



# The General Gauss Principle of Least Constraint

**Firdaus E. Udvardia**

Professor

Department of Civil and Environmental  
Engineering;

Department of Aerospace and Mechanical  
Engineering;

Information and Operations Management;  
Olin Hall of Engineering,

University of Southern California,  
Los Angeles, CA 90089

e-mails: [fudvardia@usc.edu](mailto:fudvardia@usc.edu); [feuUSC@gmail.com](mailto:feuUSC@gmail.com)

*This paper develops a general form of Gauss's Principle of Least Constraint, which deals with the manner in which Nature appears to orchestrate the motion of constrained mechanical systems. The theory of constrained motion has been at the heart of classical mechanics since the days of Lagrange, and it is used in various areas of science and engineering like analytical dynamics, quantum mechanics, statistical physics, and nonequilibrium thermodynamics. The new principle permits the constraints on any mechanical system to be inconsistent and shows that Nature handles these inconsistent constraints in the least squares sense. This broadening of Gauss's original principle leads to two forms of the General Gauss Principle obtained in this paper. They explain why the motion that Nature generates is robust with respect to inaccuracies with which constraints are often specified in modeling naturally occurring and engineered systems since their specification in dynamical systems are often only approximate, and many physical systems may not exactly satisfy them at every instant of time. An important byproduct of the new principle is a refinement of the notion of what constitutes a virtual displacement, a foundational concept in all classical mechanics. [DOI: 10.1115/1.4062887]*

*Keywords: dynamics, new understandings of Gauss's Principle and d'Alembert's Principle, constraint forces that do work under virtual displacements, inconsistent constraint specifications, the expanded notion of a virtual displacement*

## 1 Introduction

The problem of finding the acceleration of a constrained mechanical system, given the acceleration of the unconstrained system and the constraints acting on it, is one of the central problems in mechanics. It was first broached systematically by Lagrange some 225 years ago [1] and has been worked on continuously by mathematicians, physicists, and engineers, to the present day. Lagrange's development of generalized coordinates was in part a direct consequence of his attempt to handle holonomically constrained systems by permitting the generalized coordinates to be the constraints themselves. He developed and refined the idea of using the so-called Lagrange multipliers (first devised by Euler to study maxima and minima of functions) to obtain the equations of constrained motion. In the years that followed, the Lagrange multiplier approach became the preeminent way of determining the behavior of constrained mechanical systems [2–5]. However, the determination of Lagrange multipliers is problem-specific; they cannot be obtained in closed form in a general setting, and for large-scale problems involving hundreds or thousands of degrees of freedom, and numerous constraints, both holonomic and nonholonomic, the approach usually becomes unfeasible.

Among his other interests, Dirac spent a considerable amount of time exploring the development of the equations of motion of constrained systems in classical mechanics in order to use these developments later on in quantum mechanics [6,7]. He employs the Hamiltonian formulation of classical mechanics because of its

easy extension to quantum mechanics and uses Poisson brackets to write the equations of motion. He too uses the ubiquitous Lagrange multiplier approach. He takes the given constraints, called “primary constraints,” to be functions of the generalized coordinates and the corresponding momenta with no explicit dependence on time. These primary constraints are used in an iterative approach to generate sets of “secondary constraints,” which are later treated for many purposes just like the primary constraints. The iterative approach finally leads to a set of linear equations from which the Lagrange multipliers are finally obtained. While the method provides a useful approach for singular Lagrangians, at each iterative step the scheme generates secondary constraints that are problem specific. Though a general procedure, it does not therefore result in a straightforward equation that explicitly provides the acceleration of a constrained system that may be subjected to general holonomic and/or nonholonomic constraints, which may or may not explicitly depend on time.

In 1829, Gauss published a remarkable three-page paper entitled, “On a Universal Principle of Mechanics,” in which he explained a deep new principle that appears to be at the root of how Nature executes the motion of constrained mechanical systems [8]. He showed in a simple and elegant manner that when a mechanical system is constrained, Nature acts like a mathematician. It picks at each instant of time, from among all the *possible* accelerations that satisfy the constraints that particular acceleration for the constrained system that minimizes, in a weighted least squares sense, the difference between the acceleration of the unconstrained and the constrained system. A *possible* acceleration is any acceleration that (exactly) satisfies the constraints. Today, this principle is what is called “Gauss's Principle of Least Constraint.” This paper aims to provide a significant and useful extension of this principle, which was originally proposed by Gauss.

Contributed by the Applied Mechanics Division of ASME for publication in the JOURNAL OF APPLIED MECHANICS. Manuscript received June 14, 2023; final manuscript received June 26, 2023; published online August 2, 2023. Tech. Editor: Pradeep Sharma.

More specifically, consider an unconstrained mechanical system described by

$$M(q, t)a(t) = Q(q, \dot{q}, t), \quad q(0) = q_0, \quad \dot{q}(0) = \dot{q}_0 \quad (1)$$

where  $M$  is an  $n$ -by- $n$  positive definite matrix,  $Q$  is the known  $n$ -vector ( $n$  by 1 vector) of what is called the “given” or “impressed force,”  $a$  is the  $n$ -vector of acceleration at time  $t$ , and  $q$  and  $\dot{q}$  are the generalized position and generalized velocity of the system. By unconstrained we mean that the components of the  $n$ -vectors  $q_0$  and  $\dot{q}_0$  can be arbitrarily chosen. Both the matrix  $M(q, t)$  and the  $n$ -vector  $Q(q, \dot{q}, t)$  are assumed to be known functions of  $q, \dot{q}$ , and  $t$ , so that the acceleration  $a(t)$  of the unconstrained system is known to be simply  $M(q, t)^{-1}Q(q, \dot{q}, t)$ .

Assume now that this system is subjected to a given set of  $m$  smooth ( $C^1$ ) constraints described by

$$\varphi_i(q, \dot{q}, t) = 0, \quad i = 1, 2, \dots, m \quad (2)$$

where  $\varphi_i$  are known functions of  $q, \dot{q}$ , and  $t$ . We note that Eq. (2) includes both holonomic and nonholonomic constraints.

The quintessential problem is mechanics, which is referred to in the first line of this paper and can then be succinctly stated as: What is the acceleration  $\ddot{q}$  of the constrained system?

In other words, how does one find the acceleration  $\ddot{q}(t)$  of the constrained system, which Nature provides, and is now described by the equation

$$M(q, t)\ddot{q}(t) = Q(q, \dot{q}, t) + Q^C(q, \dot{q}, t), \quad q(0) = q_0, \quad \dot{q}(0) = \dot{q}_0 \quad (3)$$

in which the  $n$ -vector  $Q^C(q, \dot{q}, t)$ , called the force of constraint, is engendered by the presence of the constraints described in Eq. (2).

One immediately notices that in the presence of constraints the initial conditions in Eq. (3) can no longer be chosen arbitrarily as they were in Eq. (1), because the components of each of the  $n$ -vectors  $q_0$  and  $\dot{q}_0$  in Eq. (3) now need to satisfy the  $m$  constraints  $\varphi_i(q, \dot{q}, t) = 0$  at time  $t=0$ . Equation (2) can be rewritten upon appropriate differentiation with respect to time as

$$A(q, \dot{q}, t)\dot{q} = b(q, \dot{q}, t) \quad (4)$$

where the  $m$ -by- $n$  matrix  $A(q, \dot{q}, t)$ , which we will call the constraint matrix, has rank  $r > 0$ . Each constraint provides one row of the equation set (4). When the rank of  $A(q, \dot{q}, t)$  is  $r$ , we shall say in this paper that there are  $r$  constraints  $\varphi_{k_i}$ ,  $i = 1, \dots, r$  among the  $m$  that are independent; when  $A(q, \dot{q}, t)$  has full row rank  $m$ , we shall refer to the constraints as being independent. It is clear that the consistency condition imposed by Eq. (4) may not be satisfied by a physical system at each instant of time, since it is only an approximation of the physical situation as modeled by the investigator. One of the aims of this paper is to permit these prescribed equations of constraint to be only approximately satisfied by the actual system, even allowing for the possibility of the prescribed constraints being inconsistent, as could possibly happen when these constraints are obtained from measurements carried out from physical experiments. Thus, situations in which the prescribed constraints are improperly and/or inaccurately supplied are included.

