Proceedings of DETC'03
ASME 2003 Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
Chicago, Illinois, USA, September 2-6, 2003

DETC2003/VIB-48318

A NOVEL APPROACH FOR THE DYNAMIC ANALYSIS AND SIMULATION OF CONSTRAINED MECHANICAL SYSTEMS

József Kövecses*
Space Technologies
Canadian Space Agency
6767 Route de l'Aéroport, Saint-Hubert,
Québec, Canada J3Y 8Y9
Email: Jozsef.Kovecses@space.gc.ca

Jean-Claude Piedbœuf
Space Technologies
Canadian Space Agency
6767 Route de l'Aéroport, Saint-Hubert,
Québec, Canada J3Y 8Y9
Email: Jean-Claude.Piedboeuf@space.gc.ca

ABSTRACT

In this paper we will outline a formulation for the dynamics of constrained systems. This formulation relies on the D'Alembert-Lagrange principle and the physically meaningful decomposition of the virtual displacements and the generalized velocities for the system where a redundant, non-minimum set of variables is used. The approach is valid for general constrained system with holonomic and/or nonholonomic constraints.

We will also discuss a potential application of this formulation in improving the accuracy and stability of the simulation of constrained systems. This application will be demonstrated by an example drawn from space robotics.

REPRESENTATION OF CONSTRAINED SYSTEMS

The D'Alembert-Lagrange principle (or the extended form of the principle of virtual work) forms the starting point for our considerations. This can be written in the following form

$$\delta\mathbf{q}^T\left(\mathbf{Q}_{inertial}-\mathbf{Q}\right)=\delta\mathbf{q}^T\left(\mathbf{M}\ddot{\mathbf{q}}+\mathbf{C}(\dot{\mathbf{q}},\mathbf{q},t)-\mathbf{Q}\right)=0, \qquad (1)$$

where $\mathbf{Q}_{inertial}$ represents an $n\times 1$ array of the generalized inertial forces, $\mathbf{Q}$ is the $n\times 1$ array of the generalized forces acting on the system, $\mathbf{M}$ is the $n\times n$ mass matrix of the system, $\mathbf{q}$ is the $n\times 1$ array representing the generalized coordinates of the system, $\mathbf{C}$ is the $n\times 1$ array of the nonlinear inertial effects (Coriolis and centrifugal effects), $\delta\mathbf{q}$ represents the $n\times 1$ array of generalized virtual displacements. This principle provides a unifying framework for all formulations used in the dynamics of multibody systems. If there are no constraints imposed on the system, thus $\dot{\mathbf{q}}$ represents a minimum set of generalized velocities, then the generalized virtual displacements are independent of each other and their coefficients must vanish individually in order to satisfy the above variational equation. This leads to the dynamic equations of the "unconstrained" system. It should be noted that even this unconstrained system may already implicitely include constraint conditions, which were embedded through the selection of a globally valid, minimum set of generalized coordinates $\mathbf{q}$ (e.g. the open-loop mechanism of a robotic system can be described by the joint coordinates, which can represent a set of globally valid generalized coordinates where the interbody constraints are embedded through these coordinates).

The decomposition of the generalized inertial forces $\mathbf{Q}_{inertial}=\mathbf{M}\ddot{\mathbf{q}}+\mathbf{C}(\dot{\mathbf{q}},\mathbf{q},t)$ does not always have to be done. For example, in inverse dynamics it can be more advantageous to establish and deal with $\mathbf{Q}_{inertial}$ directly without forming the mass matrix and calculating Coriolis and centrifugal effects separately. From the analytical mechanics point of view, the components of the generalized inertial forces can be represented as $\mathbf{Q}_{inertial_i}=\frac{d}{dt}\left(\frac{\partial T}{\partial\dot{q}_i}\right)-\frac{\partial T}{\partial q_i}=\frac{\partial S}{\partial\ddot{q}_i}$ $(i=1,\ldots,n)$, where T is the kinetic energy function of the system and S is the Gibbs-Appell

---

*Address all correspondence to this author. As of June 1, 2003, the author will be with the Department of Mechanical Engineering, McGill University, 817 Sherbrooke St. West, Montréal, Québec, Canada H3A 2K6 function. The generalized inertial forces can however be also formed by the direct application of the principle of vitual work or Jourdain's principle.

