EQUATIONS OF MOTION FOR MECHANICAL SYSTEMS

By Firdaus E. Udwadia[1] and Robert E. Kalaba[2]

ABSTRACT: This paper deals with the description of constrained motion within the context of classical dynamics. An alternative, and simpler, proof for the recently developed new equation of motion for constrained systems is presented. The interpretation of this equation leads to new principles of analytical dynamics. We show how these results relate to Lagrange's formulation of constrained motion. New results related to the existence, uniqueness, and explicit determination of the Lagrange multipliers are provided. The approach developed herein is compared with those of Gibbs and Appell, and that of Dirac. Three examples of the application of the new equation are provided to illustrate their use.

INTRODUCTION

The concepts and methods of analytical mechanics are at the foundations of most of modern-day physics. Historically, this field has progressed by fits and starts, with major improvements in our understanding of the dynamic behavior of mechanical systems being interspersed with relatively large interludes in between. So it was that a century after Newton's *Principia* (1687), Lagrange (1787) published his monumental treatise on analytical dynamics, setting the tone and direction, in turn, for centuries of future research. The work of Gauss (1829) and Hamilton (1834) followed shortly thereafter. The general equations of motion for constrained mechanical systems were discovered independently by Gibbs (1879) and by Appell (1899). They are often referred to as the celebrated Gibbs-Appell equations and are considered by most to represent the pinnacle of our understanding of the time evolution of constrained mechanical systems.

Although the formulations of analytical mechanics by Newton, Lagrange, Gauss, and Hamilton are considered to be equivalent, they each have a slightly different perspective on the understanding of the evolution of a dynamical system. Besides adding to our understanding of the motion of mechanical systems, each perspective has certain advantages when dealing with specific situations. In this paper we shall rely on Gauss's perspective.

Though the problem of determining the evolution of a constrained mechanical system was first formulated at least as far back as Lagrange, the determination of the explicit equations of motion for such a system, especially when the constraints are nonintegrable, has since been a major hurdle in mechanics. The Lagrange multiplier method that Lagrange devised to handle this problem, is very difficult, if not impossible, to use (from both an analytical and a computational viewpoint) when the system has several tens of degrees of freedom and is subjected to many nonintegrable constraints.

More than 100 years after Lagrange, Gibbs and Appell came up with a general approach for handling constrained motion by expanding the set of coordinates to include quasi-coordinates. The equations they proposed broke a major impasse in our ability to obtain the explicit equations of motion for constrained systems; many consider the Gibbs-Appell equations to be "probably the simplest and most comprehensive equations of motion so far discovered" (Pars 1965). Yet these equations require (1) a felicitous choice of quasi-coordinates; and (2) an elimination of some of the quasi-coordinates. Both these aspects are problem-specific and depend on the actual nature of the nonintegrable constraints that the system is subjected to. These requirements make the Gibbs-Appell approach very difficult, if not impossible, to use when considering the constrained motion of systems described by several tens of degrees of freedom with several nonintegrable constraints. It is not an approach that lends itself particularly to automatization.

In 1964, Dirac considered Hamiltonian systems with constraints that were not explicitly time dependent; he once more attacked the problem of determining the Lagrange multipliers of the Hamiltonian corresponding to the constrained dynamical system. By ingeniously extending the concept of Poisson brackets, he developed a step-by-step recursive scheme for determining these multipliers through the repeated use of the consistency conditions (Dirac 1964; Sudarshan and Mukunda 1974). He used the primary constraints to generate additional secondary constraints, and these secondary constraints in turn to generate more secondary constraints in a step-by-step procedure. This eventually culminated in a set of linear equations involving the Lagrange multipliers whose solution led to their determination. However, no simple equation directly yielding the Lagrange multipliers was obtained.

For the last 200 years, work on the description of constrained motion has been carried out by several people including Gauss, Volterra, Boltzmann, Gibbs, Appell, Whittaker, and Synge, to name but a few. The Russian school of analytical mechanics has also been very active in this area. A 1968 Russian monograph on the subject lists more than 500 (more recent) publications on the topic (Neimark and Fufaev 1972). Yet, it is only recently (Udwadia and Kalaba 1992) that a simple general description of the equations of motion, applicable to complex systems that are subjected to general nonintegrable constraints, has become available.

Udwadia and Kalaba (1992) show that such a simple equation that is applicable to general nonconservative systems and that can handle a rather general class of constraints can be obtained by using Gauss's principle. In their work they provide a constructive proof for their general equation of motion.

In this paper a simpler and more direct proof of this result is presented. The equations of motion pertinent to constrained systems are obtained without the use of Lagrange multipliers. The systems we consider encompass all of Lagrangian mechanics, and then some. For the general constraints considered in this paper, we provide results related to the existence and uniqueness of the Lagrange multipliers. More importantly, an explicit expression for their determination is also provided; this leads to deeper physical insights into the character of these multipliers. The approach and the results developed herein are compared with those of Gibbs and Appell and those of Dirac. Three examples demonstrating the use of the new equations are provided.

