New General Principle of Mechanics and Its Application to General Nonideal Nonholonomic Systems

Firdaus E. Udwadia[1]

Abstract: In this paper we develop a general minimum principle of analytical dynamics that is applicable to nonideal constraints. The new principle encompasses Gauss's Principle of Least Constraint. We use this principle to obtain the general, explicit, equations of motion for holonomically and/or nonholonomically constrained systems with non-ideal constraints. Examples of a nonholonomically constrained system where the constraints are nonideal, and of a system with sliding friction, are presented.

DOI: 10.1061/(ASCE)0733-9399(2005)131:4(444)

CE Database subject headings: Constraints; Equations of motion; Mechanical systems; Friction.

Introduction

The motion of complex mechanical systems is often mathematically modeled by what we call their equations of motion. Several formalisms [Lagrange's equations (Lagrange 1787), Gibbs–Appell equations (Gibbs 1879, Appell 1899), generalized inverse equations (Udwadia and Kalaba 1992)] have been developed for obtaining the equations of motion for such structural and mechanical systems. Though these formalisms do not all afford the same ease of use in any given practical situation, they are equivalent to one another. They all rely on D'Alembert's principle which states that, at each instant of time during the motion of the mechanical system, the sum total of the work done by the forces of constraint under virtual displacements is zero. Such forces of constraint are often referred to as being ideal. D'Alembert's principle is equivalent to a principle that was first stated by Gauss (Gauss 1829) and is referred to nowadays as Gauss's principle of least constraint. In fact, like D'Alembert's principle, Gauss's principle can be thought of as a starting point from which the machinery of analytical dynamics can be developed (see Udwadia and Kalaba 1996). For example, it has been used in Udwadia and Kalaba (1992) and Kalaba and Udwadia (1993), in conjunction with the concept of the Penrose inverse of a matrix, to obtain a simple and general set of equations for holonomically and nonholonomically constrained mechanical systems when the forces of constraint are ideal.

Though these two fundamental principles of mechanics are often useful to adequately model mechanical system there are, however, numerous situations where they are not applicable since the constraint forces actually *do* do work under virtual displacements. Such systems have, to date, been left outside the perview of the Lagrangian framework. As stated by Goldstein (1981, p. 14) "This [total work done by forces of constraint equal to zero] is no longer true if sliding friction is present, and we must exclude such systems from our [Lagrangian] formulation." And Pars (1979) in his treatise on analytical dynamics writes, "There are in fact systems for which the principle enunciated [D'Alembert's principle]… does not hold. But such systems will not be considered in this book." Newtonian approaches are usually used to deal with the problem of sliding friction (Goldstein 1981). For general systems with nonholonomic constraints, the inclusion into the framework of Lagrangian dynamics of constraint forces that *do* work has remained to date an open problem in analytical dynamics, because neither D'Alembert's principle nor Gauss's principle is then applicable.

In this paper we obtain a general principle of analytical dynamics that encompasses nonideal constraints. It extends Gauss' principle to situations where the forces of constraint do *do* work under virtual displacements. It therefore brings nonideal constraints within the scope of Lagrangian mechanics. The power of the new principle is exhibited by the simple and straightforward manner in which we obtain the general, explicit equations of motion for holonomically and nonholonomically constrained mechanical systems where the constraints may not be ideal. We provide two illustrative examples. The first deals with a generalization of a problem first proposed by Appell in which we obtain the explicit equations of motion for a nonholonomic mechanical system with nonideal constraints; the second deals with sliding friction.

The paper is organized as follows. In the next section we present a statement of the problem and establish our notation. This is followed by the section in which we derive our new general principle of mechanics applicable to nonideal constraints. The section titled, "General Equations of Motion for Holonomic and Nonholonomic Systems with Nonideal Constraints," illustrates the power of this new principle to obtain the general, explicit equations of motion where the constraints may be nonideal. In the "Illustrative Examples" section we present two examples showing the simplicity with which the general equation of motion yields results for nonholonomic, nonideal constraints, and for problems with Coulomb friction. The last section gives the conclusions.

[1]Professor, Dept. of Aerospace and Mechanical Engineering, Civil Engineering, Mathematics, and Operations and Information Management, 430K Olin Hall, Univ. of Southern California, Los Angeles, CA 90089-1453. E-mail: fudwadia@usc.edu

Note. Associate Editor: Joel P. Conte. Discussion open until September 1, 2005. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on November 10, 2003; approved on September 28, 2004. This paper is part of the ***Journal of Engineering Mechanics***, Vol. 131, No. 4, April 1, 2005. ©ASCE, ISSN 0733-9399/2005/4-444–450/$25.00.

Statement of Problem of Constrained Motion with Nonideal Constraints

Consider a mechanical system comprised of n particles, of mass m_i, $i=1,2,3,\ldots,n$. We shall consider an inertial Cartesian coordinate frame of reference and describe the position of the jth particle in this frame by its three coordinates x_j, y_j, and z_j. Let the "impressed" forces on the jth mass in the X, Y, and Z directions be given F_{x_j}, F_{y_j}, and F_{z_j}, respectively. Then the equation of motion the "unconstrained" mechanical system can be written as

$$M\ddot{x} = F(x,\dot{x},t)$$

$$x(0) = x_0,\ \dot{x}(0) = \dot{x}_0 \tag{1}$$

Here the matrix M is a $3n$ by $3n$ diagonal matrix with the masses of the particles in sets of three along the diagonal; $x=[x_1,y_1,z_1,\ldots,x_n,y_n,z_n]^T$; and the dots refer to differentiation with respect to time. Similarly, the $3n$-vector ($3n$ by *1* column vector) $F=[F_{x_1},F_{y_1},F_{z_1},\ldots,F_{x_n},F_{y_n},F_{z_n}]^T$. By "unconstrained" we mean that the components of the $3n$-vector of velocity, $\dot{x}(0)$, can be independently prescribed. We note that the acceleration, $a(t)$, of the unconstrained system is then simply given by

$$a(t) = M^{-1}F(x,\dot{x},t) \tag{2}$$

The impressed force, $F(x,\dot{x},t)$, which has $3n$ components is a known vector, i.e., it is a known function of its arguments. The matrix M has positive entries along its diagonal; it is therefore symmetric and positive definite, as is the matrix M^{-1}.

