Nonholonomic Constraints

Mattias Flygare
KAU Department of Physics and Electrical Engineering
Examiner: Jürgen Fuchs
(Dated: January 15, 2012)

CONTENTS

I. INTRODUCTION

When studying mechanical systems using either the Lagrangian or the Hamiltonian formalism it is often useful to consider generalised ways to treat constraint forces acting on the system. It is well known how to do this for so called holonomic constraints (constraints expressible as a function of the generalised coordinates and time):

$$f(\mathbf{q}; t) = 0 \tag{1}$$

However, when dealing with nonholonomic constraints (constraints that are not holonomic), it is normally not possible to give general formulas to derive the equations of state. Instead we have to look at some special cases of constraints which may be useful for commonly occurring types of constraints. There are also several pitfalls to avoid when dealing with nonholonomic constraints, which can render the wrong equations of motion because of hidden mistakes in the derivation of the formulas. There has historically been a controversy as to which method produces the correct formulas, and even in recent years several publications have published what is generally considered (and often also proved) to be the wrong results, derived from a faulty principle (as pointed out by Flannery in [5]). This text has therefore become as much of a discussion on **what not to do** when it comes to taking on nonholonomic systems as it is on **what to do**.

The main ambition of this text is to clarify and classify some of the different types of constraints that are often encountered and to give some general path to follow for obtaining the solution of the system in question, or simply come to the conclusion that a particular type of constraint is outside of the scope of our formulation of the problem.

This text follows the reasonings (and will present some results from) from the articles by Flannery and Koon and Marsden.

II. NONHOLONOMIC?

Describing nonholonomic constraints as not holonomic constraints might not be very helpful (even though accurate). To be more specific, when a path integral is computed in a nonholonomic system, the value represents a deviation and is said to be an anholonomy produced by the specific path taken. In other words, a nonholonomic system is a system whose state depends on the path taken to achieve it [3]. A clarifying example of what this means is given below, inspired by the example given by [7]. These types of constraints often arise in mechanical systems with rolling or sliding contact, but can also occur in less obvious ways [2].

II.1. Example: The rolling sphere

The rolling sphere is a very simple example that demonstrates what a nonholonomic system might be.

Take a sphere and place it on a level surface. The x-y position of the center of the sphere describes the position and together with two points on the sphere (describing the rotation of the sphere) this describes the state of the system. Take the constraints to be:

1. The sphere rolls on the surface without slipping
2. The sphere is never rotated about its z-axis

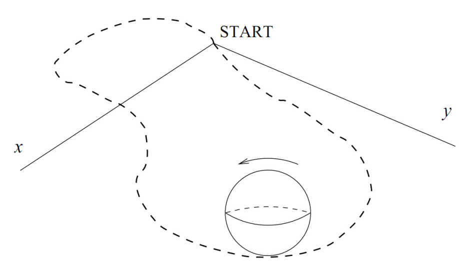

FIG. 1. A rolling sphere on a flat surface

It can now be shown that by rolling the sphere in any closed path (that is, back to the same x-y position as it started) you can obtain any and all of the systems possible rotation-states. Since the sphere can be rolled into any position we see that any state can be obtained without violating the constraints. Hence the system state depends on the path taken to achieve it, which is by the above definition a nonholonomic system.

II.2. Other constraints

A nonholonomic system could mean:

1. A system with constraints that can not be given by an equation at all
2. A system where line integrals depend on not just the start- and endpoints (as with holonomic systems) but also the path taken, that is, the system is nonintegrable
3. A system with constraint on the generalised velocities that are not derivable from position constraints

$$g(\mathbf{q}, \dot{\mathbf{q}}; t) = 0 \tag{2}$$

III. D'ALEMBERT'S AND HAMILTON'S PRINCIPLES

The results following will all rest on **d'Alembert's basic principle** which is derived staring from newtons second law and the assumption that **the constraining forces performs no virtual work**. This is the first restriction that is imposed on the constraining forces that is treated in this text. The steps to follow to arrive at d'Alembert's basic principle are standard, and can be found in [5]. The principle is:

$$\sum_{j=1}^{n} \left[\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} - Q^{NP} \right] \delta q_j = 0 \tag{3}$$

Where L is

$$L(\mathbf{q}, \dot{\mathbf{q}}; t) = T(\mathbf{q}, \dot{\mathbf{q}}; t) - V(\mathbf{q}, \dot{\mathbf{q}}; t) \tag{4}$$

Here T is the total kinetic energy, V is the potential for all conservative forces and Q^{NP} are the non-potential forces. This principle can also be written in δL version:

$$\delta L = \frac{d}{dt}(p_j \delta q_j) - Q^{NP} \delta q_j \tag{5}$$

This new equation is equivalent to (3) and can be used to derive **Hamilton's variational principle**

$$\delta S = \delta \int_1^2 L \; dt = 0 \tag{6}$$

by only considering systems where $Q^{NP} = 0$.

