

Nonholonomic dynamics as limit of friction an introduction by example

Jaap Eldering

27th July 2015

Abstract

Nonholonomic dynamics is a principle to extend Lagrangian mechanics to incorporate constraints on velocities. Such “no slipping” constraints often show up in mechanical systems and it is natural to think of them as enforced by contact friction forces.

Think for example of a car with rolling tires: it can only move forward and turn by steering, but not directly slip sideways. Physically, one can view this slipping as being prevented by a strong contact friction force, which could be overcome when the force is smaller, for example on an icy road surface.

I will introduce the principle of nonholonomic dynamics and then show how it indeed can be viewed as a limit of infinite (viscous) friction forces. I will do so by explicitly treating a simple example system, the Chaplygin sleigh, to illustrate this.

Reference literature:

- A good advanced undergraduate level introduction to classical mechanics is [Arn89]; for a more advanced and geometrical text, see [MR99].
- For a graduate level introduction to nonholonomic systems, see [Blo03].
- For more background material on the theory needed to prove this limit, see [Eld13] chapter 1, especially example 1.7 and section 1.6 (up to subsection 1.6.1). Available at <http://www.jaapeldering.nl/includes/get.php?file=NHIM-noncompact-book.pdf>

In these lectures, I will show how a mechanical system with nonholonomic constraints can be viewed as the limit of an unconstrained system with additional friction forces. When the friction forces are scaled to infinity, the nonholonomic system is recovered.

The exposition is aimed at advanced undergraduate and beginning graduate students. I will assume little to no prior knowledge in differential geometry and stick to explicit coordinate formulations; at a few times these are put in an abstract geometric context, but it should be possible to ignore these minor abstractions. I will assume a basic knowledge of mechanics and briefly review Lagrangian mechanics. Then we shall define nonholonomic dynamics and treat the Chaplygin sleigh in detail.

Finally, I will show how this system can be obtained from a freely moving sleigh, but with friction damping sliding motion normal to the single skate of the sleigh. Time permitting, I will indicate how this idea can be generalized to arbitrary nonholonomic systems and be rigorously proven.

1 Lagrangian mechanics

Lagrange's formulation of mechanics is an alternative to Newton's equations. This method has the advantage that the equations of motion can be derived with any choice of coordinates, so one can choose the coordinates that best match the system. For example polar coordinates in a rotationally symmetric problem.

Lagrange mechanics starts with a Lagrangian, which is a real-valued function that depends on the position and velocity of the system (and possibly on time too). This function, normally denoted by L , expresses the kinetic minus the potential energy of the system. That is, for a simple mechanical system of a particle in a potential, we have

$$L(x, v) = \frac{1}{2}m\|v\|^2 - V(x), \quad (1)$$

Here, $x \in \mathbb{R}^n$ is the position of the particle, $v = \dot{x} \in \mathbb{R}^n$ its velocity, and V is the potential energy. For a single particle, we would typically have $n = 2$ or $n = 3$, but we could also describe many particles, say N , in three dimensions by setting $n = 3N$ and using $(x_{3i+1}, x_{3i+2}, x_{3i+3}) \in \mathbb{R}^3$ to describe the position of the i -th particle.

Now Lagrange's equations of motion are given by

$$\frac{d}{dt} \frac{\partial L}{\partial v_i} - \frac{\partial L}{\partial x_i} = 0, \quad \text{for } 1 \leq i \leq n. \quad (2)$$

Example 1.1. For a particle in the Earth's gravitational field we have $V(x) = -mgx_3$, where x_3 measures the height of the particle. Then we obtain

$$\begin{aligned} \frac{d(mv_1)}{dt} &= 0, \\ \frac{d(mv_2)}{dt} &= 0, \\ \frac{d(mv_3)}{dt} + mg &= m\dot{v}_3 + mg = 0. \end{aligned}$$

Thus, we see that the horizontal components of the velocity, (v_1, v_2) , are constant and in the vertical direction we have $\ddot{x}_3 = -g$, so we obtain a parabola, just as expected. $\circ$

The nice thing about Lagrange's formalism is that it works for any choice of coordinate system: simply construct the Lagrangian L as function of kinetic minus potential energy with respect to these coordinates and calculate the equations of motion (2). This even works for time-dependent coordinate systems. The next exercise shows how one can relatively easily derive the apparent forces in a non-inertial, rotating coordinate frame.