Let this system now be subjected to a further set of $m=h+s$ constraints of the form

$$\varphi(x,t) = 0 \tag{3}$$

and

$$\psi(x,\dot{x},t) = 0 \tag{4}$$

where φ is an h-vector and ψ is an s-vector. We shall assume that the initial conditions x_0 and $\dot{x}_0$ satisfy these constraint equations at time $t=0$, i.e., $\varphi(x_0,0)=0$, $\dot{\varphi}(x_0,0)=0$, and $\psi(x_0,\dot{x}_0,0)=0$.

Assuming that Eqs. (3) and (4) are sufficiently smooth, we differentiate Eq. (3) twice with respect to time, and Eq. (4) once with respect to time, to obtain the equation

$$A(x,\dot{x},t)\ddot{x} = b(x,\dot{x},t) \tag{5}$$

where the matrix A is m by $3n$, and b is the m-vector that results from carrying out the differentiations. We place no restrictions on the rank of the matrix A.

This set of constraint equations includes among others, the usual holonomic, nonholonomic, scleronomic, rheonomic, catastatic, and acatastatic varieties of constraints; combinations of such constraints may also be permitted in Eq. (5). It is important to note that Eq. (5), together with the initial conditions $[\varphi(x_0,0)=0$, and $\psi(x_0,\dot{x}_0,0)=0]$, is equivalent to Eqs. (3) and (4). In what follows we shall, for brevity, drop the arguments of the various quantities, unless needed for clarification.

Consider the mechanical system at any instant of time t. Let us say we know its position, $x(t)$, and its velocity, $\dot{x}(t)$, at that instant. The presence of the constraints whose kinematic description is given by Eqs. (3) and (4) causes the acceleration, $\ddot{x}(t)$, of the constrained system to differ from its unconstrained acceleration, $a(t)$, so that the acceleration of the constrained system can be written as

$$\ddot{x}(t) = a(t) + \ddot{x}^c(t) \tag{6}$$

where $\ddot{x}^c$ is the *deviation* of the acceleration of the constrained system from what it would have been had there been no constraints imposed on it at the instant of time t. Alternatively, upon premultiplication of Eq. (6) by M, we see that at the instant of time t

$$M\ddot{x} = Ma + M\ddot{x}^c = F + F^c \tag{7}$$

and so a force of constraint F^c is brought into play that causes the deviation in the acceleration $\ddot{x}$ of the constrained system from what it might have been (i.e., a) in the absence of the constraints.

Thus the constrained mechanical system is described so far by the matrices M and A, and the vectors F and b. As recognized by Lagrange (1787), the determination from Eqs. (7) and (5) of the acceleration $3n$-vector, $\ddot{x}$, of the constrained system, and of the constraint force $3n$-vector, F^c, constitutes an underdetermined problem and cannot, in general, be solved; for, there are $6n$ unknowns and $3n+m$ relations. Additional information related to the nature of the force of constraint F^c is required, and this information is *situation specific*. Thus, to obtain an equation of motion for a given mechanical system under consideration, additional information—beyond that contained in the four quantities M, A, F, and b—needs to be provided by the mechanician who is modeling the motion of that *specific* system.

Let us assume that we have this additional information (for some specific mechanical system) regarding the constraint force $3n$-vector F^c at each instant of time t in the form of the work done by this force under virtual displacements of the mechanical system at time t. A virtual displacement of the system at time t is defined as any nonzero $3n$-vector, $v(t)$, that satisfies the relation (see Udwadia et al. 1997)

$$A(x,\dot{x},t)v(t) = 0 \tag{8}$$

The mechanician modeling the motion of the given system then provides the work done, $W^c(t)$, under virtual displacements by F^c through knowledge of a vector C at each instant of time t, so that

$$v^TF^c \equiv W^c(t) = v^TC(x(t),\dot{x}(t),t) \tag{9}$$

At any given instant of time t, W^c may be *positive, zero, or negative*. The determination of C is left to the mechanician, and is most often done by inspection of, and/or experimentation with, the specific mechanical system that (s)he is attempting to mathematically model.

For example, upon examination of a given mechanical system, the mechanician could decide that $C(t)\equiv 0$ (for all time t) is a good enough approximation to the behavior of the actual force of constraint F^c in a specific system under consideration. In that situation, Eq. (9) reduces to

$$v^TF^c \equiv W^c(t) = 0 \tag{10}$$

which is, of course, D'Alembert's principle, and the constraints are now referred to as being ideal. Though this approximation is a useful one in several practical situations, it is most often, still, only an approximation at best. More generally, the mechanician would be required to provide the $3n$-vector $C(x,\dot{x},t)$.

