On Motion

by FIRDAUS E. UDWADIA

Mechanical Engineering, Civil Engineering and Decision Systems, University of Southern California, Los Angeles, CA 90089-1451, U.S.A.

and ROBERT E. KALABA

Economics, Biomedical Engineering and Electrical Engineering, University of Southern California, Los Angeles, CA 90089-1451, U.S.A.

ABSTRACT: *Starting with the principle of virtual work and D'Alembert's principle, new general equations of motion for constrained mechanical systems are obtained. These equations lead to a new fundamental principle of analytical mechanics.*

I. Introduction

One of the central problems in the field of analytical mechanics is the determination of the equations of motion pertinent to constrained systems. This problem dates at least as far back as Lagrange (1), who devised in 1787 the method of Lagrange multipliers specifically to handle constrained motion. Yet the Lagrange multiplier approach is extremely difficult, if not impossible to use, both analytically and computationally, even when dealing with relatively small systems that have several tens of degrees of freedom and several nonintegrable constraints. Realizing, in addition, that this approach is suitable at best to problem-specific situations, the basic problem of constrained motion has since been worked on intensively by numerous people, including Volterra, Boltzmann, Hamel, Novozhilov, Whittaker and Synge, to name but a few.

About 100 years after Lagrange, Gibbs in 1879, and Appell in 1899, independently devised what is today known as the Gibbs–Appell method (2, 3). The method requires a felicitous choice of quasi-coordinates and, like the Lagrange multiplier approach, is amenable to problem-specific situations. This approach is likewise extremely difficult, if not impossible to use, when dealing with systems having several tens of degrees of freedom and several nonintegrable constraints. Yet, since their discovery more than a century ago, the Gibbs–Appell equations have remained the pinnacle of our understanding of constrained motion; they have been referred to by Pars (4) in his magnificent opus on analytical mechanics as "...probably the simplest and most comprehensive equations of motion so far discovered." The problem of constrained motion continues to be worked on. A 1968 monograph on the subject lists more than 500 more recent papers (5). Yet, at the present time there exists no simple explicit set of equations of motion describing constrained systems.

In this paper we present a new fundamental set of "equations of motion" that describe constrained systems. By equations of motion, we mean here the *explicit* second-order differential equations of motion for systems that are holonomically or non-holonomically constrained. These new equations lead us to a new fundamental principle of Lagrangian mechanics.

II. Equations of Motion

Consider first an unconstrained system of particles, each particle having a constant, but different, mass. By "unconstrained" we mean that the number of generalized coordinates, n, needed to describe the configuration of the system at any time, t, equals the number of degrees of freedom of the system. We can write down the equations of motion for such a system, using either Lagrange's equations or Newtonian mechanics, in the form

$$M(\mathbf{q}, t)\ddot{\mathbf{q}} = \mathbf{Q}(\mathbf{q}, \dot{\mathbf{q}}, t) \tag{1}$$

where $\mathbf{q}(t)$ is the n-vector (i.e. n by 1 vector) of generalized coordinates, M is an n by n matrix, Q is the known n-vector of impressed (or "given") forces, and the dots refer to differentiation with respect to time. Elementary Lagrangian mechanics informs us that the matrix M is always symmetric and positive definite. We note from Eq. (1) that the quantity $\mathbf{a}(t) = M^{-1}\mathbf{Q}$ gives the acceleration corresponding to the unconstrained motion of the system at time t.

Next, let this system be subjected to a set of m given consistent Pfaffian equality constraints of the form

$$\sum_{j=1}^{n} a_{ij}(\mathbf{q}, t)\, \mathrm{d}q_j(t) + g_i(\mathbf{q}, t)\, \mathrm{d}t = 0, \quad i = 1, 2, \ldots, m \tag{2}$$

where q_j is the jth component of the n-vector q. These m equations need not be linearly independent. Holonomic constraints can be differentiated once with respect to time, and put in the form of Eq. (2). Nonholonomic equality constraints are Pfaffian in form. Thus the equation set (2) comfortably accommodates all the constraints that fall within the usual framework of Lagrangian mechanics.