Equation (3) shows that at each instant of time  $t$ , we have two unknown  $n$ -vectors— $\ddot{q}$  and  $Q^C$ —and therefore a total of  $2n$  unknowns that need to be found. However, we have only the  $n$  equations given in Eq. (3) and  $m$  equations in Eq. (2) (or in Eq. (4)), a total of  $(n+m)$  equations at hand, of which  $(n+r)$  are independent. To determine the  $2n$  unknowns, we need  $2n - (n+r) = n-r$  additional independent relations, and this is where the main problem of understanding the way in which Nature orchestrates constrained motion resides.

The general approach for obtaining these additional equations was provided by considering an assumption first introduced by d’Alembert [9] and later described with precision by Lagrange [1]. It invokes two fundamental notions called virtual displacements and virtual work that are at the bedrock of mechanics. A virtual

displacement is defined as any nonzero  $n$ -vector  $v(t)$  that satisfies the relation

$$A(q, \dot{q}, t)v(t) = 0 \quad (5)$$

Thus, a virtual displacement vector  $v$  at time  $t$  belongs to the null space of the matrix  $A(q, \dot{q}, t)$ . This notion of a virtual displacement was initially applied to nonholonomic constraints that had the so-called Pfaffian form in which the generalized velocities appear only linearly [3–5]. Equation (5) shown here uses the expanded notion of virtual displacements developed more recently for general nonholonomic constraints in which the generalized velocities (see Eq. (2)) can appear nonlinearly [10].

Lagrange’s assumption (called d’Alembert’s Principle) prescribes that the force of constraint  $Q^C(q, \dot{q}, t)$  does no work under any (every) virtual displacement, i.e., that

$$v^T(t)Q^C(q, \dot{q}, t) = 0 \quad (6)$$

In this paper, we will include an extension of d’Alembert’s Principle and consider situations in which the right-hand side of Eq. (6) is not zero, but is prescribed by a known  $n$ -vector  $C(q, \dot{q}, t)$  for a given specific constrained system so that

$$v^T(t)Q^C(q, \dot{q}, t) = v^T(t)C(q, \dot{q}, t) \quad (7)$$

As before, by “known,” we mean a known function of  $q, \dot{q}$  and  $t$ . The force  $n$ -vector  $C$ , could for example, represent the force of friction on a surface. The right-hand side of Eq. (7) can be chosen to be positive, zero, or negative, to account for forces that do work on the system in a virtual displacement. Thus, energy can be added and/or extracted through the constraints as in, for example, modeling friction provided by a constraining surface. When  $C \equiv 0$ , Eqs. (6) and (7) become identical and we obtain d’Alembert’s Principle (assumption). In the interest of brevity, from here on, we will drop the arguments of the various entities mentioned earlier, except when necessary for clarity.

In contrast to d’Alembert’s Principle (Eq. (6)), Gauss’s Principle of Least Constraint, which chronologically came about 40 years after Lagrange’s book [1], provides a different approach to understanding the manner in which Nature picks the acceleration  $\ddot{q}$  of a constrained system at every instant of time  $t$ . It does not use notions like virtual displacement and virtual work. It states that at each instant of time, the acceleration of the constrained system  $\ddot{q}$  is obtained so that [8]

$$\begin{aligned} \ddot{q} = \min \|(\ddot{z} - a)\|_M^2 := \min \|(\ddot{z} - a)^T M(\ddot{z} - a)\|, \\ \text{from among all possible acceleration } n\text{-vectors} \\ \ddot{z} \text{ that (exactly) satisfy the constraint equation } A\ddot{z} = b \end{aligned} \quad (8)$$

The subscript on the square of the Euclidean norm refers to the positive definite mass matrix  $M$  in Eq. (1) that weights the norm. We shall refer from here on to  $\|(\ddot{z} - a)\|_M^2$  as the Gaussian,  $G(t)$ .

Gauss’s principle thus says that at each instant of time Nature: (1) first, obtains the set,  $Z$ , of all the possible accelerations  $\ddot{z}$  that satisfy the (consistent) constraints, i.e., that exactly satisfy the consistent equation  $A\ddot{z} = b$ , (2) and then picks from this set,  $Z$ , of possible accelerations the one acceleration vector,  $\ddot{q}$ , that minimizes  $G = \|(\ddot{z} - a)\|_M^2$ . Nature, according to Gauss, behaves like a mathematician and resorts to a weighted least squares solution to get  $\ddot{q}$  at each instant of time. Gauss assumed that the mass matrix  $M$  is positive definite, which is what is assumed in this paper too [3,8].

Gauss’s approach to understanding constrained motion is markedly different from that of d’Alembert’s. Specifically, it is totally innocent of notions like virtual displacements and Lagrange multipliers. It is this radically different view of Gauss regarding the way Nature operates that permits the development in classical mechanics of a single explicit equation of motion for constrained mechanical systems subjected to holonomic and/or nonholonomic constraints, in which time may or may not be explicitly involved [11]. It is also noteworthy that unlike most other so-called “minimum-

principles,” Gauss’s Principle is perhaps the only fundamental principle in classical mechanics in which the minimum is a unique *global* minimum. Thus, though complementary to one another in the sense that they each, of course, lead to the same acceleration of the constrained mechanical system at each instant of time, Gauss’s Principle and d’Alembert’s Principle (Eqs. (6) and (8)) are conceptually quite different from one another.

In this paper, we provide a more general form of Gauss’s principle wherein we

- (1) include the extended form of the d’Alembert Principle, in which we consider systems in which Eq. (7) rather than Eq. (6) is valid; that is, we assume that the total work done by the constraint forces is positive, negative, or zero, under virtual displacements [12]; and
- (2) simultaneously relax the condition in Eq. (8) by requiring that all the *possible* acceleration  $n$ -vectors  $\ddot{z}$  eligible in the minimization of  $G(t)$  belong to the set  $Z_g$  of  $n$ -vectors that satisfy the constraint equation  $A\ddot{z} \simeq b$ , in which the sign ‘ $\simeq$ ’ is understood to mean throughout this paper “in the least squares sense.” This extension has the important consequence of allowing the constraints to be inconsistent, often a result of (a) experimental measurement errors when describing the constraints placed on complex mechanical systems that could bring about errors in the prescription of the  $m$ -vector  $b$  in Eq. (4), and (b) numerical errors brought about during computational procedures.

We note that the prescribed constraint equation  $A\ddot{x} = b$ , which must be satisfied at every instant of time in the evolving motion of a constrained system, is at best an approximate description of the physically occurring constraints in a mechanical system and is unlikely to be exactly satisfied in a mathematical sense by any actual system. The actual physical constraints acting on a system may be better modelled by the equation  $A\ddot{x} = b + \varepsilon(q(t), \dot{q}(t), t)$ , where the  $n$ -vector  $\varepsilon(t)$  denotes errors (modeling and/or computational) in the specification of the constraints. The question then arises: what is the extent to which the form of the equation of motion (orchestrated by Nature) is sensitive to the accuracy with which the constraints are specified? In what follows we show that Nature does not necessarily appear to look for the exact satisfaction of all the constraints when devising the motion of constrained systems but only considers their satisfaction in the least squares sense. For example, for a wheel rolling down a flat surface, the specified constraint that the surface is “flat” only approximates the actual physical constraint, since at each instant of time along the wheel’s path there may be small surface undulations which are not specified in the prescribed constraint equation. Another way of stating this question is: why is Nature robust (forgiving) in its response with respect to miss-statements in the specification of the constraints that may be prevalent at every instant of time throughout the duration of motion of a constrained mechanical system? This question is answered by the General Gauss Principle obtained herein that Nature seems to abide by.