At those instants of time at which $C(t)\neq 0$, D'Alembert's principle is no longer valid, because the constraint forces do positive or negative work. The constraints are then said to be nonideal, and Eq. (9) provides a generalization of D'Alembert's principle in such situations. We note that from a dimensional analysis standpoint, the units of C are those of force. Yet, this force vector, C, is not a "given" or "impressed force." It arises by virtue of the presence of the constraints on the system. It would disappear altogether if the constraints, which are kinematically described by Eqs. (3) and (4), are removed. Furthermore, though Eq. (9) gives the work done by the constraint forces under virtual displacements, we shall see that the force vector C in that relation does *not*, in general, directly equal to the additional constraint force actually acting on the mechanical system. In Udwadia and Kalaba (2000) and Udwadia and Kalaba (2002a,b) we explain in greater detail the general nature of the specification of the nonideal constraint force F^c given by Eq. (9).

When the constraints are ideal, as pointed out in Udwadia and Kalaba (1992) and Udwadia and Kalaba (1996) the equations of motion can be deduced from only a knowledge of the matrices M and A, and the vectors F and b. But the specification of constrained motion of a mechanical system where the constraints are nonideal requires in addition to the knowledge of these four quantities also a knowledge of the vector C, which is situation specific. By "knowledge" we mean, as before, knowledge of these quantities as known functions of their respective arguments.

A General Principle of Mechanics

We again consider the mechanical system at time t, and assume that we know $x(t)$ and $\dot{x}(t)$ at that time. The objective in analytical dynamics is to determine the acceleration $\ddot{x}$ of the system at time t. Since by Eq. (7), $F^c=M\ddot{x}-F$, Eq. (9) can be rewritten at time t as

$$v^T(M\ddot{x}-F-C)=0 \tag{11}$$

where v is any virtual displacement at time t and $\ddot{x}$ is the acceleration of the constrained system.

Now we consider a *possible* acceleration $\hat{\ddot{x}}(t)$ of the mechanical system at time t. A *possible* acceleration is defined as *any* $3n$-vector $\hat{\ddot{x}}(t)$ that satisfies the constraint equation $A\hat{\ddot{x}}=b$ at that time. Since the actual acceleration of the constrained system $\ddot{x}$ at time t must also satisfy the same equation, we must have

$$A(\hat{\ddot{x}}-\ddot{x})=Ad=0 \tag{12}$$

and so by virtue of relation (8), the $3n$-vector $d=\hat{\ddot{x}}-\ddot{x}$ at the time t, then qualifies as a 'virtual displacement'! Hence by Eq. (11) at time t, we must have

$$d^T(M\ddot{x}-F-C)=0 \tag{13}$$

We now present two results that we will use later on.

Lemma 1: For any symmetric k by k matrix Y, and any set of k vectors e, f, and g

$$(e-g,e-g)_Y-(e-f,e-f)_Y=(g-f,g-f)_Y-2(e-f,g-f)_Y \tag{14}$$

where we define $(a,b)_Y\equiv a^TYb$ for the two k vectors a and b.

Proof: This identity can be verified directly. □

For short, in what follows, we shall call a^TYa the Y norm of the vector a (actually it is the square of the Y norm). Relation (14) may be viewed as a generalization of the so-called "cosine rule" for a triangle.

Lemma 2: Any vector $d=\hat{\ddot{x}}-\ddot{x}$ satisfies at time t, the relation

$$\begin{aligned}&[M\hat{\ddot{x}}-(F+C),M\hat{\ddot{x}}-(F+C)]_{M^{-1}}\\&\quad-[M\ddot{x}-(F+C),M\ddot{x}-(F+C)]_{M^{-1}}\\&=(d,d)_M+2[M\ddot{x}-(F+C),d]\end{aligned} \tag{15}$$

Proof: Set $k=3n$, $Y=M$, $e=M^{-1}(F+C)$, $f=\ddot{x}$, and $g=\hat{\ddot{x}}$ in relation (14). The result follows. □

A New General Principle of Mechanics

We are now ready to state the general minimum principle of analytical dynamics.

A constrained mechanical system subjected to nonideal constraints evolves in time in such a manner that its acceleration $3n$-vector, $\ddot{x}$, minimizes at each instant of time the quadratic form

$$G_{ni}(\hat{\ddot{x}})=[M\hat{\ddot{x}}-(F+C),M\hat{\ddot{x}}-(F+C)]_{M^{-1}} \tag{16}$$

where the minimization is done over all "possible" acceleration $3n$-vectors $\hat{\ddot{x}}$ that satisfy the equation $A\hat{\ddot{x}}=b$ at that instant of time.

Proof: For the constrained mechanical system described by Eqs. (1)–(4) and (9), the $3n$-vector d satisfies relation (13); hence the last member on the right hand side of Eq. (15) becomes zero. Since M is positive definite, the scalar $(d,d)_M$ on the right hand side of Eq. (15) is always positive for $d=\hat{\ddot{x}}-\ddot{x}\neq 0$. So the first member on the left hand side of Eq. (15) always exceeds the second member on the left hand side of Eq. (15) by a positive number, unless $\hat{\ddot{x}}=\ddot{x}$, when the difference between them vanishes. Thus by virtue of Eq. (15), the minimum of Eq. (16) must therefore occur when $\hat{\ddot{x}}=\ddot{x}$. □

Our proof also shows that the acceleration that minimizes the quadratic form $G_{\text{ni}}(\hat{\ddot{x}})$, is unique. We use the subscript ni to refer to nonideal constraints. To find the acceleration vector $\ddot{x}$ of the constrained system at time t, the general principle then says that Nature appears, as it were, to be doing the following. She starts with the set of all imaginable $3n$-vectors, $\{\ddot{z}_i\}_{i=1}^{\infty}$. Of these she chooses only those vectors that are *possible* acceleration vectors, i.e., the set of vectors $\{\hat{\ddot{x}}_i\}$ that satisfy the relation $A\hat{\ddot{x}}_i=b$ at the time t. For each such vector $\hat{\ddot{x}}_i$ in this set the quantity $G_{\text{ni}}(\hat{\ddot{x}}_i)$ as given in Eq. (16) is determined. Then the actual (unique) acceleration $\ddot{x}$ of the mechanical system at time t is given by that vector from this set of possible acceleration vectors that minimizes the quantity G_{ni}.