Let us consider the case when the system characterized by equation (1) is subjected to m constraints, which at the velocity level can be given in the form

$$\mathbf{A}(\mathbf{q},t)\dot{\mathbf{q}}+\mathbf{b}(\mathbf{q},t)=\mathbf{0}, \tag{2}$$

where $\mathbf{A}$ is the $m\times n$ constraint Jacobian matrix. These can be either holonomic or nonholonomic constraints (also scleronomic or rheonomic). If these are holonomic constraints, then (2) can be "integrated" and expressed in the form of position constraints as

$$\mathbf{\Phi}(\mathbf{q},t)=\mathbf{0}, \tag{3}$$

where $\mathbf{A}=\frac{\partial\mathbf{\Phi}}{\partial\mathbf{q}}$ and $\mathbf{b}=\frac{\partial\mathbf{\Phi}}{\partial t}$. Both holonomic and nonholonomic constraints can also be expressed at the acceleration level as

$$\mathbf{A}\ddot{\mathbf{q}}+\dot{\mathbf{A}}\dot{\mathbf{q}}+\dot{\mathbf{b}}=\mathbf{0}, \tag{4}$$

If the constraints are holonomic and $\mathbf{\Phi}$ is linear (or can be brought to a linear form through some transformation) in the generalized coordinates $\mathbf{q}$, then the system can again be transformed to an "unconstrained" form through the introduction of a new, globally valid, minimum set of generalized coordinates, the number of which is $n-m$. However, for many practical cases, the constraint equations are either holonomic but nonlinear, or nonholonomic thus form (3) does not exist at all. Examples can include the loop closure constraints of mechanisms (nonlinear, holonomic constraints), or the steering of a vehicle where the axes of the vehicle cannot move transversally (linear, nonholonomic constraints). For such situations, we need to consider the constraint conditions explicitely to carry out analysis, modeling and simulation for the system.

The virtual displacements must be admissible with the constraints, thus, the constraints also introduce interdependence for the generalized virtual displacements as

$$\mathbf{A}\delta\mathbf{q}=\mathbf{0}. \tag{5}$$

This equation is of key importance in the dynamics investigations. It can be derived from equations (2) and (3) based on the interpretation of the virtual displacements [9].

For this constrained mechanical system, the generalized forces can be decomposed to generalized applied forces ($\mathbf{Q}_{applied}$) and generalized constraint forces ($\mathbf{Q}_{constraint}$), $\mathbf{Q}=\mathbf{Q}_{applied}+\mathbf{Q}_{constraint}$. This decomposition is one of the fundamental characteristics of Lagrangian dynamics as opposed to the Newton-Euler description. A more precise term for the generalized applied force is generalized impressed forces. In this paper, we keep the applied force term beacuse it is more widespread in the literature. The generalized applied forces can include for example the actuator forces or torques, generalized gravitational forces, external loads and potential friction effects in the joints. The sole purpose of the generalized constraint forces is to maintain the constraints. Using this decomposition, equation (1) becomes

$$\delta\mathbf{q}^T\left(\mathbf{Q}_{inertial}-\mathbf{Q}_{applied}-\mathbf{Q}_{constraint}\right)=0. \tag{6}$$

The individual terms of this equation can be interpreted as the totality of the virtual work of generalized inertial forces ($\delta\mathbf{q}^T\mathbf{Q}_{inertial}$), generalized applied forces ($\delta\mathbf{q}^T\mathbf{Q}_{applied}$), and the generalized constraint forces ($\delta\mathbf{q}^T\mathbf{Q}_{constraint}$). A large class of practically important problems can be characterized by so-called ideal constraints, where the totality of the virtual work of the constraint forces vanishes,

$$\delta\mathbf{q}^T\mathbf{Q}_{constraint}=0. \tag{7}$$

For such systems,

$$\mathbf{Q}_{constraint}=\mathbf{A}^T\boldsymbol{\lambda} \tag{8}$$

where $\boldsymbol{\lambda}$ represents the array of Lagrangian multipliers [9].

