CDS 205 Final Project:
Incorporating Nonholonomic Constraints in Basic Geometric Mechanics Concepts

Michael Wolf
wolf@caltech.edu

6 June 2005

1 Introduction

1.1 Motivation

While most mechanics and dynamics course material concerns itself with unconstrained or holonomically constrained systems, nonholonomic constraints play a crucial role in numerous systems of interest, including vehicles of all types. Understanding how to generalize fundamental concepts in geometric mechanics for analysis of nonholonomic systems provides great academic benefits and preparation for deeper work in control, robotics, and other fields.

1.2 Goals and Outline

The following lists the key aims of this project and serves as a rough outline of this resulting paper:

1. Gain a deeper understanding of what being nonholonomic means from both analytical and geometric perspectives. What are the essential properties of these constraints and how are they best described?

2. Restate key aspects of course material with the addition of nonholonomic constraints. The focus of this work regards the derivation of the equations of motion in a Lagrangian perspective — essentially the re-evaluation of Hamilton's principle and the Euler-Lagrange equations. The purpose here is not only to accomplish the project's

key aim but also to gain new depth into the already familiar (holonomic) subject matter.

3. Elucidate the above material via simple examples, illustrating the various methods by which the resulting equations of motion may or may not be obtained for different types of systems.

Generally, a slow and careful approach to ensure depth of understanding and all derivations is emphasized over covering a broad range of material. (Hopefully, already-knowledgeable readers will not mind bearing with the paper through these details.)

2 Constraints in Dynamical Systems

2.1 Holonomic and Nonholonomic Constraints

The classification of a constraint as holonomic or nonholonomic holds key consequences for the analysis of the dynamical system. Roughly, holonomic constraints limit the possible positions of system while nonholonomic constraints limit its allowable types of motions. Various perspectives on describing this classification more precisely are below.

The key characterization of a holonomic constraint is that it can be written as a function of the position only, and a constraint that does not possess this property is termed nonholonomic. Clearly, either type can be written as a function of velocity (by simply differentiating a holonomic position constraint, or in its natural form for a nonholonomic constraint). Assume, for this paper, constraints are time-independent and linear in the velocity. (Most interesting cases appear to be linear in velocity – I do not even know of a nonlinear velocity constraint, and the procedures can be modified to account for affine constraints.) We can thus write all such constraints of interest as

$$a_i^j(q)\dot{q}^i = 0 \tag{1}$$

for the jth constraint where the coefficients of each velocity component are functions of the configuration q. Expressing all constraints together in matrix form gives

$$A(q)\dot{q} = 0 \tag{2}$$

where $\dot{q}$ is a column n-vector (n-dimensional configuration space) and $A(q)$ is an $m \times n$ matrix (m constraints).

Thus the defining property of a holonomic constraint is, given in form (1), it may be integrated to a relation on q rather than $\dot{q}$ (and hence such constraints are sometimes termed *integrable*). That is, for the constraint $a_i(q)\dot{q}^i = 0$ there exists a real-valued function $h(q)$ such that

$$a_i(q) = \frac{\partial h(q)}{\partial q^i}, \forall i$$

and so

$$\begin{aligned} \frac{\partial h(q)}{\partial q^i}\dot{q}^i &= 0 \\ h(q) &= \text{constant.} \end{aligned}$$

Note that $a(q)$ actually being a function of q is a necessary (but not sufficient) condition for the constraint to be nonholonomic, since if a is constant then the constraint can obviously be integrated.

From a geometric perspective, holonomic constraints define a submanifold of the configuration manifold (or, alternatively, they can be written on the tangent bundle as discussed below). Indeed, the configuration space of a holonomically constrained dynamical system is often taken to be the "sub-configuration manifold" defined by the constraints. An example of this is choosing S^1 as the configuration space for a planar pendulum rather than using $\mathbb{R}^2$ with the appropriate length constraint, e.g. $\|\mathbf{x}\| = \text{constant}$.

By contrast, nonholonomic constraints define a (nonintegrable!) distribution on the configuration manifold Q. This makes sense since for each point $q \in Q$, the constraints limit the permitted tangent vectors (velocities), defining an allowable subspace of the tangent space T_qQ. Over the entire tangent bundle TQ this creates a vector subbundle (the constraint distribution). The constraint distribution's integrability is equivalent to the integrability of the constraints as above, as the integrability of a subbundle precisely means there exists a submanifold $\tilde{Q} \subset Q$ whose tangent bundle is the constraint distribution restricted to $\tilde{Q}$. That is, for an integrable (holonomic) constraint distribution, one can think of transferring the constraint on TQ to a constraint on Q, as mentioned in the pendulum example above. Because integrability is the defining property of this classification, Frobenius's Theorem serves as an instrumental tool in showing whether constraints are holonomic.