We note from the proof of our result that at each instant of time t the minimum in Eq. (16) is a *global minimum*, since we do not restrict the possible accelerations in magnitude, as long as they satisfy the relation $A\hat{\ddot{x}}=b$. Comparing this principle with other fundamental principles of analytical dynamics [like Hamilton's principle (Pars 1979; Goldstein 1981), which is an extremal principle], this then appears to be the *only* general global minimum principle in analytical dynamics.

When $C(t)\equiv 0$, all the constraints are ideal, and the general principle stated above reduces to Gauss's principle of least constraint. The general minimum principle given in Eq. (16) could then be viewed as encompassing Gauss's principle (Gauss 1829), for it is valid for general constraints, ideal and nonideal.

Alternative Forms of General Principle of Mechanics

As mentioned before, the units of C are those of force; it needs to be prescribed (at each instant of time) by the mechanician, based upon examination of the given specific mechanical system whose equations of motion are desired. From Eq. (7), we have $F^c=M\ddot{x}-F$, where $\ddot{x}$ is the acceleration of the constrained system. Were we to replace $\ddot{x}$ on the right hand side of this relation by any particular *possible* acceleration $\hat{\ddot{x}}$, we would obtain the corresponding *possible* force of constraint relevant to this possible acceleration as $\hat{F}^c = M\hat{\ddot{x}} - F$. Thus the quadratic form Eq. (16) can be rewritten as

$$G_{\mathrm{ni}}(\hat{F}^c) = (\hat{F}^c - C, \hat{F}^c - C)_{M^{-1}} \tag{17}$$

And hence the minimization of $G_{\mathrm{ni}}(\hat{\ddot{x}})$ over all possible vectors $\hat{\ddot{x}}$ implies the minimization of $G_{\mathrm{ni}}(\hat{F}^c)$ and leads to the following alternative statement:

A constrained mechanical system evolves in time so that the M^{-1} norm of the force of constraint that is generated less the prescribed vector C is minimized, at each instant of time, where the minimization is done over all "possible" forces of constraint $\hat{F}^c$ at that time.

It should be noted that, in general, $F^c \neq C$. In fact as seen from Eq. (9) the quantity $(F^c - C)$ at time t is that part of the total force of constraint that does no work under virtual displacements v at time t, since $v^T (F^c - C) = 0$. Likewise, the quantity $(\hat{F}^c - C)$ that appears in the quadratic form Eq. (17) may be thought of as that part of the possible force of constraint that does no work under a virtual displacement at time t.

The constraint force that Nature seems to utilize is such that the M^{-1} norm of $(\hat{F}^c - C)$ is minimized. Nature thus determines the constraint forces on each of the masses paying more "attention (weight) to" minimizing the force $(\hat{F}^c - C)$ on the smaller masses than on the larger ones when it comes to satisfying the constraints imposed on the motion of the masses in a mechanical system.

There is yet another alternative form of the general principle that is useful and that we shall employ in the next section. Thinking of the vector C as a force that is prescribed by the mechanician at time t, $c = M^{-1}C$ is the acceleration that this force would engender in the unconstrained system at that time; similarly $a = M^{-1}F$ is the acceleration that the impressed force F would engender in the unconstrained system at the time t. Denoting at time t

$$a_{\mathrm{ni}} = a + c \tag{18}$$

we can rewrite Eq. (16) as

$$G_{\mathrm{ni}}(\hat{\ddot{x}}) = (\hat{\ddot{x}} - a_{\mathrm{ni}}, \hat{\ddot{x}} - a_{\mathrm{ni}})_M \tag{19}$$

Hence we have the following alternative understanding of constrained motion:

A constrained mechanical system evolves in time in such a way that at each instant of time the M norm of the deviation of its acceleration from a_{ni} is a minimum. The acceleration a_{ni} at any time t is the acceleration of the unconstrained system under the combined action of force F(t) that is impressed on it and the prescribed force C(t) that describes the nature of the nonideal constraints.

When $C(t) \equiv 0$, all the constraints are ideal, and $c(t) = 0$ so that $a_{\mathrm{ni}} = a$. From Eq. (19) we then see that the general principle stated above reduces, when the constraints are ideal, to Gauss's principle (Gauss 1829) of least constraint. The general principle given in Eq. (16) [and its alternative forms given in Eqs. (17) and (19)] can therefore be viewed as a generalization of Gauss's principle (Gauss 1829), which is valid only when the constraints are nonideal.

We next show the power of this new principle by obtaining in a very simple way the general equation of motion of constrained mechanical systems when the forces of constraint are not ideal.