Equations (2), (5), (6) and (7) give the basis for the various dynamics investigations of constrained systems, while equations (2), (3) and (4) are necessary for the kinematics analysis. In this form, the system is described by a redundant, non-minimum set of variables. The methods for the dynamic investigation of the constrained system can basically be grouped based on the virtual displacements and their interdependence. This determines the methods how constraint forces and kinematic constraints are handled and incorporated in the investigations. Two major groups of approaches can be distinguished:

Analysis using a new, independent set of virtual displacements, where the new set is formed based on equations (5) - (7). This group of approches include for example the independent projections [3], [10], and the complete constraint embedding at the velocity level via the introduction of independent generalized velocities. It should be noted that these approaches, as presented in the various publications, are often not derived from the extended form of the principle of virtual work directly. However, this principle (or any other differential variational principle, e.g., Jourdain's or Gauss' principles) provides a unifying root framework for the various approaches [6].

Analysis using the original, redundant set of virtual displacements. This group includes the most commonly applied method for ideally constrained systems, where the constraint forces are directly incorporated in the analysis through the Lagrangian multipliers, and the constraints at the acceleration level are also considered together with the dynamic equations. This leads to the well-known set of equations

$$\begin{bmatrix} \mathbf{M} & -\mathbf{A}^T \\ -\mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \ddot{\mathbf{q}} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{applied} - \mathbf{C} \\ \dot{\mathbf{A}}\dot{\mathbf{q}} + \dot{\mathbf{b}} \end{bmatrix}, \tag{9}$$

which can be directly solved for $\ddot{\mathbf{q}}$ and $\boldsymbol{\lambda}$ in the case of independent constraints. Another approach in this group is the penalty function based description where, the constraint forces are replaced by artificial applied forces through the introduction of penalty systems [3]. The penalty systems depends on the monitoring of the satisfaction of the constraints at the position, velocity and acceleration levels.

In the following we will describe another, novel approach with the use of the original redundant virtual displacements.

DECOMPOSITION OF THE GENERALIZED VIRTUAL DISPLACEMENTS

Using the constraints, it is possible to find a kinematically admissible set of virtual displacements using the original redundant set. Based on equation (5), we can split the array of generalized virtual displacements, $\delta\mathbf{q}$, into two parts

$$\delta\mathbf{q} = \delta\mathbf{q}_c + \delta\mathbf{q}_a \tag{10}$$

where $\delta\mathbf{q}_c$ is necessary to satisfy the constraints, and $\delta\mathbf{q}_a$ is an array admissible with the constraints, i.e., from the purely kinematic point of view it can take non-zero values without violating the constraints. The elements of these two arrays are basically linear combinations of the original virtual displacements, $\delta\mathbf{q}$, and these linear combinations can be determined based on the constraints (particularly the constraint Jacobian).

For determining the actual decomposition, it is useful to approach the problem using some geometric representations. We can think of $\delta\mathbf{q}$ as an array containing the coordinates of a vector interpreted in an n dimensional vector space. This vector space is often called the tangent space of the configuration manifold [1]. Each row of the constraint Jacobian matrix $\mathbf{A}$ also represents coordinates of a vector in this vector space. Equation (5) can be represented as a result of scalar products between the vectors in the rows of $\mathbf{A}$ and the vector represented by $\delta\mathbf{q}$. Thus, based on the constraint Jacobian matrix it is possible to decompose the tangent space to two mutually orthogonal subspaces: where one subspace called the space of admissible motion is spanned by the generalized directions in the tangent space that are admissible with the constraints and the other subspace called the space of constrained motion represents the constrained directions. All the vectors represented by the rows of $\mathbf{A}$ belong to this second subspace. We can find the admissible set of virtual displacements $\delta\mathbf{q}_a$ by projecting $\delta\mathbf{q}$ onto the two subspaces of the tangent space. However, neither $\mathbf{A}$ nor $\delta\mathbf{q}$ carry directly information about the basis of the vector space relative to which the vector coordinates are interpreted. The knowledge of such a basis is necessary to establish the decomposition between the two subspaces.

Were the elements of $\delta\mathbf{q}$ and $\mathbf{A}$ dimensionless quantities, numbers only, then the basis of this vector space can be the usual Eucledian basis [11] known from linear algebra, resulting in the $n \times n$ identitity matrix as the metric tensor representation. Also leading to the conclusion that the basis and the dual basis of the vector space are basically the same (covariant and contravariant bases). In this case, the singular value decomposition of matrix $\mathbf{A}$ [5] can be used to define subspaces of the n dimensional vector space and determine projector operators [4] to find the decomposition of $\delta\mathbf{q}$.