[1]Prof., Mech. Engrg., Civ. Engrg., and Decision Sys., Olin Hall 430K, Univ. of Southern California, Los Angeles, CA 90089-1453.
[2]Prof., Biomedical Engrg., Electrical Engrg., and Economics, Olin Hall 430K, Univ. of Southern California, Los Angeles, CA.

Note. Discussion open until December 1, 1996. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on October 2, 1995. This paper is part of the ***Journal of Aerospace Engineering***, Vol. 9, No. 3, July, 1996. ©ASCE, ISSN 0893-1321/96/0003-0064–0069/$4.00 + $.50 per page. Paper No. 11744.

EQUATIONS OF MOTION

For simplicity, consider first a discrete mechanical system consisting of n particles of masses $m_1, m_2, m_3, \ldots, m_n$. With reference to an inertial Cartesian coordinate frame of reference, let us represent the configuration of the system, at any time t, by the $3n$ vector of coordinates $\mathbf{x} = [x_1\ x_2\ x_3 \ldots x_{3n}]^T$. The unconstrained equations of motion for the system can be obtained using Newton's equations in the form

$$\mathbf{M}\ddot{\mathbf{x}} = \mathbf{F}(\mathbf{x}, \dot{\mathbf{x}}, t) \tag{1}$$

In this paper we shall use the term "unconstrained system" to mean a system for which the number of degrees of freedom equals the minimum number of coordinates required to describe the system. The matrix $\mathbf{M}$ is diagonal and positive definite and the vector $\mathbf{F}$ is made up of the "given" or "impressed" forces acting on the system. The acceleration of the unconstrained system, which we shall denote by $\mathbf{a}(t)$, is obtained from (1) as $\mathbf{a}(t) = \mathbf{M}^{-1}\mathbf{F}$.

We next impose a set of m consistent constraints on the system, described by the equations

$$\phi_i(\mathbf{x}, \dot{\mathbf{x}}, t) = 0 \quad i = 1, 2, \ldots, m \tag{2}$$

We shall assume that the functions ϕ_i are sufficiently smooth to allow differentiation with respect to their arguments. Differentiating with respect to time, equation set (2) can then be expressed as the consistent set of equations

$$\mathbf{A}(\mathbf{x}, \dot{\mathbf{x}}, t)\ddot{\mathbf{x}} = \mathbf{b}(\mathbf{x}, \dot{\mathbf{x}}, t) \tag{3}$$

where $\mathbf{A}$ = an m by $3n$ matrix. In what follows we will see that the matrix $\mathbf{AM}^{-1/2}$ plays a significant role. We shall call this matrix the Constraint Matrix. [The Constraint Matrix is defined here a bit differently from that in Udwadia and Kalaba (1992).]

The set of constraint equations (2) are general enough to include both integrable and nonintegrable equality constraints. Often, the nonintegrable constraints in analytical dynamics are expressed in Pfaffian form. Differentiating these constraints once with respect to time would yield equations of the form described by (3). Similarly, holonomic constraints would require two differentiations to be put into the form of (3).

The presence of these constraints now causes the number of degrees of freedom of the system to be less than the minimum number of coordinates needed to specify the system's configuration. The central problem of constrained motion is the determination of the equations of motion that describe the time evolution of this constrained system, given that the system's position, $\mathbf{x}(t_0)$, and its velocity, $\dot{\mathbf{x}}(t_0)$, are known and compatible with the constraint equation set (2) at some time t_0. It should be noted that given $\mathbf{x}(t_0)$ and $\dot{\mathbf{x}}(t_0)$, (3) is equivalent to the equation set (2). By virtue of the imposed constraints the accelerations of the constrained system differ from those of the unconstrained system. Thus our primary objective becomes the determination of the acceleration, $\ddot{\mathbf{x}}$, of the constrained system at time t_0.

The explicit equation of motion developed by Udwadia and Kalaba (1992) that specifies the time evolution of the constrained mechanical system that has been described previously is

$$\mathbf{M}\ddot{\mathbf{x}} = \mathbf{F} + \mathbf{M}^{1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{AM}^{-1}\mathbf{F}) \tag{4}$$

where $(\mathbf{AM}^{-1/2})^+$ = Moore-Penrose generalized inverse (Moore 1920; Penrose 1955) of the constraint matrix $(\mathbf{AM}^{-1/2})$. For brevity, we will omit writing the arguments of the various quantities explicitly, except where such information becomes important to our understanding.

An alternative form of this result can be obtained by premultiplying both sides of (4) by $\mathbf{M}^{-1}$ to yield at any time t

$$\ddot{\mathbf{x}} = \mathbf{a} + \mathbf{M}^{-1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{Aa}) \tag{5}$$

where we have made use of the fact that $\mathbf{a}(t) = \mathbf{M}^{-1}\mathbf{F}$.

