

Realizing Nonholonomic Dynamics as Limit of Friction Forces

Jaap Eldering*

*Universidade de São Paulo — ICMC
Avenida Trabalhador Sao-carlense 400,
CEP 13566-590, Sao Carlos, SP, Brazil*

Received April 27, 2016; accepted May 8, 2016

Abstract—The classical question whether nonholonomic dynamics is realized as limit of friction forces was first posed by Carathéodory. It is known that, indeed, when friction forces are scaled to infinity, then nonholonomic dynamics is obtained as a singular limit.

Our results are twofold. First, we formulate the problem in a differential geometric context. Using modern geometric singular perturbation theory in our proof, we then obtain a sharp statement on the convergence of solutions on infinite time intervals. Secondly, we set up an explicit scheme to approximate systems with large friction by a perturbation of the nonholonomic dynamics. The theory is illustrated in detail by studying analytically and numerically the Chaplygin sleigh as an example. This approximation scheme offers a reduction in dimension and has potential use in applications.

MSC2010 numbers: 37J60, 70F40, 37D10, 70H09

DOI: 10.1134/S156035471604002X

Keywords: nonholonomic dynamics, friction, constraint realization, singular perturbation theory, Lagrange mechanics

1. INTRODUCTION

Nonholonomic dynamics is a classical subject in mechanics that has seen an increase of activity in the last decades with people studying conserved quantities, symmetries and integrability, numerical integrators, as well as various toy problems with intricate behavior — the rattleback and tippe top being the most famous ones — and has many applications in engineering sciences such as robotics, see, e.g., [4, 13].

We consider the fundamental question whether nonholonomic dynamics is realized as the idealization, or limit, of large friction forces. Our main results are twofold. First, we reprove previous results by Brendeleev [8] and Karapetian [20] that answer the question in the affirmative, but we treat the problem in a differential geometric setting, see Theorem 1. Our use of geometric singular perturbation theory provides an improved statement on the convergence of solution curves on infinite time intervals. The examples in Section 3 show that our result in general is sharp. Secondly, we obtain an approximation scheme for the large friction dynamics by an expansion in the singular perturbation parameter. This allows modeling this nonideal, large friction dynamics on the reduced nonholonomic phase space with a modification term added to the nonholonomic system. This has potential applications for simplified modeling of systems where some slippage occurs, e.g., in explaining tippe top inversion [9] as well as in engineering sciences, such as control of robots or submerged vehicles where slipping/drift may become important, see, e.g., [17, 33]. As an example, we apply this approximation scheme to the Chaplygin sleigh, see Section 7.1 and Figs. 7 and 8.

*E-mail: jaap@jaapeldering.nl

1.1. *Realizing Constraints as Idealizations*

Constraints in mechanical dynamical systems should often be seen as simplifying idealizations of more intricate underlying models. Examples are ubiquitous. A pendulum can be modeled as a point mass moving on a circle — or a sphere in 3D — under gravity, but in physical reality, the constraint of staying on the circle is only approximately realized by a very stiff rod that connects the mass to the point of suspension. A rigid body can be considered an approximation of a large number of atoms, or as a continuous medium with a strong interaction potential that keeps the body rigid. Finally, a smooth, convex body rolling on a surface without slipping, or a figure skater sliding without sideways movement are examples of the idealization of friction forces, and are modeled by nonholonomic mechanics.

The examples above show how both holonomically and nonholonomically constrained systems can be considered idealizations of larger, unconstrained systems under strong potentials or friction forces. On the other hand, the dynamics for constrained systems is conventionally postulated to be given by Lagrange’s variational principle (which is equivalent to Newton’s laws in the unconstrained case) together with d’Alembert’s principle that reaction forces do no work along virtual displacements that satisfy the constraints. Thus, a natural question is whether these two different viewpoints yield corresponding and experimentally correct equations of motion for a given system.

In this paper we revisit the question whether friction forces can realize nonholonomically¹⁾ constrained systems. That is, when we consider a mechanical system with friction forces acting in only some of the directions and then scale this friction to infinity, do we recover nonholonomic dynamics in an appropriate limiting sense? This question dates back at least to Carathéodory, who considered it for the Chaplygin sleigh [10], see Section 1.2 below for a brief history.

A second, related question is whether indeed the Lagrange–d’Alembert principle for nonholonomic dynamics is “correct”. The relevance of this question is indicated by the confusion in the late 19th century about the correct formulation of nonholonomic dynamics, see [11, pp. 46–47] and [29, Section 3.2.3] for a discussion. An alternative method to obtain equations of motion for nonholonomically constrained systems, called “vakonomic mechanics”, was proposed by Kozlov [25]. This method produces different dynamics²⁾, and hence raises the question which method is “correct”. The quotes are used since such questions ultimately have to be decided by physical experiment, and the answer may depend on the kind of system studied. For rolling mechanical systems the general consensus seems to be that the d’Alembert principle is correct. Indeed, Lewis and Murray verified experimentally that for a ball rolling on a turntable, nonholonomic dynamics gives a better description than vakonomic mechanics [28]. On the other hand, vakonomic dynamics, satisfying a variational principle, is applicable in control theory and the motion of rigid bodies in fluids [23, 24] as cited in [17]. Alternative methods to realize nonholonomic dynamics besides friction have been proposed too. Bloch and Rojo [7] (see also [5]) used coupling to an external field to obtain nonholonomic dynamics³⁾, while Ruina [32] shows that mechanical systems with intermittent contact, i.e., with piecewise holonomic constraints, can be viewed as nonholonomic systems in the limit as the contact switches more frequently.

Although these “correctness” questions should be decided by experiment, the idea of trying to realize constraint dynamics from the unconstrained system using an underlying principle can offer theoretical insights. For example, whereas nonholonomic dynamics based on d’Alembert’s principle can be viewed as a limit of friction forces, vakonomic mechanics can be viewed as letting the mass in the mechanical metric go to infinity along constrained directions [25]. See also [26] for a discussion of various methods to realize constraints in dynamics and [3, Section 0.3] for a review of the applicability of nonholonomic dynamics to sliding and rolling bodies.

¹⁾The realization of holonomic constraints by strong potentials has been studied in [22, 31, 34] and other papers. In that case the limit is somewhat subtle, and depends on the energy in oscillatory modes normal to the constrained manifold.

²⁾Vershik and Gershkovich showed that in a generic system with nonholonomic constraints the vakonomic and Lagrange–d’Alembert solutions are incompatible [36, Section 4.3].

³⁾Although the nonholonomic system appears as a slow manifold of this infinite-dimensional coupled system, the third line in [7, Eq. (9)] seems to indicate that the fast dynamics of θ around $\alpha(x, t)$ is oscillatory, i.e., the normal directions are elliptic. Thus, convergence of the limit dynamics is not clear to me.

Showing that a fundamental principle such as friction — or large mass terms in the vakonomic principle — can realize the constrained system provides a fundamental justification for the constrained equations of motion as an idealized limit. But conversely, given constrained equations of motion, it also allows one to study under precisely which conditions the underlying principle leads to this realization. For example, in the case of friction realizing nonholonomic dynamics: instead of linear (viscous) friction forces, do Coulomb-like forces also realize nonholonomic dynamics? In [27] this is considered for the specific case of a homogeneous rolling ball. Finally, when the limit is sufficiently well-behaved, e.g., in the case of linear friction, one can moreover calculate correction terms to the idealized dynamics. For the friction limit, these represent the effect of large, but finite friction forces, see Section 7.

1.2. A Brief Historic Overview

To our knowledge, Carathéodory [10] first explicitly addressed the question whether friction forces can realize nonholonomic dynamics, although his starting sentence that “nonholonomic motion in mechanics is known to be caused by friction forces” indicates that the idea is older. His study of the Chaplygin sleigh (without attributing it to Chaplygin [12]) includes a complete solution in quadratures and finally addresses the question whether the dynamics can be realized by a viscous friction force. Carathéodory’s conclusion is negative, but this is due to an error in his reasoning as discussed thoroughly in [18] (see also [30, pp. 233–237]). That is, Carathéodory fixes initial conditions for the velocity perpendicular to the skate blade, but these are not on the slow manifold associated to the nonholonomic dynamics. See also the exposition in Section 4, but note that compared to [10, 18] we retain v , the velocity perpendicular to the sleigh’s skate blade, instead of rewriting it into a second-order equation for ω , the angular velocity of the sleigh.

The realization of nonholonomic constraints by friction forces was independently shown by Brendeleev [8] and Karapetian [20]. They proved that indeed an unconstrained system with viscous friction forces added to it will in the limit converge to the nonholonomic system whose constraint distribution is defined by the friction forces being zero on it. The limit is understood in the sense that solution curves of the unconstrained system converge to solutions of the nonholonomically constrained system, uniformly on each bounded time interval $[t_1, T]$ with $0 < t_1 < T$, as the friction is scaled to infinity. Both authors used Tikhonov’s theorem [35] for singularly perturbed systems to obtain this result in a local coordinate setting. Kozlov combined this friction limit into a larger framework including both friction and inertial terms, obtaining both nonholonomic and vakonomic dynamics as limits in special cases [26].

1.3. Outline and Contributions of the Paper

After the introduction, we first briefly review nonholonomic dynamics according to the Lagrange–d’Alembert principle. In Section 3 we show how different approaches using potentials or friction to realize the same constrained system yield qualitatively different solutions, arbitrarily close to their respective limits. Then we start with the main part of the paper and present how friction forces realize nonholonomic dynamics. First with the Chaplygin sleigh as an example. In Section 4 we derive its equations of motion as a nonholonomic system and in Section 5 as an unconstrained system, but with friction added. Secondly we treat the general theory in Section 6. We show how the Lagrange–d’Alembert equations are recovered as a limit, using modern geometric singular perturbation theory. In keeping the exposition clear, we shall not try to attain the utmost generality. That is, we restrict to Lagrangian systems of mechanical type and consider only time-independent, linear friction forces. This means that we only recover linear nonholonomic constraints⁴). Instead, we shall present various simple examples to illustrate some of the issues discussed above.

The first aim of this paper is to provide an accessible exposition of the problem of realizing nonholonomically constrained systems as a limit of infinite friction forces, using the Chaplygin sleigh as a canonical example. Secondly, we present the problem in a more geometric setting

⁴) Affine constraints do appear naturally, for example, for a ball rolling on a turntable. However, fully nonlinear constraints in mechanical systems seem scarce: according to [30, pp. 213, 223–233] all known examples are variations of the one due to Appell [1]. See also [29] for a discussion of nonholonomic constraint principles, including nonlinear constraints.

than [8, 20]. That is, we express the problem in a differential geometric formulation and use geometric singular perturbation theory, as developed by Fenichel [16] and based on the theory of normally hyperbolic invariant manifolds (abbreviated as NHIM), see, e. g., [14, 15, 19]. This allows us to phrase the results independently of coordinate charts and provide improvements on the results obtained in [8, 20]; in particular, we can describe more precisely in what sense the solution curves converge on positively unbounded time intervals. Moreover, persistence of the NHIM allows us to rigorously study a perturbation expansion of the nonholonomic system away from the infinite friction limit and obtain corrections to the nonholonomic dynamics as an expansion for small $\varepsilon > 0$, where $1/\varepsilon$ is the scale factor multiplying the friction forces. This can be used to describe systems with large, but finite friction forces as nonholonomic systems with small correction terms. We derive general formulas for this expansion and illustrate their use with an analytical and numerical study of the approximation of the Chaplygin sleigh with large friction in Section 7.1. Efforts towards obtaining such a general expansion were made in [37], which studies “creep dynamics” for a few example systems. For the Chaplygin sleigh they find the same first-order correction term (our $h^{(1)}$) to the invariant manifold \mathcal{D}_ε , but their modified equations of motion are not specified.