However, the elements of $\delta\mathbf{q}$ and $\mathbf{A}$ are not numbers only, they are physical quantities carrying the number value as one part of the quantity, and the other part is being the physical unit. Also both $\delta\mathbf{q}$ and $\mathbf{A}$ are usually not homogeneous in units, thus it is not possible to transfer them into a dimensionless form. Therefore, the use of the usual Eucledian basis does not lead to a vector norm that would be invariant under coordinate transformations (including transformations in units too), i.e., it is not suitable to consider the system on a geometric basis. Furthermore, it can also involve physically meaningless operations, e.g., "addition" of two quantities with different physical units. However, if the usual Eucledian basis assumption is not correct then we should conclude based on equation (5) that the vector coordinates contained in the rows of $\mathbf{A}$ are not expressed in the same basis as the elements of $\delta\mathbf{q}$. The rows of $\mathbf{A}$ are represented relative to the (contravariant) basis of the tangent space while the elements of $\delta\mathbf{q}$ are expressed relative to the dual (covariant) basis. This is a necessary condition for equation (5) to be understood on a geometric ground if the assumption of the usual Eucledian basis is not applicable.

Thus, we need to find the physically meaningful bases for the vector space, which can result in a positive definite matrix for the metric tensor, and would assure that the norm of a vector in the vector space is invariant under coordinate transformations. This has to be related to a scalar invariant of the system interpreted in the tangent space. The mass matrix is a positive definite, symmetric matrix, which is related to such a scalar invariant (kinetic energy). Let us consider the Cholesky decomposition of the mass matrix: $\mathbf{M} = \mathbf{L}^T\mathbf{L}$, where $\mathbf{L}$ is an upper triangular, nonsingular matrix. The algorithm of the Cholesky decomposition of the mass matrix involves only physically meaningful algebraic manipulations [7]. Based on this decomposition, the rows of $\mathbf{L}^T$ (columns of $\mathbf{L}$) can be seen as a possible set of the dual base vectors for the tangent space. The inverse of the mass matrix can be written as $\mathbf{M}^{-1} = \mathbf{L}^{-1}\mathbf{L}^{T^{-1}}$, and the columns of $\mathbf{L}^{T^{-1}}$ will contain the base vectors of the tangent space. We can introduce the following coordinate transformation for the generalized virtual displacements

$$\delta\tilde{\mathbf{q}} = \mathbf{L}\delta\mathbf{q}, \qquad \delta\mathbf{q} = \mathbf{L}^{-1}\delta\tilde{\mathbf{q}} \tag{11}$$

where the physical units of $\delta\mathbf{q}$ and $\delta\tilde{\mathbf{q}}$ are different: $\delta\mathbf{q}$ is in general nonhomogeneous in units while $\delta\tilde{\mathbf{q}}$ will always be homogeneous as a result of the above transformation. Based on this transformation, equation (5) can also be rewritten as

$$\tilde{\mathbf{A}}\delta\tilde{\mathbf{q}} = \mathbf{0} \tag{12}$$

where

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{L}^{-1} \tag{13}$$

is the modified constraint Jacobian matrix, where each of the rows are now homogeneous in units. The matrix product $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$ does make physical sense, hence for $\tilde{\mathbf{A}}$, it is possible to perform a singular value analysis (which is related to the eigenvalue problem of $\tilde{\mathbf{A}}\tilde{\mathbf{A}}^T$). We can now define projector operators related to equation (12), and decompose $\delta\tilde{\mathbf{q}}$ as

$$\delta\tilde{\mathbf{q}} = \tilde{\mathbf{P}}_c\delta\tilde{\mathbf{q}} + \tilde{\mathbf{P}}_a\delta\tilde{\mathbf{q}}, \tag{14}$$

where

$$\tilde{\mathbf{P}}_c = \tilde{\mathbf{A}}^\dagger\tilde{\mathbf{A}}, \qquad \tilde{\mathbf{P}}_a = (\mathbf{I} - \tilde{\mathbf{A}}^\dagger\tilde{\mathbf{A}}) \tag{15}$$

