JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 127, No. 3, pp. 475–484, December 2005 (© 2005)
DOI: 10.1007/s10957-005-7496-7

Reflections on the Gauss Principle of Least Constraint

Y. Y. Fan,[1] R. E. Kalaba,[2] H. H. Natsuyama,[3] and F. E. Udwadia[4]

Abstract. The Gauss principle of least constraint is derived from a new point of view. Then, an extended principle of least constraint is derived to cover the case of nonideal constraints. Finally, a version of the principle for general underdetermined systems is adumbrated. Throughout, the notion of generalized inverses of matices plays a prominent role.

Key Words. Gauss principle of least constraint, nonideal nonholonomic constraints, generalized inverses of matrices.

1. Introduction

In his epochal paper of 1829, Gauss (Ref. 1) began by remarking that the D'Alembert principle reduced all of dynamics to statics and that the principle of virtual works reduced all of statics to a mathematical problem. Thus, there could be no new principle of mechanics that is not included already in those two. Yet, he observed that every new principle is not without merit, especially if it can shed new light on mechanical processes and perhaps render the solution of certain problems simpler to obtain. He then went on to state his own new principle, the principle of least constraint reducing all of mechanics, dynamics, and statics, to a single principle. Since its enunciation, it has been a cornerstone of analytical dynamics. His own derivation, which relies on the aforementioned two

[1]Assistant Professor, Civil and Environmental Engineering Department, School of Engineering, University of California, Davis, California.
[2]Professor, Biomedical Engineering and Economics Departments, University of Southern California, Los Angeles, California.
[3]Professor Emeritus, System Engineering Department, California State University, Fullerton, California.
[4]Professor, Aerospace and Mechanical Engineering, Civil Engineering, Mathematics, and Information and Operations Management Departments, University of Southern California, Los Angeles, California.

principles and the law of cosines, has often been presented but never surpassed for brevity and clarity.

This paper, besides being a tribute to Gauss in the 175th anniversary of the principle bearing his name, will present a rederivation of the Gauss principle and then an extension of his principle to cases in which the standard principle of virtual works is not applicable. These cases include those in which sliding friction (for example) is significant and cannot be neglected.

2. Gauss Principle of Least Constraint

Consider a system of n particles. The ith particle has mass m_i and its position vector is x_i in an inertial Cartesian frame of reference. Its velocity is $\dot{x}_i$ and its acceleration is $\ddot{x}_i$. It is subjected to an impressed force f_i; if no constraints are present, its free motion acceleration is $a_i = f_i/m_i$.

The system mass matrix M is of dimension $3n$ by $3n$ and is a nonsingular diagonal matrix with the masses down the main diagonal in sets of three and zeroes elsewhere. The system displacement vector x_i, of dimension $3n$ by 1, has the vectors x_i stacked in the usual fashion. Similarly, we construct the system velocity and acceleration vectors $\dot{x}$ and $\ddot{x}$. Likewise for the system free acceleration vector a.

It is assumed that the system is subjected to m consistent equality constraint conditions of the form

$$f_i(x, \dot{x}, t) = 0, \quad i = 1, 2, \dots, m.$$

Note that this constraint form is general and includes both holonomic or nonholonomic constraints. Through use of the chain rule of differentiation, these constraints assume the form

$$A\ddot{x} = b,$$

where A is a matrix of dimension m by $3n$ and b is a vector of dimension m by 1. Both may depend upon $x,\ \dot{x},\ t$. Thus, the constraint on the acceleration is linear. Together with the initial conditions $x(t_0)$ and $\dot{x}(t_0)$, this constraint is equivalent to the original form $f_i = 0,\ i = 1, 2, \dots, m$.

For some years, it has been known (Ref. 2) that the actual system acceleration vector is given by the explicit formula

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

where $(AM^{-1/2})^{+}$ denotes the usual pseudoinverse of the matrix $AM^{-1/2}$. At the present moment, we need not enter into the details of this explicit formula for the system acceleration, because later, in Eq. (17), we shall produce a more comprehensive formula from a more general point of view. Our aim now is to derive the Gauss principle of least constraint from formula (1)