2.2 Examples of Nonholonomic Constraints

Rolling without slipping represents a commonly encountered nonholonomic constraint. For rolling without slipping, the point of contact on the rolling body has zero velocity. Another set of nonholonomic constraints arises from steered vehicles whose motion must be in the direction of their heading, for example automobiles, sleds, ships, and missiles. A skate or knife edge exhibits such a constraint as well.

The Knife Edge. A simple example of a nonholonomic constraint is exhibited by the knife edge on an inclined plane as in Figure 1. As this is planar motion of an oriented rigid body, the configuration space is the planar Special Euclidean group, SE(2). The key constraint is that the knife edge does not slide sideways but rather moves only in the direction of its heading. This can be represented as

$$\dot{x} \sin \varphi = \dot{y} \cos \varphi. \tag{3}$$

We see this constraint cannot be integrated into a configuration constraint.

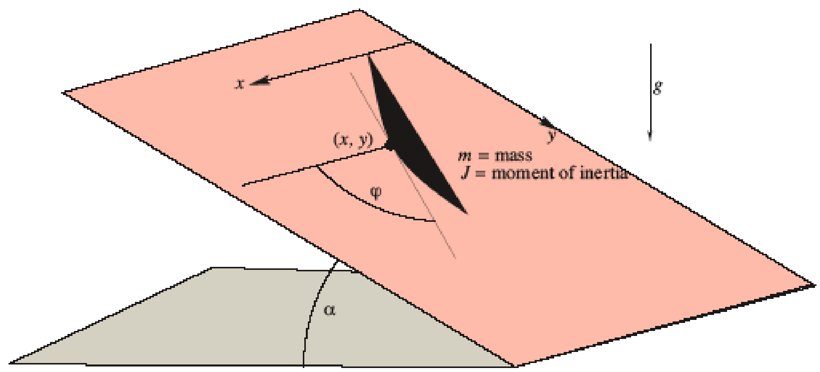

Figure 1: Knife Edge on an Inclined Plane (from [2])

Simple Pendulum (A Counterexample). A simple pendulum in $\mathbb{R}^2$ may serve as a clarifying (if pedantic) example for the structure of holonomic constraints. In cylindrical coordinates (r, θ), the constraint (length of the pendulum rod) is trivially written

$$\begin{aligned} \dot{r} &= 0 \\ r &= \text{constant.} \end{aligned}$$

In Cartesian coordinates (x, y), we can obtain geometrically the velocity form constraint

$$x\dot{x} + y\dot{y} = 0. \tag{4}$$

(Note this agrees with differentiating $\|q\| = R$ for $q^T \dot{q} = 0$.) This equation can be integrated by parts to obtain $\int x\dot{x}dt = \frac{1}{2}x^2$ for the familiar $x^2 + y^2 = \text{constant} = R^2$.

The Veritcal Rolling Disk. Consider a coin or disk rolling freely on a planar surface. For simplicity, we restrict the disk to a vertical position so it does not tilt. The configuration space of this system is $Q = \mathrm{SE}(2) \times S^1$ defining the position and orientation in the plane and the roll angle, with coordinates $((x, y, \varphi), \theta)$ (see Figure 2).

The rolling without slipping constraint can be written, for disk radius R and coordinates as shown in the figure,

$$\dot{x} = R(\cos \varphi)\dot{\theta} \tag{5}$$

$$\dot{y} = R(\sin \varphi)\dot{\theta}. \tag{6}$$

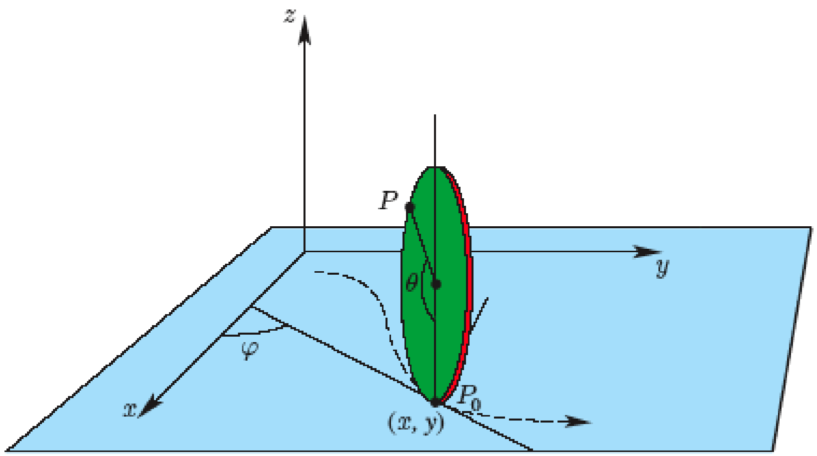

Figure 2: Vertical Rolling Disk on Horizontal Surface (from [2])

In the notation of (2), we have

$$A(q) \quad = \quad \begin{pmatrix} 1 & 0 & 0 & -R\cos\varphi \\ 0 & 1 & 0 & -R\sin\varphi \end{pmatrix}, \; q = \begin{pmatrix} x \\ y \\ \varphi \\ \theta \end{pmatrix}.$$

Note that we could also consider the larger configuration space that includes the z-direction, but as we are assuming the holonomic constraint $z = \text{constant} = 0$, a trivial reduction of the space to planar motion results.