Eq. (5) states that the acceleration, $\ddot{\mathbf{x}}$, of the constrained system at time t equals the acceleration of the unconstrained system, $\mathbf{a}(t)$, at that time, plus a correction term, which is brought into play by virtue of the constraints. This correction term is explicitly provided by the second member on the right-hand side of (5).

Eq. (4) may also be viewed as an extended use of Newton's second law of motion, which is directly applicable to constrained dynamical systems—the additional term on the right-hand side of that equation takes into account the effect of the imposed constraints in an explicit manner.

We provide now a simpler and more expository derivation of this result than that previously given by Udwadia and Kalaba (1992).

We begin by defining the Moore-Penrose (MP) inverse of any m by n matrix $\mathbf{B}$, and give some of its important properties. The MP-inverse of the matrix $\mathbf{B}$ is the (unique) matrix $\mathbf{B}^+$, which satisfies the following three conditions: $\mathbf{BB}^+\mathbf{B} = \mathbf{B}$; $\mathbf{B}^+\mathbf{BB}^+ = \mathbf{B}^+$; and $\mathbf{BB}^+$ and $\mathbf{B}^+\mathbf{B}$ are both symmetric matrices. The matrix $\mathbf{B}^+$ has the following four basic properties, which can be easily proved using the aforementioned definition (Graybill 1983).

P1. If the singular value decomposition of the matrix $\mathbf{B}$ is $\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$, where $\mathbf{\Lambda}$ = r by r diagonal matrix of singular values whose elements are all positive, then $\mathbf{B}^+ = \mathbf{V}\mathbf{\Lambda}^{-1}\mathbf{U}^T$. Thus the singular value decomposition of $\mathbf{B}$ provides a way of determining $\mathbf{B}^+$.

P2. A necessary and sufficient condition for the matrix equation $\mathbf{Bx} = \mathbf{b}$ to be consistent, where $\mathbf{b}$ is an m vector, is

$$\mathbf{BB}^+\mathbf{b} = \mathbf{b} \tag{6}$$

P3. If $\mathbf{Bx} = \mathbf{0}$, where $\mathbf{x} \neq \mathbf{0}$ is an n vector, then

$$\mathbf{x}^T\mathbf{B}^+ = \mathbf{0} \tag{7}$$

P4. The consistent linear set of equations $\mathbf{Bx} = \mathbf{b}$ has the solution

$$\mathbf{x} = \mathbf{B}^+\mathbf{b} + (\mathbf{I} - \mathbf{B}^+\mathbf{B})\mathbf{h} \tag{8}$$

where $\mathbf{h}$ = an arbitrary n vector.

Now appealing to Gauss's principle, we simply need to prove two things: first, that the acceleration $\ddot{\mathbf{x}}$ of the constrained system provided by (5) satisfies the constraint equation (3), and second, that this acceleration $\ddot{\mathbf{x}}$ given by (5) minimizes the Gaussian

$$G_1 = (\ddot{\mathbf{y}} - \mathbf{M}^{-1}\mathbf{F})^T\mathbf{M}(\ddot{\mathbf{y}} - \mathbf{M}^{-1}\mathbf{F}) \tag{9}$$

among all those $3n$ vectors $\ddot{\mathbf{y}}$ that satisfy (3). During this minimization we treat the $3n$ vectors $\mathbf{x}(t)$ and $\dot{\mathbf{x}}(t)$ to be fixed, as prescribed by Gauss.

To prove that $\ddot{\mathbf{x}}$ satisfies (3), we note that using $\ddot{\mathbf{x}}$ from (5) in (3) we obtain

$$\mathbf{A}\ddot{\mathbf{x}} = (\mathbf{AM}^{-1/2})(\mathbf{AM}^{-1/2})^+\mathbf{b} = \mathbf{b} \tag{10}$$

where the last equality follows, by property P2, from the consistency of (3).

We next prove that any acceleration vector other than $\ddot{\mathbf{x}}$ that satisfies the constraints (3) increases the value of the Gaussian G_1. Consider the vector $\ddot{\mathbf{y}} = \ddot{\mathbf{x}} + \mathbf{u}$, where $\ddot{\mathbf{x}}$ is as defined by (5), and $\mathbf{u}$ = an arbitrary nonzero $3n$ vector. For $\ddot{\mathbf{y}}$ to satisfy the constraint equation (3), we then require $\mathbf{Au} = \mathbf{AM}^{-1/2}\mathbf{M}^{1/2}\mathbf{u} = 0$, so that by property P3 we have

$$(\mathbf{M}^{1/2}\mathbf{u})^T(\mathbf{AM}^{-1/2})^+ = 0 \tag{11}$$

Denote the value of the Gaussian G_1 when $\ddot{\mathbf{y}} = \ddot{\mathbf{x}}$, by G. Hence

$$G = (\ddot{\mathbf{x}} - \mathbf{M}^{-1}\mathbf{F})^T\mathbf{M}(\ddot{\mathbf{x}} - \mathbf{M}^{-1}\mathbf{F}) \quad (12)$$

Substituting for $\ddot{\mathbf{y}}$ and using (5), G_1 can therefore be expressed as

$$G_1 = [(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{AM}^{-1}\mathbf{F}) + \mathbf{M}^{1/2}\mathbf{u}]^T \cdot [(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{AM}^{-1}\mathbf{F}) + \mathbf{M}^{1/2}\mathbf{u}] \quad (13)$$

Making use of relation (11), we find that $G_1 = G + (\mathbf{M}^{1/2}\mathbf{u})^T(\mathbf{M}^{1/2}\mathbf{u})$. From this it follows that the absolute minimum of the Gaussian G_1 occurs when the acceleration is as given by (5). We note that the vector $\ddot{\mathbf{x}}$ given by (5) provides a global minimum of G_1.