IV. HOLONOMIC CONSTRAINTS

For simplicity we will only look at systems for which $Q^{NP} = 0$ although all below derived formulas from d'Alembert's basic principle can easily be extended to include Q^{NP} (except formulas that rely on Hamilton's variational principle). When trying to solve a system with n coordinates

$$\mathbf{q} = \mathbf{q}(q_1, q_2, ..., q_n) \tag{7}$$

and c holonomic constraints

$$f_k(\mathbf{q}; t) = 0, \qquad (k = 1, 2, ..., c) \tag{8}$$

there are basically two paths you can take:

1. Use the constraint equations to directly eliminate the c dependent coordinates and end up with a system of n-c independent coordinates. Since all terms in (3) are now independent, we get:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q_j}}\right)-\frac{\partial L}{\partial q_j}=0, \qquad (j=1,2,...,n-c) \tag{9}$$

2. Include the constraints in the Lagrangian, thereby extending the system to n + c dependent variables (where the new c variables are the Lagrange multipliers). This might be useful if you do not want to (or can not) directly eliminate the dependant variables or if you want to know what the forces of constraint are. These will now be given by the new variables.

To begin, take

$$L^* = L(\mathbf{q},\dot{\mathbf{q}};t)+\sum_{k=1}^{c}\lambda_k(t)f_k(\mathbf{q};t) \tag{10}$$

The new terms may be added since

$$\frac{\partial f_k}{\partial \dot{q_j}}=0, \qquad (k=1,2,...,c) \tag{11}$$

and by differentiating (8), and setting dt to zero we get

$$\sum_{j=1}^{n}\frac{\partial f_k}{\partial q_j}\delta q_j=0, \qquad (k=1,2,...,c) \tag{12}$$

The new terms are thus invariant to an infinitesimal displacement and can be added to the Lagrangian. (3) now becomes

$$\sum_{j=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q_j}}\right)-\frac{\partial L}{\partial q_j}-\sum_{k=1}^{c}\lambda_k(t)\frac{\partial f_k}{\partial q_j}(\mathbf{q};t)\right]\delta q_j=0 \tag{13}$$

Once the λ_k's are chosen cancel all the c terms of Lagrangian derivatives that are connected to the c dependent q's, then the n equations of motion for all dependent and independent variables are given by

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q_j}}\right)-\frac{\partial L}{\partial q_j}=\sum_{k=1}^{c}\lambda_k(t)\frac{\partial f_k}{\partial q_j}(\mathbf{q};t), \qquad (j=1,2,...,n) \tag{14}$$

The remaining c equations required to solve the n + c system are simply the constraint equations (8).

Equivalent equations of state are obtained by taking L^* and applying (6).

Since we have used the multiplier theorem to obtain these equations of state we are now restricted only to constraints where all possible displaced paths δq are geometrically possible and $\delta g = 0$. This is no problem for holonomic constraints since $\delta f = 0$ but as we shall see, it will be a problem when considering nonholonomic constraints.

V. NONHOLONOMIC CONSTRAINTS

Letting the constraints be a part of the Lagrangian and thus letting them go through the variational part of the derivation of equations means that we have to guarantee that the variation of the constraint is zero, not just for the correct path, but for ALL possible displacement paths. When it comes to constraints involving the velocities this is not impossible but far from the standard case. In fact, when the constraints are not linear in the velocities it is impossible to use the linear arguments used in all the principles discussed here, so in these cases we can not say anything about the forces from constraints.

V.1. Linear constraints

When the constraints are linear in the velocities (commonly occurring in rolling without slipping for instance) they fit better into the picture. By first arriving at d'Alembert's basic principle (3) and now introducing constraint forces we can get some results.