## 2 The General Gauss Principle of Least Constraint

As seen earlier, the dynamical system is specified by the mass matrix  $M(q, t)$ , the constraint matrix  $A(q, \dot{q}, t)$ , and the vectors  $Q(q, \dot{q}, t)$ ,  $C(q, \dot{q}, t)$ , and  $b(q, \dot{q}, t)$ ; these quantities are assumed to be known functions of the  $q$ ’s,  $\dot{q}$ ’s, and  $t$ .  $M$  is an  $n$ -by- $n$  positive definite matrix,  $A$  is the  $m$ -by- $n$  matrix,  $Q$  and  $C$  are  $n$ -vectors, and  $b$  is an  $m$ -vector. The  $n$ -vector  $a := M^{-1}Q$ . In what follows, the constrained system shall refer to the system described in the previous section using these entities.

Consider the set of  $n$ -vectors  $Z_g = \{\ddot{z}: A\ddot{z} \simeq b\}$ , where  $\ddot{z}$  is the least squares solution of the equation  $A\ddot{z} \simeq b$ . Our aim is to show that the equation of motion of the constrained system obtained by doing the minimization of  $G_g$  (described below in Eq. (10)) when using  $n$ -vectors from the set  $Z$  described earlier has a form identical to

that obtained by using  $n$ -vectors from the set  $Z_g$ . We shall therefore assume in what follows that the specified constraint equation  $A\ddot{q} \simeq b$  may or may not be consistent. In contrast to the set  $Z_g$ , recall that the  $n$ -vectors  $\ddot{z}$  that belong to the set  $Z$  (used in Gauss’s original principle) are required to exactly satisfy the consistent constraint equation  $A\ddot{z} = b$ , i. e.,  $Z = \{\ddot{z}: A\ddot{z} = b\}$ .

We begin by stating the new principle, which we refer to as the First Form of the General Gauss Principle of Least Constraint. In what follows, we use the notation established in Sec. 1.

**First Form of the General Gauss Principle of Least Constraint.** *At each instant of time  $t$ , Nature picks the unique acceleration  $\ddot{q}(t)$  of a constrained mechanical system such that*

$$\ddot{q} = \min \|\ddot{z} - M^{-1}(Q + C)\|_M^2 \text{ from among all acceleration}$$

$$n\text{-vectors } \ddot{z} \text{ for which } \|(A\ddot{z} - b)\|^2 \text{ is a minimum.}$$

$$\text{The minimum attained at } \ddot{q}, \text{ is a global minimum.} \quad (9)$$

At every instant of time, we define the generalized Gaussian (with subscript  $g$  for generalized) as

$$G_g := \|\ddot{z} - M^{-1}(Q + C)\|_M^2 \quad (10)$$

Comparison with  $G$  given in statement (8) shows that  $G_g$  contains the vector  $C$ , so the force of constraint  $Q^C$  can do positive, negative, or zero, work in a virtual displacement, when viewed from the viewpoint of d’Alembert’s Principle (see Eq. (7)) [12,13]. More importantly, in statement (9), the  $n$ -vectors  $\ddot{z}$  used in the minimization of  $G_g$  belong to the set  $Z_g$  and not to the set  $Z$ ; that is, we permit inconsistent constraints at each instant of time, which take the form

$$A\ddot{q} - b = \varepsilon(t) \neq 0, \text{ or } A\ddot{q} \simeq b \quad (11)$$

where  $b$  may not be in the range space of  $A$ , thereby allowing for errors in the specification/prescription of the constraints, which may arise either from computational procedures and/or from inaccurate descriptions of one or more of the physically extant constraints.

Our discussion below shows that it appears to be helpful in analytical dynamics to expand the notion of a *possible* acceleration vector. We define, in our expanded view, a *possible* acceleration at any given instant of time  $t$ , as any acceleration  $n$ -vector,  $\ddot{z}$ , which satisfies the constraint equation  $A\ddot{z} \simeq b$  imposed on the system *in the least squares sense*, at that instant of time. We thus define the set  $Z_g$  that contains all such *possible* acceleration vectors, at a given instant of time, by

$$Z_g(t) = \{\ddot{z}(t): \varepsilon(t)^T \varepsilon(t) = (A\ddot{z} - b)^T (A\ddot{z} - b) \text{ is a minimum}\} \quad (12)$$

The First Form of the General Gauss Principle of Least Constraint then says that at each instant of time Nature picks out from the set,  $Z_g$ , the one unique vector,  $\ddot{z}^*$ , that minimizes the Gaussian  $G_g$ , and it thereby obtains the acceleration of the constrained mechanical system,  $\ddot{q} := \ddot{z}^*$ , at that instant of time. (That the vector  $\ddot{z}^*$  is unique, will be shown below.) This can be then stated as

$$\ddot{q}(t) := \ddot{z}^*(t) = \min_{\ddot{z} \in Z_g(t)} [G_g(t)] = \min_{\ddot{z} \in Z_g} [\|\ddot{z} - M^{-1}(Q + C)\|_M^2] \quad (13)$$

Our approach to establish this result is to find  $\ddot{q}$  explicitly using Eq. (13) and show that it results in an equation that has exactly the same form as the equation of motion obtained by using the  $n$ -vectors  $\ddot{z}$  that belong to the set  $Z$  to minimize  $G_g$ , which is known to be [13,14]

$$\ddot{q} = a + M^{-1/2} B^+ (b - Aa) + M^{-1/2} (I - B^+ B) M^{-1/2} C \quad (14)$$

where the  $m$ -by- $n$  matrix  $B := AM^{-1/2}$ ,  $X^+$  is the Moore–Penrose (MP) inverse [15–17] of the matrix  $X$ , and  $a := M^{-1}Q$ . We will refer to Eq. (14) as the *fundamental equation of constrained motion* (FEMC).

Equation (14) not only describes the manner in which Nature orchestrates constrained motion of mechanical systems, but also

opens up new ways of performing exact control of mechanical systems and has been widely used in aerospace, civil, and mechanical engineering; in robotics and control designs for nonlinear systems like unmanned aerial vehicles, and in the area of local and decentralized control of dynamical systems.

We begin by considering the following slightly more general result, which will lead us later on to a Second Form of the General Gauss Principle of Least Constraint.

**Result 1.** At each instant of time  $t$ , the unique acceleration  $n$ -vector  $\ddot{q} = \min [G_g] = \min \|\ddot{z} - M^{-1}(Q + C)\|_M^2$  from among all  $n$ -vectors  $\ddot{z}$  for which  $\|(A\ddot{z} - b)\|_W^2 := (A\ddot{z} - b)^T W(A\ddot{z} - b)$  is a minimum, is given by

$$\ddot{q} = a + M^{-1/2}P^+(\tilde{b} - PM^{-1/2}Q) + M^{-1/2}(I - P^+P)M^{-1/2}C \quad (15)$$

where the  $m$ -by- $n$  matrix  $P = W^{1/2}B = W^{1/2}AM^{-1/2}$ ,  $\tilde{b} = W^{1/2}b$ ,  $a = M^{-1}Q$ , and the  $m$ -by- $m$  matrix  $W > 0$ . The acceleration  $n$ -vector  $\ddot{q}$  provides the global minimum of  $G_g$ . ■

*Remark 1.* Result 1 above differs from the statement of the First Form of the General Gauss Principle in Eq. (13) in that it replaces the set of candidate  $n$ -vectors  $\ddot{z} \in Z_g$  (see Eq. (12)) that are to be used in the minimization of  $G_g$  by the set

$$Z_W(t) = \{\ddot{z}(t): \varepsilon(t)^T W \varepsilon(t) = (A\ddot{z} - b)^T W(A\ddot{z} - b) \text{ is a minimum}\} \quad (16)$$

where the weighting matrix  $W > 0$ .

Obviously, when  $W = I_m$ , the sets  $Z_g$  and  $Z_W$  become identical, and we recover from Result 1 (see Eq. (15)) the First Form of the General Gauss Principle of Least Constraint, since  $P = B = AM^{-1/2}$ ,  $\tilde{b} = b$ , and so Eq. (14) becomes identical to Eq. (15). ■

We provide two different proofs for Result 1, since each provides different insights into the manner in which Eq. (15) results. The first shows explicitly the various quantities of interest from an engineering standpoint. The second is algebraic, and it provides the requisite result in a succinct manner, but it does not lend itself to deeper physical insights.