Exercise 1.2. Express the Lagrangian of a free particle in the plane (i.e. (1) with $n = 2$ and $V \equiv 0$) in a rotating frame with angular velocity ω . That is, substitute

$$x = R(t) \cdot y \quad \text{with } R(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{pmatrix}$$

and calculate and substitute the associated velocity $w = \dot{y}$ to obtain the transformed, time-dependent(!) Lagrangian $\tilde{L}(y, w)$. Next, show that the Lagrange equations (2) for $\tilde{L}$ are given by

$$\dot{w} = -2\Omega \cdot w + \omega^2 y \quad \text{with } \Omega = \begin{pmatrix} 0 & -\omega \\ \omega & 0 \end{pmatrix}.$$

Here Ω is the generator of rotation, that is, $R(t) = e^{t\Omega}$. Note that the first term on the right-hand side is the Coriolis force and the second term the centrifugal force.

That Lagrange's equations (2) give the correct equations of motion in any coordinate system stems from the fact that they can be seen as a 'variational derivative' of a functional. That is, let $x(t) \in \mathbb{R}^n$ be any C^2 curve on the interval $[a, b]$, and define the so-called action functional

$$S(x) = \int_a^b L(x(t), \dot{x}(t)) dt. \quad (3)$$

Lagrange showed that equations (2) appear as the variational equations for this functional. That is, a curve $x(t)$ extremizes S precisely if it satisfies (2). Since this property is coordinate invariant, it follows that the Lagrange equations yield the correct equations of motion in any coordinate system. For more details, see e.g. [Arn89, Chap. 3].

Since Lagrangian mechanics can be expressed with respect to any coordinate system, one often talks of using 'generalized coordinates' instead of traditional Euclidean coordinates. Such coordinates are conventionally written with the letter ' q '; we shall adopt this convention from now on.

The following theorem is an important result in physics.

Theorem 1.3. *When the Lagrangian L does not explicitly depend on time, then the energy function*

$$H(q, v) = \frac{\partial L}{\partial v}(q, v) \cdot v - L(q, v) \quad (4)$$

is a conserved quantity. That is, it is constant along solution curves of the Lagrange equations.

Proof. Let $q(t)$ be a solution curve of (2). We do a direct calculation:

$$\begin{aligned} \frac{d}{dt} h(q(t), \dot{q}(t)) &= \frac{d}{dt} \frac{\partial L}{\partial v} \cdot \dot{q}(t) + \frac{\partial L}{\partial v} \cdot \ddot{q}(t) - \frac{d}{dt} L(q(t), \dot{q}(t)) \\ &= \frac{d}{dt} \frac{\partial L}{\partial v} \cdot \dot{q}(t) + \frac{\partial L}{\partial v} \cdot \ddot{q}(t) - \frac{\partial L}{\partial q} \cdot \dot{q}(t) - \frac{\partial L}{\partial v} \cdot \ddot{q}(t) \\ &= \left(\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial q} \right) \cdot \dot{q}(t) = 0. \end{aligned}$$

In the last step we used that $q(t)$ satisfies the Lagrange equations, thus the term in parentheses is zero. $\square$

The Lagrangian can only incorporate potential forces, i.e. force functions F that are (minus) the gradient of a scalar potential V , that is,

$$F(q) = -\frac{\partial V}{\partial q}.$$

Other, non-potential forces can be included in the Lagrange equations as

$$\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial q} = F(q, v). \quad (5)$$

These forces can also depend on velocities, for example in case of friction. However, Theorem 1.3 does not apply anymore when extra forces are added.

Remark 1.4. Lagrangian mechanics can also be defined on manifolds, that is, on spaces which cannot globally be viewed as $\mathbb{R}^n$, only locally. Examples of such spaces are the sphere, S^2 , and the space of rotation matrices, $SO(3)$. On a manifold Q , the Lagrangian is a function

$$L: TQ \rightarrow \mathbb{R},$$

where TQ is the tangent bundle of Q , consisting of points and all possible velocities at these points. The Lagrange equations can then naturally viewed as a map

$$\left(\frac{d}{dt} \frac{\partial L}{\partial v^i} - \frac{\partial L}{\partial q^i} \right) dq^i: TTQ \rightarrow T^*Q,$$

where the dq^i are elements dual to velocity vectors (they measure their components) that live in the cotangent bundle T^*Q . For the main part of these lectures we shall stick to explicit coordinates, where these details can mostly be ignored. $\diamond$

2 Nonholonomic dynamics

Lagrangian (as well as Newtonian) mechanics describes the motion of a system on a certain manifold Q where all possible velocities are allowed. However, in real life this is not always true. For example, consider a ball rolling without slipping on a flat surface. Unless the ball moves horizontally, it can only rotate around its vertical axis. However, by rolling it along a small square on the surface, we can return it to its original position but with a different orientation that is not a rotation around its vertical axis. Another example is parallel parking. A car can only move forward, backward and turn, but not sideways. Still, by maneuvering we can effectively move it sideways to park it into a parking bay.