What happens if we further constrain this system to move only in the x-direction? This is equivalent to adding the holonomic constraint $\varphi = 0$, which makes (6)

$$\begin{aligned} \dot{y} &= 0 \\ y &= \text{constant}, \end{aligned}$$

now a holonomic constraint. Likewise (5) becomes

$$\begin{aligned} \dot{x} &= R\dot{\theta} \\ x &= R\theta + \text{constant}, \end{aligned}$$

so that all constraints are now holonomic. Note that the x-θ relationship only makes sense for the first roll (in $\theta \in [0, 2\pi]$) or if θ is considered as a non-cyclic variable, but regardless the constraint makes physical sense. Now the configuration space may be taken as $\tilde{Q} = \mathbb{R}$, and with knowledge of the constraints and initial conditions the system is fully specified.

This addition of further constraints illustrates that a nonintegrable subbundle of the tangent bundle can be further restricted to a smaller distribution that is actually integrable.

3 The Lagrange-d'Alembert Equations

The primary means of deriving equations of motion for nonholonomic systems appears to be a Lagrangian approach. This section discusses the Lagrange-d'Alembert principle and the Lagrange-d'Alembert equations as generalizations of Hamilton's principle and the Euler-Lagrange equations, respectively, and thus explains the key differences from CDS 205 course material. Since the systems of concern are nonholonomic, I write constraints in velocity form; note, however, that this formulation can include also holonomic systems and is therefore general (given other assumptions such as scleronomic constraints). Definitions, derivations, and several observations herein are provided through Bloch [2].

3.1 The Lagrange-d'Alembert Principle

The Lagrange-d'Alembert variational principle determines the equations of motion for a constrained system. Let Q be the (n-manifold) configuration space let the constraint distribution D be determined by the system's m kinematic constraints, $m < n$. That is, D is a collection of linear subspaces $D_q \subset T_qQ$ such that a curve $q(t) \in Q$ satisfies the system's constraints iff $\dot{q} \in D_{q(t)}$ for all t. Finally, consider a Lagrangian $L : TQ \to \mathbb{R}$, $L(q^i, \dot{q}^i)$ for generalized coordinates q^i, $i = 1, ..., n$, on Q. Then the Lagrange-d'Alembert principle can be stated as follows:

The equations of motion for the system are determined by

$$\delta \int_a^b L(q^i, \dot{q}^i) dt = 0 \tag{7}$$

where we choose variations $\delta q(t)$ *of the curve* $q(t)$ *that satisfy* $\delta q(t) \in D_q(t)$ *for each* t, $a \leq t \leq b$, *and* $\delta q(a) = \delta q(b) = 0$.

Also, the curve $q(t)$ must satisfy the constraints to be a valid trajectory of the constrained system.

The statement of Hamilton's principle differs in the restrictions on the variations taken (here, they must be in the distribution whereas in Hamilton's principle they are arbitrary).

Bloch elaborates on the somewhat subtle but important point of *when* the constraints are imposed relative to taking the variations. For this principle, the variations δq are taken *before* imposing the constraints; so, although we restrict variations to lie in the distribution D in the principle, this is not equivalent to the curve $q(t)$ satisfying constraints.

On the other hand, imposing the constraints first results is equivalent to taking arbitrary variations on a Lagrangian that includes the constraints, multiplied by Lagrange multipliers. This approach, which is a common technique for finding the equations of motion for holonomically constrained systems, results in equations that are *not* associated with physical dynamics in general (but do play a role for optimal control problems). More on this topic is discussed later.

Taking variations gives us

$$\int_a^b \left(\frac{\partial L}{\partial q^i} \delta q^i + \frac{\partial L}{\partial \dot{q}^i} \delta \dot{q}^i \right) dt = 0.$$

Integrating by parts with fixed endpoints:

$$\int_a^b \left(\frac{\partial L}{\partial q^i} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} \right) \delta q^i dt = 0$$

Which gives us, from the arbitrariness of time intervals,

$$\left(\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} \right) \delta q^i = 0. \tag{8}$$

However, we *cannot* conclude $\frac{d}{dt} \frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0$ (the Euler-Lagrange equations) since the variations δq^i are not arbitrary but rather are stated to lie in the distribution. Thus, further analysis is required to develop the equations of motion from the principle.