Equation set (1) pertains to the description of the motion of the unconstrained system; equation set (2) pertains to the further constraints imposed on the system. Together, these two sets of equations encompass all discrete dynamical systems that go under the usual rubric of analytical mechanics (4). Our purpose is to obtain the *explicit* equations of motion for the constrained system described above, given that at some time t_0, the quantities $\mathbf{q}(t_0)$ and $\dot{\mathbf{q}}(t_0)$ are known, and are compatible with the prescribed constraints.

We note that for the constrained system, the number of generalized coordinates, n, exceeds the number of degrees of freedom of the system. If these are k linearly independent equations in the set (2), then the number of degrees of freedom of the system now becomes $(n-k)$. We shall, however, resist the customary temptation to eliminate the redundant generalized coordinates—a strategy that has been customarily utilized, albeit with little success, for the last 200 years. Instead, the underlying theme of our approach will be to explicitly determine (an equation for) the acceleration $\ddot{\mathbf{q}}(t)$ of the constrained system at any time t, given the vectors of generalized displacement, $\mathbf{q}(t)$, and generalized velocity, $\dot{\mathbf{q}}(t)$, and given that they satisfy the constraints (2).

Assuming that $a_{ij}(\mathbf{q}, t)$ and $g_i(\mathbf{q}, t)$ are sufficiently smooth functions of t, we begin by differentiating the constraint set (2) with respect to time to yield the consistent matrix equation

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

where $A = [a_{ij}]$ is an m by n matrix and $\mathbf{b}$ is a suitably defined m-vector that results from carrying out the differentiation of Eq. (2). We will find that the m by n, matrix $B = AM^{-1/2}$ plays an important role in the characterization of constrained motion; accordingly, we shall refer to it as the *Constraint Matrix*. For brevity, in what follows, we will usually omit the explicit arguments of the various matrices and vectors.

Now a virtual displacement at an instant of time t, is defined as any non-zero, infinitesimal n-vector, $\mathbf{v}$, that satisfies the relation (4)

$$A\mathbf{v} = \mathbf{0}. \tag{4}$$

Define $\mathbf{v} = M^{-1/2}\mathbf{u}$. Hence Eq. (4) is equivalent to the equation

$$B\mathbf{u} = 0, \tag{5}$$

where $\mathbf{u}$ is any non-zero n-vector, and $B = AM^{-1/2}$ is the Constraint Matrix.

The acceleration of the constrained system must satisfy Eq. (3). Defining this acceleration as $\ddot{\mathbf{q}} = M^{-1/2}\ddot{\mathbf{r}}$, Eq. (3) can be rewritten as,

$$B\ddot{\mathbf{r}} = \mathbf{b}, \tag{6}$$

whose explicit solution is

$$\ddot{\mathbf{r}} = B^{+}\mathbf{b} + (I - B^{+}B)\mathbf{y}. \tag{7}$$

Here, the n by m matrix B^{+} is the usual Moore–Penrose (MP) generalized inverse (6, 7) of the matrix B, and $\mathbf{y}$ is any arbitrary n-vector (see the Appendix, items 1 and 2).

The first term on the right-hand side of Eq. (7) is known since both the vector $\mathbf{b}$ and the matrix B^{+} are known; hence our aim is to determine the vector, $(I - B^{+}B)\mathbf{y}$, based on the principles of mechanics.

The equation of motion describing the constrained system can be expressed as

$$M(\mathbf{q}, t)\ddot{\mathbf{q}} = \mathbf{Q}(\mathbf{q}, \dot{\mathbf{q}}, t) + \mathbf{Q}_c(\mathbf{q}, \dot{\mathbf{q}}, t) \tag{8}$$

where the additional "constraint force" $\mathbf{Q}_c$ arises by virtue of the constraints (2) imposed on the system. Our purpose would therefore be accomplished if we could determine $\mathbf{Q}_c$ explicitly.