It is convenient to make the substitutions

$$D = AM^{-1/2}, \tag{2}$$

$$e = b - Aa, \tag{3}$$

$$\ddot{x}_s = M^{1/2}\ddot{x}, \tag{4}$$

$$a_s = M^{1/2}a. \tag{5}$$

Then, Eq. (1) becomes

$$\ddot{x}_s - a_s = D^+e. \tag{6}$$

But D^+e is the solution of the variational problem

$$\min_y \quad y^T y, \tag{7}$$

$$\text{s.t.} \quad Dy = e. \tag{8}$$

Thus, we have

$$\min_{\ddot{x}_s} \quad (\ddot{x}_s - a_s)^T(\ddot{x}_s - a_s), \tag{9}$$

$$\text{s.t.} \quad (AM^{-1/2})(\ddot{x}_s - a_s) = b - Aa. \tag{10}$$

Reverting now to the original variables, we see that the variational problem becomes

$$\min_{\ddot{x}} \quad (\ddot{x} - a)^T M(\ddot{x} - a), \tag{11}$$

$$\text{s.t.} \quad A\ddot{x} = b. \tag{12}$$

Equations (11) and (12) constitute a form of the Gauss principle of least constraint.

3. General Equation of Motion

Next, we wish to view matters from a more general point of view. In fact, we wish to determine all the equations of motion that are compatible with the constraint $A\ddot{x} = b$ with no physical assumptions being made at all. Experience has shown the importance of the matrix $AM^{-1/2}$, so we shall rewrite the above equation in the form

$$AM^{-1/2}(M^{-1/2}\ddot{x}) = b. \tag{13}$$

The theory of generalized inverses shows us that the general solution of this constraint set of linear algebraic equations for $\ddot{x}$ is

$$M^{1/2}\ddot{x} = (AM^{-1/2})^{+}b + [I - (AM^{-1/2})^{+}(AM^{-1/2})]z, \tag{14}$$

where z is an arbitrary vector of dimension $3n$ by 1. Furthermore, it is most revealing to write the arbitrary vector z in the form

$$z = M^{1/2}a + M^{-1/2}C, \tag{15}$$

where a is the free motion acceleration vector and C is an arbitrary vector, both being of dimension $3n$ by 1. It follows that

$$\begin{aligned}\ddot{x} = a &+ M^{-1/2}(AM^{-1/2})^{+}(b - Aa)\\ &+ M^{-1/2}[I - (AM^{-1/2})^{+}(AM^{-1/2})]M^{-1/2}C,\end{aligned} \tag{16}$$

or

$$\begin{aligned}M\ddot{x} = Ma &+ M^{1/2}(AM^{-1/2})^{+}(b - Aa)\\ &+ M^{1/2}[I - (AM^{-1/2})^{+}(AM^{-1/2})]M^{-1/2}C.\end{aligned} \tag{17}$$

Thus, Equation (17) is the most general possible equation of motion that is compatible with the constraint relation $A\ddot{x} = b$. No physical assumptions were employed in the derivation of this equation.

We notice that Eq. (1) is a special case of Eq. (17). It is convenient to rewrite Eq. (17) in the form

$$M\ddot{x} = F^{N} + F^{L} + F^{C}, \tag{18}$$

where

$$F^{N} = Ma, \tag{19}$$

$$F^{L} = M^{1/2}(AM^{-1/2})^{+}(b - Aa), \tag{20}$$

$$F^{C} = M^{1/2}\left[I - (AM^{-1/2})^{+}AM^{-1/2}\right]M^{-1/2}C. \tag{21}$$

The notation bears the names of Newton, Lagrange, and Coulomb. Later, we shall elaborate on this. Note that only two essential mathematical ideas have entered the analysis: the chain rule of differentiation and generalized inverses of matrices. Modern computing environments, such as Matlab, have built-in commands for calculating the generalized inverse of a matrix, so it makes the approach highly suitable for numerical studies.

4. Extended Gauss Principle

Once again, we emphasize that Eq. (17) is the most general possible equation of motion that is compatible with the constraint condition $A\ddot{x} = b$, of course assuming that the matrix M is nonsingular. Now, we wish to obtain an extended Gauss principle that leads to this equation of motion. Using the same notation as earlier, plus the additional notation

$$c_s = M^{-1/2}C,$$

we find that Eq. (17) becomes

$$\ddot{x} = a_s + D^+ e + [I - D^+ D]c_s. \tag{22}$$

We rearrange this to yield

$$\ddot{x}_s - a_s - c_s = D^+(e - Dc_s). \tag{23}$$

Recalling Eqs. (7) and (8), we find that $\ddot{x}_s$ solves the problem

$$\min_{\ddot{x}_s} \quad [\ddot{x}_s - (a_s + c_s)]^T [\ddot{x}_s - (a_s + c_s)], \tag{24}$$

$$\text{s.t} \quad AM^{-1/2}[\ddot{x}_s - (a_s + c_s)] = b - Aa - AM^{-1}C. \tag{25}$$

In terms of the original variables, this problem is

$$\min_{\ddot{x}} \quad [\ddot{x} - a - M^{-1}C]^T M[\ddot{x} - a - M^{-1}C], \tag{26}$$

$$\text{s.t.} \quad A\ddot{x} = b. \tag{27}$$

Equations (26) and (27) constitute an extended Gauss principle of least constraint. This principle covers the cases of nonideal and of course nonholonomic constraints.