2. NONHOLONOMIC DYNAMICS

Let us here briefly recall the Lagrange–d’Alembert formalism for nonholonomic dynamics and establish notation. A Lagrangian system is given by a Lagrangian function $L: TQ \rightarrow \mathbb{R}$, where Q is an n -dimensional smooth manifold. We shall assume that the Lagrangian is of mechanical type, i. e., of the form

$$L(q, \dot{q}) = \frac{1}{2}\kappa_q(\dot{q}, \dot{q}) - V(q), \quad (q, \dot{q}) \in TQ, \tag{2.1}$$

where $\frac{1}{2}\kappa_q(\dot{q}, \dot{q})$ is the kinetic energy, which is given by a Riemannian metric κ on Q , and $V: Q \rightarrow \mathbb{R}$ is the potential energy. This implies that L is hyperregular and that the Euler–Lagrange equations of motion have unique solutions. Let q be local coordinates on Q and let (q, \dot{q}) denote induced coordinates on TQ . Then the Euler–Lagrange equations are given by⁵⁾

$$[L]_i := \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0. \tag{2.2}$$

A linear nonholonomic constraint is imposed on this system by specifying a regular distribution, i. e., a smooth, constant rank k vector subbundle, $\mathcal{D} \subset TQ$. Kinematically, this constraint specifies the allowed velocities of the system, that is, a velocity vector $\dot{q} \in T_qQ$ is allowed precisely if $\dot{q} \in \mathcal{D}_q$.

To specify the nonholonomic dynamics with this constraint, an extra principle has to be imposed, as solutions of (2.2) would generally violate the kinematic constraint imposed by \mathcal{D} . To guarantee that the solutions satisfy the constraint, we add a constraint reaction force F_c to the right-hand side of (2.2) according to d’Alembert’s principle, see, e. g., [11, pp. 4–6].

Hypothesis 1 (d’Alembert’s principle). *The constraint forces F_c do no work along movements that are compatible with the constraints. That is, for any $(q, \dot{q}) \in \mathcal{D}$ we have that $F_c(q, \dot{q}) \cdot \dot{q} = 0$.*

From now on we shall implicitly assume this principle for nonholonomically constrained systems. This leads to the Lagrange–d’Alembert equations, which again have unique solutions. Abstractly, these state that a smooth curve $c: (t_0, t_1) \rightarrow Q$ is a solution of the equations if for all $t \in (t_0, t_1)$

$$[L](\ddot{c}(t)) \in \mathcal{D}^0 \quad \text{and} \quad \dot{c}(t) \in \mathcal{D}. \tag{2.3}$$

Here $\mathcal{D}^0 \subset T^*Q$ denotes the annihilator of \mathcal{D} , that is, all covectors $\alpha \in T^*Q$ such that $\alpha(\dot{q}) = 0$ for any $\dot{q} \in \mathcal{D}_{\pi(\alpha)}$.

⁵⁾Note that $[L] := [L]_i dq^i$ can intrinsically be viewed as a map from $J_0^2(\mathbb{R}; Q)$, the space of second-order jets of curves in Q , to T^*Q , covering the identity map on Q .

Let \mathcal{D} locally be given by a set of $n - k$ independent constraint one-forms $\zeta^a = C_i^a(q) dq^i$, i.e., $\mathcal{D} = \{(q, \dot{q}) \in TQ \mid \forall a \in [1, n-k]: \zeta_q^a(\dot{q}) = 0\}$. Then the Lagrange–d’Alembert equations in local coordinates are of the form

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = \lambda_a C_i^a(q) \quad \text{and} \quad \zeta_q^a(\dot{q}) = 0, \quad (2.4)$$

where λ_a are Lagrange multipliers that have to be solved for, and the constraint force is given by $F_c = \lambda_a \zeta^a$.

3. AN EXAMPLE OF DIFFERENT CONSTRAINT REALIZATIONS

The following toy example illustrates how different choices for realizing the same constrained system lead to qualitatively different dynamics close to the limit. These qualitative differences are essentially due to the fact that the limits of sending time to infinity or a singular perturbation parameter to zero do not commute. A nice illustration of this simple fact is given by Arnold [2, p. 65]. Consider a bucket of water with a hole of size ε in the bottom. After any fixed, finite time t the water level remains unchanged as $\varepsilon \rightarrow 0$, but for any $\varepsilon > 0$ the bucket becomes empty as $t \rightarrow \infty$.

We consider the mathematical pendulum in the vertical plane. We view it as a system constrained to the circle in a holonomic, nonholonomic and vakonomic sense. That is, we view the constraint as generated by a strong potential, strong friction forces and large inertia, and compare the dynamics when the respective constraining terms are large but finite.

The unconstrained system is given by the Lagrangian

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) - g y$$

on $T\mathbb{R}^2$, where g is the gravitational acceleration and a unit mass has been chosen.

Viewing the pendulum as a mass attached to the origin by a perfectly stiff rod, we can choose as constraining potential

$$V_\varepsilon(x, y) = \frac{1}{2\varepsilon}(\sqrt{x^2 + y^2} - 1)^2,$$

that is, the square distance from the constrained submanifold $S^1 \subset \mathbb{R}^2$, modeling a perfectly stiff rod as $\varepsilon \rightarrow 0$. We can now easily apply the results from [34] as S^1 is a codimension one manifold and the constraining potential has a constant second derivative in the normal direction. It follows that the limit motion is precisely given by the standard pendulum on S^1 . Energy conservation also makes it immediately clear that for small ε , solutions stay close to the constraint manifold S^1 . See the top image in Fig. 1.

Secondly, the constrained system can be obtained from a “nonholonomic” limit of adding friction forces in the radial direction. This can be thought of as a model for a leaf falling or “swirling down” under gravity where the leaf surface is tangential to circles around the origin and there is a large air resistance against perpendicular movement. We model this air resistance by the friction force defined by the Rayleigh function

$$R_\varepsilon(x, y, \dot{x}, \dot{y}) = \frac{1}{2\varepsilon} \left(\frac{x\dot{x} + y\dot{y}}{\sqrt{x^2 + y^2}} \right)^2,$$

or $R_\varepsilon = \dot{r}^2/(2\varepsilon)$ in polar coordinates. The friction force is given by $F = -\frac{\partial R}{\partial \dot{q}}$. In the limit of $\varepsilon \rightarrow 0$ this gives rise to a nonholonomic system whose constraint distribution is actually integrable; the associated foliation consists of concentric circles. Although the limit dynamics is the same as with the constraining potential, the dynamics for any finite $\varepsilon > 0$ is qualitatively different: the leaf will not stay close to the original submanifold S^1 , but slowly drift down. This can most easily be seen for a leaf with initial conditions $x = \dot{x} = 0$. Under the combined forces of gravity and friction, it will settle to a vertical speed of $\dot{y} = -\varepsilon g$. Hence, over long times the constraint manifold S^1 is not even approximately preserved. See the middle image in Fig. 1.

Finally, we consider the system as a limit of adding a large inertia term in the radial direction. That is, we add to the Lagrangian a term

$$I = \frac{1}{2\varepsilon} \left(\frac{x\dot{x} + y\dot{y}}{\sqrt{x^2 + y^2}} \right)^2,$$

or in polar coordinates $I = \dot{r}^2/(2\varepsilon)$. In the limit of $\varepsilon \rightarrow 0$ this realizes vakonomic dynamics, see [25]. Since the constraint distribution is integrable, the resulting limit dynamics is equal to the nonholonomic limit with friction and the holonomic limit with potential forces on S^1 . The dynamics for finite ε resembles most closely that of the friction model, but there is a difference that can most clearly be observed by considering again the initial conditions $x = \dot{x} = 0$. In this case the equation for y takes the form

$$\left(1 + \frac{1}{\varepsilon} \right) \ddot{y} = -g.$$

Thus, instead of settling on a slow descend, the particle accelerates downward, but with a slowed acceleration due to the added inertial term. See the bottom image in Fig. 1.

The conclusion is that even though for all three constraint realization methods, solutions converge on a fixed time interval to the solutions of the constraint system — which *in this case* is identical for the three methods — this does not hold true on unbounded time intervals. Indeed, for these three different methods, we see three different behaviors when $t \rightarrow \infty$ for fixed $\varepsilon > 0$. With the constraining potential the system stays uniformly⁶⁾ close to the initial leaf S^1 of the constraint distribution $\mathcal{D} = \{\dot{r} = 0\}$. With the friction force, this is not true anymore, but the solution does stay close to \mathcal{D} as a submanifold of TR^2 . This is intuitively clear in the current example as friction prevents \dot{r} from growing large, and can be proven in general under reasonable conditions, see Section 6. Finally, with the inertial term, the pendulum example shows that solutions can even diverge arbitrarily far from \mathcal{D} .

4. THE CHAPLYGIN SLEIGH

The Chaplygin sleigh [12] is a simple, yet interesting example of a mechanical nonholonomic system. The sleigh is a body that can move on the plane, but one of its contact points is a skate, see Fig. 2. Alternatively, one can think of the contact point as a wheel that is fixed to the body. The contact point can only move in the direction along the skate blade/wheel, not in the perpendicular direction. The other two ground contacts can move freely without constraint. The center of mass is located a distance a away from the skate contact point along the line of the blade⁷⁾.

An interesting feature of this nonholonomic system is that it exhibits center-stable and center-unstable relative equilibria, something that cannot occur in purely symplectic⁸⁾ Hamiltonian systems, where eigenvalues must always occur in pairs or quadruples symmetric about the real and imaginary axes.

We shall explicitly recover the equations of motion for this system in the Lagrangian setting⁹⁾ and then illustrate in detail its realization through friction.

We describe the system with coordinates $(x, y) \in \mathbb{R}^2$ for the skate contact point and $\varphi \in S^1$ for the angle of the skate blade with the x axis. Assume that the sleigh has total mass m and moment

⁶⁾However, solutions do not uniformly in time converge to the constrained solutions. Consider as a counterexample the pendulum without gravity and an initial velocity along S^1 : in the case of a finite potential, the particle will oscillate slightly away from S^1 , hence trade some of its kinetic for potential energy and thus have on average a lower angular velocity than the perfectly constrained particle. This builds up over time to arbitrary large errors in position.

⁷⁾An offset of the center of mass perpendicular to the skate blade does not intrinsically change the system, since a shift of the contact point perpendicular to the skate blade does not alter the constraint.

⁸⁾This does not hold anymore on Poisson manifolds; the Chaplygin sleigh can actually be realized with an adapted Poisson bracket $\{u, \omega\} = \frac{a\omega}{I+ma^2}$ with respect to the moving frame coordinates introduced below.

⁹⁾See [11, Chapter 5] for a thorough account how to derive the equations in various different settings, and the end of the chapter for some interesting notes.