are projector operators, $\mathbf{I}$ is an $n \times n$ identity matrix, and $\mathbf{A}^\dagger$ represents the pseudo-inverse of $\tilde{\mathbf{A}}$. Applying the coordinate trasformation (11) to equations (14) and (15), and using the definition of $\tilde{\mathbf{A}}$, we can obtain the physically meaningful decomposition of the generalized virtual displacements $\delta\mathbf{q}$ as

$$\delta\mathbf{q} = \mathbf{P}_c\delta\mathbf{q} + \mathbf{P}_a\delta\mathbf{q}, \tag{16}$$

where

$$\mathbf{P}_c = \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^\dagger\mathbf{A}, \qquad \mathbf{P}_a = (\mathbf{I} - \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^\dagger\mathbf{A}), \tag{17}$$

and $\delta\mathbf{q}_c = \mathbf{P}_c\delta\mathbf{q}$, $\delta\mathbf{q}_a = \mathbf{P}_a\delta\mathbf{q}$, $()^\dagger$ indicates the pseudo-inverse of $()$. Operator $\mathbf{P}_a$ projects onto the space of admissible motion and operator $\mathbf{P}_c$ projects onto the space of constrained motion. It can be shown that these operators are symmetric, i.e., $\mathbf{P}_c = \mathbf{P}_c^T$ and $\mathbf{P}_a = \mathbf{P}_a^T$ (in the following we will still indicate the transpose operation to make it easier to follow the derivations).

In our discussion, we did not use the assumption that the constraints are independent of each other, i.e., the constraint Jacobian matrix has a full row rank. The above expressions are also valid for systems where the constraint Jacobian does not have a full row rank (e.g. over-constrained systems, singular systems). For the case when the rows of the constraint Jacobian are independent, i.e., we have independent constraint equations, it can be shown that

$$\mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^\dagger = \mathbf{L}^{-1}\mathbf{L}^{-1^T}\mathbf{A}^T(\mathbf{A}\mathbf{L}^{-1}\mathbf{L}^{-1^T}\mathbf{A}^T)^{-1} =$$

$$\mathbf{M}^{-1}\mathbf{A}^T\left(\mathbf{A}\mathbf{M}^{-1}\mathbf{A}^T\right)^{-1}, \tag{18}$$

which in linear algebraic terms can be interpreted as a weighted generalized inverse of $\mathbf{A}$.

Array $\delta\mathbf{q}_a$ gives the admissible set of virtual displacements without the introduction of new, independent variations. Thus based on equation (7),

$$\delta\mathbf{q}_a^T\mathbf{Q}_{constraint} = 0, \tag{19}$$

which, based on equations (8), (10) and (16), leads to

$$\mathbf{P}_a^T\mathbf{Q}_{constraint} = \mathbf{P}_a^T\mathbf{A}^T\boldsymbol{\lambda} = \mathbf{0}, \tag{20}$$

By simple subsitution, it can be seen that $\mathbf{P}_a^T$ given by equation (17) satisfies the physical criteria of equation (20). Projector $\mathbf{P}_a$ will eliminate the constraint forces from the formulation. Using the above considerations and equations, the following identities can also be derived

$$\mathbf{P}_a^T\mathbf{M}\mathbf{P}_c = \mathbf{0}, \quad \mathbf{P}_c^T\mathbf{M}\mathbf{P}_a = \mathbf{0}, \tag{21}$$

$$\mathbf{P}_a^T\mathbf{M}\mathbf{P}_a = \mathbf{P}_a^T\mathbf{M}, \tag{22}$$

and

$$\mathbf{P}_a\mathbf{M}^{-1}\mathbf{Q}_{constraint} = \mathbf{0}. \tag{23}$$

The role of the virtual displacements is to characterize admissible and constrained directions for the motion of the system. The actual physical quantities associated with the directions interpreted in the tangent space are the generalized velocities. Therefore, the decomposition based on the projector operators can be done for $\dot{\mathbf{q}}$ using equations (2) as

$$\dot{\mathbf{q}} = \mathbf{P}_c\dot{\mathbf{q}} + \mathbf{P}_a\dot{\mathbf{q}} = \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}\mathbf{A}\dot{\mathbf{q}} + (\mathbf{I} - \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}\mathbf{A})\dot{\mathbf{q}}. \tag{24}$$