3.2 Equations of Motion: Structure

Let $\omega^a(q)$ be a one-form whose vanishing describes the a^{th} system constraint, so that the constraints on $\delta q \in TQ$ are defined by $\omega^a \cdot v = 0$, $(a = 1, ..., m)$. We can then find local coordinates $q = (r, s) \in \mathbb{R}^{n-m} \times \mathbb{R}^m$ such that we can write the one-forms as

$$\omega^a(q) = ds^a + A^a_\alpha(r, s) dr^\alpha$$

So the constraints on δq become

$$\delta s^a + A^a_\alpha(r, s) \delta r^\alpha \tag{9}$$

One might think of varying the $\{r^\alpha\}$ coordinates $(\alpha = 1, ..., n - m)$ freely and then the above relation defines the corresponding variations on each of the m s coordinates so that the resultant variation is allowable (i.e. is in D_q). We use this arbitrariness of δr and combine equations (8) and (9) to get a complete description of the system's equations of motion:

$$\begin{aligned} \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{r}^\alpha} - \frac{\partial L}{\partial r^\alpha}\right) &= A^a_\alpha \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{s}^a} - \frac{\partial L}{\partial s^a}\right) && (10)\\ \dot{s}^a &= -A^a_\alpha(r,s)\dot{r}^\alpha && (11) \end{aligned}$$

Equation (10) gives $n-m$ second-order equations and (10) gives in m first-order equations representing the constraints.

3.3 Equations of Motion: Derivation with Lagrange Multipliers

A more common form of the equations of motion is derived here from the Lagrange-d'Alembert principle, using some of the intuition gained in the preceding subsection.

Using the form of (1) to describe the constraints, the constraints on the variations ($\delta q^i \in D$) can be written as

$$a^j_i(q)\delta q^i = 0.$$

Recall $i = 1, .., n$ and $j = 1, ..., m$. We will see it is useful to multiply the constraints by constants λ_j and combine the constraint equations into

$$\lambda_j a^j_i(q)\delta q^i = 0$$

or

$$\lambda A \delta q = 0$$

where $\lambda \in \mathbb{R}^m$ (row vector), $\delta q \in \mathbb{R}^n$, and $A = a^j_i(q) \in \mathbb{R}^{m\times n}$.

Since these terms are zero-valued, we can append them to equation (8) to obtain

$$\sum_{i=1}^{n}\left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} - \sum_{j=1}^{n}\lambda_j a^j_i(q)\right)\delta q^i = 0, \qquad (12)$$

adding the summation notation to be explicit. Although these steps have been trivial, their usefulness soon comes to light. In a similar fashion to the previous subsection, we attempt to make some directions of the variation arbitrary and use the constraints to specify what the other directions of the variations must be. Independence of the constraints implies at least one $m\times m$ minor of a^i_j is nonzero. Take this to correspond to the first m columns. The result: the variations $\delta q^{m+1}, ..., \delta q^n$ may be treated as arbitrary and variations $\delta q^1, ..., \delta q^m$ are fully determined by the constraints.

As noted earlier, we cannot say the first m terms of equation (12) (that is, the coefficients of the first m variations in the equation) vanish since these (dependent) variations are not arbitrary. However, we can make them vanish by appropriately choosing values of the Lagrange multipliers λ_i, a freedom we introduced. Finding the λ_i is feasible via a linear system of algebraic equations since we have independence of constraints. Then equation (12) reduces to

$$\sum_{i=m+1}^{n} \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} - \sum_{j=1}^{n} \lambda_j a_i^j(q) \right) \delta q^i = 0. \tag{13}$$

Now, however, we *do* have independent variations so each coefficient must vanish separately in the usual way. Thus our equations for the dependent and independent variables look identical and are written concisely as

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = \sum_{j=1}^{n} \lambda_j a_i^j(q) \tag{14}$$

or

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \lambda A(q) \tag{15}$$

where each side of the equation is a row vector – a cotangent vector to Q at the point $q(t)$. These equations are referred to as the *Lagrange-d'Alembert equations* or the *dynamic nonholonomic equations of motion* and are equivalent to Newton's law $F = ma$ with reaction forces.

Note that we have introduced the Lagrange multipliers and so have additional variables to solve for as part of the solution to the problem. There are $2n + m$ unknowns, $q, \dot{q}, \lambda$. The above equation gives n second-order equations (which is equivalent to $2n$ first-order equations). The remaining m equations are given by the constraints $A(q)\dot{q} = 0$.

3.4 Equations of Motion: Classical Derivation

For completeness, I briefly discuss here a common but flawed derivation of the Lagrange-d'Alembert equations found in many mechanics textbooks.