We call such systems ‘nonholonomically constrained’: even though all positions q in the system are reachable, not all velocities are allowed.

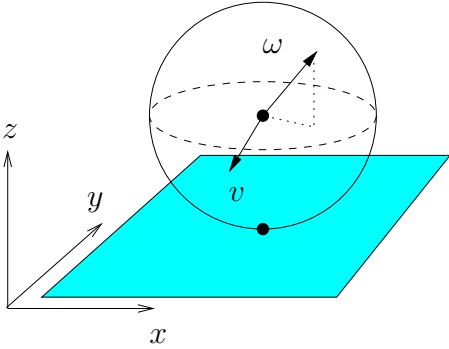


Figure 1: A rolling ball.

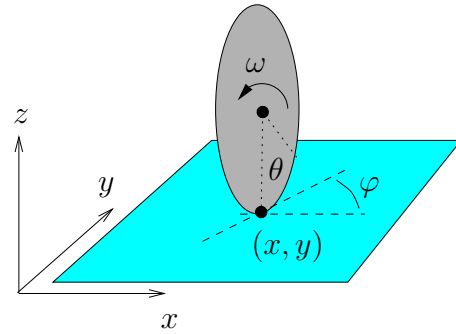


Figure 2: A rolling disc.

Example 2.1 (Rolling ball).

A ball rolling on a flat surface is described by the position of the point of contact with the surface and the orientation of the ball, that is, $Q = \mathbb{R}^2 \times SO(3)$. The rotational velocity (which can be viewed as a vector $\omega \in \mathbb{R}^3$) fixes the horizontal velocity ($v \in \mathbb{R}^3$ with $v_3 = 0$) via the constraint

$$v = r \omega \times e_3.$$

Here e_3 is the vertical unit vector and r the radius of the ball. $\circ$

Example 2.2 (Rolling disc).

A vertical disc on a surface is described by the position of its point of contact, the direction along which it can roll and its rotation about its center. Here we have natural coordinates $q = (x, y, \varphi, \theta) \in Q = \mathbb{R}^2 \times S^1 \times S^1$ and the constraints are given by

$$(\dot{x}, \dot{y}) = (\cos(\varphi), \sin(\varphi))\dot{\theta}. \quad \circ$$

In general a nonholonomic system is a system which has constraints that depend on the velocities (and positions) of the system. For simplicity we shall only consider constraints that are linear in the velocities — this is also the most common case in mechanical systems. This can be expressed geometrically by saying that the pair of position and velocity of the system must lie in a subspace, a ‘distribution’ $\mathcal{D}$ within the tangent bundle:

$$(q, v) \in \mathcal{D} \subset TQ. \quad (6)$$

More explicitly, the constraint set $\mathcal{D}$ can be described by a k -tuple of (local) constraint functions $C: TQ \rightarrow \mathbb{R}^k$ that are linear in the velocities. In the rolling disc example we would have

$$C(q, \dot{q}) = C(x, y, \varphi, \theta, \dot{x}, \dot{y}, \dot{\varphi}, \dot{\theta}) = \begin{pmatrix} \dot{x} - \cos(\varphi)\dot{\theta} \\ \dot{y} - \sin(\varphi)\dot{\theta} \end{pmatrix} \in \mathbb{R}^2$$

and $\mathcal{D} = \{C(q, \dot{q}) = 0\}$.

For systems with nonholonomic constraints we can still write down a Lagrangian L , but we cannot simply use the Lagrange equations anymore. In general these would give rise to solution curves that violate the constraints, i.e.

$$(q(t), \dot{q}(t)) \in \mathcal{D} \quad (7)$$

would not need to hold all the time. Thus, we need to modify the Lagrange equations to guarantee that the solutions satisfy the constraints. To this end, we add a constraint force F_c to right-hand side of the Lagrange equations that enforces (7) to hold. This does not completely determine F_c as we can always add a term that is tangential to $\mathcal{D}_q$, the allowed velocities at the current position q . Thus we also impose d’Alembert’s principle.