NEW PRINCIPLES OF ANALYTICAL DYNAMICS

We have proved that (4) and (5) provide explicitly the equations of motion pertaining to general constrained systems, where the constraints can be expressed in the form of (2). Both, the Lagrange multiplier method and the Gibbs-Appell method rely on problem-specific approaches. They rely on the ingenuity of the analyst to eliminate the Lagrange multipliers or to design the quasi-coordinates in a manner that is dependent on the specific constraints to which the dynamical system is subjected. Dirac's approach (Dirac 1964; Sudarshan and Mukunda 1974) is applicable to a smaller subset of constrained systems, and provides a method for obtaining the Lagrange multipliers: it does not directly yield the explicit equations of motion of general constrained systems as considered here. Also, Dirac's primary constraints generate secondary constraints, which in turn may generate further secondary constraints through the application of the consistency conditions. The proliferation of secondary constraints in Dirac's approach and the repeated use of extended Poisson brackets may make this approach difficult to use when dealing with large numbers of degrees of freedom and several tens of constraints, as is common in many engineering and physical systems.

The simplicity of the equation of motion obtained above lends itself to deeper physical interpretations. Consider a constrained system whose position and velocity are known at time t to be $\mathbf{x}(t)$ and $\dot{\mathbf{x}}(t)$. The central issue in the determination of the motion of the constrained system (Gauss 1829) can be thought of as the determination of its acceleration at time t. As noted previously, the acceleration of the unconstrained system is provided by (1) and can be easily obtained by the use of Newton's laws or Lagrangian mechanics. Hence the central issue centers around finding in what way, and by how much, the acceleration pertinent to the constrained system deviates from that of the unconstrained system. At time t, the acceleration pertinent to the constrained system can be written, using (5), as

$$\ddot{\mathbf{x}} - \mathbf{a} = \mathbf{M}^{-1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{Aa}) \quad (14)$$

The left-hand side of (14) represents the deviation of the acceleration vector of the constrained system from what it would be were the system to have no constraints. The difference vector $\mathbf{e}(t) = (\mathbf{b} - \mathbf{Aa})$ on the right-hand side of (14) measures the extent to which the acceleration corresponding to the unconstrained motion does not satisfy the constraint (3) at time t. Eq. (14) can then be restated as

$$\Delta\ddot{\mathbf{x}} = \mathbf{M}^{-1/2}(\mathbf{AM}^{-1/2})^+\mathbf{e} \quad (15)$$

where the deviation $\Delta\ddot{\mathbf{x}} = \ddot{\mathbf{x}} - \mathbf{a}$. The matrix $\mathbf{K}_1 = \mathbf{M}^{-1/2}(\mathbf{AM}^{-1/2})^+$ can be thought of as a weighted Moore-Penrose generalized inverse of the Constraint Matrix; the weighting is done through premultiplication of the Constraint Matrix by $\mathbf{M}^{-1/2}$. This equation explicitly exposes the linear relation between $\Delta\ddot{\mathbf{x}}$ and $\mathbf{e}$, and leads to the following fundamental principle of analytical dynamics:

> The motion of a discrete dynamical system subjected to constraints evolves, at each instant of time, in such a manner that the deviation of its acceleration from that which it would have had, at that instant, if there were no constraints on it, is directly proportional to the extent to which the acceleration corresponding to the unconstrained motion, at that instant, does not satisfy the constraints; the matrix of proportionality $\mathbf{K}_1$ is the weighted Moore-Penrose generalized inverse of the Constraint Matrix, and the measure of dissatisfaction of the constraints is the vector $\mathbf{e}$.

An alternative approach to determining the equations of motion for constrained systems is to realize that by virtue of the constraints that the system has to satisfy, additional forces are brought into play that act on the system. Thus while the motion of the unconstrained system is described by (1), the motion of the constrained system is described by the equation

$$\mathbf{M}\ddot{\mathbf{x}} = \mathbf{F} + \mathbf{F}^c \quad (16)$$

where $\mathbf{F}^c$ = force of constraint that ensures that its action in conjunction with the "given" force vector, $\mathbf{F}$, causes the system to satisfy the constraints. Comparing (4) and (16) we obtain the force of constraint explicitly as

$$\mathbf{F}^c = \mathbf{M}^{1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{Aa}) = \mathbf{M}^{1/2}(\mathbf{AM}^{-1/2})^+\mathbf{e} = \mathbf{Ke} \quad (17)$$

leading to the following alternative form of the fundamental principle:

> At each instant of time, the force of constraint acting on a constrained mechanical system is directly proportional to the extent to which the acceleration of the unconstrained system, at that time, does not satisfy the constraints; the matrix of proportionality $\mathbf{K}$ is the weighted Moore-Penrose inverse of the Constraint Matrix and the measure of the dissatisfaction of the constraint is given by the vector $\mathbf{e}$.

