horse. When the goal is to isolate actions, the contribution of these nuisance motions should be eliminated as much as possible before comparing videos. It therefore makes sense to seek the optimal path $g_i(\tau) \in G$ (a group of nuisance motions) such that $Y_i(\tau) = g_i(\tau) \cdot X_i(\tau)$ is minimally varying. Minimizing $\int_0^1 \|dY_i/d\tau\|^2 d\tau$ results in an Euler–Poincaré problem on G . After nuisance motions are removed, the resulting $Y_i^*(\tau) = g_i^*(\tau) \cdot X_i(\tau)$ can be optimally reparameterized as described above. The resulting motion-corrected and temporally reparameterized videos $Y_i^*(\tau_i^*(t))$ can then be compared.

In this way a problem reduces to something analogous to the joint θ - s problem in curve framing (with motion in the image plane being like $\theta(t)$, and compression or expansion of the temporal variable being like $s(t)$). This joint problem was addressed in [58].

6 Conclusions

A number of variational calculus problems with globally optimal solutions are reviewed and used to construct new globally optimal solutions to variational problems on larger spaces.

This “bootstrapping” approach is demonstrated in the context of simultaneous twist adjustment relative to the Frenet frames of a space curve to give the Bishop frames, together with the reparametrization of the framed curve starting with arclength to result in a framing that evolves more gradually. These ideas are then generalized to classes of coordinate-dependent variational problems and coordinate-free Lie-group problems with globally minimal solutions. In many cases of interest one underlying space can be endowed with multiple different group operations (e.g., direct product or semidirect product), and it is shown how different globally minimal trajectories result from bootstrapping from subgroups to these larger product groups.

Acknowledgements This work was performed under National Research Foundation, Singapore, under its Medium Sized Centre Programme - Centre for Advanced Robotics Technology Innovation (CARTIN), subaward R-261-521-002-592, MOE Tier 2 grant REBOT A-8000424-00-00, SMI Maritime R+D Grant A-8000081-02-00, CDE Board funds E-465-00-0009-001, and National University of Singapore Startup grants A-0009059-02-00 and A-0009059-03-00. The author would like to thank Mr. Jikai Ye, Mr. Yuwei Wu, and Ms. Yi Xu for checking equations.

Data Availability There is no data in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Brechtken-Manderscheid, U.: Introduction to the Calculus of Variations. Chapman and Hall, New York (1991)
2. Ewing, G.M.: Calculus of Variations with Applications. W.W. Norton and Co., New York (1969)
3. Cheeger, J., Ebin, D.G.: Comparison Theorems in Riemannian Geometry, vol. 9. North-Holland, Amsterdam (1975)
4. Bridson, M.R., Haefliger, A.: Metric Spaces of Non-positive Curvature, vol. 319. Springer, New York (2013)
5. Gruver, W.A., Sachs, E.: Algorithmic Methods in Optimal Control. Pitman Publishing Ltd, Boston (1980)
6. Kamien, M.I., Schwartz, N.L.: Dynamic Optimization: The Calculus of Variations and Optimal Control in Economics and Management. North-Holland, New York (1991)
7. Chirikjian, G.S., Shiffman, B.: Applications of convex geometry to Minkowski sums of ellipsoids in \mathbb{R}^N : closed-form parametric equations and volume bounds. *Int. J. Math.* **32**, 2140009 (2021)
8. Brockett, R.: Notes on stochastic processes on manifolds. In: Byrnes, C.I., Datta, B.N., Martin, C.F. (eds.) *Systems and Control in the Twenty-First Century*, pp. 75–100. Birkhäuser, Boston (1997)
9. Do Carmo, M.P.: *Differential Geometry of Curves and Surfaces: Revised and Updated*, 2nd edn. Courier Dover Publications, Mineola (2016)
10. Dubrovin, B.A., Fomenko, A.T., Novikov, S.P.: *Modern Geometry-Methods and Applications: Part I: The Geometry of Surfaces, Transformation Groups, and Fields*, vol. 93. Springer, New York (1984)
11. Lavrentiev, M.: Sur quelques problèmes du calcul des variations. *Ann. Matem. Pura Appl.* **4**, 7–28 (1926)
12. Maniá, B.: Sopra un esempio di Lavrentieff. *Boll. Un. Matem. Ital.* **13**, 147–153 (1934)
13. Lowen, P.D.: On the Lavrentiev phenomenon. *Can. Math. Bull.* **30**(1), 102–108 (1987)
14. Buttazzo, G., Belloni, M.: A survey on old and recent results about the gap phenomenon in the calculus of variations. In: Lucchetti, R., Revalski, J. (eds.) *Recent Developments in Well-Posed Variational Problems*, pp. 1–28. Springer, Berlin (1995)
15. Ball, G.M., Mizel, V.J.: Singular minimizers for regular one-dimensional problems in the calculus of variations. *Bull. Am. Math. Soc.* **11**, 143–146 (1984)