General Explicit Equations of Motion for Holonomic and Nonholonomic Systems with Nonideal Constraints

Let us denote $r = M^{1/2}(\ddot{x} - a - c)$, so that

$$\ddot{x} = M^{-1/2}r + a + c = M^{-1/2}r + a_{\mathrm{ni}} \tag{20}$$

and the relation $A\ddot{x} = b$ becomes

$$Br = (AM^{-1/2})r = b - Aa - Ac \tag{21}$$

where $B = AM^{-1/2}$. Then the general principle Eq. (19) reduces to minimizing $\|r\|^2$, subject to the condition $Br = b - Aa - Ac$. But the solution of this problem is simply (Udwadia and Kalaba 1996)

$$r = B^{+}(b - Aa - Ac) \tag{22}$$

where B^+ stands for the Moore–Penrose inverse of the matrix B. [Instead of the Moore–Penrose inverse we could use any so-called {1,4} generalized inverse (Udwadia and Kalaba 1996)]. Substituting for r in Eq. (20) yields the explicit equation of motion for the system as

$$\begin{aligned}\ddot{x} &= M^{-1/2}B^{+}(b - Aa - Ac) + a + c \\ &= a + M^{-1/2}B^{+}(b - Aa) - M^{-1/2}B^{+}BM^{1/2}M^{-1}C + M^{-1}C \\ &= a + M^{-1/2}B^{+}(b - Aa) + M^{-1/2}(I - B^{+}B)M^{-1/2}C \end{aligned} \tag{23}$$

Premultiplying Eq. (23) by M, one obtains the general form for the equation of motion of a constrained system with nonideal constraints as

$$M\ddot{x} = F + M^{1/2}B^{+}(b - Aa) + M^{1/2}(I - B^{+}B)M^{-1/2}C = F + F_i^c + F_{\mathrm{ni}}^c \tag{24}$$

When $C(t) \equiv 0$, all the constraints are ideal and we obtain the results given in Udwadia and Kalaba (1992).

From the right hand side of Eq. (24) we notice that at each instant of time t the total force acting on the constrained system is made up of three members: (1) the impressed force F; (2) the force $F_i^c = M^{1/2}B^{+}(b - Aa)$ which would exist were all the constraints ideal, i.e., $C(t) \equiv 0$; and (3) the force $F_{\mathrm{ni}}^c = M^{1/2}(I - B^{+}B)M^{-1/2}C$ that is brought into play solely because the constraints are nonideal. The neat separation of the total force acting on the system as a sum of the above-mentioned three components highlights the simplicity and elegance with which Nature seems to operate.

One last point, what does the new general principle of mechanics given in the "General Principle of Mechanics" section look like in generalized Lagrangian coordinates, q? Let us say we have k generalized coordinates describing the unconstrained system, and a total of p (holonomic and nonholonomic) constraint equations. As will be seen from their proofs, to obtain the general principle of mechanics and the general explicit equation of motion for constrained systems with nonideal constraints in Lagrangian coordinates all one has to do is make the substitutions: $x \to q$, $\dot{x} \to \dot{q}$, $\ddot{x} \to \ddot{q}$, $M \to M(q,t)$, $F \to Q$, $F^c \to Q^c$, and $\hat{\ddot{x}} \to \hat{\ddot{q}}$ in Eqs. (1)–(24). Eq. (1) now becomes Lagrange's equation of motion for the unconstrained system. Since q is a k-vector, so is the "given" force vector $Q(q, \dot{q}, t)$ and the positive definite matrix $M(q,t)$ will then be a k by k matrix. Appropriate differentiation of the constraints $\varphi(q,t)=0$ and $\psi(q,\dot{q},t)=0$, which total p in number, will lead to the equation $A\ddot{q}=b$ where $b(q,q,t)$ is a p-vector, and the matrix $A(q,\dot{q},t)$ is accordingly a p by k matrix.

The general principle of analytical dynamics in Lagrangian coordinates then becomes:

A constrained mechanical system subjected to nonideal constraints evolves in time in such a manner that its acceleration, $\ddot{q}$, at each instant of time minimizes the quadratic form

$$G_{\mathrm{ni}}(\hat{\ddot{q}}) = [M(q,t)\hat{\ddot{q}} - Q(q,\dot{q},t) - C(q,\dot{q},t), M(q,t)\hat{\ddot{q}} - Q(q,\dot{q},t) - C(q,\dot{q},t)]_{M^{-1}} \tag{25}$$

where the minimization is done over all "possible" accelerations $\hat{\ddot{q}}$ that satisfy the equation $A(q,\dot{q},t)\hat{\ddot{q}}=b(q,\dot{q},t)$ at that instant of time. For clarity, we have shown the arguments of the various quantities explicitly.

As in the "Alternative Forms of General Principle of Mechanics" section, alternate forms of the general principle can be obtained in the obvious manner by using relations (17) and (19) and making the necessary substitutions described above.

Furthermore, the general explicit equation of motion of the constrained system with nonideal constraints in generalized coordinates becomes

$$M\ddot{q} = Q + M^{1/2}B^{+}(b - Aa) + M^{1/2}(I - B^{+}B)M^{-1/2}C = Q + Q_i^c + Q_{\mathrm{ni}}^c \tag{26}$$

where, for clarity, we have suppressed the arguments of the various quantities. (Again, we can use *any* {1,4}-inverse instead of the Moore–Penrose inverse.) As before, the matrix $B=AM^{-1/2}$. The total generalized force on the constrained system is again seen to be made up of three components: (1) the k-vector Q, which is the impressed, or given, force; (2) the k-vector Q_i^c, which is the force of constraint that would be caused were all the constraints ideal; and (3) the k-vector Q_{ni}^c which arises because of the nonideal nature of the constraints. This nonideal character of the constraints is prescribed, as before, by the mechanician through the work done (which may be positive, zero, or negative) under virtual displacements by the constraint force as $W^c(t) = v^T(t)C(q,\dot{q},t)$.