It should be noted that the weighting matrices involved in the two fundamental principles just stated are different.

EXISTENCE OF LAGRANGE MULTIPLIERS AND THEIR EXPLICIT DETERMINATION

We now show that for the n-particle system whose unconstrained motion is described by (1) and which is further constrained by the equation set (2), there exists, at each instant of time t, a Lagrange multiplier m vector $\boldsymbol{\lambda}$ such that

$$\mathbf{M}\ddot{\mathbf{x}} = \mathbf{F} + \mathbf{A}^T\boldsymbol{\lambda} \quad (18)$$

where $\ddot{\mathbf{x}}$ = acceleration of the constrained system. The components of the m vector $\boldsymbol{\lambda}$ are often referred to as the Lagrange multipliers (Rosenberg 1972). Comparing (4) with (18), we therefore need to show that the equation

$$\mathbf{A}^T\boldsymbol{\lambda} = \mathbf{M}^{+1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{Aa}) \quad (19)$$

is consistent. Premultiplying (19) by $\mathbf{M}^{-1/2}$, we get

$$\mathbf{C}^T\boldsymbol{\lambda} = \mathbf{C}^+\mathbf{e} \quad (20)$$

where we have denoted the constraint matrix $\mathbf{AM}^{-1/2}$ by $\mathbf{C}(\mathbf{x}, \dot{\mathbf{x}}, t)$. By property P2, the consistency of (20) is equivalent to demonstrating that

$$\mathbf{C}^T[\mathbf{C}^T]^+\mathbf{C}^+\mathbf{e} = \mathbf{C}^+\mathbf{e} \quad (21)$$

But $\mathbf{C}^T[\mathbf{C}^T]^+\mathbf{C}^+ = \mathbf{C}^T[\mathbf{C}^+]^T\mathbf{C}^+ = [\mathbf{C}^+\mathbf{C}]^T\mathbf{C}^+ = \mathbf{C}^+\mathbf{CC}^+ = \mathbf{C}^+$, and therefore the existence of the Lagrange multiplier vector $\boldsymbol{\lambda}$ defined by (18) is established.

We next proceed to determine this vector. Using property P4, the solution of the linear equation (19) is given by

$$\boldsymbol{\lambda} = (\mathbf{A}^T)^+\mathbf{M}^{+1/2}(\mathbf{AM}^{-1/2})^+(\mathbf{b} - \mathbf{Aa}) + [\mathbf{I} - (\mathbf{A}^T)^+\mathbf{A}^T]\mathbf{h} \quad (22)$$

where $\mathbf{h}$ = any arbitrary m vector. It should be observed that at each instant of time t, the right-hand side of (22) is known, for it is a function of $\mathbf{x}$, $\dot{\mathbf{x}}$, and t.

We have thus obtained an explicit expression that directly yields the time-dependent Lagrange multipliers, as functions of $\mathbf{x}$, $\dot{\mathbf{x}}$, and t, should they be required. Furthermore, because the m vector $\mathbf{h}$ is arbitrary, in general, the Lagrange multipliers are not unique. However, when the rank of the matrix $\mathbf{A}$ equals m, then, by using property P1, we can see that $(\mathbf{A}^T)^+\mathbf{A}^T = \mathbf{I}$, and the second member on the right-hand side of (22) vanishes; the Lagrange multipliers are then unique. In that case, they are given by the relation

$$\boldsymbol{\lambda} = (\mathbf{A}^T)^+\mathbf{M}^{+1/2}(\mathbf{A}\mathbf{M}^{-1/2})^+\mathbf{e} = \mathbf{K}_2\mathbf{e} \qquad (23)$$

Eq. (23) provides physical insight into the character of the Lagrange multiplier vector; for, when the constraints (3) are linearly independent, the Lagrange multiplier, $\boldsymbol{\lambda}(t)$, is directly proportional to $\mathbf{e}(t)$, the extent to which the accelerations of the unconstrained system do not satisfy the constraints (3). The matrix of proportionality is then given by $\mathbf{K}_2 = (\mathbf{A}^T)^+ \mathbf{M}^{+1/2}(\mathbf{A}\mathbf{M}^{-1/2})^+ = (\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T)^{-1}$.

ILLUSTRATIVE EXAMPLES

In this section we demonstrate the ease with which the new equations can be used by considering three examples. The first is chosen for its familiarity in the field of physics; the second for its historical value; and the third to illustrate the use of these equations when dealing with nonconservative systems, subjected to nonintegrable constraints that explicitly depend on time.