*Proof 1.* We divide the proof of Eq. (15) into two parts: the first deals with the minimization of  $\|A\ddot{z} - b\|_W^2$ , and the second deals with the minimization of the Gaussian  $G_g$ .

*Part 1 of Proof 1.* We start by considering the well-known least squares solution (LSS),  $\ddot{z}$ , of the possibly inconsistent equation  $A\ddot{z} \simeq b$ , in which one finds

$$\min \|A\ddot{z} - b\|_W^2 = \min \|W^{1/2}(A\ddot{z} - b)\|^2 \quad \text{over all } n\text{-vectors } \ddot{z} \text{ belonging to } \mathfrak{R}^n \quad (17)$$

While this is a well-studied problem (e.g., see Refs. [16] and [17]), we will consider some important aspects of the LSS solution that will be useful to us later on. Several remarks are provided along the way that give immediate insights into the equations obtained at various stages.

We begin by noting that

$$\begin{aligned} \|A\ddot{z} - b\|_W^2 &= \|W^{1/2}(A\ddot{z} - b)\|^2 = \|(W^{1/2} \underbrace{AM^{-1/2}}_B \underbrace{M^{1/2}\ddot{z}}_{\ddot{z}_s} - W^{1/2}b)\|^2 \\ &= \|P\ddot{z}_s - \tilde{b}\|^2 \end{aligned} \quad (18)$$

where the  $m$ -by- $n$  matrix  $P = W^{1/2}B = W^{1/2}AM^{-1/2}$ ,  $\ddot{z}_s = M^{1/2}\ddot{z}$ , and  $\tilde{b} = W^{1/2}b$ . Hence, the minimization problem in Eq. (17) can be alternatively solved by finding the least squares solution,  $\ddot{z}_s$ , to  $P\ddot{z}_s \simeq \tilde{b}$ , and then setting  $\ddot{z} = M^{-1/2}\ddot{z}_s$ . Since  $\ddot{z}$  can be (uniquely) obtained from knowledge of  $\ddot{z}_s$  (since  $M$  is known), it is useful to think of the acceleration  $n$ -vector  $\ddot{z}_s$  as a transformed (scaled) version of  $\ddot{z}$ . In what follows, we shall be dealing primarily with  $\ddot{z}_s$ .

The determination of the least squares solution (LSS), which minimizes  $\|P\ddot{z}_s - \tilde{b}\|$ , leads to the well-known consistent so-called normal equation, given by

$$P^T P \ddot{z}_s = P^T \tilde{b} \quad (19)$$

whose solution can be explicitly obtained as [16]

$$\ddot{z}_s = (P^T P)^+ P^T \tilde{b} + [I_n - (P^T P)^+ P^T P] h \quad (20)$$

where  $h$  is an arbitrary  $n$ -vector, and  $P^+$  is the Moore–Penrose (MP) inverse of  $P$ . Noting that  $(P^T P)^+ P^T = P^+$  [15], Eq. (20) simplifies to

$$\ddot{z}_s = P^+ \tilde{b} + (I_n - P^+ P) h, \quad \text{where } h \text{ is arbitrary} \quad (21)$$

The first component,  $P^+ \tilde{b}$ , of the  $n$ -vector  $\ddot{z}_s$  shown on the right-hand side belongs to  $\text{Range}(P^T)$  (*Range* stands for range space) and is uniquely defined since the MP inverse of a matrix is unique, while the second component,  $(I_n - P^+ P)h$ , belongs to  $\text{Null}(P)$  (*Null* stands for null space) and is arbitrary, since  $h$  is arbitrary. These two components are orthogonal to one another as seen from  $(I - P^+ P)^T P^+ = (I - P^+ P)P^+ = 0$ , and the two subspaces are orthogonal complements of one another.

Having found  $\ddot{z}_s$ , the acceleration  $n$ -vector  $\ddot{z}$  that solves the equation  $A\ddot{z} \simeq b$  (in the least squares sense) is then explicitly given by

$$\ddot{z} = M^{-1/2}\ddot{z}_s = M^{-1/2}[P^+ \tilde{b} + (I - P^+ P)h] \quad (22)$$

and the set  $Z_W$  of all  $n$ -vectors that solve  $A\ddot{z} \simeq b$  (with the weighting matrix  $W$ ) is given by

$$Z_W = \{\ddot{z} = M^{-1/2}[\underbrace{P^+ \tilde{b} + (I - P^+ P)h}_{\ddot{z}_s}]: h \text{ is an arbitrary } n\text{-vector}\} \quad (23)$$

Note that elements of the set  $Z_W$  differ from one another solely because of the arbitrary component,  $(I - P^+ P)h$ , of  $\ddot{z}_s$  that belongs to  $\text{null}(P)$ .

*Remark 2.* The normal equation  $P^T P \ddot{z}_s = P^T \tilde{b}$  has a unique solution if and only if the  $n$ -by- $n$  matrix  $P^T P$  is invertible, i.e.,  $n = \text{rank}(P^T P) = \text{rank}(P)$ . Hence, a unique solution of this normal equation results if and only if the  $m$ -by- $n$  matrix  $P$  has full column rank  $n$ . Furthermore, when  $P^T P$  is invertible, then  $P^+ = (P^T P)^+ P^T = (P^T P)^{-1} P^T$  so that  $P^+ P = I_n$ . Thus, we have shown that when  $P$  has full column rank,  $P^+ P = I_n$ ; the converse is also easy to show. Hence, from Eq. (23), we see that the set  $Z_W$  is comprised of just a single element,  $\ddot{z} = M^{-1/2}P^+ \tilde{b}$ , if and only if  $P^+ P = I_n$ . ■

*Remark 3.* Since  $P = W^{1/2}AM^{-1/2}$  and  $M, W > 0$ , then  $\text{rank}(P) = \text{rank}(A)$  and therefore  $\text{rank}(P^T P) = \text{rank}(A^T A)$ . In that case, as shown in Remark 2,  $P^+ P = A^+ A = I_n$  if and only if  $\text{rank}(P) = \text{rank}(A) = n$ , that is, the constraint matrix  $A$  has full column rank. ■

*Remark 4.* For each different value of the arbitrary  $n$ -vector  $h$ , Eq. (22) gives an LSS to the problem of minimizing  $\|A\ddot{z} - b\|_W^2$ , yet for each of these different solutions,  $\ddot{z}$ , the minimal value of  $\|A\ddot{z} - b\|_W^2$  is independent of  $h$  and equals  $\|(PP^+ - I)\tilde{b}\|^2$ . To show this, we simply substitute for  $\ddot{z}$  from Eq. (22) in  $\|A\ddot{z} - b\|_W^2$ . Thus

$$\begin{aligned} \|A\ddot{z} - b\|_W^2 &= \|P\ddot{z}_s - \tilde{b}\|^2 = \|P[P^+ \tilde{b} + (I_n - P^+ P)h] - \tilde{b}\|^2 \\ &= \|(PP^+ - I_m)\tilde{b}\|^2 := \|\delta\|^2 \end{aligned} \quad (24)$$

We see that every element of the set  $Z_W$  in Eq. (23) attains the same minimal value of  $\|A\ddot{z} - b\|_W^2$  defined earlier as  $\|\delta\|$ . ■

*Remark 5.* If the constraints were consistent and  $A\ddot{q} = b$  (i.e.,  $b$  belongs to the range space of  $A$ ) that would mean that  $\underbrace{W^{1/2}AM^{-1/2}}_P M^{1/2}\ddot{q} = W^{1/2}b$ , which is just  $P\ddot{q}_s = \tilde{b}$ , where  $\ddot{q}_s := M^{1/2}\ddot{q}$ . The necessary and sufficient condition for the equation

$P\ddot{q}_s = \tilde{b}$  to be consistent is known to be  $(PP^+ - I_m)\tilde{b} = 0$  [15]. Hence, a measure of the extent to which the equation  $P\ddot{q}_s = \tilde{b}$  is inconsistent is the  $m$ -vector  $\delta := (PP^+ - I_m)\tilde{b}$ , whose norm is the minimum value of  $\|A\ddot{z} - b\|_W$  that is attained, as seen in Eq. (24). Thus, the norm of the error  $\sqrt{\varepsilon^T W \varepsilon}$  for each acceleration  $n$ -vector  $\ddot{z}$  that minimizes  $\|A\ddot{z} - b\|_W^2$  is the norm of the extent of the inconsistency,  $\delta$ , in the constraint equation  $P\ddot{q}_s \simeq \tilde{b}$ . ■