5. Physical Significance of the Terms in Eq. (17)

Let us now comment further on the terms in the general equation of motion, Eq. (17) or Eq. (18). The special vector a is chosen to be the acceleration that the system would have if there were no constraints. This is why in Eq. (19) we have denoted Ma as being F^N, the Newtonian impressed force vector. Notice that Eq. (17) involves only the concepts of mass, distance, and time. It is in the physical interpretation of F^N as Ma that we introduce a further physical concept, that of force.

Now, we consider Eq. (20). We introduce the notion of a virtual displacement vector v to be any vector v that satisfies the homogeneous equation $Av = 0$. Thus, if $\ddot{x}$ is the actual acceleration at time t of the system, then $\ddot{x} + v$ is a possible acceleration, since

$$A(\ddot{x} + v) = A\ddot{x} + Av = A\ddot{x} = b. \tag{28}$$

The equation $Av = 0$ is homogeneous, so that the vector v can be multiplied by any dimensional constant and it remains a solution. Thus, v is not necessarily a displacement.

Finally, we introduce the notion of the work done by the force F^L in a virtual displacement ν to be $\nu^T F^L$. Since ν satisfies the equation

$$(AM^{-1/2})M^{1/2}v = 0, \tag{29}$$

we have

$$M^{1/2}v = [I - (AM^{-1/2})^+ AM^{-1/2}]q, \tag{30}$$

where q is an arbitrary vector of dimension $3n$. Thus,

$$v^T = q^T [I - (AM^{-1/2})^+ AM^{-1/2}]M^{-1/2}. \tag{31}$$

It follows that

$$\begin{aligned} v^T F^L &= q^T [I - (AM^{-1/2})^+ AM^{-1/2}]M^{-1/2} M^{1/2} (AM^{-1/2})^+ (b - Aa) \\ &= 0. \end{aligned} \tag{32}$$

Thus, F^L is a constraint force that does no work on the system in a virtual displacement ν. We see that the other constraint force F^C would be the null vector if $C = 0$, so F^L must be the constraint force that maintains the constraints while doing no work on the system in any virtual displacement. It is the force that Lagrange denoted $A^T\lambda$, where λ is an m-dimensional vector of Lagrange multipliers. It should be remembered that the columns of A^+ and A^T span the same space.

There remains the third term on the right in Eq. (17),

$$F^C = M^{1/2}[I - (AM^{-1/2})^+ AM^{-1/2}]M^{-1/2}C. \tag{33}$$

How much work does it do in a virtual displacement? We have

$$\begin{aligned} v^T F^C &= q^T [I - (AM^{-1/2})^+ AM^{-1/2}]M^{-1/2} \\ &\quad \times M^{1/2}[I - (AM^{-1/2})^+ (AM^{-1/2})]M^{-1/2}C \\ &= q^T [I - (AM^{-1/2})^+ AM^{-1/2}]M^{-1/2}C \\ &= v^T C. \end{aligned} \tag{34}$$

From the last equation, we see that a specification of the vector C is a specification of the work done by the constraint force F^C in a virtual displacement v. Since the total force of constraint is $F^L + F^C$ and since the force F^L does no work in a virtual displacement, we have

$$v^T(F^L + F^C) = v^T C. \tag{35}$$

This equation is a generalization of the classical equation for the principle of virtual work. That principle, a cornerstone of analytical mechanics since it was enunciated by Lagrange, states that the constraint forces do no work in any virtual displacement. This is referred to as an ideal constraint. A constraint force such as sliding friction is ruled out of consideration, as Pars and Goldstein have stated in their textbooks. The new principle in Eq. (35) allows the constraint forces to do work in virtual displacements.