Hypothesis 2.3 (d’Alembert’s principle).

The constraint forces F_c do no work along movements that are compatible with the constraints. That is, for any $(q, v) \in \mathcal{D}$ we have that $F_c \cdot v = 0$.

With d’Alembert’s principle we obtain unique, well-defined equations of motion that generate solutions that satisfy the constraints. These extended equations are called the Lagrange–d’Alembert equations, and can be written in abstract form as

$$\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial q} \in \mathcal{D}^0, \quad (q, \dot{q}) \in \mathcal{D}, \quad (8)$$

or in more explicit form as

$$\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial q} = \lambda^i \frac{\partial C_i}{\partial v}, \quad C(q, \dot{q}) = 0. \quad (9)$$

Here the λ_i are Lagrange multipliers that must be solved for together with solution $q(t)$, and the constraint force is given by $F_c = \lambda^i \frac{\partial C_i}{\partial v}$.

Remark 2.4. When the constraint functions C only depend on the position q , the constraint is called ‘holonomic’. In that case, the same equations of motion can also be obtained by restricting the Lagrangian to the submanifold $\tilde{Q} = \{C = 0\} \subset Q$ and solving the normal Lagrange equations (2) on $T\tilde{Q}$. $\diamond$

Also in nonholonomic systems we have energy preservation.

Theorem 2.5. *Let $(L, \mathcal{D})$ describe a nonholonomically constrained Lagrangian system. When the Lagrangian L does not explicitly depend on time, then the energy function*

$$H(q, v) = \frac{\partial L}{\partial v}(q, v) \cdot v - L(q, v)$$

is a conserved quantity.

Proof. We can repeat most of the proof of Theorem 1.3 verbatim. There we found that

$$\frac{d}{dt}h(q(t), \dot{q}(t)) = \left(\frac{d}{dt} \frac{\partial L}{\partial v} - \frac{\partial L}{\partial q} \right) \cdot \dot{q}(t).$$

Now we use that $q(t)$ satisfies the Lagrange–d’Alembert equations (9), which means that we can substitute

$$\frac{d}{dt}h(q(t), \dot{q}(t)) = F_c \cdot \dot{q}(t),$$

but $\dot{q}(t)$ must satisfy the constraints and by the d’Alembert principle, the constraint force F_C does no work on such velocities. $\square$

3 The Chaplygin sleigh

The Chaplygin sleigh is a simple, but 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 Figure 3. The skate contact point can only move in the direction along the blade of the skate, not in the perpendicular direction. The other 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 blade.

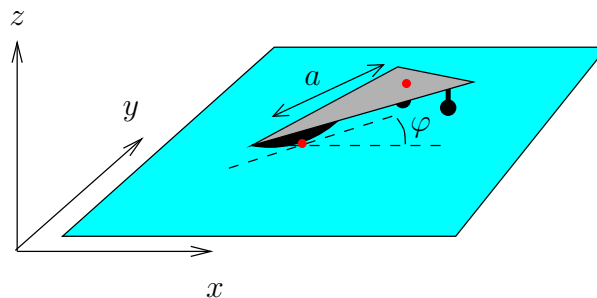


Figure 3: A Chaplygin sleigh: the left red dot is the point of contact of the skate, the right red dot is the center of mass.

We describe the system with coordinates $(x, y) \in \mathbb{R}^2$ for the skate contact point¹ and $\varphi \in S^1$ the angle of the skate blade with the x axis. Assume the sleigh has total mass m

¹It might seem more natural to choose the (x, y) coordinates for the center of mass. However, this choice turns out to yield simpler equations, so we use this foresight.

and moment 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

$$\begin{aligned} L &= \frac{1}{2}m(\dot{x}_c^2 + \dot{y}_c^2) + \frac{1}{2}I\dot{\varphi}^2 \\ &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) + \frac{1}{2}(I + ma^2)\dot{\varphi}^2 + ma\dot{\varphi}(-\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi)). \end{aligned} \quad (10)$$

The velocity parallel and perpendicular to the skate are given by²

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

Thus the constraint is described by the function

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

since $C(q, \dot{q}) = w = 0$ says that the skate cannot move sideways.