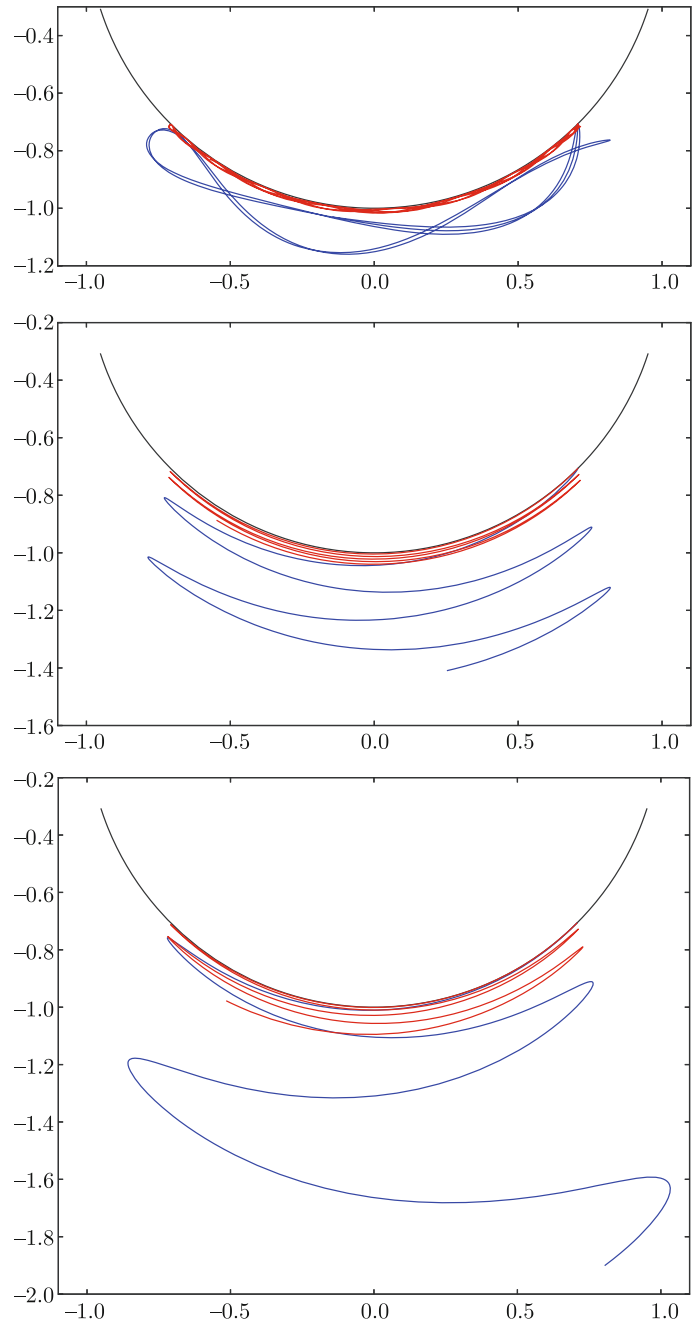


Fig. 1. Top to bottom: numerical simulations of potential, frictional and inertial constraint approximations. The particle initially starts on S^1 at 45° angle and with zero velocity. Its trajectory is shown for 10 seconds with $g = 9.81$ and values $\varepsilon = 0.01$ in blue and $\varepsilon = 0.001$ in red.

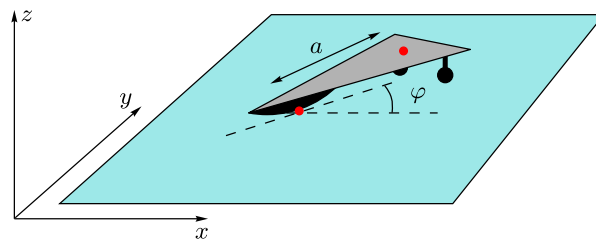


Fig. 2. A Chaplygin sleigh: the left red dot is the point of contact of the skate, the right red dot is the center of mass.

of inertia I about the center of mass. To construct the Lagrangian, we first express the center of mass point as

$$(x_c, y_c) = (x, y) + a(\cos(\varphi), \sin(\varphi)).$$

Then the Lagrangian is given by

$$\begin{aligned} L &= \frac{1}{2}m(\dot{x}_c^2 + \dot{y}_c^2) + \frac{1}{2}I\dot{\varphi}^2 \\ &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{1}{2}(I + ma^2)\dot{\varphi}^2 + ma\dot{\varphi}(-\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi)). \end{aligned} \tag{4.1}$$

We now switch to moving frame coordinates (u, v, ω) , where u and v are the velocities parallel and perpendicular to the skate, respectively, and ω is the angular velocity. They are given by

$$\begin{aligned} u &= \dot{x}\cos(\varphi) + \dot{y}\sin(\varphi), \\ v &= -\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi), \\ \omega &= \dot{\varphi}. \end{aligned} \tag{4.2}$$

This moving frame is aligned with the constraint distribution \mathcal{D} in the sense that the constraint is described by the function

$$\zeta_q(\dot{q}) := v = -\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi),$$

i. e., the constraint $v = 0$ precisely expresses that the skate cannot move sideways.

It is now a straightforward exercise to calculate the Lagrange–d’Alembert equations (2.4). We find for each of the coordinates x, y, φ and their associated velocities:

$$\begin{aligned} \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= m\ddot{x} + ma(-\ddot{\varphi}\sin(\varphi) - \dot{\varphi}^2\cos(\varphi)) = -\lambda\sin(\varphi), \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} &= m\ddot{y} + ma(\ddot{\varphi}\cos(\varphi) - \dot{\varphi}^2\sin(\varphi)) = \lambda\cos(\varphi), \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{\varphi}} - \frac{\partial L}{\partial \varphi} &= (I + ma^2)\ddot{\varphi} + ma\frac{d}{dt}[-\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi)] \\ &\quad + ma\dot{\varphi}(\dot{x}\cos(\varphi) + \dot{y}\sin(\varphi)) = 0. \end{aligned} \tag{4.3}$$

Next, we switch to moving frame coordinates by substituting in (4.2) and taking linear combinations of the first two equations with factors $\cos(\varphi)$ and $\sin(\varphi)$. This yields

$$\begin{aligned} m(\dot{u} - v\omega - a\omega^2) &= 0, \\ m(\dot{v} + u\omega + a\dot{\omega}) &= \lambda, \\ (I + ma^2)\dot{\omega} + ma(\dot{v} + u\omega) &= 0. \end{aligned} \tag{4.4}$$

Finally, we use the constraint $v = 0$, which implies that $\dot{v} = 0$. Inserting this, we obtain the usual Chaplygin sleigh equations

$$\dot{u} = a\omega^2, \quad \dot{\omega} = -\frac{mau\omega}{I + ma^2}, \tag{4.5}$$

with the Lagrange multiplier $\lambda = m(u\omega + a\dot{\omega})$ giving rise to the constraint force

$$F_c = m(u\omega + a\dot{\omega}) \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}.$$

These equations give rise to the phase plot in Fig. 3. A typical trajectory of the skate point of contact is shown in Fig. 4. We see that the solutions are half-ellipses in the (u, ω) -plane starting from the negative u axis and converging onto the positive u axis. When u is positive and $\omega = 0$, then the sleigh is moving in a straight line with the center of mass forward of the skate. This turns out to be the stable solution, while the opposite direction where $u < 0$ is unstable.

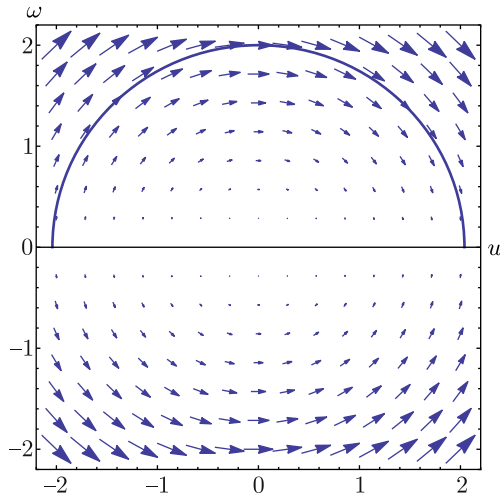


Fig. 3. Phase plot of the (u, ω) coordinates with $m = I = 1$ and $a = 1/5$.

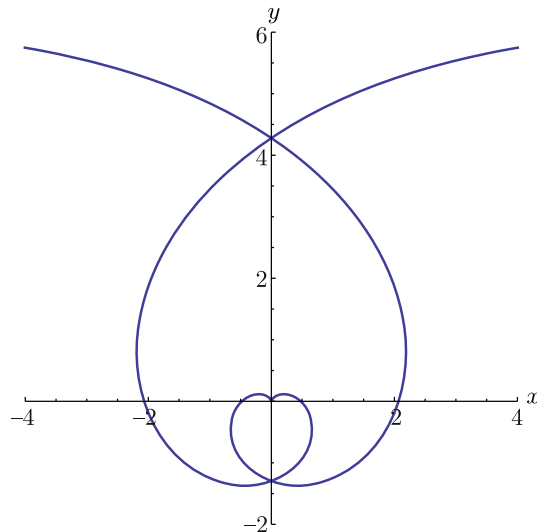


Fig. 4. Trajectory plot associated to the orbit in Fig. 3. The sleigh approaches from the right.

5. THE SLEIGH WITH SLIDING FRICTION

Now we shall obtain the Chaplygin sleigh equations (4.5) in a different way. Instead of enforcing the no-slip constraint $v = 0$ with the constraint reaction force F_c , we consider the same Lagrangian without constraint, but now we add a friction force F_f . Then we scale the friction force to infinity and obtain the nonholonomic Chaplygin sleigh equations in the limit. Note that the friction force F_f is defined also outside the constraint distribution \mathcal{D} — and actually zero on it — while the reaction force F_c is only defined on \mathcal{D} !

This idea can be physically motivated in the following way. A nonholonomic system with a no-slip constraint is a mathematical idealization of a physical system where there is a (very) strong force that prevents the system from going into slip. In this case, we assume that a strong friction force prevents the skate from slipping sideways. By showing that scaling the friction force to infinity leads to nonholonomic dynamics, we provide a fundamental physics argument for the d'Alembert principle of nonholonomic dynamics. We shall first argue heuristically how this limit is obtained, and then make it rigorous using geometric singular perturbation theory. Secondly, we can consider what happens when the skate is turning too quickly at high speed and the friction may not be able to generate the force necessary to (nearly) keep the sleigh from slipping. Singular perturbation theory also provides us with modifications to the nonholonomic dynamics through a series expansion in the scaling parameter.