Also, a decomposition similar to the above can be performed for the generalized acceleration $\ddot{\mathbf{q}}$ using equation (4) as

$$\ddot{\mathbf{q}} = \mathbf{P}_c\ddot{\mathbf{q}} + \mathbf{P}_a\ddot{\mathbf{q}} = \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}\mathbf{A}\ddot{\mathbf{q}} + (\mathbf{I} - \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}\mathbf{A})\ddot{\mathbf{q}}, \tag{25}$$

However, we have to note that geometrically, the elements of $\ddot{\mathbf{q}}$ do not represent the coordinates of the generalized acceleration vector [13].

Based on these identities and decompositions, it is possible to see that projectors $\mathbf{P}_a$ and $\mathbf{P}_c$ are separating the dynamics characteristics of the admissible and constrained motions of the system. Equation (21) shows that for a natural system where the kinetic energy is a quadratic function of the generalized coordinates, projector operators $\mathbf{P}_a$ and $\mathbf{P}_c$ exactly separates the kinetic energy into two isolated parts (without coupling): one that characterizes constrained motion ($\frac{1}{2}\dot{\mathbf{q}}^T\mathbf{P}_c^T\mathbf{M}\mathbf{P}_c\dot{\mathbf{q}}$), and the other characterizing admissible motion ($\frac{1}{2}\dot{\mathbf{q}}^T\mathbf{P}_a^T\mathbf{M}\mathbf{P}_a\dot{\mathbf{q}}$) of the system.

Based on these considerations, by substituting the decomposition (10) into (6) we obtain the constraint force free equations as

$$\mathbf{P}_a^T\left(\mathbf{Q}_{inertial} - \mathbf{Q}_{applied}\right) = \mathbf{0}. \tag{26}$$

This equation can be seen as the counterpart of the independent projections where a minimum set of virtual displacements is introduced leading to $n-m$ projected equations, but here the formulation is expressed in terms of n equations, using the non-minimum set of variables. The major physical idea and contribution is expressed by equation (26), which is a consequence of the D'Alembert-Lagrange principle, as fundamental idea of constrained systems. Considering the above expressions, equation (26) can be rewritten as

$$\mathbf{P}_a^T\left(\mathbf{M}\mathbf{P}_a\ddot{\mathbf{q}} + \mathbf{C} - \mathbf{Q}_{applied}\right) = \mathbf{0}. \tag{27}$$

This formulation expresses the dynamics of the system admissible with the constraints but using the original redundant set of variables. Since the number of degrees of freedom of the constrained system is less then the number of variables used, the "generalized mass matrix" $\mathbf{P}_a^T\mathbf{M}\mathbf{P}_a$ is singular. However, this form can be useful in inverse dynamics and control related applications, where the user can keep the original (larger) set of variables and can solve the inverse dynamics and study the effects of control actions on the behavior of the system. Matrix $\mathbf{P}_a^T\mathbf{M}\mathbf{P}_a$ also gives information about how well the admissible motion of the system can be influenced by control actions. We believe that the formulation represented by equations (26) and (27) has not been derived and discussed before.

Considering the expression of $\mathbf{P}_a$ and substituting it into equation (27), we obtain

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{L}^T(\mathbf{A}\mathbf{L}^{-1})^{\dagger}(\dot{\mathbf{A}}\dot{\mathbf{q}} + \dot{\mathbf{b}}) + \mathbf{P}_a^T\mathbf{C} - \mathbf{P}_a^T\mathbf{Q}_{applied} = \mathbf{0}. \tag{28}$$

This result is a new, general form of the equations of motion for constrained systems for both forward and inverse dynamics investigations. It applies to both redundantly ($\mathbf{A}$ does not have a full row rank) and nonredundantly ($\mathbf{A}$ has a full row rank) constrained systems, where the constraints can be holonomic and/or nonholonomic. We can consider that term $\mathbf{L}^T(\mathbf{A}\mathbf{L}^{-1})^{\dagger}(\dot{\mathbf{A}}\dot{\mathbf{q}} + \dot{\mathbf{b}})$ characterizes the nonlinearity and time dependence of the constraints. If the system is subjected to linear, stationary constraints only, then this term vanishes. The other terms in the equations are present for any type of constraints. This feature of equation (28) can be useful to assess the constraints and the effect of their nonlinearity and time dependence on the dynamic behavior of mechanical systems.