Lagrange's equations with generalized forces are

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = F, \tag{16}$$

where $F = [F_1, ..., F_n]$, similarly a cotangent vector, represents the nonconservative external forces. The generalized forces F_i can be found by $F_i = \left\langle F, \frac{\partial}{\partial q^i} \right\rangle$, where $\left\{ \frac{\partial}{\partial q^i} \right\}$ are the standard basis for the tangent space at q.

The following argument is then made to relate F to the system constraints $A(q)\dot{q} = 0$. Considering virtual displacements $\delta q \in D$, we have $A(q)\delta q = 0$. We then assume F must lie in the annihilator of the space of virtual displacements so that $F = \lambda A(q)$, a linear combination of the rows of $A(q)$. This assumption, referred to as the *nonholonomic principle*, results immediately in the Lagrange-d'Alembert equations

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}} - \frac{\partial L}{\partial q} = \lambda A(q).$$

As asserted in Bloch [2], the nonholonomic principle is not adequately justified, since no sufficient reasoning states F has to annihilate *all* possible virtual displacements. A conservation of energy argument does require $\langle F, \dot{q}\rangle = 0$, but this by itself does not provide the principle. Thus, the derivation in the preceding subsection is preferable.

3.5 Conservation of Energy

Energy of a nonholonomic constrained system can be written in terms of the Lagrangian in the familiar (holonomic) fashion,

$$E(q^i, \dot{q}^i) = \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L(q^i, \dot{q}^i). \tag{17}$$

Taking the time derivative of energy gives

$$\begin{aligned}
\frac{d}{dt}E(q^i, \dot{q}^i) &= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\dot{q}^i - L(q^i, \dot{q}^i)\right) \\
&= \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\dot{q}^i + \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i - \frac{\partial L}{\partial q^i}\dot{q}^i - \frac{\partial L}{\partial \dot{q}^i}\ddot{q}^i \\
&= \left(\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i}\right)\dot{q}^i \\
&= \lambda_j a_i^j(q)\dot{q}^i \\
&= \lambda_j(0) \\
&= 0.
\end{aligned}$$

So energy is conserved for nonholonomic systems, using equations of motion (14) and constraints (1).

This conservation is to be expected since the prescribed constraints do no work – there is no displacement in the direction of the constraint forces, as the constraints are preventing motion in those directions.

Despite the above conservation discussion, which also may be found as Proposition 1.3.3 in Bloch [2], Bloch later appears to contradict this proposition. For example, even in [2, page 213]:

> With this assumption, the total energy of the system is conserved, and conservation of energy indeed holds for many systems with nonholonomic constraints — for example, systems involving constraints of rolling without slipping.

This statement implies that conservation of energy does *not* hold for all systems with nonholonomic constraints. Additionally, it is stated in [1, page 225]

> In general, the constraint force associated with a nonholonomic constraint performs work. A special case when this is not valid is rolling without slipping....

The clearest resolution to this apparent conflict would be that the conservation (and accompanying rationale) I derived above apply to constraints of the type I have assumed, but not necessarily to all nonholonomic constraints. Recall I have assumed constraint equations to be purely linear in the velocity (and also time-independence, but this seems a safer assumption.) Comments and illustrative examples from the reader are welcome.

3.6 Variational Nonholonomic Equations

As noted earlier, the order in which variations are taken in the Lagrange-d'Alembert principle relative to the application of the constraints plays an important role for nonholonomic systems. Because of the method of introducing the constraints, it is stated the Lagrange-d'Alembert equations are not literally variational. A true variational approach is as follows: First construct the Lagrangian

$$L_\mu(q, \dot{q}) = L(q, \dot{q}) + \mu A \dot{q} \tag{18}$$

where μ are the Lagrange multipliers. (Aside: I use L_μ to distinguish the *Lagrangian with multipliers* from L_c the *constrained Lagrangian*, a term sometimes used to mean the same thing. The constrained Lagrangian, by contrast, does not contain Lagrange multipliers but rather substitutes in the constraint equations for the appropriate constrained velocities. In the language of section 3.2, $L_c(r, s, \dot{r}) = L(r, s, \dot{r}, -A_\alpha(r, s)\dot{r}^\alpha)$. The constrained Lagrangian is useful when focusing connections and momentum — an approach that will unfortunately not be discussed in this paper.)

Next, take variations as in Hamilton's principle with L_μ (or alternatively just form the Euler-Lagrange equations using this Lagrangian). Lagrange multipliers μ_j ($j = 1, ..., m$) may be determined by the constraints and initial conditions. Thus we impose the constraints on the velocity vectors of the class of *allowable curves*. By contrast, when deriving the Lagrange-d'Alembert equations, we imposed the constraints only on the variations. The resulting *variational nonholonomic equations* do not give the correct mechanical dynamical equations (which are provided by the dynamic nonholonomic equations) but address rather a problem in optimal control. However, for holonomic constraints, this variational approach does indeed result in equations identical to the dynamical equations — the reasoning for this is investigated via later examples. Indeed, to my understanding, the formulation of the Lagrangian with multipliers L_μ for use in the Euler-Lagrange equations appears the standard method for deriving equations of motion for holonomic systems.