Consider a system of n generalised coordinates with c linear constraint forces acting on them, described by the equations

$$g(\mathbf{q},\dot{\mathbf{q}};t)=\sum_{j=1}^{n}A_{kj}(\mathbf{q};t)\dot{q}_j+B_k(\mathbf{q};t)=0,\qquad (k=1,2,...,c) \tag{15}$$

By rewriting (15) in differential form

$$g(\mathbf{q},\dot{\mathbf{q}};t)dt=\sum_{j=1}^{n}A_{kj}(\mathbf{q};t)dq_j+B_k(\mathbf{q};t)dt \tag{16}$$

and displacing with dq=δq in zero time, dt = 0, we get

$$\sum_{j=1}^{n}A_{kj}(\mathbf{q};t)\delta q_j=0,\qquad (k=1,2,...,c) \tag{17}$$

This allows us to multiply each $A_{kj}\delta q_j$ with λ_k and adding them to the right side of (3). After the same reasonings as above we end up with

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right)-\frac{\partial L}{\partial q_j}=\sum_{k=1}^{c}\lambda_k(t)A_{kj}(\mathbf{q};t),\qquad (j=1,2,...,n) \tag{18}$$

These n equations together with the constraint equations (15) give the correct equations of motion.

V.2. Finding a general formula

So what will happen if we try to follow the same path as with holonomic constraints in the section above? We take a new set of multipliers μ_k, $(k=1,2,...,c)$ and add the constraints to the Lagrangian:

$$L^*=L(\mathbf{q},\dot{\mathbf{q}};t)+\sum_{k=1}^{c}\mu_k(t)g_k(\mathbf{q},\dot{\mathbf{q}};t) \tag{19}$$

$$\frac{d}{dt}\left(\frac{\partial L^*}{\partial \dot{q}_j}\right)=\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right)+\sum_{k=1}^{c}\mu_k(t)\frac{d}{dt}\left(\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial \dot{q}_j}\right)+\sum_{k=1}^{c}\dot{\mu}_k(t)\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial \dot{q}_j} \tag{20}$$

$$\frac{\partial L^*}{\partial q_j}=\frac{\partial L}{\partial q_j}+\sum_{k=1}^{c}\mu_k(t)\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial q_j} \tag{21}$$

The three g-derivatives are:

$$\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial \dot{q}_j}=A_{kj}(\mathbf{q};t) \tag{22}$$

$$\frac{d}{dt}\left(\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial \dot{q}_j}\right)=\frac{d}{dt}\left(A_{kj}(\mathbf{q};t)\right)=\sum_{i=1}^{n}\frac{\partial A_{kj}(\mathbf{q};t)}{\partial q_i}\dot{q}_i+\frac{\partial A_{kj}(\mathbf{q};t)}{\partial t} \tag{23}$$

$$\frac{\partial g_k(\mathbf{q},\dot{\mathbf{q}};t)}{\partial q_j} = \sum_{i=1}^{n} \frac{\partial A_{ki}(\mathbf{q};t)}{\partial q_j}\dot{q}_i + \frac{\partial B_k(\mathbf{q};t)}{\partial q_j} \tag{24}$$

Putting (22), (23) and (24) into (20) and (21) and using this new Lagrangian in d'Alembert's basic principle (3) we arrive at

$$\sum_{j=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} + \sum_{k=1}^{c}\dot{\mu}_k A_{kj} + \sum_{k=1}^{c}\mu_k\left(\sum_{i=1}^{n}\left(\frac{\partial A_{kj}}{\partial q_i} - \frac{\partial A_{ki}}{\partial q_j}\right)\dot{q}_i + \frac{\partial A_{kj}}{\partial t} - \frac{\partial B_k}{\partial q_j}\right)\right]\delta q_j = 0 \tag{25}$$

Comparing this to equation (18) we see that, since the constraint forces are some linear combination of the A_{jk}'s, then $\dot{\mu}_k = -\lambda_k$ but there are extra terms preventing us from getting the correct equations of motion, as in (18).

V.3. Determining the restrictions

For the approach of including the constraints together with a multiplier in the Lagrangian to work we must now have these extra terms cancel each other. In other words, we need that

$$\sum_{i=1}^{n}\left(\frac{\partial A_{kj}}{\partial q_i} - \frac{\partial A_{ki}}{\partial q_j}\right)\dot{q}_i + \frac{\partial A_{kj}}{\partial t} - \frac{\partial B_k}{\partial q_j} = 0, \qquad (j = 1,2,...,n),(k = 1,2,...,c) \tag{26}$$