The approach described so far also opens up another possibility to establish further methods for dynamics investigations. If we are not applying the decompostion of the virtual displacements (as given in equation (10)), then from equation (6) with the explicit incorporation of the constraint forces, the generalized accelerations can be obtained as

$$\ddot{\mathbf{q}} = \mathbf{M}^{-1}(\mathbf{Q}_{applied} + \mathbf{Q}_{constraint} - \mathbf{C}). \tag{29}$$

From this, using projector $\mathbf{P}_a$ and equation (23), we can obtain

$$\mathbf{P}_a\ddot{\mathbf{q}} = \mathbf{P}_a\mathbf{M}^{-1}(\mathbf{Q}_{applied} - \mathbf{C}), \tag{30}$$

where the Lagrangian multipliers are eliminated by employing the projector operator. The generalized accelerations can then be recovered based on equations (4) and (25) as

$$\ddot{\mathbf{q}} = \mathbf{P}_a\ddot{\mathbf{q}} - \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}\left(\dot{\mathbf{A}}\dot{\mathbf{q}} + \dot{\mathbf{b}}\right). \tag{31}$$

Equations (30) and (31) represent another way of solving equation (9) when constraints can also be dependent. This method given by equations (29) - (31) can be used for the forward dynamics problem for example. Udwadia and Kalaba [14], [15] also arrived to a formulation suitable to handle redundant constraints based on a different line of thought.

The formulation discussed can be used in various applications (e.g., inverse dynamics, control analysis, forward dynamics). In the following, we would like to discuss some applications to the simulation of constrained systems. We will use the notation that $\mathbf{W} = \mathbf{L}^{-1}(\mathbf{A}\mathbf{L}^{-1})^{\dagger}$.

SIMULATION OF CONSTRAINED SYSTEMS

The formulation discussed in the preceding part of the paper can be used to improve the simulation of constrained systems, i.e., the solution of the time-discretized form of the equations of motion. Algorithms can be developed which are based on the detailed formulation presented in the previous section. However, here we would like to describe another potential of the approach. This originates from the decomposition described and can be used to develop a toolset to support any simulation environment and method for constrained system simulation. Before explaining this, let us look at briefly the simulation of a constrained system. In this section, we will primarily deal with systems where the constraint equations are independent, hence, it is possible to model them using the traditional techniques too.

The usual preliminary requirement for a simulation that the accelerations need to be determined from the dynamic equations. The two main approaches to consider the problem are to either use a minimum set of independent quantities or a non-minimum set of variables. One thing that should be emphasized, is that each method of these approaches is suitable to determine the exact accelerations at each time step if the input values of the velocities and coordinates are also accurate.

The second step in the simulation process is to integrate the accelerations and velocities to obtain the velocity and coordinate values for the next time step. This is where one of the main difficulties in the simulation of constrained systems lies: satisfying the constraint conditions during the simulation process. The problem dealt with here can basically be stated as follows. Given the time-discretized system, we consider two subsequent time steps t_i and t_{i+1} and assume that the generalized velocities and coordinates (either in minimum or non-minimum form) are "exactly" (down to the level of machine precision) known at t_i. Based on this, no matter what method we use at the acceleration level, we can determine the exact accelerations for the constrained motion problem for t_i. However, by integrating the accelerations and velocities numerical errors can be introduced, and also solutions not admissible with the constraints are possible (e.g. time integration is indeterminate in terms of adding a constant term). These problems can de-stabilize the solution of the constrained system. The problem then we have to deal with is that how to ensure that the velocities and coordinates determined for time step t_{i+1} using a numerical time integration process will meet the constraint conditions.

The other main problem in simulations is that even if the constraints are satisfied how we know that the solution obtained for the generalized coordinates and velocities is actually the correct one, i.e., the one which could be observed during the motion of the real physical system.