*Remark 6.* The above discussion underlines the importance of viewing the constraints imposed on a mechanical system as limiting at each instant of time the acceleration  $n$ -vectors that the system can have to only those that satisfy the equation  $A\ddot{z} \simeq b$ . It should be noted that this equation in *linear* in the  $n$ -vectors of accelerations,  $\ddot{z}$ , that the constrained system can access, no matter the nonlinear nature of the constraints in Eq. (2), which are, in general, functions of the  $n$ -vectors  $q$ ,  $\dot{q}$ , and of  $t$ . Our notion of consistency of the imposed constraints is then the easy conformity with that of the consistency of a linear set of equations. Similarly, we will see below that when the constraints are functionally independent the rank of the  $m$ -by- $n$  matrix  $A$  is  $m$ . ■

*Part 2 of Proof 1.* Having obtained the set  $Z_W$  of vectors that minimize  $\|A\ddot{z} - b\|_W^2$ , we now go on to choose from this set containing the  $n$ -vectors  $\ddot{z} = M^{-1/2}\ddot{z}_s$ —which, as seen earlier, differ from one another in the component of  $\ddot{z}_s$  which belongs *only* to  $\text{Null}(P)$ —the one (unique) component in this  $\text{Null}(P)$  that minimizes the Gaussian  $G_g$ .

*Remark 7.* We note that if the set  $Z_W$  of vectors that minimize  $\|A\ddot{z} - b\|_W^2$  is limited and contains just one (a single)  $n$ -vector,

then finding a vector that minimizes the Gaussian  $G_g$  from among all the vectors in the set  $Z_W$  becomes trivial, since the set  $Z_W$  now contains only one element in it! One could indeed exclude this trivial case from consideration when dealing with the minimization of the Gaussian  $G_g$  over the set of  $n$ -vectors,  $\ddot{z}$ , belonging to  $Z_W$ . This case arises if and only if the  $\text{rank}(P) = n$ , making  $P^+P = I$  so that  $\ddot{z}_s = P^+\tilde{b}$  (see Eq. (21)) and  $\ddot{q} = \ddot{z}^* = M^{-1/2}\ddot{z}_s = M^{-1/2}P^+\tilde{b}$  (see Remark 2). Furthermore,  $P^+P = I$  implies that  $\text{rank}(P) = n$ . Thus, when  $P^+P = I$ , the Gaussian has the value  $G_g = \|P^+\tilde{b} - M^{-1}(Q + C)\|^2$ , and the acceleration of the constrained system is then given by  $\ddot{q} = M^{-1/2}P^+W^{1/2}\tilde{b}$ . ■

Having taken care of the trivial case in which the set  $Z_W$  is made up of a singleton, we now consider the minimization of  $G_g$  when the set  $Z_W$  contains more than one  $n$ -vector in it. Rewriting the Gaussian as

$$\begin{aligned} G_g(\ddot{z}) &= \|\ddot{z} - M^{-1}(Q + C)\|_M^2 = [\ddot{z} - M^{-1}(Q + C)]^T M [\ddot{z} - M^{-1}(Q + C)] \\ &= [M^{1/2}\ddot{z} - M^{-1/2}(Q + C)]^T [M^{1/2}\ddot{z} - M^{-1/2}(Q + C)] \\ &= [\ddot{z}_s - M^{-1/2}(Q + C)]^T [\ddot{z}_s - M^{-1/2}(Q + C)] \\ &= \|\ddot{z}_s - M^{-1/2}(Q + C)\|^2 \end{aligned} \quad (25)$$

we see that we need to solve a second least squares problem to find that specific acceleration  $n$ -vector  $\ddot{z}_s$  that minimizes this  $G_g$  from among all the  $n$ -vectors that are given in Eq. (21). Furthermore, Eq. (25) shows that  $G_g(\ddot{z})$  is in fact a function of  $\ddot{z}_s$  and can as well be denoted, using its last line, by  $G_g(\ddot{z}_s)$ .

Using Eq. (21) in the last line of Eq. (25), we get

$$\begin{aligned} G_g(\ddot{z}) &= [P^+\tilde{b} + (I - P^+P)h - M^{-1/2}(Q + C)]^T [P^+\tilde{b} + (I - P^+P)h - M^{-1/2}(Q + C)], \\ &= [P^+\tilde{b} + (I - P^+P)h]^T [P^+\tilde{b} + (I - P^+P)h] - 2[M^{-1/2}(Q + C)]^T (I - P^+P)h + \dots \\ &= [h^T(I - P^+P)(I - P^+P)h] - 2[M^{-1/2}(Q + C)]^T (I - P^+P)h + \dots \\ &= [h^T(I - P^+P)h] - 2[M^{-1/2}(Q + C)]^T (I - P^+P)h + \dots \end{aligned} \quad (26)$$

In the equalities above, the three dots at the end refer to terms that do not contain  $h$ . In the third equality, we have used the fact that the vectors  $(I - P^+P)h$  and  $P^+\tilde{b}$  are orthogonal because  $h^T(I - P^+P)^T P^+\tilde{b} = h^T(I - P^+P)P^+\tilde{b} = 0$ ; in the last equality, the fact that the matrix  $I - P^+P$  is idempotent.

Our aim then is to minimize this quadratic form  $G_g$  over all  $n$ -vectors  $h$ . Noting that  $I - P^+P$  is symmetric, differentiating the last line in Eq. (26) with respect to  $h$ , and setting it to zero, we then get

$$(I - P^+P)h = (I - P^+P)M^{-1/2}(Q + C) \quad (27)$$

Thus, we have found the component of  $\ddot{z}_s$  in the null space of  $P$  which will make  $G_g$  a minimum, and the vector  $\ddot{z}_s^*$  that minimizes the Gaussian  $G_g$  in Eq. (25) is explicitly given by

$$\ddot{z}_s^* = P^+\tilde{b} + (I - P^+P)M^{-1/2}(Q + C) \quad (28)$$

The acceleration that minimizes the Gaussian  $G_g$  from among all  $n$ -vectors  $\ddot{z}$  that minimize  $\|A\ddot{z} - b\|_W^2$  is therefore

$$\begin{aligned} \ddot{q} &:= \ddot{z}^* = M^{-1/2}\ddot{z}_s^* \\ &= M^{-1/2}[(P^+\tilde{b} + (I - P^+P)M^{-1/2}(Q + C))] \\ &= M^{-1}Q + M^{-1/2}P^+(\tilde{b} - PM^{-1/2}Q) + M^{-1/2}(I - P^+P)M^{-1/2}C \end{aligned} \quad (29)$$

which is the same as the equation given in Eq. (15). That the acceleration  $n$ -vector  $\ddot{q}$  is unique, is obvious since each

of the members on the right-hand side of Eq. (29) are uniquely defined.

Using Eq. (28) in Eq. (25), the value of the Gaussian for the  $\ddot{q}$  given in Eq. (29) is then given by

$$\begin{aligned} G_g^{\min}(\ddot{z}^*) &= \|\ddot{z}_s^* - M^{-1/2}(Q + C)\|^2 \\ &= \|P^+\tilde{b} + (I - P^+P)M^{-1/2}(Q + C) - M^{-1/2}(Q + C)\|^2 \\ &= \|P^+[\tilde{b} - PM^{-1/2}(Q + C)]\|^2 \end{aligned} \quad (30)$$

Our last task is to show that  $G_g^{\min}(\ddot{z}^*)$  which is attained when  $\ddot{q} := \ddot{z}^*$  as given in Eq. (29) indeed provides the minimum of  $G_g$  and, in fact, provides its *global* minimum. We show this through reasoning by contradiction.