Now, if the constraint function g can be written as the total time derivative of another function, f

$$g_k(\mathbf{q},\dot{\mathbf{q}};t) = \sum_{j=1}^{n} A_{kj}(\mathbf{q};t)\dot{q}_j + B_k(\mathbf{q};t) = \frac{d}{dt}f_k(\mathbf{q};t) = \sum_{j=1}^{n}\frac{\partial f_k(\mathbf{q};t)}{\partial q_j}\dot{q}_j + \frac{\partial f_k(\mathbf{q};t)}{\partial t} \tag{27}$$

we see that

$$A_{kj} = \frac{\partial f_k}{\partial q_j} \tag{28}$$

and

$$B_k = \frac{\partial f_k}{\partial t} \tag{29}$$

so the left hand side of equation (26) becomes

$$\sum_{i=1}^{n}\left(\frac{\partial^2 f_k}{\partial q_j \partial q_i} - \frac{\partial^2 f_k}{\partial q_i \partial q_j}\right)\dot{q}_i + \frac{\partial^2 f_k}{\partial q_j \partial t} - \frac{\partial^2 f_k}{\partial t \partial q_j} = \sum_{i=1}^{n}\left(\frac{\partial^2 f_k}{\partial q_i \partial q_j} - \frac{\partial^2 f_k}{\partial q_i \partial q_j}\right)\dot{q}_i + \frac{\partial^2 f_k}{\partial q_j \partial t} - \frac{\partial^2 f_k}{\partial q_j \partial t} = 0,\ (j = 1,...,n),(k = 1,...,c) \tag{30}$$

The conclusion is that if $g = g(\mathbf{q},\dot{\mathbf{q}};t)$ is the total time derivative of some function $f = f(\mathbf{q};t)$, i.e. that equation (15) is **exact**, then the problem can be solved (even if only the function g is given and f is unknown) by including the constraints in the Lagrangian and proceeding as above. Linear nonholonomic constraints that are expressed in exact form are integrable and therefore actually holonomic. The term **semiholonomic** is commonly used for such constraints and they also include constraints that are not in exact form but can be put in one by multiplying with an integrating factor.

This interesting result shows that the generalisation of formulas for equations of motion with constraint forces are restricted to holonomic and semiholonomic constraints (which are really holonomic but in another form).

VI. HAMILTONIAN APPROACH

In their article "The Hamiltonian and Lagrangian Approaches to the Dynamics of Nonholonomic Systems" Koon and Marsden [8] also review the Hamiltonian formulation of the problem. The reasonings they use include some concepts from differential geometry that are outside the scope of this text. The Hamiltonian equations of motion are however derived below, taking into account a linear nonholonomic constraint of the form (15). We start from the correct equations of state,(18) and:

$$p_j = \frac{\partial L}{\partial \dot{q}_j} \tag{31}$$

so that (18) becomes

$$\dot{p}_j - \frac{\partial L}{\partial q_j} = \sum_{k=1}^{c} \lambda_k(t) A_{kj}(\mathbf{q};t) \tag{32}$$

We know that

$$dL = \sum_{j=1}^{n} \left(\frac{\partial L}{\partial q_j} dq_j + \frac{\partial L}{\partial \dot{q}_j} d\dot{q}_j \right) + \frac{\partial L}{\partial t} dt \tag{33}$$

and inserting p_j and $\dot{p}_j$ from (31) and (32) we get

$$dL = \sum_{j=1}^{n} \left[\left(\dot{p}_j - \sum_{k=1}^{c} \lambda_k(t) A_{kj}(\mathbf{q};t) \right) dq_j + p_j d\dot{q}_j \right] + \frac{\partial L}{\partial t} dt \tag{34}$$

Now, the Hamiltonian is defined in the usual way as

$$H = H(\mathbf{q}, \dot{\mathbf{q}}; t) = \sum_{j=1}^{n} p_j \dot{q}_j - L(\mathbf{q}, \dot{\mathbf{q}}; t) \tag{35}$$

so we get

$$dH = \sum_{j=1}^{n} \left[\dot{q}_j dp_j + \left(\sum_{k=1}^{c} \lambda_k(t) A_{kj}(\mathbf{q};t) - \dot{p}_j \right) dq_j \right] - \frac{\partial L}{\partial t} dt \tag{36}$$

After the substitution of variables in H (the $\dot{q}$'s are replaced by the new p's) we have

$$H = H(\mathbf{q}, \mathbf{p}; t) \tag{37}$$

and

$$dH = \sum_{j=1}^{n} \left[\frac{\partial H}{\partial p_j} dp_j + \frac{\partial H}{\partial q_j} dq_j \right] + \frac{\partial H}{\partial t} dt \tag{38}$$

So by linearity in dp_j and dq_j we get the new Hamilton equation of motion:

$$\frac{\partial H}{\partial p_j} = \dot{q}_j \tag{39a}$$

$$\frac{\partial H}{\partial q_j} = -\dot{p}_j + \sum_{k=1}^{c} \lambda_k(t) A_{kj}(\mathbf{q};t) \tag{39b}$$

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \tag{39c}$$

VII. CONCLUSIONS

We have seen that general constraints on the velocity are not compatible with the principles used to derive the equations of motion from some constraint equation. Even when these equations are linear in the velocities some problems may occur and the generalised formulas that handle constraints by including them in the Lagrangian are totally unsuited for all true nonholonomic constraints.