We start with the same Lagrangian (4.1), but now insert a friction force F_f on the right-hand side. The friction force is supposed to suppress sideways sliding of the skate blade, which is given by the velocity $v = -\dot{x} \sin(\varphi) + \dot{y} \cos(\varphi)$. We take the friction force linear in this slipping velocity and pointing in the opposite direction to dampen it. Thus we have¹⁰⁾

$$F_f = -v \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}.$$

Note that this force acts at the point of contact of the skate, so there is no associated torque around that point, i.e., the ω -component of the force is zero. Secondly, F_f is zero when $v = 0$ (as opposed to F_c), so this finite force does *not* enforce the nonholonomic constraint. We insert F_f/ε into the

¹⁰⁾We could have obtained this friction force more geometrically from a Rayleigh dissipation function $R(q, \dot{q}) = \frac{1}{2} \nu_q(\dot{q}, \dot{q})$, where ν is a family of quadratic forms on TQ with kernel \mathcal{D} . Then F_f is given by minus the fiber derivative of R , that is, $F_f = -\frac{\partial R}{\partial \dot{q}}: TQ/\mathcal{D} \rightarrow \mathcal{D}^0 \subset T^*Q$. Using the moving frame coordinate v to coordinatize TQ/\mathcal{D} , we find $R = \frac{1}{2} \bar{\nu}(q) v^2$. In our example we choose the friction coefficient $\bar{\nu}(q) = 1$.

right-hand side of (4.3), so that we can scale the friction force to infinity by taking the limit $\varepsilon \rightarrow 0$. This yields

$$\begin{aligned} m\ddot{x} + ma(-\ddot{\varphi} \sin(\varphi) - \dot{\varphi}^2 \cos(\varphi)) &= \frac{v}{\varepsilon} \sin(\varphi), \\ m\ddot{y} + ma(\ddot{\varphi} \cos(\varphi) - \dot{\varphi}^2 \sin(\varphi)) &= -\frac{v}{\varepsilon} \cos(\varphi), \\ (I + ma^2)\ddot{\varphi} + ma\frac{d}{dt}[-\dot{x} \sin(\varphi) + \dot{y} \cos(\varphi)] \\ &+ ma\dot{\varphi}(\dot{x} \cos(\varphi) + \dot{y} \sin(\varphi)) = 0. \end{aligned}$$

As in the previous section, we transform these equations to moving frame coordinates, cf. (4.4), and obtain

$$\begin{aligned} \dot{v} - v\omega - a\omega^2 &= 0, \\ \dot{v} + u\omega + a\dot{\omega} &= -\frac{v}{m\varepsilon}, \\ (I + ma^2)\dot{\omega} + ma(\dot{v} + u\omega) &= 0, \end{aligned} \tag{5.1}$$

but now we cannot insert the constraint condition $v = 0$. Instead, we shall analyze the dynamics and see that when $\varepsilon > 0$ is small, then $v(t)$ will quickly converge to near zero. This will imply that the other two equations effectively behave as if the constraint $v = 0$ is active. In other words, there is a slow manifold described by the variables (u, ω) and on that manifold the fast variable v is small and slaved to (u, ω) . Secondly, the friction force naturally takes values in \mathcal{D}^0 , the annihilator of \mathcal{D} . This is preserved in the limit to realize a constraint reaction force according to d’Alembert’s principle 1. All of this indicates that we can expect to recover nonholonomic dynamics.

To obtain a differential equation for \dot{v} , we subtract $\frac{a}{I+ma^2}$ times the third equation from the second to make the term with $\dot{\omega}$ cancel. This gives

$$\frac{I}{I + ma^2}(\dot{v} + u\omega) = -\frac{v}{m\varepsilon}. \tag{5.2}$$

When ε is small, this gives fast exponential decay of v . We shall now assume that the typical rate at which $v(t)$ changes is much faster than that of $u(t)$ and $\omega(t)$. That is, we consider v as a fast variable and u, ω as slow variables. Further conclusions based on this can be made rigorous using singular perturbation theory, but we postpone these arguments to the general theory in Section 6 and focus on obtaining the result.

Thus, we assume in (5.2) that u, ω are approximately constant and obtain as solution

$$v(t) = \left(v(0) + \frac{u\omega}{\rho}\right)e^{-\rho t} - \frac{u\omega}{\rho} \quad \text{with } \rho = \frac{I + ma^2}{Im\varepsilon}.$$

This quickly settles to $v = -\frac{u\omega}{\rho}$, so for the slow dynamics of u, ω we insert this relation, which gives

$$\dot{u} = a\omega^2 + \varepsilon\frac{Im}{I + ma^2}u\omega^2, \quad \dot{\omega} = -\frac{mau\omega}{I + ma^2} + \varepsilon Im^2 a\frac{d}{dt}[u\omega].$$

In these differential equations ε only appears in the numerator and the ε multiplying the time derivative of $(u\omega)$ does not lead to singular equations, so we can sensibly take the limit: we simply insert $\varepsilon = 0$ to arrive at the original Eqs. (4.5) for the nonholonomically constrained Chaplygin sleigh. This is also confirmed by numerical integration of this system: with decreasing values of ε , the trajectories converge to the trajectory of the nonholonomic system, see Fig. 5. It is clearly visible though that the convergence is not uniform in time, as already illustrated in Section 3. In Section 7.1 we will analyze the perturbation to the slow system in more detail.

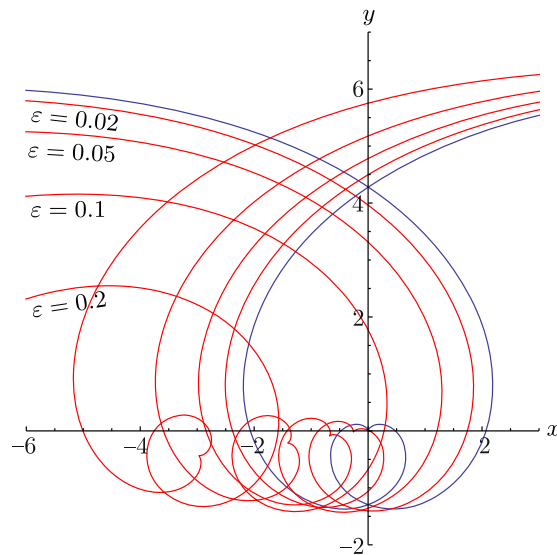


Fig. 5. The red orbits are trajectories of the sleigh with friction with indicated parameter values of ε . These clearly converge to the nonholonomic trajectory in blue.

6. THE GENERAL THEORY

We shall finally prove the statement that linear friction forces realize nonholonomic dynamics in a more general setting. This rephrases the results in [8, 20] in a more geometric formulation, and specifies in what sense the solution curves converge, also on positively unbounded time intervals.

Recall that we consider Lagrangian systems of mechanical type,

$$L(q, \dot{q}) = \frac{1}{2}\kappa_q(\dot{q}, \dot{q}) - V(q), \quad (q, \dot{q}) \in TQ. \quad (2.1)$$

Furthermore, to state our result, we need the concept of a pseudo-solution (or pseudo-orbit) of a dynamical system.

Definition 1 (Pseudo-solution). *Let (M, g) be a smooth Riemannian manifold and let $X \in \mathfrak{X}(M)$ be a C^1 vector field on it. We say that a C^1 curve $x(t) \in M$ is a δ -pseudo-solution of X when*

$$\|\dot{x}(t) - X(x(t))\| \leq \delta \quad \text{for all } t.$$

Furthermore, we consider linear friction forces F that are specified by a Rayleigh function R . That is, given a smooth family of positive semidefinite bilinear forms $\nu_q: T_qQ \times T_qQ \rightarrow \mathbb{R}$, we define the Rayleigh function $R(q, \dot{q}) = \frac{1}{2}\nu_q(\dot{q}, \dot{q})$ and the friction force as its negative fiber derivative,

$$F: TQ \rightarrow T^*Q, \quad F(q, \dot{q}) = -\frac{\partial R}{\partial \dot{q}}(q, \dot{q}) = -\nu_q^b(\dot{q}).$$

Then we have the following result.¹¹⁾

Theorem 1. *Let Q be a compact manifold, let L be a Lagrangian of mechanical type (2.1), and let F be a linear friction force with kernel $\ker(F) = \mathcal{D}$ a regular distribution. Let $X_\varepsilon \in \mathfrak{X}(TQ)$ denote the vector field of the Lagrangian system with Rayleigh friction force F/ε as above, and let $X_{\text{NH}} \in \mathfrak{X}(\mathcal{D})$ denote the vector field of the Lagrange–d’Alembert nonholonomic system (Q, L, \mathcal{D}) .*

Then solution curves of X_ε converge to solutions of X_{NH} as $\varepsilon \rightarrow 0$ in the following sense. Fix an energy level $\bar{E} > \inf V$ and consider initial conditions $x_0 \in TQ$ with energy less than \bar{E} . Then for all $\varepsilon > 0$ sufficiently small we have the following results.

¹¹⁾We shall assume that everything has sufficient smoothness, say, C^r for a large r . Note that the perturbed manifolds \mathcal{D}_ε will generally only have a finite degree of smoothness, see, e.g., [14, Section 1.2.1].

- i. All solution curves $x_\varepsilon(t)$ of X_ε converge at a uniform exponential rate to an invariant manifold¹²⁾ \mathcal{D}_ε that is C^1 -close and diffeomorphic to \mathcal{D} . The manifold \mathcal{D}_ε depends smoothly on ε near zero, with $\mathcal{D}_0 = \mathcal{D}$.
- ii. There exists a family of curves $\bar{x}_\varepsilon(t) \in \mathcal{D}$ that are $\mathcal{O}(\varepsilon)$ -pseudo-solution curves of X_{NH} , such that¹³⁾

$$\forall t_1 > 0 \quad \sup_{t \in [t_1, \infty)} d(x_\varepsilon(t), \bar{x}_\varepsilon(t)) \rightarrow 0$$

as $\varepsilon \rightarrow 0$, uniformly in x_0 .

The compactness condition of Q can be replaced by uniformity assumptions on all of Q, L, F, \mathcal{D} . One crucial assumption that has to be added is that V is bounded from below. Moreover, one should require (Q, κ) to have bounded geometry and dV, F , and \mathcal{D} to be uniformly C^1 bounded and F also bounded away from zero. This is needed if one wants to apply normal hyperbolicity theory in a noncompact context, see [14, Theorem 3.1]. In the Chaplygin sleigh example $Q = \mathbb{R}^2 \times S^1$ is not compact, but this generalization does apply: the system is symmetric under the Euclidean group $\text{SE}(\mathbb{R}^2)$ acting transitively on Q , hence (Q, κ) has bounded geometry (see [14, Example 2.3]) and the remaining uniformity conditions are fulfilled as well. Another viewpoint is to compactify \mathbb{R}^2 to the two-torus, or to consider the closely related Suslov problem of a nonholonomically constrained rigid body, so that Theorem 1 does apply.

Alternatively, one can consider the convergence of solution curves only on finite time intervals, which effectively allows one to restrict to a compact set covering the nonholonomic solution curve. This is essentially the setting of the results in [8, 20], which are given in local coordinates. It does not automatically yield uniformity of convergence of solution curves (even on finite time intervals) when Q is noncompact though, as illustrated by the following example.