4 Examples Illustrating Various Approaches

4.1 The Vertical Rolling Disk

Consider the vertical rolling disk example from section 2.2, and let us attempt an analysis from the different perspectives discussed above. Recall the constraints in form $A(q)\dot{q} = 0$ and chosen coordinates on $Q = \mathrm{SE}(2) \times S^1$ are

$$A(q) = \begin{pmatrix} 1 & 0 & 0 & -R\cos\varphi \\ 0 & 1 & 0 & -R\sin\varphi \end{pmatrix}, \ q = \begin{pmatrix} x \\ y \\ \varphi \\ \theta \end{pmatrix}.$$

4.1.1 Dynamic Nonholonomic Equations

We take the Lagrangian as the kinetic energy (note gravitational potential energy is constant):

$$L(x, y, \varphi, \theta, \dot{x}, \dot{y}, \dot{\varphi}, \dot{\theta}) = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) + \frac{1}{2}I\dot{\theta}^2 + \frac{1}{2}J\dot{\varphi}^2 \tag{19}$$

where m is the disk mass and I and J are the appropriate moments of inertia of the disk (perpendicular to the place of their respective corresponding angles). We have:

$$\begin{aligned} \lambda A(q) &= (\lambda_1, \lambda_2, 0, -\lambda_1 R\cos\varphi - \lambda_2 R\sin\varphi), \\ \frac{\partial L}{\partial \dot{q}} &= \left(m\dot{x}, m\dot{y}, J\dot{\varphi}, I\dot{\theta}\right). \end{aligned}$$

Then, using the dynamic nonholonomic equations (15) directly, we obtain:

$$m\ddot{x} = \lambda_1 \tag{20}$$

$$m\ddot{y} = \lambda_2 \tag{21}$$

$$J\ddot{\varphi} = 0 \tag{22}$$

$$I\ddot{\theta} = -\lambda_1 R\cos\varphi - \lambda_2 R\sin\varphi \tag{23}$$

Now we can eliminate the Lagrange multipliers for explicit equations of motion. Substituting the constraints into (20) and (21) gives

$$\begin{aligned} m\frac{d}{dt}\left(R\cos\varphi\,\dot{\theta}\right) &= \lambda_1 \\ m\frac{d}{dt}\left(R\sin\varphi\,\dot{\theta}\right) &= \lambda_2 \end{aligned}$$

and then using these expressions in (23) results in

$$\begin{aligned} I\ddot{\theta} &= -mR^2\cos\varphi\left(-\sin\varphi\,\dot{\theta}+\cos\varphi\,\ddot{\theta}\right)-mR^2\sin\varphi\left(\cos\varphi\,\dot{\theta}+\sin\varphi\,\ddot{\theta}\right) \\ &= -mR^2\ddot{\theta}\left(\cos^2\varphi+\sin^2\varphi\right) \\ &= -mR^2\ddot{\theta}. \end{aligned}$$

Therefore (22) and (23) can be written

$$J\ddot{\varphi} = 0 \tag{24}$$

$$\left(I+mR^2\right)\ddot{\theta} = 0, \tag{25}$$

and so $\dot{\varphi}$ and $\dot{\theta}$ are constant. Letting these constants be ω and Ω, respectively, we have a set of first-order equations

$$\begin{aligned} \dot{x} &= R(\cos\varphi)\dot{\theta} \\ \dot{y} &= R(\sin\varphi)\dot{\theta} \\ \dot{\varphi} &= \omega \\ \dot{\theta} &= \Omega, \end{aligned}$$

and so the equations of motion can be solved explicity, denoting initial conditions with $q(0)=q_0$, for

$$\begin{aligned} x &= \frac{\Omega}{\omega}R(\cos\varphi)+x_0 \\ y &= \frac{\Omega}{\omega}R(\sin\varphi)+y_0 \\ \varphi &= \omega t+\varphi_0 \\ \theta &= \Omega+\theta_0. \end{aligned}$$

4.1.2 Dynamic Equations without Using Lagrange Multipliers

To see what "falls out," let us the structure we derived earlier in section 3.2. Namely, start with equations (10) and (11). Choose $r=\left(r^1,r^2\right)=(\varphi,\theta)$ as the $n-m=4-2=2$

directions whose variations we consider unconstrained. Then $s = \left(s^1, s^2\right) = (x, y)$ are the $m = 2$ directions whose variations we ensure the resultant variation remains in the constraint distribution. Then the constraints are specified with

$$\begin{aligned} \dot{x} = R(\cos\varphi)\dot{\theta} \quad &\Rightarrow \quad A_1^1 = 0,\ A_2^1 = -R\cos\varphi \\ \dot{y} = R(\sin\varphi)\dot{\theta} \quad &\Rightarrow \quad A_1^2 = 0,\ A_2^2 = -R\sin\varphi. \end{aligned}$$