It is now a straightforward exercise to calculate the Lagrange–d’Alembert equations (9). We find for each of the coordinates x, y, φ and their associated velocities:

$$\begin{aligned} \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} - \frac{\partial L}{\partial x} &= m\ddot{x} + ma(-\ddot{\varphi}\sin(\varphi) - \dot{\varphi}^2\cos(\varphi)) = -\lambda\sin(\varphi), \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{y}} - \frac{\partial L}{\partial y} &= m\ddot{y} + ma(\ddot{\varphi}\cos(\varphi) - \dot{\varphi}^2\sin(\varphi)) = \lambda\cos(\varphi), \\ \frac{d}{dt} \frac{\partial L}{\partial \dot{\varphi}} - \frac{\partial L}{\partial \varphi} &= (I + ma^2)\ddot{\varphi} + ma\frac{d}{dt}[-\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi)] \\ &\quad + ma\dot{\varphi}(\dot{x}\cos(\varphi) + \dot{y}\sin(\varphi)) = 0. \end{aligned} \quad (11)$$

Next, we take linear combinations of the first two equations with factors $\cos(\varphi)$ and $\sin(\varphi)$ such that the term with $\ddot{\varphi}$ cancels in one of the equations, and remains without sine or cosine in the second. Also recognizing v and w in the third equation, we get

$$\begin{aligned} m(\ddot{x}\cos(\varphi) + \ddot{x}\sin(\varphi) - a\dot{\varphi}^2) &= 0, \\ m(-\ddot{x}\sin(\varphi) + \ddot{y}\cos(\varphi) + a\ddot{\varphi}) &= \lambda, \\ (I + ma^2)\ddot{\varphi} + ma(\dot{w} + \dot{\varphi}v) &= 0. \end{aligned}$$

Furthermore, we calculate

$$\begin{aligned} \dot{v} &= \ddot{x}\cos(\varphi) + \ddot{y}\sin(\varphi) + (-\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi))\dot{\varphi} = \ddot{x}\cos(\varphi) + \ddot{y}\sin(\varphi) + w\dot{\varphi}, \\ \dot{w} &= -\ddot{x}\sin(\varphi) + \ddot{y}\cos(\varphi) - (\dot{x}\cos(\varphi) + \dot{y}\sin(\varphi))\dot{\varphi} = -\ddot{x}\sin(\varphi) + \ddot{y}\cos(\varphi) - v\dot{\varphi} \end{aligned}$$

and recognize these terms in the first two equations. We conventionally denote $\omega = \dot{\varphi}$ and rewrite our equations as

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

²In abstract terms, we are constructing a ‘moving frame’ aligned with the constraint distribution $\mathcal{D}$.

Finally, $w = 0$ also implies that $\dot{w} = 0$. Inserting this, we obtain the differential equations

$$\dot{v} = a\omega^2, \quad \dot{\omega} = -\frac{mav\omega}{I + ma^2}, \quad (12)$$

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

$$F_c = m(v\omega + a\dot{\omega}) \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}. \quad (13)$$

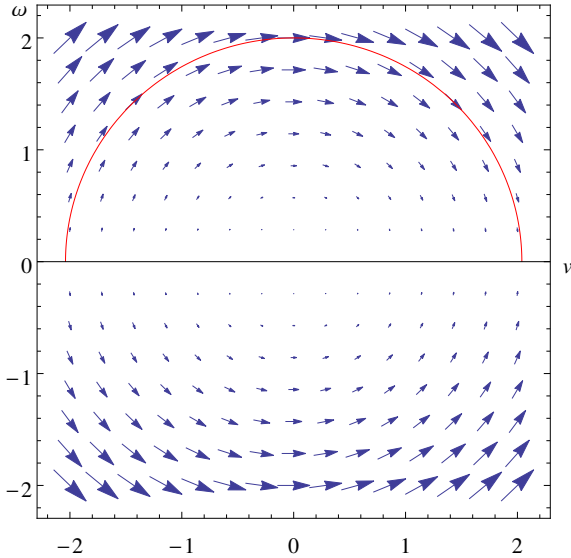


Figure 4: Phase plot of the (v, ω) coordinates with $m = I = 1$ and $a = 1/5$.