Assume that  $\ddot{z}_s^*$  given in Eq. (28) does not yield the global minimum of  $G_g$ . Assume that  $\ddot{z}_s$  has a different component in the null space of  $P$  from that shown on the right-hand side of Eq. (27) and that the minimum of  $G_g$  (see Eq. (25)) occurs when

$$\ddot{z}_s = \ddot{z}_s^* + (I - P^+P)\Delta \quad (31)$$

with  $(I - P^+P)\Delta \neq 0$ . Using Eq. (25), the Gaussian whose value is assumed to be a minimum now is then given by

$$G_g(\ddot{z}) = \|\ddot{z}_s^* + (I - P^+P)\Delta - M^{-1/2}(Q + C)\|^2 \quad (32)$$

As shown in Eq. (30),  $\ddot{z}_s^* - M^{-1/2}(Q + C) = P^+[\tilde{b} - PM^{-1/2}(Q + C)]$ , so Eq. (32) becomes

$$\begin{aligned} G_g(\ddot{z}) &= \|P^+[\tilde{b} - PM^{-1/2}(Q + C)] + (I - P^+P)\Delta\|^2 \\ &= \|P^+[\tilde{b} - PM^{-1/2}(Q + C)]\|^2 + \|(I - P^+P)\Delta\|^2 \end{aligned}$$

The second equality above follows because the  $n$ -vectors  $P^+[\tilde{b} - PM^{-1/2}(Q + C)]$  and  $(I - P^+P)\Delta$  are orthogonal. Noting Eq. (30), we find that

$$G_g(\ddot{z}) = G_g^{\min}(\ddot{z}^*) + \|(I - P^+P)\Delta\|^2 > G_g^{\min}(\ddot{z}^*) \quad (33)$$

which contradicts our assumption that  $G_g$  is a minimum when  $\ddot{z}_s$  is given by Eq. (31). Hence,  $G_g$  attains its minimum at  $\ddot{z}_s^*$  given in Eq. (28). Moreover we have shown that for any other  $\ddot{z}$  that differs from the  $\ddot{z}^*$  that is given in Eq. (29) the Gaussian,  $G_g$ , will increase. We have thus proved that  $\ddot{q} = \ddot{z}^*$  in Eq. (15) gives the unique global minimum of  $G_g$ . As noted before, global minimum principles are rare in classical mechanics.

Note that when  $P^+P = I$  the set  $Z_W$  of  $n$ -vectors to be used for minimizing  $G_g$  shrinks to just a singleton, thereby making the minimization of  $G_g$  trivial; then too the  $n$ -vector  $\ddot{q}$  and  $G_g^{\min}$  are correctly given by Eqs. (29) and (30), respectively (see Remark 6). ■

We now provide a second proof of Result 1, which is much shorter, relies on the somewhat abstract algebraic properties of the Moore–Penrose (MP) inverse of a matrix, and involves mainly algebraic manipulations.

*Proof 2.* Let us define the  $n$ -vector

$$s := M^{1/2}\ddot{z} - M^{-1/2}(Q + C) \quad (34)$$

so that the Gaussian can be rewritten as (see Eq. (25))

$$G_g(s) = \|s\|^2 \quad (35)$$

and the  $m$ -vector  $p$  as

$$p := W^{1/2}[b - AM^{-1}(Q + C)] \quad (36)$$

Then noting that  $P = W^{1/2}B = W^{1/2}AM^{-1/2}$

$$\begin{aligned} Ps - p &= W^{1/2}AM^{-1/2}[M^{1/2}\ddot{z} - M^{-1/2}(Q + C)] \\ &\quad - W^{1/2}[b - AM^{-1}(Q + C)] \\ &= W^{1/2}(A\ddot{z} - b) \end{aligned}$$

so that

$$\|Ps - p\|^2 = (A\ddot{z} - b)^T W(A\ddot{z} - b) = \|A\ddot{z} - b\|_W^2 \quad (37)$$

From Eq. (35) we see that minimizing the Gaussian is then tantamount to minimizing the norm of the  $n$ -vector  $s$ , and minimizing  $\|A\ddot{z} - b\|_W^2$  is tantamount to minimizing  $\|Ps - p\|^2$ . Result 1 then requires us to find the  $n$ -vector  $s$  that minimizes  $G_g$  from among all vectors  $s$  that minimize  $\|Ps - p\|$ . This can be restated as the problem of finding the minimum-norm solution  $s^*$  that solves the possibly inconsistent equation  $Ps \simeq p$ .

It is known that the minimum-norm least squares solution to the equation  $Ps \simeq p$  is unique and is given simply by [18]

$$s^* = P^+p \quad (38)$$

Using Eqs. (34) and (36), Eq. (38) can then be rewritten as

$$M^{1/2}\ddot{z}^* - M^{-1/2}(Q + C) = P^+W^{1/2}[b - AM^{-1}(Q + C)] \quad (39)$$

from which it follows that

$$\ddot{q} := \ddot{z}^* = M^{-1}Q + M^{-1/2}P^+(\tilde{b} - PM^{-1/2}Q) + M^{-1/2}(I - P^+P)M^{-1/2}C \quad (40)$$

which is Eq. (29) (and Eq. (15)). ■

As stated before in Remark 1, by setting  $W = I_m$ , so that  $P = B = AM^{-1/2}$ , we obtain the equation of motion given in Eq. (14), thus validating the First Form of the General Gauss Principle of Least Constraint as a global minimum principle.

*Remark 8.* It should be pointed out that the  $m$ -by- $n$  matrix  $A(P)$  need not have full row rank and the set of constraints need not be functionally independent. This is important since in many complex mechanical systems there can be many nonholonomic constraints, which are themselves differential equations. Since the use of a multiplier can change the appearance of such equations it would be tedious, at best, to discern which constraints in the set are the independent constraints. That there is no need to do this, is an advantage in the development of the equations of motion of such constrained systems.

We investigate next the manner in which the acceleration  $\ddot{q}$  depends on the weighting matrix  $W$  in Result 1. Specifically, we consider the influence of using a positive definite weighting matrix  $W > 0$ , with  $W \neq I_m$ . To assess the significance of the weighting matrix  $W$ , we first prove the following two lemmas, the first of which does not appear to be widely known in the literature on generalized inverses. ■

*LEMMA 1.* If  $P = UB$  where  $B$  is any  $m$ -by- $n$  matrix and  $U$  is a nonsingular  $m$ -by- $m$  matrix, then

$$P^+P = B^+B$$

*Proof.*  $P = UB \Rightarrow PB^+B = UBB^+B = UB = P$ . Hence,  $PB^+B = P$ , and  $PB^+B = P \Rightarrow P^+PB^+B = P^+P$ . Therefore,  $P^+P = (P^+P)^T = (P^+PB^+B)^T = (B^+B)^T(P^+P)^T = B^+BP^+P$ . Hence, we find that

$$P^+P = B^+(BP^+P) \quad (41)$$

Also,  $B = U^{-1}P \Rightarrow BP^+P = U^{-1}PP^+P = U^{-1}P = B$ . Hence,  $BP^+P = B$ . Using this in Eq. (41), we get  $P^+P = B^+B$ , which is the desired result. ■

*LEMMA 2.* For an  $m$ -by- $m$  matrix  $W$  and an  $m$ -by- $n$  matrix  $B$

$$P^+ = (W^{1/2}B)^+ = B^+(W^{1/2})^+ \quad (42)$$

if and only if

$$[I_m - (W^{1/2})^+W^{1/2}]BB^TW^{1/2} = 0 \text{ and } [I_m - BB^+]WB = 0 \quad (43)$$

*Proof.* See Ref. [19], and Eqs. (5) and (6) therein. ■

*Remark 7.* When  $W > 0$ , then  $W^{1/2}$  is nonsingular. Hence, Lemma 1 yields the relation (recall  $P = W^{1/2}B = W^{1/2}AM^{-1/2}$ )

$$P^+P = B^+B \quad (44)$$

Furthermore, because  $W^{1/2}$  is nonsingular,  $(W^{1/2})^+ = W^{-1/2}$  so that the first equality in Eq. (43) is always satisfied.