6. Toward a Theory of Underdetermined Systems

As general as the previous discussion has been, it is possible to go even beyond the bounds of mechanics. In fact, we can even begin to see the outlines of a theory of underdetermined systems. Now, let $x = x(t)$ be the state vector of a system S and let the dimension of x be n. Next, suppose that the theory suggests that certain relations must be satisfied. We may also desire certain relations to be fulfilled. Thus, at least in the formative states of a theory, we are led to m relations of the form

$$f_i(x, \dot{x}, t) = 0, \quad i = 1, 2, \ldots, m. \tag{36}$$

Notice that the number of relations (constraints) may be less than, equal to, or greater than the dimension of x. Through use of the chain rule of differentiation, we arrive at the equation

$$A\ddot{x} = b, \tag{37}$$

where the matrix A is of dimension m by n and the vector b is of dimension m by 1.

Again, the theory of generalized inverses of matrices lets us write the general solution of this consistent system of linear algebraic equations in the form

$$\ddot{x} = A^+ b + (I - A^+ A)z, \tag{38}$$

where A^+ is the generalized inverse of A and z is an arbitrary n-dimensional vector. We may write the arbitrary vector z in the form $z = s + c$, where s is a special vector that is up to the modeler choice and c is an arbitrary vector. Then, Eq. (38) becomes

$$\ddot{x} = A^+ b + (I - A^+ A)(s + c) \tag{39}$$

or

$$\ddot{x} = s + A^+(b - As) + (I - A^+ A)c. \tag{40}$$

This equation may be written as

$$\ddot{x} = s + v_1 + v_2, \tag{41}$$

where

$$v_1 = A^+(b - As), \tag{42}$$

$$v_2 = (I - A^+ A)c. \tag{43}$$

The second-order ordinary differential equation above becomes the equation of motion of system S, based on the partial knowledge that we possess, as expressed in the m equations of constraint. The terms s, v_1, v_2 may now be investigated somewhat more.

If there were no constraints, then $A = b = 0$, so that $\ddot{x} = z$. This indicates that the arbitrary vector z in Equation (38) is what $\ddot{x}$ would be if there were no constraints present, as is intuitively obvious. When there are constraints, we could follow the hints provided by mechanics. We let v be a vector in the null space of A, for which $Av = 0$. The vector v must have the form

$$v = (I - A^+ A)w,$$

where w is an arbitrary vector. Since $I - A^+ A$ is symmetric and

$$A^+ A A^+ = A^+,$$

it follows that

$$v^T v_1 = 0. \tag{44}$$

Lastly, with regard to the vector ν_2, we see that

$$\begin{aligned} v^T \nu_2 &= w^T (I - A^+ A)(I - A^+ A)c \\ &= w^T (I - A^+ A)c, \end{aligned} \tag{45}$$

since the matrix $I - A^+A$ is also idempotent. But this implies that

$$\nu^T \nu_2 = \nu^T c. \tag{46}$$

In addtion, from their definitions, we see that the vectors ν_1 and ν_2 are orthogonal,

$$\nu_2^T \nu_1 = 0. \tag{47}$$

In summary, we may say that the underdetermined system S satisfies the second-order ordinary differential equation (33). The arbitrary vector z is what $\ddot{x}$ would be if there were no constraints. The matrix A and the vector b come from the constraints. This shows the nature of the vector ν_l. Finally, a specification of the vector c is a prescription of

$$\nu^T (\nu_l + \nu_2) = \nu^T \nu_2 \tag{48}$$

or

$$\nu^T (\ddot{x} - s) = \nu^T c, \tag{49}$$

where

$$z = s + c.$$

If c is in the null space of A, so that $Ac = 0$, then $\nu_2 = c$. If c is in the range space of A^T, then $\nu_2 = 0$.

7. Discussion

The entire analysis presented in this paper was started directly with the constraints on the system, which emphasizes again the importance of the constraints in a mechanical system. In fact, it is the constraints that make a set of point masses and rigid bodies into a system. Our previous analysis shows that the Lagrange principle of virtual works is a sagacious hypothesis that $F^c = 0$. This assumption is what makes analytical dynamics into pure mathematics and makes it possible to eliminate all thermodynamic considerations. This assumption works well because most practical mechanical systems, through design and the use of lubricants, do minimize the effects of constraint forces that do work on a system. However, to be more precise, nonideal constraint forces must also be considered.

This paper considers mechanical systems including constraints that are holonomic or nonholonomic and also ideal or nonideal. It also looks into more general underdetermined systems. This widens the applicability of the magnificent contribution of the Gauss principle to human thought.

References

1. Gauss, C. F., *Ueber ein neues algemeines Grundgesetz der Mechanik*, Zeitschrift fuer die reine und angewandte Mathematik, Vol. 4, pp. 232–235, 1829.
2. Udwadia, F. E, and Kalaba, R. E., *Analytical Dynamics*, Cambridge University Press, New York, NY, 1996.