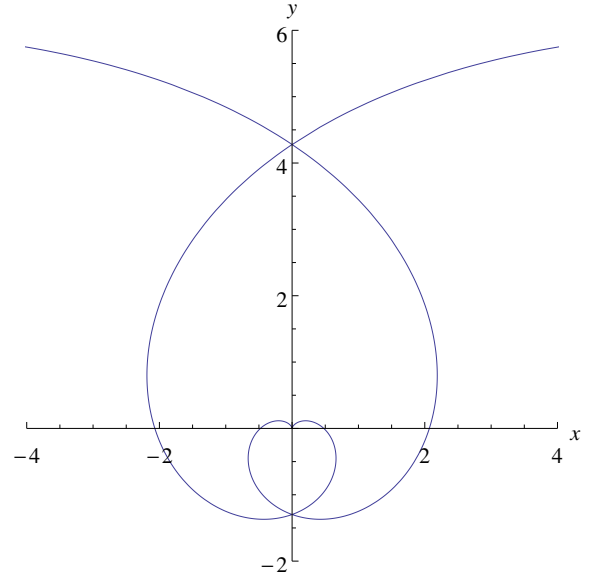


Figure 5: Trajectory plot associated to the red orbit in Figure 4.

These equations give rise to the phase plot in Figure 4. A typical trajectory of the skate point of contact is shown in Figure 5. We see that the solutions look like half-ellipsoids starting from the negative v axis and converging onto the positive v axis. When v 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 is unstable. Note that this cannot happen in ‘normal’ Lagrangian systems. It can be shown that in that case the eigenvalues of the linearization about a fixed point must always appear in pairs or quadruples that are symmetric about the real and imaginary axis. Here, the fixed points on the v axis have one zero eigenvalue and one negative *or* positive eigenvalue when $v < 0$ or $v > 0$, respectively.

4 The sleigh with sliding friction

Now we shall obtain the Chaplygin sleigh equations (12) in a different way. Instead of enforcing the no-slip constraint $w = 0$ with the constraint reaction force F_c , we consider the same Lagrangian, now without constraint, but 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.

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. Think of a car driving with rubber tires on tarmac: when you steer the car into a turn, there is sufficient friction to keep the tires from slipping sideways. However, when you drive the same car over an icy road, there may not be enough friction, and you can easily start slipping. By showing that scaling the friction force to infinity leads to nonholonomic dynamics, we provide a fundamental physical/mathematical argument for the d'Alembert principle of nonholonomic dynamics.

We start with the same Lagrangian (10), 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 $w = -\dot{x}\sin(\varphi) + \dot{y}\cos(\varphi)$. We take the friction force linear in the slipping velocity and pointing in the opposite direction to dampen it. Thus we have³

$$F_f = -w \begin{pmatrix} -\sin(\varphi) \\ \cos(\varphi) \end{pmatrix}. \quad (14)$$

Note that this force acts at the point of contact of the skate, so there is no associated torque around that point. Secondly, note that F_f is zero when $w = 0$ (as opposed to F_c), so this finite force as it is does not enforce the nonholonomic constraint. We insert F_f/ε into the right-hand side of (11), so that we can scale the friction force to infinity by taking the limit $\varepsilon \rightarrow 0$. This yields

$$\begin{aligned} m\ddot{x} + ma(-\ddot{\varphi}\sin(\varphi) - \dot{\varphi}^2\cos(\varphi)) &= -\frac{w}{\varepsilon}\sin(\varphi), \\ m\ddot{y} + ma(\ddot{\varphi}\cos(\varphi) - \dot{\varphi}^2\sin(\varphi)) &= -\frac{w}{\varepsilon}\cos(\varphi), \\ (I + ma^2)\ddot{\varphi} + ma\dot{\varphi}(\dot{x}\cos(\varphi) + \dot{y}\sin(\varphi)) &= 0. \end{aligned} \quad (15)$$

As in the previous section, we can rewrite these equations as

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

but now we cannot insert the constraint condition $w = 0$. Instead, we shall analyze the dynamics and see that when $\varepsilon > 0$ is small, $w(t)$ will quickly converge to near zero. This will imply that the other two equations will effectively behave as if the constraint $w = 0$ is active.

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

$$\frac{I}{I + ma^2}(\dot{w} + v\omega) = -\frac{w}{m\varepsilon}. \quad (16)$$

Note that when ε is small, this gives fast exponential decay of w . We shall now assume that the typical rate at which $w(t)$ changes is much faster than that of $v(t)$ and $\omega(t)$. That is, we consider w as fast variable and v, ω as slow variables. Further conclusions based

³We could have obtained this friction force more geometrically from a Rayleigh dissipation function $R(q, v) = \frac{1}{2}g_q(v^\perp, v^\perp)$, where $v^\perp$ is the projection of v onto $\mathcal{D}^\perp$, the distribution orthogonal to $\mathcal{D}$. Then F_f is given by minus the fiber derivative of R .

on this can be made rigorous using singular perturbation theory, but we postpone these arguments and focus on obtaining the result.