Though not in any detail, our approach has been inspired by the central idea used by Gauss (1829) in developing his results. One now sees why Gauss, in his original paper, did not bother to use Lagrangian coordinates, despite the fact that he used angle coordinates all the time for his astronomical measurements of comet motions.

Illustrative Examples

To illustrate the simplicity with which we can write out the equations of motion for nonholonomic systems with nonideal constraints we consider here two examples: a generalization of a well-known (and controversial) problem that was first introduced by Appell (1911); and, a block moving down an inclined plane with Coulomb friction.

Example 1

Consider a particle of unit mass moving in a Cartesian inertial frame subjected to the known impressed (given) forces $F_x(x,y,z,t)$, $F_y(x,y,z,t)$, and $F_z(x,y,z,t)$ acting in the X, Y, and Z directions. Let the particle be subjected to the nonholonomic constraint $\dot{x}^2+\dot{y}^2-\dot{z}^2=2\alpha g(t)$, where α is a given scalar constant and g is a given, known function of time. Appell (1911) takes $\alpha=0$, and he describes a physical mechanism that would yield his constraint.

Furthermore, Appell assumed that the constraint is ideal. Let us generalize his example and say that the mechanician (who has supposedly examined the physical mechanism which is being modeled here) ascertains that this nonholonomic constraint subjects the particle to a force that is proportional to the square of its velocity and opposes its motion, so that the virtual work done by the force of constraint (under a virtual displacement ν) on the particle is prescribed (by the mechanician) as

$$W^c(t) = -a_0 v^T(t)\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix}\frac{u^2}{|u|} = -a_0 v^T(t)\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix}|u| \tag{27}$$

where $u(t)$ is the speed of the particle; and a_0 is a given constant. It should be pointed out that this force arises because of the presence of the constraint $\dot{x}^2+\dot{y}^2-\dot{z}^2=2\alpha g(t)$. It cannot therefore be considered as a given, or impressed, force. Were this constraint to be removed, this force [whose nature is described by relation (27)] would disappear. Thus the constraint is noholonomic and nonideal. We shall obtain the explicit equations of motion of this system.

On differentiating the constraint equation with respect to time, we obtain

$$[\dot{x} \quad \dot{y} \quad -\dot{z}]\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} = \alpha\dot{g} \tag{28}$$

where $\dot{g}=dg/dt$. Thus we have $A=[\dot{x} \quad \dot{y} \quad -\dot{z}]$, and the scalar $b=\alpha\dot{g}$. Since $M=I_3$, $B=A$. Hence $B^{+}=1/u^2[\dot{x} \quad \dot{y} \quad -\dot{z}]^T$, and the vector $C=-(a_0u^2/|u|)[\dot{x} \quad \dot{y} \quad \dot{z}]^T$. The force C thus acts in a direction opposite to the velocity of the particle and is proportional to the square of its speed. The unconstrained acceleration $a=[F_x \quad F_y \quad F_z]^T$.

The equation of motion for this constrained system can now be written down directly by using Eq. (24). It is given by

$$\begin{bmatrix}\ddot{x}\\ \ddot{y}\\ \ddot{z}\end{bmatrix} = \begin{bmatrix}F_x\\ F_y\\ F_z\end{bmatrix} + \frac{(\alpha\dot{g} - \dot{x}F_x - \dot{y}F_y + \dot{z}F_z)}{u^2}\begin{bmatrix}\dot{x}\\ \dot{y}\\ -\dot{z}\end{bmatrix} - \frac{a_0}{|u|}\begin{bmatrix}(\dot{y}^2+\dot{z}^2) & -\dot{x}\dot{y} & \dot{x}\dot{z}\\ -\dot{y}\dot{x} & (\dot{x}^2+\dot{z}^2) & \dot{y}\dot{z}\\ \dot{x}\dot{z} & \dot{y}\dot{z} & (\dot{x}^2+\dot{y}^2)\end{bmatrix}\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{z}\end{bmatrix} \tag{29}$$

which simplifies to

$$\begin{bmatrix}\ddot{x}\\ \ddot{y}\\ \ddot{z}\end{bmatrix} = \begin{bmatrix}F_x\\ F_y\\ F_z\end{bmatrix} + \frac{(\alpha\dot{g} - \dot{x}F_x - \dot{y}F_y + \dot{z}F_z)}{u^2}\begin{bmatrix}\dot{x}\\ \dot{y}\\ -\dot{z}\end{bmatrix} - \frac{a_0}{|u|}\begin{bmatrix}2\dot{x}\dot{z}^2\\ 2\dot{y}\dot{z}^2\\ 2\dot{z}(\dot{x}^2+\dot{y}^2)\end{bmatrix} \tag{30}$$

The first term on the right hand side is the impressed force. The second term on the right hand side is the constraint force F_i^c that would prevail were all the constraints ideal so that they did no work under virtual displacements. This term ensures that under the combined influence of the impressed force and the ideal force of constraint, the motion of the particle satisfies the kinematical constraint equation $\dot{x}^2+\dot{y}^2-\dot{z}^2=2\alpha g(t)$ and satisfies Gauss's dynamical principle (Gauss 1829) for ideal constraints (or, equivalently, D'Alembert's dynamical principle for ideal constraints). The third term on the right hand side of Eq. (30) is the contribution F_{ni}^c to the total constraint force generated by virtue of the fact that the constraint force is not ideal, and its nature in the given physical situation is specified by the vector C, which gives the work done by this constraint force under virtual displacements. We thus obtain an intrinsic and qualitative picture of the evolution of the dynamical system in time. Note that $F_{\mathrm{ni}}^c \neq C$.