Then using equation (10) we write

$$\begin{aligned} \alpha = 1: \quad J\ddot{\varphi} \ &= 0 \\ \alpha = 2: \quad I\ddot{\theta} \ &= -mR\cos\varphi\, \ddot{x} - mR\sin\varphi\, \ddot{y} \\ &= -mR\cos\varphi \frac{d}{dt}\left(R\cos\varphi\, \dot{\theta}\right) - mR\sin\varphi \frac{d}{dt}\left(R\cos\varphi\, \dot{\theta}\right) \\ &= -mR^2\ddot{\theta} \end{aligned}$$

simplifying as before. So indeed we can obtain the same equations (somewhat more directly, as the substitutions served by the Lagrange multipliers have "already been carried out") by using the equations of this form.

4.1.3 Variational Nonholonomic Equations

To derive the variational equations, we start with the Lagrangian with multipliers as in equation (18)

$$L_\mu(x, y, \varphi, \theta, \dot{x}, \dot{y}, \dot{\varphi}, \dot{\theta}) = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) + \frac{1}{2}I\dot{\theta}^2 + \frac{1}{2}J\dot{\varphi}^2 + \mu_1\left(\dot{x} - R\dot{\theta}\cos\varphi\right) + \mu_2\left(\dot{y} - R\dot{\theta}\sin\varphi\right).$$

Variations of this Lagrangian are allowed over the whole tangent bundle rather than limiting to the constraint distribution. As this is equivalent to the Euler-Lagrange equations $\frac{d}{dt}\frac{\partial L}{\partial \dot{q}^i} - \frac{\partial L}{\partial q^i} = 0$, I will just use that result.

$$\begin{aligned} \frac{\partial L}{\partial \dot{q}} &= \left(m\dot{x} + \mu_1, m\dot{y} + \mu_2, J\dot{\varphi}, I\dot{\theta} - R\mu_1\cos\varphi - R\mu_2\sin\varphi\right) \\ \frac{\partial L}{\partial q} &= \left(0, 0, R\mu_1\dot{\theta}\sin\varphi - R\mu_2\dot{\theta}\cos\varphi, 0\right). \end{aligned}$$

Thus Euler-Lagrange gives the equations

$$m\ddot{x} + \dot{\mu}_1 = 0 \tag{26}$$

$$m\ddot{y} + \dot{\mu}_2 = 0 \tag{27}$$

$$J\ddot{\varphi} - R\mu_1\dot{\theta}\sin\varphi - R\mu_2\dot{\theta}\cos\varphi = 0 \tag{28}$$

$$I\ddot{\theta} - R\frac{d}{dt}\left(\mu_1\cos\varphi - \mu_2\sin\varphi\right) = 0 \tag{29}$$

Integrating equations (26) and (26) gives

$$\begin{aligned}\mu_1 &= -m\dot{x} + C_1 = -mR\dot{\theta}\cos\varphi + C_1 \\ \mu_2 &= -m\dot{y} + C_2 = -mR\dot{\theta}\sin\varphi + C_2\end{aligned}$$

for integration constants C_1 and C_2 and where the constrain equations have been used. Substituting into equations (28) and (29) and simplifying gives

$$\begin{aligned}J\ddot{\varphi} &= R\dot{\theta}\left(C_1\sin\varphi - C_2\cos\varphi\right) \\ \left(I + mR^2\right)\ddot{\theta} &= R\dot{\varphi}\left(-C_1\sin\varphi + C_2\cos\varphi\right).\end{aligned}$$

By comparing with equations (24) and (25) from the dynamic equations, we see that indeed the variational method has given different equations, provided the integration constants are nonzero. Interestingly, C_1 and C_2 cannot be determined from initial conditions and constraints; we interpret this as indicating the multiplicity of possible *variational* nonholonomic trajectories $q(t)$ for given initial conditions. The fact that choosing $C_1 = C_2 = 0$ yields the dynamic nonholonomic equations is not true in general but is particular to a class of problems to which the vertical rolling disk belongs.