Let $Q = \mathbb{R}^2$ with the Euclidean metric and consider a particle of unit mass under the potential $V(x, y) = -x^2$. Let the constraint distribution be

$$\mathcal{D} = \text{span } \chi(x - y)\partial_x + (1 - \chi(x - y))\partial_y,$$

where $\chi \in C^\infty(\mathbb{R}; [0, 1])$ such that $\chi(x) = 0$ for $x \leq 0$ and $\chi(x) = 1$ for $x \geq 1$. Finally, let the friction force F be arbitrary, but with kernel \mathcal{D} and uniformly bounded. Now consider the unconstrained system with friction and initial conditions $x(0) > 0$ small, $y(0) \geq 1$ arbitrarily large and $\dot{x}(0) = \dot{y}(0) = 0$. Then the particle will first accelerate in the positive x direction due to the potential; once it hits the line $x = y$, the friction force kicks in to make the particle follow the changing constraint direction of \mathcal{D} . However, for any fixed $\varepsilon > 0$, the particle’s velocity component \dot{x} will overshoot the constraint $\mathcal{D} = \text{span } \partial_y$ in the region $x \geq y + 1$ by an arbitrary amount as the initial condition $y(0)$ is made large, since \dot{x} then is arbitrary large at “impact” with the changing constraint. Even worse, since the potential force increases with x , even in the region $x \geq y + 1$, the velocity component \dot{x} will diverge away from zero, that is, from \mathcal{D} .

Proof (of Theorem 1). The system with friction described by X_ε is given by the second-order equations

$$\kappa^b(\nabla_{\dot{q}}\dot{q}) = -dV(q) + \frac{1}{\varepsilon}F(q, \dot{q}), \tag{6.1}$$

or, in explicit local coordinates, it is expressed as

$$\dot{q}^i = v^i, \quad \dot{v}^i = -\Gamma_{jk}^i v^j v^k + \kappa^{ij} \left(-\frac{\partial V}{\partial q^j} + \frac{1}{\varepsilon} F_j \right). \tag{6.2}$$

¹²⁾The manifold \mathcal{D}_ε is unique up to the choice of a cutoff function at the boundary of the energy \bar{E} in $\text{T}Q$. Different choices lead to differences of order ε that decay exponentially away from the boundary. However, the limit to the reduced vector field X_{NH} will not depend on this choice.

¹³⁾Here d is some appropriate distance on $\text{T}Q$, for example, the distance induced by the Sasaki metric coming from (Q, κ) .

The limit of the vector field X_ε is singular, so we switch to a rescaled version $Y_\varepsilon = \varepsilon X_\varepsilon$. This can be viewed as changing to a fast time variable $\tau = t/\varepsilon$ (whose derivative we denote by a prime) for any $\varepsilon > 0$, but the vector field Y_0 is also well-defined and given in local coordinates as

$$q'^i = 0, \quad v'^i = \kappa^{ij} F_j = -\kappa^{ij} \nu_{jk} v^k.$$

Both κ^{ij} and ν_{jk} are (semi)positive definite matrices with $\ker(\nu_q) = \mathcal{D}_q$. This shows that the smooth manifold $\mathcal{D} \subset \text{T}Q$ is the fixed point set of Y_0 and that Y_0 is linear in the normal direction, and the linear term given by

$$-\kappa^\# \circ \nu^\flat|_{\text{T}Q/\mathcal{D}}$$

has strictly negative eigenvalues when we identify $\text{T}Q/\mathcal{D} \cong \mathcal{D}^\perp$. Thus, \mathcal{D} is a normally hyperbolic invariant manifold (NHIM) for Y_0 and it is globally exponentially attractive.

We now restrict attention to the compact subset

$$\mathcal{E} = \{(q, v) \in \text{T}Q \mid E(q, v) \leq \bar{E}\}$$

below the energy level \bar{E} , which is nonempty when $\bar{E} > \inf V$. The dynamics of X_ε (and thus also of Y_ε) is dissipative, and hence leaves \mathcal{E} forward invariant. This implies that we can use a cutoff argument¹⁴⁾ outside \mathcal{E} which will not affect forward solutions with initial conditions inside \mathcal{E} . Using the persistence of NHIMs (see [14, 15, 19]), we find that for $\varepsilon > 0$ sufficiently small there is a unique perturbed manifold \mathcal{D}_ε that is C^1 -close and diffeomorphic to \mathcal{D} and invariant under Y_ε . Moreover, \mathcal{D}_ε depends smoothly on ε and is a NHIM again with perturbed stable fibration. In particular, \mathcal{D}_ε is locally exponentially attractive, while the exponential attraction globally inside \mathcal{E} follows from the uniform smooth dependence of flows on parameters over compact domains and time intervals, see, e. g., [14, Theorem A.6]. Hence the solution curves $\tilde{x}_\varepsilon(\tau)$ of Y_ε converge at a uniform exponential rate to \mathcal{D}_ε , and thus the solution curves $x_\varepsilon(t) = \tilde{x}_\varepsilon(t/\varepsilon)$ of X_ε do so too, and at increasing rates as $\varepsilon \rightarrow 0$. This proves the first part of the claims.

To prove the remaining claims, we need to analyze the dynamics of Y_ε restricted to the invariant manifold \mathcal{D}_ε . Let us denote by

$$\text{pr}: \text{T}Q \rightarrow \mathcal{D} \tag{6.3}$$

the projection along \mathcal{D}^\perp , that is, orthogonal according to the kinetic energy metric κ . Note that (6.3) can be viewed as a vector bundle where we forget about the linear structure on \mathcal{D} . The Lagrange–d’Alembert equations (2.3) can then be expressed as

$$\text{pr} \circ [L](\ddot{q}(t)) = 0 \quad \text{with } \dot{q}(t) \in \mathcal{D}.$$

Furthermore, let $f: U \times \mathbb{R}^n \rightarrow \text{T}Q$ be a frame on an open set $U \subset Q$, such that the first k components span \mathcal{D} and the remaining $n-k$ span \mathcal{D}^\perp . Let $\omega \in \Omega^1(Q; \text{End}(\mathbb{R}^n))$ denote the connection one-form with respect to the frame f and let $(\xi, \eta) \in \mathbb{R}^n$ denote coordinates associated to this frame, where $\xi \in \mathbb{R}^k$ and $\eta \in \mathbb{R}^{n-k}$, respectively, coordinatize the fibers of \mathcal{D} and \mathcal{D}^\perp . In these frame coordinates, the projection pr simply maps (ξ, η) to $(\xi, 0)$. Let $\text{pr}_\xi: \mathbb{R}^n \rightarrow \mathbb{R}^k$ denote this projection, $\text{pr}_\eta: \mathbb{R}^n \rightarrow \mathbb{R}^{n-k}$ its complementary projection. See Appendix for a brief overview of using a connection one-form to express equations of motion in a moving frame and its relation to using structure functions.

In frame coordinates the Lagrange–d’Alembert equations are then

$$X_{\text{NH}} \iff \begin{cases} \dot{q} = f_q \cdot (\xi, 0), \\ \dot{\xi} = \text{pr}_\xi \left[-\omega(\dot{q}) \cdot (\xi, 0) - f_q^{-1} \cdot \kappa_q^\# \cdot dV \right], \end{cases} \tag{6.4}$$

since $\dot{q} \in \mathcal{D}_q$ is equivalent to $\eta = 0$. This is the vector field X_{NH} on \mathcal{D} .

¹⁴⁾Without going into full details, we modify Y_ε at the boundary of \mathcal{E} such that $\mathcal{D} \cap \mathcal{E}$ is an overflowing invariant (see [15]) for the modified vector field.

Next, we show that there is a well-defined limit of the family of vector fields

$$X_\varepsilon|_{\mathcal{D}_\varepsilon} \in \mathfrak{X}(\mathcal{D}_\varepsilon)$$

and that the limit is precisely X_{NH} . To make sense of this limit, however, we first have to re-express these vector fields on a fixed manifold, \mathcal{D} . We do this using the projection $\text{pr}: \text{T}Q \rightarrow \mathcal{D}$. The invariant manifolds are C^1 -close and diffeomorphic to \mathcal{D} , so they can be written as the graphs of functions $h_\varepsilon: \mathcal{D} \rightarrow \mathcal{D}^\perp$ covering the identity in Q or, in other words, as graphs of sections $h_\varepsilon \in \Gamma(\text{pr}: \text{T}Q \rightarrow \mathcal{D})$. That is, in frame coordinates we have

$$\mathcal{D}_\varepsilon = \{ \eta = h_\varepsilon(q, \xi) \}.$$

Since \mathcal{D}_ε depends smoothly on ε , we can consider a Taylor expansion of h_ε . We denote this as

$$h_\varepsilon(q, \xi) = \sum_{i=0}^r \frac{\varepsilon^i}{i!} h^{(i)}(q, \xi) + \mathcal{O}(\varepsilon^{r+1}). \tag{6.5}$$

Note that $h^{(0)}(q, \xi) = h_0(q, \xi) = 0$ since $\mathcal{D}_0 = \mathcal{D}$. We now consider Y_ε in frame coordinates, cf. (6.1) and (6.4). Let us denote by $(\kappa_q^\# \nu_q^b)_f$ the linear operator $\kappa_q^\# \nu_q^b: \mathcal{D}_q^\perp \rightarrow \mathcal{D}_q^\perp$ with respect to the frame f , then

$$Y_\varepsilon \iff \begin{cases} q' = \varepsilon f_q \cdot (\xi, \eta), \\ \xi' = \varepsilon \text{pr}_\xi \left[-\omega(f_q \cdot (\xi, \eta)) \cdot (\xi, \eta) - f_q^{-1} \cdot \kappa_q^\# \cdot dV \right], \\ \eta' = \varepsilon \text{pr}_\eta \left[-\omega(f_q \cdot (\xi, \eta)) \cdot (\xi, \eta) - f_q^{-1} \cdot \kappa_q^\# \cdot dV \right] - (\kappa_q^\# \nu_q^b)_f \cdot \eta. \end{cases}$$

Note that $\omega(f_q \cdot \cdot)$ are the connection coefficients as defined in Appendix, but written without indices. Since Y_ε leaves \mathcal{D}_ε invariant, we can consider the restriction $Y_\varepsilon|_{\mathcal{D}_\varepsilon}$. Secondly, $\text{pr}: \mathcal{D}_\varepsilon \rightarrow \mathcal{D}$ is a diffeomorphism (its inverse is the map $\text{Id}_\mathcal{D} + h_\varepsilon$), so we can push forward the vector field $Y_\varepsilon|_{\mathcal{D}_\varepsilon}$ along pr to

$$\tilde{Y}_\varepsilon := \text{pr}_*(Y_\varepsilon|_{\mathcal{D}_\varepsilon}) \in \mathfrak{X}(\mathcal{D}).$$

In frame coordinates this amounts to projecting the vector field onto the coordinates (q, ξ) , while inserting $\eta = h_\varepsilon(q, \xi)$, see Fig. 6. That is, in frame coordinates we have

$$\tilde{Y}_\varepsilon \iff \begin{cases} q' = \varepsilon f_q \cdot (\xi, h_\varepsilon(q, \xi)), \\ \xi' = \varepsilon \text{pr}_\xi \left[-\omega(f_q \cdot (\xi, h_\varepsilon(q, \xi))) \cdot (\xi, h_\varepsilon(q, \xi)) - f_q^{-1} \cdot \kappa_q^\# \cdot dV \right]. \end{cases} \tag{6.6}$$

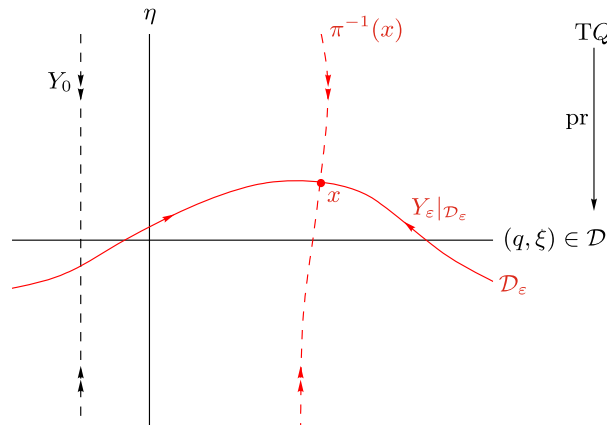


Fig. 6. The invariant manifolds \mathcal{D}_ε and vector fields living on them in frame coordinates (q, ξ, η) . The dashed lines indicate stable fibers.