VII.1. Classification of systems with constraint forces

Type of constraint	**Approach**	**Advantages**
Holonomic	Use the constraint equations to eliminate the dependent variables directly	If you are not interested in the constraint forces this is usually the most straightforward approach
	Include the constraints in the Lagrangian with multipliers	If you want to solve for the forces of constraint or if you can't (or don't want to) eliminate the constraints directly
Semiholonomic (linear and exact)	If you can find a function f from your given function g, proceed as with holonomic constraints	Depending on what you are after, in this way you can treat the system as holonomic
	Using multipliers you can include the constraints in the Lagrangian	You can now solve for the constraint forces
Nonholonomic with linear velocities	No generalised principle works, only d'Alembert's basic principle give the proper equations of motion	
Nonholonomic with nonlinear velocities	None of the above principles work	

VII.2. A counter-intuitive example

As mentioned before, which of the equations of motion are really the correct ones have been debated and the question still gives rise to some controversy. One of the reasons for this is that some of the results are counter intuitive. The example below (from [2]) has been presented in more detail and simulated by [4]. A video showing the curious behaviour for real can also be seen on YouTube: [1].

The example is similar to the example from II.1 but now the sphere is rolling without slipping (the sphere is assumed to have enough velocity to roll without slipping) inside of a vertical hollow cylinder (a ping-pong ball inside a tennis ball container for instance). As the ball rolls around the tube it initially rolls downward (as would be expected) but the surprising result is that after some distance it changes direction and moves back up. Finally it oscillates between two heights of the cylinder in a simple harmonic manner. The effect might be the same effect that makes a golf ball initially roll down in the hole but then pop back up.

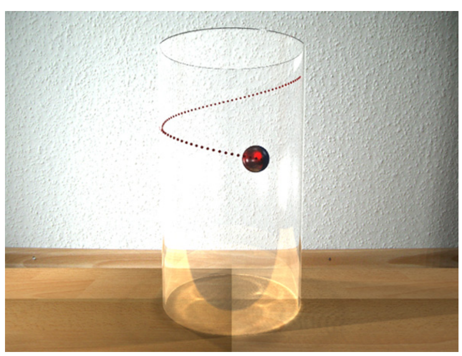

FIG. 2. A rolling sphere inside a cylinder

VII.3. Summary

The difference in the equations of motion rendered by the different approaches was the start of the confusion as to whether or not the equations of motion can be derived from a variational principle. Since including the constraints before taking the variations works for some systems (holonomic and semiholonomic) it would be easy to assume it works for all constraints, but as we have seen this is not the case.

Imposing the constraints after taking the variation is the proper way to proceed when dealing with nonholonomic constraints and to quote A.M. Bloch: [2, *"there is no doubt that the correct equations of motion for nonholonomic* *mechanical systems are given by the Lagrange-d'Alembert principle"*].

Even in the recent years there have been serious publications giving generalised formulas by variational principle for nonholonomic constraints without pointing out the restrictions these generalisations need and the confusion when it comes to nonholonomic constraints and the terminology that come with it is still strong.

Hopefully this text has not added to the confusion but instead cleared up some of the common unclarities.

---

[1] Ball in cylinder. URL http://www.youtube.com/watch?v=1t1grbgT5pE.

[2] A.M. Bloch. *Nonholonomic Mechanics and Control*. Springer Science, 2003.

[3] Wikipedia contributors. Nonholonomic system, January 2012. URL http://en.wikipedia.org/wiki/Nonholonomic_system.

[4] Dan Curtis, Alexi Radovinsky, and Stan Wagon. Rolling ball inside a cylinder (from the wolfram demonstrations project). URL http://demonstrations.wolfram.com/RollingBallInsideACylinder/.

[5] M. R. Flannery. The enigma of nonholonomic constraints. *American Journal of Physics*, Vol. 73, Issue 3:pages 265–272, 2005.

[6] H. Goldstein, C. Poole, and J. Safko. *Classical Mechanics*. Pearson Education, 2002.

[7] B.D. Johnson. The nonholonomy of the rolling sphere. URL http://euler.slu.edu/~johnson/public/maths/bdjsphere.pdf.

[8] W. S. Koon and J. E. Marsden. The hamiltonian and lagrangian approaches to the dynamics of nonholonomic systems. *Reports on Mathematical Physics*, Vol. 40:pages 21–62, 1997.