By D'Alembert's principle (8), the motion of the constrained system proceeds in such a way that at each instant of time, the total work done by the forces of constraint under virtual displacements equals zero. Using relation (4) and Eq. (8), this implies that for all nonzero vectors $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$,

$$\mathbf{v}^{T}\mathbf{Q}_c = \mathbf{v}^{T}[M\ddot{\mathbf{q}} - \mathbf{Q}] = 0. \tag{9}$$

We now express condition (9) in terms of the previously defined vectors $\mathbf{u}$ and $\ddot{\mathbf{r}}$, noting the equivalence between relations (4) and (5). The acceleration $\ddot{\mathbf{r}}$

corresponding to the constrained motion must then be such that for all nonzero vectors $\mathbf{u}$ which satisfy the condition $B\mathbf{u} = 0$, we must have

$$\mathbf{u}^T M^{-1/2}\mathbf{Q}_c = \mathbf{u}^T M^{-1/2}(M^{1/2}\ddot{\mathbf{r}} - \mathbf{Q}) = \mathbf{u}^T(\ddot{\mathbf{r}} - M^{-1/2}\mathbf{Q}) = 0. \tag{10}$$

Since the acceleration $\ddot{\mathbf{r}}$ must satisfy the constraints, it must have the form given in Eq. (7); hence, from (10), for all nonzero vectors $\mathbf{u}$ which satisfy the condition $B\mathbf{u} = 0$, D'Alembert's principle demands that

$$\mathbf{u}^T[B^+\mathbf{b} + (I - B^+B)\mathbf{y} - M^{-1/2}\mathbf{Q}] = 0. \tag{11}$$

However, $B\mathbf{u} = \mathbf{0}$ implies that $\mathbf{u}^+B^+ = \mathbf{0}$, and this in turn implies that $\mathbf{u}^TB^+ = \mathbf{0}$ since $\mathbf{u} \neq \mathbf{0}$ (see the Appendix, items 3 and 4). Condition (11) thus requires the vector $\mathbf{y}$ to be such that

$$\mathbf{u}^T(\mathbf{y} - M^{-1/2}\mathbf{Q}) = 0, \quad \text{for all nonzero vectors } \mathbf{u} \text{ such that } \mathbf{u}^TB^T = \mathbf{0}. \tag{12}$$

Noting (12), and the condition that $\mathbf{u}^TB^T = \mathbf{0}$, we see that $\mathbf{y}$ must therefore be given by

$$\mathbf{y} = M^{-1/2}\mathbf{Q} + \mathbf{z} \tag{13}$$

where $\mathbf{z}$ is any n-vector belonging to the column space of B^T, i.e. $\mathbf{z}$ is any n-vector that is some linear combination of the columns of B^T. Hence $\mathbf{z}$ can be expressed as $\mathbf{z} = B^T\mathbf{w}$, where $\mathbf{w}$ is some m-vector. Using Eq. (13) in (7), we obtain

$$\ddot{\mathbf{r}} = B^+\mathbf{b} + (I - B^+B)(M^{-1/2}\mathbf{Q} + B^T\mathbf{w}). \tag{14}$$

However, $(I - B^+B)$ is a symmetric matrix (see the Appendix, item 5). Therefore, using the MP conditions, $[(I - B^+B)B^T\mathbf{w}]^T = \mathbf{w}^TB(I - B^+B) = \mathbf{w}^T(B - B) = \mathbf{0}$. Hence $(I - B^+B)B^T\mathbf{w} = \mathbf{0}$, and Eq. (14) becomes

$$\ddot{\mathbf{r}} = B^+\mathbf{b} + (I - B^+B)M^{-1/2}\mathbf{Q}. \tag{15}$$

Noting that $\ddot{\mathbf{q}} = M^{-1/2}\ddot{\mathbf{r}}$, we obtain the following new fundamental explicit equation of motion of the constrained system:

$$M\ddot{\mathbf{q}}(t) = \mathbf{Q} + M^{1/2}(AM^{-1/2})^+(\mathbf{b} - AM^{-1}\mathbf{Q}). \tag{16}$$

Our purpose is thus accomplished.