Also, when the  $m$ -by- $n$  constraint matrix  $A$  has full row rank  $m$ , i.e., the constraints are independent, then  $B = AM^{-1/2}$  has full row rank, and  $BB^+ = BB^T(BB^T)^+ = BB^T(BB^T)^{-1} = I$  so that the second equality in Eq. (43) is then also satisfied, and by Lemma 2, we have

$$P^+ = (W^{1/2}B)^+ = B^+(W^{1/2})^+ = B^+W^{-1/2} \quad (45)$$

Since  $\tilde{b} = W^{1/2}b$ , we get

$$P^+\tilde{b} = B^+W^{-1/2}(W^{1/2}b) = B^+b \quad (46)$$

Hence, when  $W > 0$  and the matrix  $B(A)$  has full row rank  $m$ , using Eqs. (44) and (46), Eq. (40) becomes

$$\begin{aligned}\ddot{q} &= M^{-1}Q + M^{-1/2}P^+\tilde{b} - P^+PM^{-1/2}Q + M^{-1/2}(I - B^+B)M^{-1/2}C \\ &= M^{-1}Q + M^{-1/2}B^+b - B^+BM^{-1/2}Q + M^{-1/2}(I - B^+B)M^{-1/2}C \\ &= M^{-1}Q + M^{-1/2}B^+(b - Aa) + M^{-1/2}(I - B^+B)M^{-1/2}C\end{aligned}\quad (47)$$

where  $a = M^{-1}Q$ . The last expression on the right is just the fundamental equation of constrained motion (FECM) given in Eq. (14). This leads to our next result.

**Result 2.** For a constraint matrix  $A$  that has full row rank and for every(any) weighting matrix  $W > 0$ , at each instant of time  $t$  the acceleration  $n$ -vector  $\ddot{q}$  of a constrained mechanical system is obtained by finding

$\ddot{q} = \min(G_g) := \min \|\ddot{z} - M^{-1}Q - M^{-1}C\|_M^2$  from among all  $n$ -vectors  $\ddot{z}$  for which  $\|(Az - b)\|_W^2$  is a minimum. The acceleration of the constrained system is explicitly given by

$$\ddot{q} = M^{-1}Q + M^{-1/2}B^+(b - BM^{-1/2}Q) + M^{-1/2}(I - B^+B)M^{-1/2}C$$

where  $B = AM^{-1/2}$ . ■

Somewhat remarkably, we have shown that the correct form of the equation of motion is obtained by minimizing the Gaussian no matter what positive definite weighting matrix  $W \neq I_m$  one uses in finding the least squares solution of the inconsistent equations  $A\ddot{q} \simeq b$ , when the matrix  $A$  has full row rank, so that the constraints are independent. This leads us to the second form of the General Gauss Principle of Least Constraint. Nature appears to discard the information on the positive definite weighting matrix  $W$ ; it picks the same acceleration vector  $\ddot{q}$  at each instant of time as though  $W$  were simply  $I_m$ .

**Second Form of the General Gauss Principle of Least Constraint.** *Nature selects the acceleration  $n$ -vector  $\ddot{q}$  of the constrained system at each time instant  $t$  by minimizing the Gaussian  $G_g$  among all acceleration  $n$ -vectors that satisfy the independent constraints  $A\ddot{q} \simeq b$  acting on the system in the weighted least squares sense, irrespective of the positive definite weighting matrix  $W$  used in getting the least squares solutions.*

### 3 Virtual Displacements

Section 2 shows that the same form of the equation of motion is obtained for a constrained system without requiring that the constraint equation (Eq. (4)) be consistent, with Nature appearing to consider all possible acceleration  $n$ -vectors,  $\ddot{z}$ , that satisfy the equation  $A\ddot{z} \simeq b$  in the least squares sense among the set of vectors that are eligible for minimizing the Gaussian  $G_g$ . This expansion in Gauss's Principle that permits inconsistent constraints in the description of constrained mechanical systems must provide *pari passu* an expansion of d'Alembert's principle too. This is because the two principles yield the same expression for the acceleration  $n$ -vector of any constrained system. Any change (expansion) in scope in either one of these principles must bring about a commensurate change in the scope of applicability of the other.

We therefore explore in this section the question: how does the increased scope of the General Gauss's Principle of Least Constraint obtained in the previous section alter our understanding of d'Alembert's principle?

The notion of a virtual displacement is at the core of d'Alembert's principle, and it is one of the notions that all of Lagrangian mechanics rests on. Though this notion was first used by d'Alembert [9], it was carefully refined in a general manner by Lagrange [1]. Later on, it was defined for constraints in Pfaffian form in which the generalized velocities appear only linearly. Even today, more than two centuries after Lagrange, many texts still consider only Pfaffian

constraints [5]. The inclusion in analytical dynamics of more general constraints in which the generalized velocities can appear nonlinearly, and the corresponding notion of virtual displacements for such constraints (as given in Eq. (5)) appears to be a recent development [10]. The expansion of d'Alembert's principle to take account of forces of constraint that do work under virtual displacements [13,14] led to a *pari passu* expansion/extension in Gauss's Principle of Least Constraint in which the Gaussian was modified from  $G$  to  $G_g$  [12]. The new understanding obtained in this paper of Gauss's Principle of Least Constraint points to the need for a further conceptual change in the notion of a virtual displacement, a quintessential element of Lagrangian mechanics.

Currently, we consider a virtual displacement  $n$ -vector to be any nonzero vector which is the difference between any two distinct  $n$ -vectors  $\ddot{q}$  (called possible accelerations) that satisfy the consistent constraint equation  $A(q, \dot{q}, t)\ddot{q}(t) = b(q, \dot{q}, t)$  [10]. Thus, if  $\ddot{z}_1$  and  $\ddot{z}_2$  are any two distinct possible accelerations such that  $A(q, \dot{q}, t)\ddot{z}_i(t) = b(q, \dot{q}, t)$ ,  $i = 1, 2$ , then

$$\tilde{v}(t) = \ddot{z}_1 - \ddot{z}_2 \quad (48)$$

qualifies as a virtual displacement at time  $t$ . A virtual displacement therefore satisfies the relation  $A\tilde{v} = 0$ , that is, it is any  $n$ -vector (not necessarily infinitesimal) that lies in the null space of the matrix  $A$  at time  $t$ . We now extend the notion of a virtual displacement on the basis of the results obtained in this paper as follows.

Consider a mechanical system subjected to the constraint equation set

$$A(q, \dot{q}, t)\ddot{q}(t) \simeq b(q, \dot{q}, t) \quad (49)$$

that may or may not be consistent. The constraint matrix  $A$  is an  $m$ -by- $n$  matrix of rank  $r$  and arises due to the presence of a total of  $m$  holonomic and/or nonholonomic constraints.

Let  $\ddot{z}_1$  and  $\ddot{z}_2$  be any two distinct possible acceleration  $n$ -vectors that provide least squares solutions (with the weighting matrix  $I_m$ ) to  $A(q, \dot{q}, t)\ddot{z}_i(t) \simeq b(q, \dot{q}, t)$ ,  $i = 1, 2$ . Then, the notion of a virtual displacement,  $v(t)$ , can be expanded to mean any nonzero  $n$ -vector which is the difference between these two possible least squares solutions, that is

$$v(t) = \ddot{z}_1(t) - \ddot{z}_2(t) \quad (50)$$

We note that the least squares solution  $n$ -vectors,  $\ddot{z}_i$ ,  $i = 1, 2$ , solve the normal equation

$$A^T A \ddot{z}_i = A^T b, \quad i = 1, 2$$

As shown in Eqs. (19)–(21), (any) two distinct, explicit solutions of this equation are given by

$$\ddot{z}_i = A^+ b + (I - A^+ A)u_i, \quad i = 1, 2,$$

where  $u_i$ ,  $i = 1, 2$ , are any two distinct arbitrary  $n$ -vectors. The (nonzero)  $n$ -vector

$$v(t) = \ddot{z}_1(t) - \ddot{z}_2(t) = (I - A^+ A)(u_1 - u_2) := (I - A^+ A)w(t) \quad (51)$$

which is the difference between any two such possible (least squares solution)  $n$ -vectors gives a virtual displacement "compatible" with the inconsistent constraints given in Eq. (49). The vector  $w = u_1(t) - u_2(t) \neq 0$  is an arbitrary  $n$ -vector.