Thus we assume in (16) that v, ω are approximately constant and obtain as solution

$$w(t) = \left(w(0) + \frac{v\omega}{\rho}\right)e^{-\rho t} - \frac{v\omega}{\rho} \quad \text{with } \rho = \frac{I + ma^2}{Im\varepsilon}. \quad (17)$$

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

$$\dot{v} = a\omega^2 + \varepsilon \frac{Im}{I + ma^2} v\omega^2, \quad \dot{\omega} = -\frac{mav\omega}{I + ma^2} + \varepsilon Im^2 a \frac{d}{dt}[v\omega]. \quad (18)$$

In these differential equations ε only appears in the numerator and the ε multiplying the time derivative of $(v\omega)$ does not lead to singular equations, so we can sensibly take the limit: we simply insert $\varepsilon = 0$ to arrive at the original equations (12) for the nonholonomically constrained Chaplygin sleigh. This is also confirmed by the numerical integration of this system: with decreasing values of ε , the trajectories converge to the trajectory of the nonholonomic system, see Figure 6.

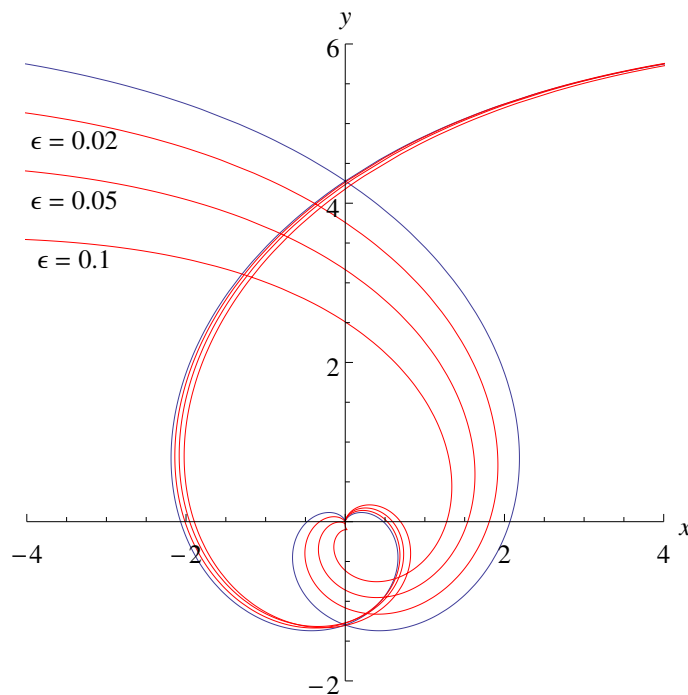


Figure 6: The red orbits are trajectories of the sleigh with friction with indicated parameter values ε . These clearly converge to the nonholonomic trajectory in blue.

4.1 Making the limit rigorous

In the previous section we argued that the dynamics of w is fast relative to the dynamics of (v, ω) and therefore we can first solve for the equilibrium of w as a function of (v, ω) and then the dynamics of (v, ω) with the equilibrium of w substituted. This can be made rigorous using singular perturbation theory and the underlying theory of normally

hyperbolic invariant manifolds (abbreviated as NHIM). For an introduction to this theory see e.g. [Kap99; Jon95].

Let us analyze the situation a bit more abstractly. We have a system of differential equations of the form

$$\begin{aligned}\dot{x} &= f_\varepsilon(x, y), \\ \dot{y} &= \frac{1}{\varepsilon} g_\varepsilon(x, y).\end{aligned}\tag{19}$$

In our case, we have $x = (v, \omega)$ as ‘slow variables’ and $y = w$ as ‘fast variable’. Note that f and g are assumed to depend smoothly on the parameter ε , but the factor $1/\varepsilon$ in the second equation generates a singularity at $\varepsilon = 0$.

To deal with this singularity, we simply multiply the right-hand side of both equations by ε . This can be seen as describing the system with a fast time variable $\tau = t/\varepsilon$. We indicate differentiation by τ with a prime instead of a dot: $x' = \frac{dx}{d\tau}$. Note that the original system (19) and the rescaled version

$$\begin{aligned}x' &= \varepsilon f_\varepsilon(x, y), \\ y' &= g_\varepsilon(x, y)\end{aligned}\tag{20}$$

generate solution curves that are equivalent up to a time rescaling for any $\varepsilon > 0$. However, the rescaled system is well-defined (i.e. *regularized*) for $\varepsilon = 0$.