16. Ball, J.M., Mizel, V.J.: One-dimensional variational problems whose minimizers do not satisfy the Euler–Lagrange equation. *Arch. Ration. Mech. Anal.* **90**, 325–388 (1985)
17. Foss, M.: Examples of the Lavrentiev phenomenon with continuous Sobolev exponent dependence. *J. Convex Anal.* **10**(2), 445–464 (2003)
18. Cellina, A., Ferriero, A., Marchini, E.M.: Reparametrizations and approximate values of integrals of the calculus of variations. *J. Differ. Equ.* **193**(2), 374–384 (2003)
19. Ferriero, A.: The approximation of higher-order integrals of the calculus of variations and the Lavrentiev phenomenon. *SIAM J. Control. Optim.* **44**(1), 99–110 (2005)
20. Lee, T., Leok, M., McClamroch, N.H.: *Global Formulations of Lagrangian and Hamiltonian Dynamics on Manifolds*. Springer, New York (2017)
21. Chirikjian, G.S.: *Stochastic Models, Information Theory, and Lie Groups: Volumes I + II*. Birkhäuser, Boston (2009/2012)
22. Chirikjian, G.S., Kyatkin, A.B.: *Harmonic Analysis for Engineers and Applied Scientists*. Dover, Mineola (2016)
23. Eisenhart, L.P.: *Riemannian Geometry*. Princeton University Press, Princeton (1950)
24. Burago, D., Burago, I.D., Burago, Y., Ivanov, S., Ivanov, S.V., Ivanov, S.A.: *A Course in Metric Geometry*, vol. 33. American Mathematical Soc., Providence (2001)
25. Marsden, J.E., Ratiu, T.S.: *Introduction to Mechanics and Symmetry*, 2nd edn. Springer, New York (1998)
26. Millman, R.S., Parker, G.D.: *Elements of Differential Geometry*. Prentice-Hall Inc., Englewood Cliffs (1977)
27. Bishop, R.: There is more than one way to frame a curve. *Am. Math. Mon.* **82**, 246–251 (1975)
28. Chirikjian, G.S.: *Theory and Applications of Hyper-Redundant Robotic Manipulators*. Division of Engineering and Applied Science, California Institute of Technology (1992). http://etd.caltech.edu/etd/available/etd-11082006-132210/unrestricted/Chirikjian_gs_1992.pdf
29. Chirikjian, G.S., Burdick, J.W.: Kinematically optimal hyper-redundant manipulator configurations. *IEEE Trans. Robot. Autom.* **11**, 794 (1995)
30. Poincaré, H.: Sur une forme nouvelle des equations de la mecanique. *Cr. Hebd. Acad. Sci.* **132**, 369 (1901)
31. Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler–Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.* **137**, 1 (1998)
32. Bloch, A., Krishnaprasad, P.S., Marsden, J.E., Ratiu, T.S.: The Euler–Poincaré equations and double bracket dissipation. *Commun. Math. Phys.* **175**, 1 (1996)
33. Kim, J.-S., Chirikjian, G.S.: Conformational analysis of stiff chiral polymers with end constraints. *Mol. Simul.* **32**(14), 15 (2006)
34. Rucker, C., Webster, R.J., III, Chirikjian, G.S., Cowan, N.J.: Equilibrium conformations of concentric-tube continuum robots. *Int. J. Robot. Res.* **29**(10), 1263–1280 (2010)
35. Junkins, J.L., Turner, J.D.: *Optimal Spacecraft Rotational Maneuvers*. Elsevier, Amsterdam (2012)
36. Markley, F.L., Crassidis, J.L.: *Fundamentals of Spacecraft Attitude Determination and Control*, pp. 361–364. Springer, New York (2014)
37. Iserles, A., Munthe-Kaas, H.Z., Nørsett, S.P., Zanna, A.: Lie group methods. *Acta Numer* **9**, 215–365 (2000)
38. Baillieul, J.: Geometric methods for nonlinear optimal control problems. *J. Optim. Theory Appl.* **25**(4), 519–548 (1978)
39. Crouch, P., Leite, F.S.: The dynamic interpolation problem: on Riemannian manifolds, Lie groups, and symmetric spaces. *J. Dyn. Control Syst.* **1**(2), 177–202 (1995)
40. Bloch, A., Crouch, P.: Nonholonomic and Vakonomic control systems on Riemannian manifolds. *Fields Inst. Commun.* **1**, 25–52 (1993)
41. Jurdjevic, V.: Non-euclidean elastica. *Am. J. Math.* **117**(1), 93–124 (1995)
42. Jurdjevic, V., Sussmann, H.J.: Control systems on Lie groups. *J. Differ. Equ.* **12**(2), 313–329 (1972)
43. Helmke, U., Moore, J.B.: *Optimization and Dynamical Systems*. Springer, Berlin (2012)
44. Park, F.C., Ravani, B.: Smooth invariant interpolation of rotations. *ACM Trans. Graph.* **16**(3), 277–295 (1997)
45. Park, F.C., Ravani, B.: Bézier curves on Riemannian manifolds and lie groups with kinematics applications. *ASME J. Mech. Des.* **117**(1), 36–40 (1995)
46. Chirikjian, G.S., Mahony, R., Ruan, S., Trumpf, J.: Pose changes from a different point of view. *ASME J. Mech. Robot.* **10**(2), 021008 (2018)
47. Mahony, R., Kumar, V., Corke, P.: Multirotor aerial vehicles: modeling, estimation, and control of quadrotor. *IEEE Robot. Autom. Mag.* **19**(3), 20–32 (2012)
48. Kozakov, D., Hall, D.R., Xia, B., Porter, K.A., Padhorny, D., Yueh, C., Beglov, D., Vajda, S.: The ClusPro web server for protein-protein docking. *Nat. Protoc.* **12**(2), 255–278 (2017)