The notion of a virtual displacement that is in use today is provided by the  $n$ -vectors  $\tilde{v}$  described in Eq. (48). Equation (50) expands the scope of this notion to include constraints that may be inconsistently prescribed. This new notion of a virtual displacement continues to satisfy the (old) condition that it lies in the null space of  $A$ , since by Eq. (51)

$$Av = A(I - A^+ A)w = 0 \quad (52)$$

The main conceptual change in our understanding of a virtual displacement here is that the constraint equations need *not* be consistent, and a (nonzero) virtual displacement is not necessarily the

difference between any two exact, distinct solutions of the equation  $A\dot{q} = b$ , which is required to be necessarily consistent, but the difference between any two distinct least squares solutions of the possibly inconsistent equation  $A\dot{q} \simeq b$ .

## 4 Conclusions

This paper deals with a topic in analytical dynamics which is at its very foundations and which affects our understanding of both solid and fluid mechanics. It gives a general form of Gauss's principle of Least Constraint.

The principle Gauss gave in 1829 states that at each instant of time, Nature picks the acceleration of a constrained system from all possible accelerations that satisfy the constraints imposed on it by minimizing the Gaussian, which is the two-norm of the difference between the acceleration (column vector) of the constrained system and the acceleration (column vector) of the unconstrained system weighted by the positive definite mass matrix of the unconstrained system. This principle is applicable only to systems in which the forces of constraint do *no* work under virtual displacements.

The investigation reported in this paper began with the exploration of the following question: For a constrained mechanical system, given: (1) that holonomic and/or nonholonomic constraints act continuously on it throughout the duration of its motion, and (2) that these constraints are almost never exactly satisfied as prescribed in the modeling of the dynamical system throughout its motion, why is it that the equation of motion given by Gauss's principle for the dynamical system provides a good prediction of its dynamical behavior, despite the possibly incorrect description of the constraints at each instant of time  $t$  throughout the motion of the system. The unexpected answer to this question is that the equation of motion provided by Gauss's Principle remains unaltered when the constraints are satisfied in the least squares sense. Thus, Nature does not seem to require that the constraints be satisfied *exactly* at each instant of time as prescribed, and allows for their miss-prescription. In fact, it even allows the constraints to be possibly inconsistent, thereby showing the robustness with which it orchestrates constrained motion.

This paper thus expands Gauss's principle and expresses it in two closely related forms. Both of these forms go beyond the principle originally proposed by Gauss. They include forces of constraint that *can* do positive, zero, or negative work under virtual displacements, so that energy could be fed to (or extracted from) a mechanical system through the presence of constraints, as may happen when the constraint "surfaces" have friction. More importantly, as mentioned earlier, these general forms of Gauss's Principle of Least Constraint are applicable to systems in which the constraints imposed on them may not be exactly satisfied at each instant of time. It is shown that Gauss's principle extends to constraints that could even be inconsistent at one or more instants of time as could happen if these constraints are experimentally obtained and subject to measurement errors, and/or subject to computational errors (and/or round-off errors) caused by the use of numerical procedures in simulating the behavior of constrained physical systems that may be either engineered or found in nature.

The first form of the General Gauss Principle of Least Constraint obtained in this paper states that at each instant of time, Nature picks the acceleration of a constrained system by minimizing the Gaussian of the constrained system from among the set of all *possible* accelerations, which satisfy the (possibly inconsistent) constraints imposed on it *in the least squares sense*, using the identity matrix as the weighting matrix when finding these *possible* accelerations. The constraints that the modeler may use could be functionally dependent, as often happens when there are numerous holonomic and/or nonholonomic constraints that a complex mechanical system can be subjected to.

The second form of the General Gauss Principle of Least Constraint asserts that at each instant of time, Nature picks the

acceleration of a constrained system by minimizing the Gaussian from among all *possible* accelerations, which satisfy the (possibly) inconsistent *independent* constraints imposed on it *in the least squares sense*, with no regard to the choice of the positive definite weighting matrix used in obtaining the *possible* accelerations; the resulting acceleration of the constrained system is independent of this choice.

The new principle obtained here points to a rethinking of the notion of a virtual displacement—a concept that is at the heart of our understanding of the statics and dynamics of all mechanical systems [1]. Currently, a virtual displacement is defined as the difference between two distinct *possible* accelerations. These possible accelerations have so far been taken to be those that satisfy the constraint equations exactly at each instant of time [10]. This notion is shown here to be broadened to allow the set of *possible* accelerations to also include those that satisfy the constraint equations in the least squares sense, with a virtual displacement defined as the difference between any two such distinct possible accelerations.

Lastly, there appears to be a peculiar mathematical aesthetic with which Nature appears to contrive the motion of any constrained mechanical system: it minimizes a quadratic form, the Gaussian—by solving a least squares problem—using as eligible candidates for this minimization only those that minimize another quadratic form—by solving yet another least squares problem, thereby obtaining the global minimum of the generalized Gaussian, which is just a scalar.

## Conflict of Interest

There are no conflicts of interest.

## Data Availability Statement

No data, models, or code were generated or used for this paper.

## References

- [1] Lagrange, J. L., 1787, *Mechanique Analytique*, Paris, Mme Ve Courcier, Paris.
- [2] Clebsch, A., 2009, *Jacobi's Lectures on Mechanics, Texts in Mathematics No. 51*, Hindustan Book Agency, New Delhi, India.
- [3] Pars, L. A., 1972, *A Treatise on Analytical Dynamics*, Oxbow Press, Woodridge, CT.
- [4] Whittaker, E. T., 1988, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, New York.
- [5] Goldstein, H., Poole, C., and Safko, J., 2011, *Classical Mechanics*, Addison Wesley, San Francisco, CA.
- [6] Dirac, P. A. M., 1950, "Generalized Hamiltonian Dynamics," *Can. J. Math.*, **2**, pp. 129–148.
- [7] Dirac, P. A. M., 1964, *Lectures on Quantum Mechanics*, Academic Press, New York.
- [8] Gauss, C. F., 1829, "Über ein Neues Allgemeines Grundgesetz der Mechanik," *J. Reine Angew. Math.*, **4**, pp. 232–235.
- [9] d'Alembert, J. L., 1743, *Traite de Dynamique*, J. B. Coignard, Paris.
- [10] Udawadia, F. E., and Kalaba, R. E., 1997, "Equations of Motion for Constrained Mechanical Systems and the Extended D'Alembert's Principle," *Quart. Appl. Math.*, **55**(2), pp. 321–331.
- [11] Udawadia, F. E., and Kalaba, R. E., 1992, "A New Perspective on Constrained Motion," *Proc. R. Soc. A*, **429**, pp. 407–410.
- [12] Kalaba, R. E., Natsuyama, H., and Udawadia, F. E., 2004, "An Extension of Gauss's Principle of Least Constraint," *Int. J. Gen. Sys.*, **33**(1), pp. 63–69.
- [13] Udawadia, F. E., and Kalaba, R. E., 2002, "On the Foundations of Analytical Dynamics," *Int. J. Nonlinear Mech.*, **37**(6), pp. 1079–1090.
- [14] Udawadia, F. E., and Kalaba, R. E., 2002, "What Is the General Form of the Explicit Equations of Motion for Constrained Mechanical Systems?," *ASME J. Appl. Mech.*, **69**(3), pp. 335–339.
- [15] Udawadia, F. E., and Kalaba, R. E., 1996, *Analytical Dynamics: A New Approach*, Cambridge University Press, New York.
- [16] Moore, E. H., 1929, "On the Reciprocal of the General Algebraic Matrix, Abstract," *Bull. Am. Math. Soc.*, **26**, pp. 394–395.
- [17] Penrose, R., 1955, "A Generalized Inverse of Matrices," *Math. Proc. Cambridge Philos. Soc.*, **51**(3), pp. 406–413.
- [18] Rao, C. R., and Mitra, S. K., 1971, *Generalized Inverse of Matrices and Its Applications*, John Wiley, New York.
- [19] Greville, T. N. E., 1966, "Note on the Generalized Inverse of a Matrix Product," *SIAM Rev.*, **8**(4), p. 518521.