Now we note that \tilde{Y}_ε is of order ε , so this vector field can be rescaled back to the original slow time $t = \varepsilon\tau$, which we denote as $\tilde{X}_\varepsilon = \frac{1}{\varepsilon}\tilde{Y}_\varepsilon$. Then we can consider its limit

$$\tilde{X}_0 := \lim_{\varepsilon \rightarrow 0} \tilde{X}_\varepsilon \in \mathfrak{X}(\mathcal{D})$$

and we find $\tilde{X}_0 = X_{\text{NH}}$, that is, the Lagrange–d’Alembert equations: we have $h_\varepsilon(q, \xi) = \varepsilon h^{(1)}(q, \xi) + \mathcal{O}(\varepsilon^2)$, so after dividing (6.6) by ε and inserting $\eta = h_\varepsilon(q, \xi)$, the only terms remaining in the limit are those with $\eta = 0$. This yields exactly the nonholonomic vector field X_{NH} as in (6.4).

Using this limit, we can finally prove the last dynamical statement that solution curves $x_\varepsilon(t)$ of X_ε converge to curves $\bar{x}_\varepsilon(t) \in \mathcal{D}$ that are uniformly $\mathcal{O}(\varepsilon)$ -pseudo-solutions of X_{NH} . Using the persistence of normal hyperbolicity, we have already proved that $x_\varepsilon(t)$ converges to the invariant manifold \mathcal{D}_ε uniformly in x_0 . Moreover, \mathcal{D}_ε has a stable fibration whose projection commutes with the flow of X_ε . That is, there exists a (nonlinear) fibration $\pi: \text{T}Q \rightarrow \mathcal{D}_\varepsilon$ such that $\pi(\Phi_\varepsilon^t(x_0)) = \Phi_\varepsilon^t(\pi(x_0))$. Each single fiber $\pi^{-1}(x)$ is not invariant, but the flow does map fibers onto fibers and is exponentially contracting along them. Hence, $x_\varepsilon(t)$ projects to a solution curve $\pi(x_\varepsilon(t)) \in \mathcal{D}_\varepsilon$ to which it converges as $t \rightarrow \infty$. Now define $\bar{x}_\varepsilon(t) := \text{pr}(\pi(x_\varepsilon(t))) \in \mathcal{D}$, which is by construction a solution of \tilde{X}_ε . Since $\tilde{X}_\varepsilon - X_{\text{NH}} \in \mathcal{O}(\varepsilon)$ and $\|h_\varepsilon\|_{\text{sup}} \in \mathcal{O}(\varepsilon)$ uniformly on the compact set \mathcal{E} , it follows that $\bar{x}_\varepsilon(t)$ is a $\mathcal{O}(\varepsilon)$ -pseudo-solution of X_{NH} and that $\pi(x_\varepsilon)$ and \bar{x}_ε are $\mathcal{O}(\varepsilon)$ -close uniformly in time. Finally, since the fast-time- τ solutions $\tilde{x}_\varepsilon(\tau)$ already converge to \mathcal{D}_ε at a fixed exponential rate uniformly in $x_0 \in \mathcal{E}$, there exists a $\tau_1 > 0$ such that $d(\tilde{x}_\varepsilon(\tau), \pi(\tilde{x}_\varepsilon(\tau))) < \varepsilon$ for all $\tau \geq \tau_1$. Now consider $t_1 > 0$ arbitrary. Since $x_\varepsilon(t) = \tilde{x}_\varepsilon(t/\varepsilon)$, we see that when $\varepsilon \leq t_1/\tau_1$, also $d(x_\varepsilon(t), \pi(x_\varepsilon(t))) < \varepsilon$ for all $t \geq t_1$, and finally

$$d(x_\varepsilon(t), \bar{x}_\varepsilon(t)) \leq d(x_\varepsilon(t), \pi(x_\varepsilon(t))) + d(\pi(x_\varepsilon(t)), \bar{x}_\varepsilon(t)) \in \mathcal{O}(\varepsilon).$$

This completes the proof of the second claim. □

7. BEYOND THE LIMIT OF NONHOLONOMIC DYNAMICS

The Lagrange–d’Alembert equations were obtained as zeroth-order term in the expansion of \tilde{X}_ε in ε . However, one can continue and inductively find higher-order terms in the expansion of \tilde{X}_ε . These terms correspond to effects of large, but finite friction forces and will, for example, contribute to drift normal to the nonholonomic constraint and energy dissipation. The advantage of studying large, but finite friction in this context is that \tilde{X}_ε is still a vector field on the lower-dimensional nonholonomic phase space \mathcal{D} as compared to the full Euler–Lagrange equations on $\text{T}Q$, while normal hyperbolicity guarantees that this is a proper Taylor expansion of a truly invariant and attractive subsystem. We shall here show how these higher terms can be obtained and calculate the first-order correction term.

The “master equation” for obtaining the expansion is derived from the coordinate expression for the invariant manifold \mathcal{D}_ε , given by $\eta = h_\varepsilon(q, x)$. Let $f_\varepsilon(q, \xi, \eta)$ denote the “horizontal” components of Y_ε for (q', ξ') and let $g_\varepsilon(q, \xi, \eta)$ describe the “vertical” component for η' . Taking a fast-time- τ derivative of the invariant manifold equations, we find

$$\begin{aligned} \eta' &= \text{D}h_\varepsilon(q, \xi) \cdot (q', \xi') \\ \iff g_\varepsilon(q, \xi, h_\varepsilon(q, \xi)) &= \text{D}h_\varepsilon(q, \xi) \cdot f_\varepsilon(q, \xi, h_\varepsilon(q, \xi)). \end{aligned} \tag{7.1}$$

Now we can expand this equation in powers of ε to obtain inductively the functions $h^{(i)}(q, \xi)$. Before starting this, let us note the following: using the same expansion as in (6.5) for f_ε and g_ε , we have for all (q, ξ, η) that

$$\begin{aligned} f^{(0)}(q, \xi, \eta) &= 0, & h^{(0)}(q, \xi) &= 0, \\ g^{(0)}(q, \xi, 0) &= 0, & \text{D}_3g^{(0)}(q, \xi, 0) &= -(\kappa_q^\# \nu_q^b)_f \end{aligned}$$

and $\text{D}_3g^{(0)}(q, \xi, 0)$ is invertible as a linear map on \mathbb{R}^{n-k} . Expanding (7.1) up to first order in ε then leads to

$$g^{(0)}(q, \xi, h^{(0)}(q, \xi)) = \text{D}h^{(0)}(q, \xi) \cdot f^{(0)}(q, \xi, h^{(0)}(q, \xi))$$

at zeroth order, and is trivially satisfied as it reduces to $0 = 0$. At first order we find

$$\begin{aligned}
 g^{(1)}(q, \xi, h^{(0)}(q, \xi)) + D_3 g^{(0)}(q, \xi, h^{(0)}(q, \xi)) \cdot h^{(1)}(q, \xi) \\
 = Dh^{(1)}(q, \xi) \cdot f^{(0)}(q, \xi, h^{(0)}(q, \xi)) \\
 + Dh^{(0)}(q, \xi) \cdot \left[f^{(1)}(q, \xi, h^{(0)}(q, \xi)) + D_3 f^{(0)}(q, \xi, h^{(0)}(q, \xi)) \cdot h^{(1)}(q, \xi) \right],
 \end{aligned}$$

which greatly simplifies to

$$g^{(1)}(q, \xi, 0) + D_3 g^{(0)}(q, \xi, 0) \cdot h^{(1)}(q, \xi) = 0.$$

Using the invertibility of $D_3 g^{(0)}$, this can be solved for $h^{(1)}$ and gives

$$\begin{aligned}
 h^{(1)}(q, \xi) &= -[D_3 g^{(0)}(q, \xi, 0)]^{-1} \cdot g^{(1)}(q, \xi, 0) \\
 &= (\nu_q^\# \kappa_q^b)_f \cdot \text{pr}_\eta \cdot [-\omega(f_q \cdot \xi) \cdot \xi - f_q^{-1} \cdot \kappa_q^\# \cdot dV],
 \end{aligned} \tag{7.2}$$

where we consider ν restricted to \mathcal{D} . The term $(\kappa_q^b)_f \cdot \text{pr}_\eta \cdot [-\omega(f_q \cdot \xi) \cdot \xi - f_q^{-1} \cdot \kappa_q^\# \cdot dV]$ can be identified as minus the reaction force $F_c: \mathcal{D} \rightarrow \mathcal{D}^0$ in the nonholonomic constraint picture, while $\eta = h^{(1)}(q, \xi)$ is the drift velocity violating the nonholonomic constraint that is necessary to generate the same force as F_c , but now due to the linear friction F_f and up to order ε .

To finally obtain the first-order expansion of \tilde{X}_ε , we have to expand $f_\varepsilon(q, \xi, h_\varepsilon(q, \xi))$ up to order two. Here we find

$$\begin{aligned}
 \frac{1}{2} \left(\frac{d}{d\varepsilon} \right)^2 \Big|_{\varepsilon=0} f_\varepsilon(q, \xi, h_\varepsilon(q, \xi)) &= \frac{1}{2} f^{(2)}(q, \xi, 0) + D_3 f^{(1)}(q, \xi, 0) \cdot h^{(1)}(q, \xi) \\
 &\quad + \frac{1}{2} D_3 f^{(0)}(q, \xi, 0) \cdot h^{(2)}(q, \xi) + \frac{1}{2} D_3^2 f^{(0)}(q, \xi, 0) \cdot h^{(1)}(q, \xi)^2 \\
 &= D_3 f^{(1)}(q, \xi, 0) \cdot h^{(1)}(q, \xi),
 \end{aligned} \tag{7.3}$$

noting that also $f^{(2)} \equiv 0$. From (6.6) we read off that

$$D_3 f^{(1)}(q, \xi, 0) \cdot \eta = \left(f_q \cdot \eta, -\text{pr}_\xi [\omega(f_q \cdot \eta) \cdot \xi + \omega(f_q \cdot \xi) \cdot \eta] \right).$$

Putting everything together, we obtain $\tilde{X}_\varepsilon = X_{\text{NH}} + \varepsilon \tilde{X}^{(1)} + \mathcal{O}(\varepsilon^2)$ with

$$\tilde{X}^{(1)} \iff \begin{cases} \dot{q} = f_q \cdot h^{(1)}(q, \xi) \\ \dot{\xi} = -\text{pr}_\xi [\omega(f_q \cdot h^{(1)}(q, \xi)) \cdot \xi + \omega(f_q \cdot \xi) \cdot h^{(1)}(q, \xi)]. \end{cases} \tag{7.4}$$

Note that with this correction term, \tilde{X}_ε is not a second-order vector field on \mathcal{D} anymore since the drift velocity $h^{(1)}(q, \xi)$ violates the nonholonomic constraint \mathcal{D} .