Now, furthermore assume that

1. $g_0(x, 0) = 0$, that is, the fast system has an equilibrium at $y = 0$ for all x ,
2. the family of equilibria are hyperbolic with respect to y , that is, the Jacobian $\frac{\partial g_0}{\partial y}(x, 0)$ has all eigenvalues off the imaginary axis, and that the Jacobian depends smoothly on x .

From the first condition we see that the set $M = \{y = 0\}$ forms an invariant manifold the system (20) for $\varepsilon = 0$ (more specifically, it consists of fixed points). The second condition says that M is hyperbolic in the normal directions. Hence, we say that M is a normally hyperbolic invariant manifold.

Note that our system satisfies both assumptions: when (16) is rescaled and $\varepsilon = 0$ set, then it reduces to

$$w' = -\frac{I + ma^2}{Im}w.$$

This clearly has an equilibrium at $w = 0$ with one negative eigenvalue, hence the set $\mathcal{D} = \{w = 0\}$ is a normally attractive invariant manifold, a special case of a NHIM.

Now, a NHIM M persists under small perturbations. This means that if we change from $\varepsilon = 0$ to a small $\varepsilon > 0$, then there is a unique manifold M_ε that is still invariant under the system (20) with that ε . This manifold is of the form

$$M_\varepsilon = \{y = h_\varepsilon(x)\}$$

where h is a function that depends smoothly on x and the parameter ε . Thus, the fast variable y depends on (or as it is called: ‘is slaved to’) the slow variable x . This means that we can plug that into the slow equation and obtain

$$x' = \varepsilon f_\varepsilon(x, h_\varepsilon(x))$$

as a differential equation that is well-defined on the invariant set M_ε (which is moreover attractive in our case).

Now that we have eliminated the fast variable y , we can rescale the slow equations back to the original time by dividing by ε . This gives

$$\dot{x} = f_\varepsilon(x, h_\varepsilon(x)).$$

Finally, unlike the original full system (19), this equation has a well-defined limit for $\varepsilon \rightarrow 0$. This is

$$\dot{x} = f_0(x, 0) \tag{21}$$

since $h_0 \equiv 0$ by the assumption that $M_0 = M = \{y = 0\}$. This is precisely what we already calculated in the previous section by substituting $w = 0$ into the equations for v and ω .

References

- [Arn89] V. I. Arnol'd. *Mathematical methods of classical mechanics*. Second. Vol. 60. Graduate Texts in Mathematics. Springer-Verlag, New York, 1989, pp. xvi+508. ISBN: 0-387-96890-3. DOI: 10.1007/978-1-4757-2063-1.
- [Blo03] A. M. Bloch. *Nonholonomic mechanics and control*. Vol. 24. Interdisciplinary Applied Mathematics. New York: Springer-Verlag, 2003, pp. xx+483. ISBN: 0-387-95535-6. DOI: 10.1007/b97376.
- [Eld13] J. Eldering. *Normally hyperbolic invariant manifolds*. Vol. 2. Atlantis Series in Dynamical Systems. Atlantis Press, Paris, Sept. 2013, pp. xii+189. ISBN: 978-94-6239-002-7; 978-94-6239-003-4. DOI: 10.2991/978-94-6239-003-4.
- [Jon95] C. K. R. T. Jones. ‘Geometric singular perturbation theory’. In: *Dynamical systems (Montecatini Terme, 1994)*. Vol. 1609. Lecture Notes in Math. Berlin: Springer, 1995, pp. 44–118. DOI: 10.1007/BFb0095239.
- [Kap99] T. J. Kaper. ‘An introduction to geometric methods and dynamical systems theory for singular perturbation problems’. In: *Analyzing multiscale phenomena using singular perturbation methods (Baltimore, MD, 1998)*. Vol. 56. Proc. Sympos. Appl. Math. Providence, RI: Amer. Math. Soc., 1999, pp. 85–131.
- [MR99] J. E. Marsden and T. S. Ratiu. *Introduction to mechanics and symmetry*. Second. Vol. 17. Texts in Applied Mathematics. New York: Springer-Verlag, 1999, pp. xviii+582. ISBN: 0-387-98643-X. DOI: 10.1007/978-0-387-21792-5.