Example 2

Consider a block of mass m moving under gravity on a straight inclined plane which is inclined to the horizontal at an angle θ, $0<\theta<\pi/2$. The unconstrained motion of the block (in the absence of the constraint imposed on its motion by the inclined plane) is given by

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ -mg \end{bmatrix} \tag{31}$$

where the x direction is taken along the horizontal and the y direction is taken pointing upwards. The vector on the right hand side of Eq. (31) represents the given force Q. The constraint imposed by the inclined plane can be described by the equation

$$y = x \tan\theta \tag{32}$$

which, upon two differentiations with respect to time, yields

$$\ddot{y} = \ddot{x}\tan\theta \tag{33}$$

Since this can be written as

$$[-\tan\theta \quad 1]\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = 0 \tag{34}$$

the matrix $A=[-\tan\theta \quad 1]$,

$$B^+ = (AM^{-1/2})^+ = m^{1/2}\cos^2\theta\begin{bmatrix} -\tan\theta \\ 1 \end{bmatrix}$$

and the scalar $b=0$.

Were the constraint represented by Eq. (32) assumed to be ideal, the equation of motion for the system, by Eq. (26), would be

$$m\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ -mg \end{bmatrix} + mg\begin{bmatrix} -\sin\theta\cos\theta \\ \cos^2\theta \end{bmatrix} \tag{35}$$

The second member on the right hand side indicates explicitly the constraint force Q_i^c generated by the ideal constraint represented by Eq. (32). The magnitude of this constraint force is $mg\cos\theta$.

Were we to include Coulomb friction along the (rough) inclined plane with a coefficient of friction μ, the constraint will no longer be ideal; the work done by the constraint force under any virtual displacement v can then be represented as

$$v^T Q^c = -v^T C \equiv -v^T \frac{\mu}{\sqrt{\dot{x}^2+\dot{y}^2}}\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} mg\cos\theta \tag{36}$$

Relation (36) states that the frictional force acts along the plane, in a direction opposing the velocity of the block, and has a magnitude of $\mu|Q_i^c|$. We note that by virtue of Eq. (32), $\dot{y}=\dot{x}\tan\theta$, so that the vector

$$C = \mu mg\cos\theta\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}\mathrm{sgn}(\dot{x}) \tag{37}$$

This nonideal constraint described by Eq. (36) now provides an additional constraint force given by

$$\begin{aligned} Q_{\mathrm{ni}}^c &= -M^{1/2}\{I-(AM^{-1/2})^+(AM^{-1/2})\}M^{-1/2}C \\ &= -\left\{I - \begin{bmatrix} \sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \cos^2\theta \end{bmatrix}\right\}C \\ &= -\begin{bmatrix} \mu mg\cos^2\theta \\ \mu mg\cos\theta\sin\theta \end{bmatrix}\mathrm{sgn}(\dot{x}) \end{aligned} \tag{38}$$

so that the constrained equation of motion is given by

$$\begin{aligned} m\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} &= \begin{bmatrix} 0 \\ -mg \end{bmatrix} + mg\begin{bmatrix} -\sin\theta\cos\theta \\ \cos^2\theta \end{bmatrix} \\ &\quad - \mu mg\begin{bmatrix} \cos^2\theta \\ \cos\theta\sin\theta \end{bmatrix}\mathrm{sgn}(\dot{x}) \end{aligned} \tag{39}$$

where we have again explicitly shown on the right hand side the three different constituents of the forces acting on the system: the first term corresponds to the given forces Q; the second to the force Q_i^c generated by the presence of the constraint given by Eq. (32), were it an ideal constraint; and, the third, to the *additional* force Q_{ni}^c generated by the presence of the nonideal constraint given by Eq. (32), whose nature is further described by Eq. (36).

We observe that it is because we do not eliminate any of the x's or the $\dot{x}$'s (as is customarily done in the development of the equations of motion for constrained systems) that we can explicitly assess the effect of the "given" force, and of the components Q_i^c and Q_{ni}^c on the motion of the constrained system.

Eqs. (24) and (26) can also be used to directly obtain the equations of motion for other types of holonomic and/or nonholonomic nonideal constraints. Examples of such systems may be found in Udwadia and Kalaba (2000) and Udwadia and Kalaba (2001). Udwadia and Kalaba (2002a,b) obtain the same general Eq. (26), but by very different routes.

Conclusions

The principles of Gauss (1829) and D'Alembert (Lagrange 1787) have been the cornerstones of analytical dynamics. Though extremely useful in obtaining the equations of motion for constrained systems, they are based on a fundamental assumption—the total work done by the forces of constraint under virtual displacements is zero. Such constraint forces are called ideal. However, from a practical viewpoint there are many situations in which this assumption may not be valid (see Pars 1979 and Goldstein 1981). Generalizations of these two principles to such situations have so far been unavailable. And so when nonideal forces of constraint arise in mechanical systems, we have lacked a formulation of Lagrangian mechanics that provides us with the equations of motion for such systems.