7.1. The Chaplygin Sleigh

We shall now follow this recipe to obtain a first-order perturbation for the nonholonomic dynamics of the Chaplygin sleigh, based on the friction model used in Section 5. We cannot apply the theory right away since the coordinates u, v, ω do not correspond to an orthogonal frame. This can be seen from the Lagrangian

$$L = \frac{1}{2} m(u^2 + v^2) + \frac{1}{2} (I + ma^2) \omega^2 + ma\omega v$$

with respect to these coordinates: it is purely kinetic, but the metric is not diagonal with respect to u, v, ω . We replace ω by a new coordinate $\psi = \omega + \frac{ma}{I+ma^2} v$ which diagonalizes the metric. That is, the coordinates u, v, ψ correspond to an orthogonal frame and we have

$$L = \frac{1}{2} m u^2 + \frac{1}{2} \frac{I m}{I + ma^2} v^2 + \frac{1}{2} (I + ma^2) \psi^2.$$

With respect to these coordinates the equations of motion (5.1) become

$$\begin{aligned} \dot{u} + \frac{ma^2 - I}{I + ma^2}v\psi + \frac{maI}{(I + ma^2)^2}v^2 - a\psi^2 &= 0, \\ \dot{v} + u\psi - \frac{ma}{I + ma^2}uv &= -\frac{I + ma^2}{\varepsilon m I}v, \\ \dot{\psi} + \frac{ma}{I + ma^2}u\psi - \frac{m^2a^2}{(I + ma^2)^2}uv &= 0. \end{aligned} \tag{7.5}$$

Note that the terms without time derivatives on the left-hand side are exactly those arising from the connection coefficients, see Appendix. The friction force only appears in the equation for v since the frame component associated to v spans \mathcal{D}^\perp , while F_f takes values in \mathcal{D}^0 and hence the term $\kappa^\sharp \cdot F_f$, see (6.2), takes values in \mathcal{D}^\perp .

To obtain the first-order vector field (7.4), we first have to recover $h^{(1)}(x, y, \varphi, u, \psi) = h^{(1)}(u, \psi)$ from (7.2). We note that $\xi = (u, 0, \psi)$, $\eta = (0, v, 0)$, $V \equiv 0$, that $\omega(f_q \cdot)$ are precisely the connection coefficients, and finally that $(\nu_q^\sharp k_q^\flat)_f$ acting on η is given by

$$(\nu_q^\sharp k_q^\flat)_f = 1 \cdot \frac{Im}{I + ma^2}.$$

This yields

$$h^{(1)}(u, \psi) = -\frac{Im}{I + ma^2}u\psi$$

since $u\psi$ is the only term in the v -component of the connection one-form in (7.5) that survives when we insert ξ , i.e., $v = 0$. Then we recover for $\tilde{X}^{(1)}$:

$$\begin{aligned} (\dot{x}, \dot{y}) &= -\frac{Im}{I + ma^2}u\psi(-\sin(\varphi), \cos(\varphi)), & \dot{\varphi} &= -\frac{Im^2a}{(I + ma^2)^2}u\psi, \\ \dot{u} &= \frac{Im(ma^2 - I)}{(I + ma^2)^2}u\psi^2, & \dot{\psi} &= -\frac{Im^3a^2}{(I + ma^2)^3}u^2\psi. \end{aligned}$$

Here we used that $\dot{\varphi} = \omega = \psi - \frac{ma}{I + ma^2}v$ and that $\psi = \omega$ on \mathcal{D} .

Now we can numerically integrate the vector field $X_{\text{NH}} + \varepsilon\tilde{X}^{(1)}$. This should, on \mathcal{D} , be a first-order approximation of the singularly perturbed vector field X_ε with friction. Indeed, the phase and trajectory plots in Figs. 7 and 8 clearly show that the green curves of the first-order expansion $X_{\text{NH}} + \varepsilon\tilde{X}^{(1)}$ converge more than linearly in ε to the red curves of the singularly perturbed system X_ε .

APPENDIX. CONNECTION FORM VS. STRUCTURE FUNCTIONS

In this appendix we briefly recall the connection one-form as a method to express (geodesic) equations of motion with respect to a moving frame. We also relate this to the formulation using the structure functions associated to the frame in terms of the Lie brackets of the vector fields spanning the moving frame. The latter formulation is more common in nonholonomic dynamics and known as Hamel’s formalism, see, e. g., [6].

Let $f: Q \times \mathbb{R}^n \rightarrow \text{T}Q$ be a (local) moving frame, where $f_\alpha \in \mathfrak{X}(Q)$ denote the individual vector fields spanning the frame and $f_\alpha = f_\alpha^i \partial_i$ their decomposition with respect to a basis induced by local coordinates q^i . Let $\lambda: \text{T}Q \rightarrow Q \times \mathbb{R}^n$ denote the inverse of f , that is, a (local) trivialization of $\text{T}Q$. Roman indices are used for induced coordinates; Greek ones to index moving frame coordinates. Now a connection ∇ on Q can be expressed with respect to the frame f as

$$\nabla_{f_\gamma} f_\beta = \omega(f_\gamma)_\beta^\alpha f_\alpha = \omega_{\beta\gamma}^\alpha f_\alpha,$$

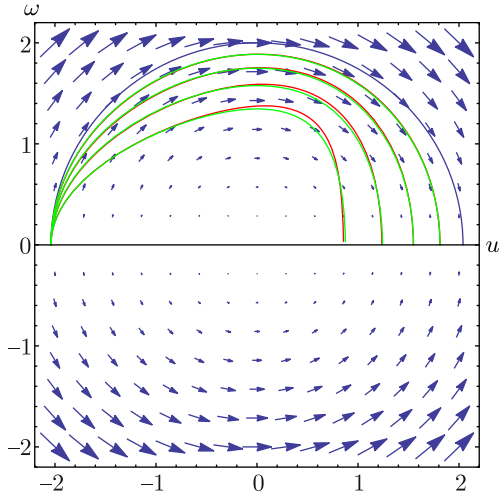


Fig. 7. Phase plot of the (u, ω) coordinates with the nonholonomic system in blue, the system with friction in red and the first-order approximation in green.

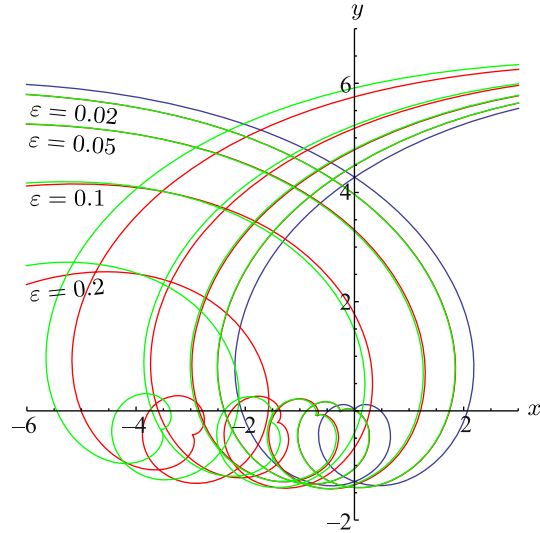


Fig. 8. Trajectory plots associated to the orbits in Fig. 7.

where $\omega \in \Omega^1(Q; \text{End}(\mathbb{R}^n))$ is the connection one-form and $\omega_{\beta\gamma}^\alpha$ are its coefficients. On the other hand, the structure functions $C_{\beta\gamma}^\alpha$ encode the Lie brackets relative to a frame as follows:

$$[f_\beta, f_\gamma] = C_{\beta\gamma}^\alpha f_\alpha.$$

Let ∇ be the Levi-Civita connection associated to a metric κ on Q . According to [21, Proposition 2.3] we have

$$2\kappa(\nabla_X Y, Z) = X \cdot \kappa(Y, Z) + Y \cdot \kappa(X, Z) - Z \cdot \kappa(X, Y) + \kappa([X, Y], Z) + \kappa([Z, X], Y) + \kappa([Z, Y], X)$$

for any vector fields $X, Y, Z \in \mathfrak{X}(Q)$. When we decompose the vector fields with respect to the frame f , i. e., write $X = X^\alpha f_\alpha$, this yields the following relation between the connection coefficients, the metric, and the structure functions:

$$\omega_{\gamma\delta}^\eta = \frac{1}{2}\kappa^{\eta\beta}(f_\delta(\kappa_{\gamma\beta}) + f_\gamma(\kappa_{\delta\beta}) - f_\beta(\kappa_{\delta\gamma})) + \frac{1}{2}(C_{\delta\gamma}^\eta + \kappa_{\alpha\gamma}\kappa^{\eta\beta}C_{\beta\delta}^\alpha - \kappa_{\alpha\delta}\kappa^{\eta\beta}C_{\beta\gamma}^\alpha). \tag{A.1}$$

Note that when f is a holonomic, coordinate-induced frame, then the structure constants are zero and the first terms reduce to the usual Christoffel symbols. Conversely, the torsion-freeness of ∇ , i.e., $[X, Y] = \nabla_X Y - \nabla_Y X$ for any $X, Y \in \mathfrak{X}(Q)$, implies that

$$C_{\beta\gamma}^\alpha = \omega_{\gamma\beta}^\alpha - \omega_{\beta\gamma}^\alpha.$$

We return to the formulation of Lagrangian mechanics. For simplicity we consider a purely kinetic¹⁵⁾ Lagrangian $L(q, \dot{q}) = \frac{1}{2}\kappa_q(\dot{q}, \dot{q})$, so its Euler–Lagrange equations correspond to the geodesic equations $\nabla_{\dot{q}}\dot{q} = 0$, or in induced coordinates, $\ddot{q}^i = -\Gamma_{jk}^i \dot{q}^j \dot{q}^k$, where Γ are the Christoffel symbols of the metric κ . Denoting the frame coordinates by v^α (these are also called quasi-velocities), we have

$$\dot{v}^\alpha = -\omega_{\beta\gamma}^\alpha v^\beta v^\gamma. \tag{A.2}$$

The coefficients $\omega_{\beta\gamma}^\alpha$ play the same role as the Christoffel symbols Γ_{jk}^i , but with respect to the moving frame f . They are related by

$$\omega_{\beta\gamma}^\alpha = \lambda_i^\alpha \Gamma_{jk}^i f_\beta^j f_\gamma^k + \lambda_i^\alpha (\partial_k f_\beta^i) f_\gamma^k \tag{A.3}$$

¹⁵⁾A potential term would add a force field that transforms covariantly, and hence is trivial to add afterwards.

since Γ represents the connection with respect to the local coordinate frame and (A.3) expresses the change to the frame f .