We next express Eq. (16) in a manner which lends considerable physical insight into the nature of constrained motion. Let us say that at some time t, we know the generalized displacement $\mathbf{q}(t)$, and the generalized velocity $\dot{\mathbf{q}}(t)$, and that these two vectors are compatible with the constraint equations (2). Since $\mathbf{q}(t)$ and $\dot{\mathbf{q}}(t)$ are known, $M(\mathbf{q}, t)$ and $\mathbf{Q}(\mathbf{q}, \dot{\mathbf{q}}, t)$ are known; the acceleration, $\mathbf{a}(t)$, of the unconstrained system can then be easily determined from Eq. (1), as noted before, and equals $M^{-1}\mathbf{Q}$. Thus determining the acceleration of the constrained system at time t, amounts to finding out in what way, and by how much, it differs from that of the unconstrained system, $\mathbf{a}(t)$, which we know.

To understand this difference, we first rewrite Eq. (16) after premultiplying on the left by M^{-1} as

$$\ddot{\mathbf{q}}(t) = \mathbf{a}(t) + M^{-1/2}(AM^{-1/2})^+(\mathbf{b} - A\mathbf{a}). \tag{17}$$

Here we have made use of the fact that the acceleration of the unconstrained system is given by $\mathbf{a}(t) = M^{-1}\mathbf{Q}$. The second term on the right-hand side of Eq. (17) can be thought of as an adjustment to the acceleration $\mathbf{a}(t)$ created by the presence of the constraints. Equation (17) describes the motion of the constrained system and is thus an alternative form of the fundamental equation (16) obtained earlier.

From this equation, we get

$$\ddot{\mathbf{q}}(t)-\mathbf{a}(t) = M^{-1/2}(AM^{-1/2})^{+}(\mathbf{b}-A\mathbf{a}) = K(\mathbf{b}-A\mathbf{a}) \tag{18}$$

where the n by m matrix $K = M^{-1/2}(AM^{-1/2})^{+}$ is just the "weighted" Moore–Penrose generalized inverse of the Constraint Matrix, $AM^{-1/2}$. The weighting matrix is $M^{-1/2}$.

The left-hand side of Eq. (18) is now the *deviation* of the acceleration of the constrained system from that of the unconstrained system at time t; we shall denote this by $\Delta\ddot{\mathbf{q}}$. The quantity $(\mathbf{b}-A\mathbf{a})$ is the extent to which the acceleration $\mathbf{a}$ of the unconstrained system does not satisfy the constraint equation set (3) at time t; we shall denote this by $\mathbf{e}$. The fundamental equation (17) can therefore be rewritten as

$$\Delta\ddot{\mathbf{q}} = K\mathbf{e} \tag{19}$$

where we have explicitly exposed the linearity between $\Delta\ddot{\mathbf{q}}$ and $\mathbf{e}$. This leads to the following new principle of Lagrangian mechanics.

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 K, *is the weighted Moore–Penrose generalized inverse of the Constraint Matrix. and the measure of dissatisfaction of the constraints is the vector* $\mathbf{e}$.

One cannot but be enchanted with the beauty and parsimony of Nature's design: 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 are not satisfied at that instant of time, much like the calculating mathematician.

III. Conclusions

This paper resolves a fundamental issue in the field of analytical mechanics—an issue which was identified more than 200 years ago, and one which has been intensively worked on since. We note that our fundamental equations (16) and (17) are obtained without the use of Lagrange multipliers. The need for quasi-coordinates is also obviated through the observation that though nonintegrable constraints cannot be integrated, they can always be differentiated if $a_{ij}(\mathbf{q}, t)$ and $g_i(\mathbf{q}, t)$ are smooth enough. It is strange that despite the highly nonlinear behavior of even the simplest of mechanical systems, it is the tools of linear algebra that are used in obtaining equations (16) and (17). Contrary to 2 centuries of conventional wisdom, no attempt is made to eliminate coordinates, even though the number of generalized coordinates exceeds the number of degrees of freedom of the constrained system. The equations of motion are consequently obtained in the same coordinates as those used to describe the unconstrained motion of the system; this, as we have shown, leads to a deeper physical understanding of the beauty and simplicity of Nature's inner workings.