49. Vakser, I.A.: Protein-protein docking: from interaction to interactome. *Biophys. J.* **107**(8), 1785–1793 (2014)
50. Lyskov, S., Gray, J.J.: The RosettaDock server for local protein–protein docking. *Nucleic Acids Res.* **36**(suppl 2), W233–W238 (2008)
51. Wang, Y., Loh, L.Y.W., Gupta, U., Foo, C.C., Zhu, J.: Bio-inspired soft swim bladders of large volume change using dual dielectric elastomer membranes. *J. Appl. Mech.* **87**(4), 041007 (2020)
52. Berndt, D.J., Clifford, J.: Using dynamic time warping to find patterns in time series. In: *KDD workshop*, vol. 10, no. (16), pp. 359–370 (1994)
53. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **7**(3), 358–386 (2005)
54. Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E.: Searching and mining trillions of time series subsequences under dynamic time warping. In: *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 262–270 (2012)
55. Mitchel, T.W., Ruan, S., Chirikjian, G.S.: Signal alignment for humanoid skeletons via the globally optimal reparameterization algorithm. In: *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*, pp. 217–223 (2018)
56. Mitchel, T.W., Ruan, S., Gao, Y., Chirikjian, G.S.: The globally optimal reparameterization algorithm: an alternative to fast dynamic time warping for action recognition in video sequences. In: *15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 1290–1297 (2018)
57. Chen, H., Chirikjian, G.S.: Curvature: a signature for action recognition in video sequences. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020)
58. Mitchel, T.W., Wuelker, C., Kim, J.S., Ruan, S., Chirikjian, G.S.: Quotienting impertinent camera kinematics for 3D video stabilization. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV* (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.