1. Consider the three-dimensional motion of a point mass m suspended from a point O by a rigid weightless rod of length L. We shall consider the motion of this pendulum with the point O as the origin of our Cartesian coordinate system, the XZ-plane horizontal, and the Y-direction pointing downwards.

The system could be thought of in terms of an unconstrained system consisting of a mass m subjected to the downward force of gravity; this unconstrained system is then further constrained because it must maintain a constant distance, L, from a fixed point O.

Denoting the position of the point mass at any time t by its coordinates (x, y, z), the equation of motion pertinent to this unconstrained system is simply

$$\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = \begin{bmatrix} 0 \\ mg \\ 0 \end{bmatrix} \qquad (24)$$

The acceleration of the unconstrained system is therefore $\mathbf{a}(t) = [0 \;\, g \;\, 0]^T$.

The weightless rod which attaches the point mass to O provides a constraint to the unconstrained motion described by (24). This constraint can be expressed as

$$x^2 + y^2 + z^2 = L^2 \qquad (25)$$

which, on two differentiations, yields

$$[x \;\, y \;\, z]\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = -(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) \qquad (26)$$

Comparing with (3), we find that the matrix $\mathbf{A} = [x \;\, y \;\, z]$, and the vector $\mathbf{b}$ = scalar $-(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)$. Since $\mathbf{M}$ = a constant diagonal matrix, $\mathbf{M}^{1/2}(\mathbf{A}\mathbf{M}^{-1/2})^+ = m\mathbf{A}^+$. Using elementary matrix algebra [the Moore-Penrose inverse of a row n-vector $\mathbf{u} = [u_1 \; u_2 \; \ldots \; u_n]$ is simply $(1/\Sigma_{i=1}^{n} u_i^2)\mathbf{u}^T$], $\mathbf{A}^+ = [1/(x^2 + y^2 + z^2)]\,[x \;\, y \;\, z]^T$. Hence the equation of motion of the constrained system can be explicitly written down, using (4), as

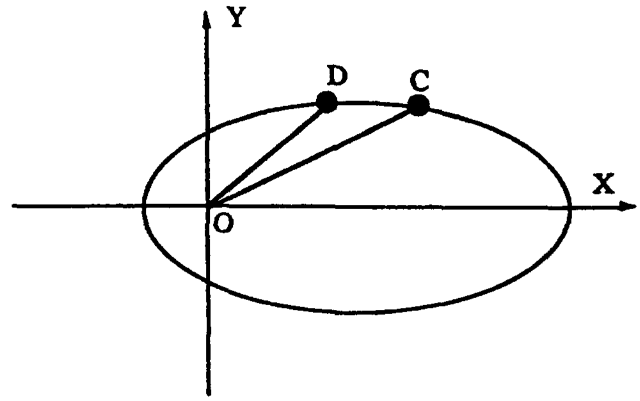

FIG. 1. Elliptic Motion of Particle

$$\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = \begin{bmatrix} 0 \\ mg \\ 0 \end{bmatrix} + \frac{m}{x^2 + y^2 + z^2}\begin{bmatrix} x \\ y \\ z \end{bmatrix}[-(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - gy] \qquad (27)$$

which, in view of (25), simplifies to

$$\begin{bmatrix} m & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & m \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = \begin{bmatrix} 0 \\ mg \\ 0 \end{bmatrix} - \frac{m\{\dot{x}^2 + \dot{y}^2 + \dot{z}^2 + gy\}}{L^2}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \qquad (28)$$

The Lagrange multiplier λ can be directly seen to be $-\{m[(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + gy]\}/L^2$. The quantity in the square brackets may be identified as the extent to which the acceleration of the unconstrained system, i.e., $[0 \;\, g \;\, 0]^T$, does not satisfy the constraint equation (26).

2. Consider the two-dimensional motion of a free particle of mass m which is constrained so as to move along an ellipse with focus at O, as well as to move so that the sector area OCD (see Fig. 1) that it traces in every unit of time is a constant. These are, of course, Kepler's first two laws of planetary motion. Our aim is to write the equations of motion of the constrained system. Alternatively, one can think of this as an inverse problem: knowing the constraints on the system, we want to determine what forces of constraint, or "control forces," Nature would apply so that the system's motion satisfies the given constraints.

Let (x, y) denote the Cartesian coordinates of the particle. The equations of motion pertinent to the unconstrained system are simply

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad (29)$$

The acceleration $\mathbf{a}$ of the unconstrained system is obviously zero.

The first constraint can be written as

$$\sqrt{x^2 + y^2} = \varepsilon x + p \qquad (30)$$

where ε and p = constants defining the ellipse. The second constraint can be expressed as

$$x\dot{y} - y\dot{x} = c \qquad (31)$$

where c = a constant. Differentiating (30) twice and (31) once, we obtain the constraint equation in the form of (3) as

$$\begin{bmatrix} (x - r\varepsilon) & y \\ y & -x \end{bmatrix}\begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = -\begin{bmatrix} c^2/r^2 \\ 0 \end{bmatrix} \qquad (32)$$

where we have denoted the radial distance of the particle from the origin by r.