The equations of motion presented in this paper appear to be the simplest and most comprehensive so far discovered. The new principle of analytical mechanics presented herein appears to be the simplest and most general way of describing constrained motion.

References

(1) J. L. Lagrange, "Mechanique Analytique", Mme. Ve Courcier, Paris, 1787.

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

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

(4) L. A. Pars, "A Treatise on Analytical Dynamics", Ox Bow Press, Woodbridge, CT, 1979.

(5) Ju. I. Neimark and N. A. Fufaev, "Dynamics of nonholonomic systems", *Am. Math. Soc.*, Vol. 33, Providence, R.I., 1968, translated from Russian in 1972.

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

(7) R. Penrose, "A generalized inverse of matrices", *Proc. Cambridge Phil. Soc.*, Vol. 51, pp. 406–413, 1955.

(8) J. le R. D'Alembert, "Traité de Dynamique", Paris, 1743.

Appendix

The results related to the Moore–Penrose (MP) Generalized Inverse used in this paper may be found in any elementary textbook dealing with the subject. However, for completeness, we provide these brief complements.

Throughout this appendix, the matrix B will be taken to be an m by n real matrix; the real vector $\mathbf{u}$ will be taken to be n by 1.

(1) Given an m by n matrix B there always exists a unique n by m matrix B^+, called the MP-generalized inverse of B, which satisfies the following four conditions, herein called the MP conditions:

$$BB^+B = B; \quad B^+BB^+ = B^+; \quad BB^+ = (BB^+)^{\mathrm{T}}; \quad \text{and,} \quad B^+B = (B^+B)^{\mathrm{T}}. \qquad \text{(A1)}$$

(2) It is straightforward to show that the solution of the consistent equation $B\mathbf{x} = \mathbf{b}$ is given by $\mathbf{x} = B^+\mathbf{b} + (I - B^+B)\mathbf{y}$, where $\mathbf{y}$ is an arbitrary n-vector. Consistency of the equations implies that $BB^+\mathbf{b} = \mathbf{b}$.

(3) The MP-inverse of a nonzero n-vector $\mathbf{u}$ is the 1 by n vector $\mathbf{u}^+ = 1/(\mathbf{u}^{\mathrm{T}}\mathbf{u})\mathbf{u}^{\mathrm{T}}$. We can easily verify that $\mathbf{u}^+$ satisfies the four MP conditions stated above in (A1).

(4) If $\mathbf{u} \neq 0$, then $B\mathbf{u} = 0$ implies $\mathbf{u}^{\mathrm{T}}B^+ = 0$.

Since $B\mathbf{u} = \mathbf{0}$, we have, $B^+B\mathbf{u}\mathbf{u}^+ = B^+\mathbf{0}\mathbf{u}^+ = \mathbf{0}$. Taking the transpose on both sides and using the MP conditions, gives

$$\mathbf{0} = [(B^+B)(\mathbf{u}\mathbf{u}^+)]^{\mathrm{T}} = (\mathbf{u}\mathbf{u}^+)^{\mathrm{T}}(B^+B)^{\mathrm{T}} = \mathbf{u}\mathbf{u}^+B^+B.$$

Again using the MP conditions, $\mathbf{u}^+B^+ = (\mathbf{u}^+\mathbf{u}\mathbf{u}^+)(B^+BB^+) = \mathbf{u}^+(\mathbf{u}\mathbf{u}^+B^+B)B^+ = \mathbf{u}^+(0)B^+ = \mathbf{0}$. Noting the expression for $\mathbf{u}^+$ obtained in item 3 above, the result now follows.

(5) The matrix $(I-B^+B)$ is symmetric because $(I-B^+B)^{\mathrm{T}} = I^{\mathrm{T}}-(B^+B)^{\mathrm{T}} = I-B^+B$ using the last MP condition in item 1 above.

Vol. 330, No. 3, pp. 571–577, 1993
Printed in Great Britain