4.2 The Simple Pendulum

I briefly include this well-known holonomic example to illustrate both the applicability of the dynamic nonholonomic equations to the holonomic case and the equivalence of the dynamic and variational approaches when the constraints are holonomic. Choose coordinates $q = (x, y) \in \mathbb{R}^2$ and write the constraint as

$$x\dot{x} + y\dot{y} = 0, \text{ or } \dot{x} = -\frac{y}{x}\dot{y}$$

The Lagrangian is (assuming unit mass for brevity) kinetic minus potential energy

$$L(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2\right) - gy.$$

4.2.1 Dynamical Equations

Using the Lagrange-d'Alembert equations with the constraint matrix

$$A(x, y) = \begin{pmatrix} x & y \end{pmatrix}$$

gives immediately

$$\left\{\begin{aligned}\ddot{x} &= \lambda x \\ \ddot{y} + g &= \lambda y,\end{aligned}\right. \tag{30}$$

or

$$\left\{ \begin{array}{rcl} \ddot{x} - \lambda x &=& 0 \\ \ddot{y} - \lambda y &=& -g. \end{array} \right. \tag{31}$$

We can show $\lambda < 0$ and thus these equations of motion are the expected harmonic oscillator with gravitational forcing. Note that the Lagrange multiplier has the physical interpretation of the frequency of oscillation.

4.2.2 Variational Equations

Construct the Lagrangian with multipliers

$$L_\mu(x, y, \dot{x}, \dot{y}) = \frac{1}{2}\left(\dot{x}^2 + \dot{y}^2\right) - gy + \mu\left(x\dot{x} + y\dot{y}\right)$$

where here μ is a scalar as there is only one constraint. Instead of simply plugging into Euler-Lagrange, we could take variations directly:

$$\begin{aligned} \delta \int L_\mu dt &= 0 \\ \int \left(\dot{x}\delta\dot{x} + \dot{y}\delta\dot{y} - g\delta y + \mu\left(\dot{x}\delta x + x\delta\dot{x} + \dot{y}\delta y + y\delta\dot{y}\right)\right) dt &= 0 \\ \int \left[-\left(\ddot{x} + \dot{\mu}x + \mu\dot{x}\right)\delta x + \mu\dot{x}\delta x - \left(\ddot{y} + \dot{\mu}y + \mu\dot{y}\right)\delta y + \mu\dot{y}\delta y - g\delta y\right] dt &= 0 \\ \int \left[-\left(\ddot{x} + \dot{\mu}x\right)\delta x - \left(\ddot{y} + \dot{\mu}y + g\right)\delta y\right] dt &= 0 \end{aligned}$$

With the assumptions that the δx and δy variations are independent, this gives

$$\left\{ \begin{array}{rcl} \ddot{x} + \dot{\mu}x &=& 0 \\ \ddot{y} + \dot{\mu}y &=& -g \end{array} \right. \tag{32}$$

Note the equivalence of these equations with equations (31) with $\lambda = -\dot{\mu}$ (a relationship related to the integrability of the constraints, or perhaps more precisely the assumption of integrability when these two approaches are used equivalently).

5 Conclusion

As we have seen, the nonholonomic nature of some constraints has several key consequences in fundamental geometric mechanics. The nonintegrability of such constraint distributions prevents us from simply taking a submanifold of the configuration space and thus we must continue to work on the tangent bundle (or actually the constraint subbundle). Thus, variational approaches must be modified to account for a smaller set of allowable motions, during which the subtleties of when and how variations are taken and constraints are imposed must be treated with care. Still, the issue is clearly not insurmountable, and concise, convenient equations of motion may be obtained.

Several areas of interest could follow this project as natural extensions. Indeed, this project has barely scratched the surface of nonholonomic system analysis. Such examples (from which time constraints prevented further investigation) include:

- a discussion at Poisson and Hamiltonian perspectives and their equivalence to the approach covered so far,
- the topic of connections, which appears to be crucial to an intrinsic (coordinate free) formulation of the above topics,
- the role of symmetry,
- momentum maps, momentum equations, and the nonconservation of momentum (apparently required for the controlled motion of some systems),
- control analysis, including the addition of control forces, optimal control and its relationship to the variational problem, and the key relationships of controllability and nonintegrability of constraints.

References

[1] Baruh, H. [1999] *Analytical Dynamics*, McGraw-Hill.

[2] Bloch, A. M., J. Baillieul, P. E. Crouch, and J. E. Marsden [2002], *Nonholonomic Mechanics and Control*, Springer-Verlag.

[3] Lewis, A. D., J. P. Ostrowski, R. M. Murray, and J.W. Burdick [1994], Nonholonomic mechanics and locomotion: the Snakeboard example, in *Proceedings of the IEEE International Conference on Robotics and Automation*, 2391-2400.

[4] Marsden, J.E. and T.S. Ratiu [1994] *Introduction to Mechanics and Symmetry*, Texts in Applied Mathematics, 17, Springer-Verlag; (Second Edition, 2002).

[5] Marsden, J.E., T.S. Ratiu, and R. Abraham [2001] *Manifolds, Tensors, Analysis, and Applications*, Springer-Verlag. (Third Edition, Draft, 2004)

I would also like to acknowledge the extensive efforts of Nawaf Bou-Rabee for subject matter clarifications, collaborative discussions, and personal encouragement.