Since the matrix $\mathbf{A}$ is nonsingular, its Moore-Penrose inverse is simply its inverse and

$$\mathbf{A}^{+} = \mathbf{A}^{-1} = \begin{bmatrix} x & y \\ y & -(x - r\varepsilon) \end{bmatrix} \frac{1}{(x^2 + y^2) - rx\varepsilon}$$

$$= \frac{1}{rp} \begin{bmatrix} x & y \\ y & -(x - r\varepsilon) \end{bmatrix} \tag{33}$$

The equation of motion of the constrained system is then directly obtained, using (4), as

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{bmatrix} \ddot{x} \\ \ddot{y} \end{bmatrix} = -\frac{c^2}{p} \frac{m}{r^2} \begin{bmatrix} x/r \\ y/r \end{bmatrix} \tag{34}$$

Since there are no impressed forces on the system, the right-hand side of (34) is simply the force of constraint. We thus find that if a free particle is to obey Kepler's first two laws, it must be subjected to a force of constraint that is central and varies inversely as the square of the radial distance—a result communicated by Newton to Edmond Halley more than 300 years ago. It is interesting to note that were we to have determined the equations of motion of a free particle using only the first constraint, we would have found that the force of constraint that keeps the particle on the ellipse is neither central nor solely varying inversely as the square of the radial distance.

3. Consider the three-dimensional motion of a particle of mass m constrained so that $\dot{y} = z\dot{x} + \alpha(t)$, where $\alpha(t)$ = a given function of time. The particle is subjected to the given impressed forces $X(t)$, $Y(t)$, and $Z(t)$, in the three Cartesian coordinate directions X, Y, and Z, respectively. We shall obtain the explicit equations of motion of the constrained system.

Let the position of the particle be denoted by its coordinates (x, y, z). Differentiating the constraint equation once, we can express it in the form of (3). Hence we obtain

$$\mathbf{A} = [-z \;\; 1 \;\; 0] \tag{35}$$

and

$$\mathbf{b} = \dot{z}\dot{x} + \dot{\alpha}(t) \tag{36}$$

The Moore-Penrose inverse of $\mathbf{A}$ is simply $\mathbf{A}^{+} = [1/(z^2 + 1)][-z \;\; 1 \;\; 0]^T$, and the acceleration of the unconstrained system is given by $\mathbf{a}(t) = (1/m)[X(t) \;\; Y(t) \;\; Z(t)]^T$. The equation of motion of the constrained system can then be directly written down, as

$$m \begin{bmatrix} \ddot{x} \\ \ddot{y} \\ \ddot{z} \end{bmatrix} = \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} + \frac{m[\dot{z}\dot{x} + \dot{\alpha}(t)] + zX - Y}{z^2 + 1} \begin{bmatrix} -z \\ 1 \\ 0 \end{bmatrix} \tag{37}$$

CONCLUSION AND COMMENTS

It is remarkable that despite the highly nonlinear behavior exhibited by even the simplest of mechanical systems (think of a pendulum undergoing large amplitude vibrations or a wheel rolling on the ground), it is the tools of linear algebra that have made it possible to determine these explicit equations of motion governing constrained mechanical systems. A key reason for this is that the constraints are linear in the accelerations.

We note in passing that it is a simple matter to generalize (4) and (5) to the situation when the system's configuration is described using generalized coordinates (Udwadia and Kalaba 1992). All we need do is replace the vector $\mathbf{x}$ by the vector $\mathbf{q}$ of generalized coordinates, the acceleration vector $\ddot{\mathbf{x}}$ by $\ddot{\mathbf{q}}$, the force $\mathbf{F}$ by the given generalized force $\mathbf{Q}$, and the constant diagonal matrix $\mathbf{M}$ by the symmetric positive definite matrix $\mathbf{M}(\mathbf{q}, t)$. The matrix $\mathbf{A}$ and the vector $\mathbf{b}$ in the constraint equation (3) would now be functions of $\mathbf{q}$, $\dot{\mathbf{q}}$ and t, as would, of course, the functions ϕ_i.

Eqs. (15) and (17) provide a strikingly simple view of the nature of constrained motion; for, at each instant of time, when the acceleration vector corresponding to the unconstrained system does not satisfy the constraints, Nature alters the acceleration in a manner directly proportional to the extent to which the constraints remain unsatisfied, much like the calculating mathematician; the matrix of proportionality is $\mathbf{M}^{-1/2}(\mathbf{AM}^{-1/2})^{+}$, the weighted Moore-Penrose inverse of the Constraint Matrix. Alternately stated, in the presence of constraints, Nature provides an additional force of constraint, $\mathbf{F}^c$, which is directly proportional to the vector $\mathbf{e}$, the extent to which the constraints are not satisfied, much like the calculating control theorist; the matrix of proportionality this time is $\mathbf{M}^{1/2}(\mathbf{AM}^{-1/2})^{+}$. This force $\mathbf{F}^c$, in conjunction with the given impressed force $\mathbf{F}$, ensures that the constraints are satisfied by the dynamical system.