On the other hand, the Euler–Lagrange equations with respect to a moving frame are given by (see, e. g., [11, Proposition 1.4.6], [6, Eq. (2.5)])

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial v^\alpha} - \frac{\partial \mathcal{L}}{\partial q^i} f_\alpha^i - \frac{\partial \mathcal{L}}{\partial v^\gamma} C_{\beta\alpha}^\gamma v^\beta = 0$$

with $\mathcal{L} = L \circ f: Q \times \mathbb{R}^n \rightarrow \mathbb{R}$ the Lagrangian with respect to the frame. In our case this boils down to

$$\begin{aligned} 0 &= \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial v^\alpha} - \frac{\partial \mathcal{L}}{\partial q^i} f_\alpha^i - \frac{\partial \mathcal{L}}{\partial v^\gamma} C_{\beta\alpha}^\gamma v^\beta \\ &= \frac{d}{dt} [\kappa_{\alpha\beta} v^\beta] - \frac{1}{2} f_\alpha \cdot \kappa_{\beta\gamma} v^\beta v^\gamma - \kappa_{\delta\gamma} C_{\beta\alpha}^\delta v^\beta v^\gamma \\ &= \kappa_{\alpha\beta} \dot{v}^\beta + f_\gamma \cdot \kappa_{\alpha\beta} v^\beta v^\gamma - \frac{1}{2} f_\alpha \cdot \kappa_{\beta\gamma} v^\beta v^\gamma - \kappa_{\delta\gamma} C_{\beta\alpha}^\delta v^\beta v^\gamma \\ \Leftrightarrow \dot{v}^\alpha &= -\kappa^{\alpha\delta} [f_\gamma \cdot \kappa_{\delta\beta} - \frac{1}{2} f_\delta \cdot \kappa_{\beta\gamma} + \kappa_{\eta\gamma} C_{\beta\alpha}^\eta] v^\beta v^\gamma. \end{aligned} \tag{A.4}$$

Note that by equating (A.2) and (A.4) we also find an explicit relation between the connection one-form ω , the metric κ and the structure functions C . It differs from (A.1) only by terms that are antisymmetric in the two lower indices, since these cancel in the geodesic equation.

ACKNOWLEDGMENTS

This topic was initially suggested to me by Hans Duistermaat as PhD research project. Although my PhD research finally went another direction, Hans' insights have been invaluable. Since then, this topic has been in the back of my mind and during my stay at PUC-Rio I started actively working on it again and gave a couple of lectures. This paper grew out of the accompanying lecture notes; I like to thank Alex Castro for his suggestion to do so. I want to thank Paula Balseiro and Luis García Naranjo for stimulating discussions and helpful remarks and Jair Koiller and the people at the Mathematics Department at PUC-Rio for their hospitality. This research was supported by the Capes grant PVE11-2012.

REFERENCES

1. Appell, P., Exemple de mouvement d'un point assujetti à une liaison exprimée par une relation non linéaire entre les composantes de la vitesse, *Rend. Circ. Mat. Palermo*, 1911, vol. 32, no. 1, pp. 48–50.
2. Arnold, V. I., *Ordinary Differential Equations*, Berlin: Springer, 2006.
3. Borisov, A. V., Karavaev, Yu. L., Mamaev, I. S., Erdakova, N. N., Ivanova, T. B., and Tarasov, V. V., Experimental Investigation of the Motion of a Body with an Axisymmetric Base Sliding on a Rough Plane, *Regul. Chaotic Dyn.*, 2015, vol. 20, no. 5, pp. 518–541.
4. Bloch, A. M., *Nonholonomic Mechanics and Control*, Interdiscip. Appl. Math., vol. 24, New York: Springer, 2003.
5. Bloch, A. M., Nonholonomic Mechanics, Dissipation and Quantization, in *Advances in the Theory of Control, Signals and Systems with Physical Modeling*, Lect. Notes Control Comput. Sci., vol. 407, Berlin: Springer, 2010, pp. 141–152.
6. Bloch, A. M., Marsden, J. E., Zenkov, D. V., Quasivelocities and Symmetries in Non-Holonomic Systems, *Dyn. Syst.*, 2009, vol. 24, no. 2, pp. 187–222.
7. Bloch, A. M. and Rojo, A. G., Quantization of a Nonholonomic System, *Phys. Rev. Lett.*, 2008, vol. 101, no. 3, 030402, 4 pp.
8. Brendeleev, V. N., On the Realization of Constraints in Nonholonomic Mechanics, *J. Appl. Math. Mech.*, 1981, vol. 45, no. 3, pp. 351–355; see also: *Prikl. Mat. Mekh.*, 1981, vol. 45, no. 3, pp. 481–487.
9. Bou-Rabee, N. M., Marsden, J. E., Romero, L. A., Dissipation-Induced Heteroclinic Orbits in Tippe Tops, *SIAM Rev.*, 2008, vol. 50, no. 2, pp. 325–344.
10. Carathéodory, C., Der Schlitten, *Z. Angew. Math. Mech.*, 1933, vol. 13, no. 2, pp. 71–76.
11. Cushman, R., Duistermaat, H., and Sniatycki, J., *Geometry of Nonholonomically Constrained Systems*, Adv. Ser. Nonlinear Dynam., vol. 26, Hackensack, N.J.: World Sci., 2010.
12. Chaplygin, S. A., On the Theory of Motion of Nonholonomic Systems. The Reducing-Multiplier Theorem, *Regul. Chaotic Dyn.*, 2008, vol. 13, no. 4, pp. 369–376; see also: *Mat. Sb.*, 1912, vol. 28, no. 2, pp. 303–314.

13. de León, M., A Historical Review on Nonholonomic Mechanics, *Rev. R. Acad. Cienc. Exactas Fís. Nat. Ser. A Math. RACSAM*, 2012, vol. 106, no. 1, pp. 191–224.
14. Eldering, J., *Normally Hyperbolic Invariant Manifolds: The Noncompact Case*, Atlantis Studies in Dynamical Systems, vol. 2, Paris: Atlantis Press, 2013.
15. Fenichel, N., Persistence and Smoothness of Invariant Manifolds for Flows, *Indiana Univ. Math. J.*, 1971/1972, vol. 21, pp. 193–226.
16. Fenichel, N., Geometric Singular Perturbation Theory for Ordinary Differential Equations, *J. Differential Equations*, 1979, vol. 31, no. 1, pp. 53–98.
17. Fedorov, Yu. N. and García-Naranjo, L. C., The Hydrodynamic Chaplygin Sleigh, *J. Phys. A*, 2010, vol. 43, no. 43, 434013, 18 pp.
18. Fufaev, N. A., On the Possibility of Realizing a Nonholonomic Constraint by Means of Viscous Friction Forces, *J. Appl. Math. Mech.*, 1964, vol. 28, no. 3, pp. 630–632; see also: *Prikl. Mat. Mekh.*, 1964, vol. 28, no. 3, pp. 513–515.
19. Hirsch, M. W., Pugh, C. C., and Shub, M., *Invariant Manifolds*, Lecture Notes in Math., vol. 583, New York: Springer, 1977.
20. Karapetian, A. V., On Realizing Nonholonomic Constraints by Viscous Friction Forces and Celtic Stones Stability, *J. Appl. Math. Mech.*, 1981, vol. 45, no. 1, pp. 30–36; see also: *Prikl. Mat. Mekh.*, 1981, vol. 45, no. 1, pp. 42–51.
21. Kobayashi, Sh. and Nomizu, K., *Foundations of Differential Geometry: Vol. 1*, New York: Wiley, 1996.
22. Kozlov, V. V. and Neishtadt, A. I., Realization of Holonomic Constraints, *J. Appl. Math. Mech.*, 1990, vol. 54, no. 5, pp. 705–708; see also: *Prikl. Mat. Mekh.*, 1990, vol. 54, no. 5, pp. 858–861.
23. Kozlov, V. V., Dynamics of Systems with Nonintegrable Constraints: 1, *Mosc. Univ. Mech. Bull.*, 1982, vol. 37, nos. 3–4, pp. 27–34; see also: *Vestn. Mosk. Univ. Ser. 1. Mat. Mekh.*, 1982, no. 3, pp. 92–100.
24. Kozlov, V. V., Dynamics of Systems with Nonintegrable Constraints: 2, *Mosc. Univ. Mech. Bull.*, 1982, vol. 37, nos. 3–4, pp. 74–80; see also: *Vestn. Mosk. Univ. Ser. 1. Mat. Mekh.*, 1982, no. 4, pp. 70–76.
25. Kozlov, V. V., Realization of Nonintegrable Constraints in Classical Mechanics, *Sov. Phys. Dokl.*, 1983, vol. 28, pp. 735–737; see also: *Dokl. Akad. Nauk SSSR*, 1983, vol. 272, no. 3, pp. 550–554.
26. Kozlov, V. V., On the Realization of Constraints in Dynamics, *J. Appl. Math. Mech.*, 1992, vol. 56, no. 4, pp. 594–600; see also: *Prikl. Mat. Mekh.*, 1992, vol. 56, no. 4, pp. 692–698.
27. Kozlov, V. V., Notes on Dry Friction and Nonholonomic Constraints, *Nelin. Dinam.*, 2010, vol. 6, no. 4, pp. 903–906 (Russian).
28. Lewis, A. D. and Murray, R. M., Variational Principles for Constrained Systems: Theory and Experiment, *Internat. J. Non-Linear Mech.*, 1995, vol. 30, no. 6, pp. 793–815.
29. Marle, Ch.-M., Various Approaches to Conservative and Nonconservative Nonholonomic Systems, *Rep. Math. Phys.*, 1998, vol. 42, nos. 1–2, pp. 211–229.
30. Neimark, Ju. I. and Fufaev, N. A., *Dynamics of Nonholonomic Systems*, Trans. Math. Monogr., vol. 33, Providence, R.I.: AMS, 1972.
31. Rubin, H. and Ungar, P., Motion under a Strong Constraining Force, *Comm. Pure Appl. Math.*, 1957, vol. 10, pp. 65–87.
32. Ruina, A., Nonholonomic Stability Aspects of Piecewise Holonomic Systems, *Rep. Math. Phys.*, 1998, vol. 42, nos. 1–2, pp. 91–100.
33. Sidek, N. and Sarkar, N., Dynamic Modeling and Control of Nonholonomic Mobile Robot with Lateral Slip, in *ICONS'08: Third International Conference on Systems (Cancún, Mexico, April 2008)*, pp. 35–40.
34. Takens, F., Motion under the Influence of a Strong Constraining Force, in *Global Theory of Dynamical Systems: Proc. Internat. Conf. (Northwestern Univ., Evanston, Ill., 1979)*, Lecture Notes in Math., vol. 819, Berlin: Springer, 1980, pp. 425–445.
35. Tikhonov, A. N., Systems of Differential Equations Containing Small Parameters in the Derivatives, *Mat. Sb. (N. S.)*, 1952, vol. 31(73), no. 3, pp. 575–586 (Russian).
36. Vershik, A. M. and Gershkovich, V. Ya., Nonholonomic Problems and the Theory of Distributions, *Acta Appl. Math.*, 1988, vol. 12, no. 2, pp. 181–209.
37. Wang, J.-Ch. and Huang, H.-P., Creep Dynamics of Nonholonomic Systems, in *Proc. of the 1996 IEEE Internat. Conf. on Robotics and Automation (Minneapolis, Minn., 22–28 Apr, 1996): Vol. 4*, pp. 3452–3457.