While this situation has been known for quite some time (see remarks in "Introduction"), a remedy for it has been difficult to propose because any extension of these fundamental principles must: (1) be general enough to encompass the behavior of every specific mechanical system where the forces of constraint do work; (2) be practical to utilize; (3) be able to give a unique equation of motion in every specific system, as is required from practical observation; (4) be able to utilize the known formalisms, and the considerable math-ware, that have been developed to date in analytical dynamics; and (5) be able to reduce to the known equations of motion, such as those given by the Gibbs, Appell, and Lagrange, when the forces of constraints are ideal.

In this paper we provide a generalization of D'Alembert's principle that satisfies these requirements (see Udwadia and Kalaba 2001, 2002a,b). We show that this leads to a new general, fundamental principle of analytical dynamics. The principle gives new insights into the physics and evolution of constrained motion where the forces of constraint are general and may do work under virtual displacements. The power of this principle is illustrated in the simplicity with which it yields a Lagrangian formulation of mechanics applicable to systems with holonomic and/or nonholonomic nonideal constraints.

Specifically, the main contributions of this paper are as follows:

1. We obtain a *minimum* principle of mechanics that is general enough to include constraint forces that may do *positive, zero, or negative* work under virtual displacements. It applies to systems in which energy can be drained from the system by (at) one or more constraints, or added to it.
2. Unlike other fundamental principles of analytical dynamics (e.g., Hamilton's principle), which are, strictly speaking, extremal principles, the principle obtained here is a global minimum principle. Furthermore, unlike principles like Hamilton's principle that involve integrals over time, this principle is satisfied at *each* instant of time as the constrained dynamical system evolves.
3. The principle is a generalization that is valid for non-ideal constraints of Gauss's principle, which was discovered by Gauss in 1829 and which he called at the time, "a universal principle of mechanics." This year marks the 175th anniversary of Gauss's paper in which he first published his principle.
4. The general principle has a greatly expanded compass of applicability for it can be used in numerous situations of practical importance where the forces of constraint *do* do work, and where, therefore, Gauss's principle becomes invalid.
5. The principle opens up the whole of Lagrangian mechanics to the inclusion of non-ideal holonomically and nonholonomically constrained systems. It offers new insights into the way Nature seems to accommodate constrained motion and explains the evolution in time of constrained dynamical systems.
6. The power of the new principle is illustrated by the simple way it allows us to obtain the general, unique equations of motion for holonomically and/or noholonomically constrained mechanical systems subjected to nonideal constraints.
7. We note that the equations of motion [see Eq. (26)] obtained on the basis of the principle do not involve the notion of Lagrange multipliers. The use of Lagrange multipliers, first introduced by Lagrange specifically to handle the problem of constrained motion, have become so entrenched in analytical dynamics over the last 250 years that many mechanicians believe them to be essential in obtaining the equations of motion for constrained mechanical systems. This is not so. In fact, Lagrange multipliers constitute only one way of dealing with constrained minimization problems; moreover they are an intermediate mathematical device used to obtain the equations of motion. The statement of the problem of constrained motion makes *no* mention of them, and, as seen from Eq. (26), neither does its solution.
8. It is interesting to note that even though the simplest systems in analytical dynamics are usually nonlinear, the general explicit equations of motion for nonideal holonomic and nonholonomic systems are obtained herein using only elementary linear algebra.

References

Appell, P. (1899). "Sur une forme generale des equations de la dynamique." *C. R. Acad. Sci. III,* 129, 459–460.

Appell, P. (1911). "Example de mouvement d'un point assujeti a une liason exprimee par une relation non lineaire entre les composantes de la vitesse." *Rend. Circ. Mat. Palermo* 32, 48–50.

Gauss, C. F. (1829). "Uber ein neues allgemeines grundgesetz der mechanik." *J. Reine Angew. Math.*, 4, 232–235.

Gibbs, J. W. (1879). "On the fundamental formulae of dynamics." *Am. J. Math.*, 2, 49–64.

Goldstein, H. (1981). *Classical mechanics*, Addison Wesley, Reading, Mass.

Kalaba, R. E., and Udwadia, F. E. (1993). "Equations of motion for nonholonomic, constrained dynamical systems via Gauss's principle." *J. Appl. Mech.*, 60, 662–668.

Lagrange, J. L. (1787). *Mecanique analytique*, Mme Ve Courcier, Paris.

Pars, L. A. (1979). *A treatise on analytical dynamics*, Oxbow, Woodridge, Conn.

Udwadia, F. E., and Kalaba, R. E. (1992). "A new perspective on constrained motion." *Proc. R. Soc. London*, 439, 407–410.

Udwadia, F. E., and Kalaba, R. E. (1996). *Analytical dynamics: A new approach*, Cambridge University. Press, Cambridge, England.

Udwadia, F. E., and Kalaba, R. E. (2000). "Nonideal constraints and Lagrangian dynamics." *J. Aerosp. Eng.*, 13(1), 17–22.

Udwadia, F. E., and Kalaba, R. E. (2001). "Explicit equations of motion for constrained systems with non-ideal constraints." *J. Appl. Mech.*, 68, 462–467.

Udwadia, F. E., and Kalaba, R. E. (2002a). "On the foundations of analytical dynamics." *Int. J. Non-Linear Mech.*, 37, 1079–1090.

Udwadia, F. E., and Kalaba, R. E. (2002b). "What is the general form of the explicit equations of motion for constrained mechanical systems?" *J. Appl. Mech.*, 69, 335–339.

Udwadia, F. E., Kalaba, R. E., and Hee-Cheng, E. (1997). "Equations of motion for constrained mechanical systems and the extended D'Alembert principle." *Q. Appl. Math.*, 56, 321–331.