Although the Moore-Penrose inverse of a matrix may be difficult to obtain analytically for large matrices $\mathbf{A}$, its numerical computation is not difficult. In fact the Moore-Penrose inverse can be readily obtained in computer environments of even modest sophistication, like MATLAB. It is the ease with which the equations of motion (4) and (5) can be directly implemented on a computer that makes them attractive for use in systems with many degrees of freedom and several constraints.

The salient features of the results developed in this paper are the following.

1. The equation of motion for the constrained system is directly obtained, bypassing the need to determine the Lagrange multipliers. Explicit expressions, if needed, for the Lagrange multipliers are provided.
2. The equation of motion (4) pertinent to the constrained system can be obtained with equal ease for both integrable and nonintegrable constraints. If anything, nonintegrable constraints appear to be somewhat easier to handle in our approach because they require only one differentiation in order to put them in the form of (3). We overcome the hurdle presented by the nonintegrability of the constraint equations through the simple observation that though such constraints cannot be integrated, they can still be differentiated, provided the functions ϕ_i are smooth enough. These differentiated equations are linear in the accelerations.
3. Contrary to conventional wisdom, we do not eliminate coordinates in an effort to obtain as many second-order differential equations as the number of degrees of freedom of the constrained system. Our equations are obtained in terms of the same coordinates as those used to specify the unconstrained system. This allows us to directly compare, at each instant of time, the constrained motion of the system with its unconstrained motion, thereby yielding additional insights into the underlying physics of constrained motion.

Finally, we would like to comment on the use of this new result in developing methods for controlling mechanical systems so that they satisfy certain constraints. The insight gathered into the way Nature creates the force of constraint, so that a given nonlinear mechanical system under the influence of a known set of impressed forces satisfies a given set of constraints, can be useful from a Baconian perspective. It gives us a method of finding the "control force" that needs to be applied to a mechanical system so that its motion follows a certain trajectory, or more generally, satisfies a certain set of constraints. For, given the trajectory, or more generally the set of constraints, we have explicitly determined the additional force, $\mathbf{F}^c$, that Nature would apply in the circumstance, so that the system satisfies the constraints [see (17)]. There may then be some merit, and consequent parsimony, in imitating Nature were we required to control a mechanical system (like a robot arm) to follow a given set of constraints.

We note that, in principle, this "control force" is explicitly determined in a simple way at each instant of time (no matter how nonlinear the mechanical system) so that in the presence of the given impressed force, **F**, the system exactly satisfies the constraints; we have provided, in principle, a simple solution that is computable in real time, to what might otherwise be a control theorist's nightmare. Our ability to do this rests on our understanding of the deep structure that governs the motion of constrained mechanical systems in Nature. Applications of this line of thinking are imminent in fields such as robotics, motion control, and control of structural and mechanical systems.

The equations of motion presented in this paper appear to be simple to use and comprehensive enough to include most physically encountered constraints. We expect that they will have numerous applications in the field of physics and engineering. Providing new spice to a mature field, they may open up new horizons in our understanding of Nature.

APPENDIX. REFERENCES

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

Dirac, P. A. M. (1964). *Lectures in quantum mechanics*. Yeshiva Univ., New York, N.Y.

Gauss, C. (1829). "Uber ein neues allgemeines grundgesetz der mechanik." *Journal fur Reine und Angewandte Mathematik*, Germany, 4, 232–235.

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

Graybill, F. A. (1983). *Matrices with applications to statistics*, 2nd Ed., Wadsworth, Belmont, Calif.

Hamilton, W. R. (1840). *On a general method in dynamics*. Collected Papers, Vol. II, Cambridge, England, 103–211.

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

Moore, E. H. (1920). "On the reciprocal of the general algebraic matrix." Abstract, *Bull., Am. Math. Soc.*, 26, 394–395.

Neimark, J. I., and Fufaev, N. A. (1972). *Dynamics of nonholonomic systems*. American Mathematical Society, Providence, RI, 33, translated from 1968 Russian version.

Newton, I. (1687). *Philosophiae naturalis principia mathematica*. Royal Society Press, London, England.

Pars, L. A. (1979). *A treatise on analytical dynamics*, Second Printing, Ox Bow Press, CT, 202.

Penrose, R. (1955). "A generalized inverse of matrices." *Proc.*, Cambridge Philosophical Society, Cambridge, England, 51, 406–413.

Rosenberg, R. M. (1972). *Analytical dynamics of discrete systems*. Plenum Press, New York, N.Y.

Sudarshan, E. C. G., and Mukunda, N. (1974). *Classical dynamics: a modern perspective*. John Wiley & Sons, New York, N.Y.

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