The technique we suggest here is an effective and simple method to improve the accuracy of simulations. An important feature is that this can be used as a stand-alone procedure or can be added to existing methods to improve the performance. Let us consider that after the time integration, we arrive to velocities $\dot{\mathbf{q}}_{i+1}$ and coordinates $\mathbf{q}_{i+1}$ for time step t_{i+1}, where the coordinate increments between the two time steps can be defined as $\Delta\mathbf{q}_{i+1} = \mathbf{q}_{i+1} - \mathbf{q}_i$. As we already discussed under Section 2, the generalized velocities at each time instant can be split into two parts using the constraint Jacobian matrix as given by equation (24). Based on this $\dot{\mathbf{q}}_{i+1}$ can also be decomposed as

$$\dot{\mathbf{q}}_{i+1} = \mathbf{P}_c(\mathbf{q}_{i+1})\dot{\mathbf{q}}_{i+1} + \mathbf{P}_a(\mathbf{q}_{i+1})\dot{\mathbf{q}}_{i+1} =$$

$$\mathbf{W}(\mathbf{q}_{i+1})\mathbf{A}(\mathbf{q}_{i+1})\dot{\mathbf{q}}_{i+1} + \left(\mathbf{I} - \mathbf{W}(\mathbf{q}_{i+1})\mathbf{A}(\mathbf{q}_{i+1})\right)\dot{\mathbf{q}}_{i+1}, \qquad (32)$$

where $\mathbf{P}_c\dot{\mathbf{q}}_{i+1}$ represents the part of the velocities that are not admissible with the constraints, i.e. constraint violations and $\mathbf{P}_a\dot{\mathbf{q}}_{i+1}$ can be considered as the filtered solution at the velocity level. This filtered solution is actually what we are looking for to construct the solution for the system velocities, it satisfies the constraints of the time-discretized system and gives the part of the velocities that are admissible with the constraints. By using the projector $\mathbf{P}_a$, the constraint violations are filtered out, and based on equation (2), the exact solution at the velocity level can be obtained as

$$\dot{\mathbf{q}}_{i+1}^{exact} = -\mathbf{W}(\mathbf{q}_{i+1})\mathbf{b}_{i+1} + \mathbf{P}_a(\mathbf{q}_{i+1})\dot{\mathbf{q}}_{i+1}, \qquad (33)$$

where $\mathbf{b}_{i+1}$ is the value of $\mathbf{b}$ at t_{i+1}. This solution expressed in the non-minimum set of variables is equivalent with the solution that could be obtained for the velocities by introducing independent velocity components. This solution for the velocities is possible because the constraints are linear at the velocity level. For obtaining the solution given by equation (33), we need to know the configuration of the system accurately at t_{i+1}. To improve the accuracy of the configuration, the correction scheme described below can be used. Alternatively, it can be accurate enough to use the admissible and constrained directions determined for time instant t_i ($\mathbf{P}_a(\mathbf{q}_i)$, $\mathbf{P}_c(\mathbf{q}_i)$) to filter the velocities.

Consider now the configuration of the system, $\mathbf{q}_{i+1}$ obtained as a result of time integration and assume that the configuration level constraints, equation (3), are violated, thus $\mathbf{\Phi}(\mathbf{q}_{i+1}, t_{i+1}) =$

$\mathbf{h} \neq \mathbf{0}$. If the constraints are linear at the configuration level, i.e. $\mathbf{\Phi}$ can be expressed as a linear function of the non-minimum set of generalized coordinates, then the exact elimination of this constraint violation is possible based on a methodology similar to the one described for the velocities. However, the configuration level constraints can be highly nonlinear in terms of the coordinates. As a result of the constraint nonlinearity the generalized directions representing admissible and constrained motions, and defined by the projector operators $\mathbf{P}_a$ and $\mathbf{P}_c$ are changing from one time step to the other. Figure 1 provides a simple illustration of the geometry involved. However, we can still consider that the increment $\Delta\mathbf{q}_{i+1}$ can be calculated as a sum of two "linear increments": a coordinate increment along the directions admissible with the constraints at t_i and another increment along the directions constrained at t_i as

$$\Delta\mathbf{q}_{i+1} = \mathbf{P}_c(\mathbf{q}_i)\Delta\mathbf{q}_{i+1} + \mathbf{P}_a(\mathbf{q}_i)\Delta\mathbf{q}_{i+1}, \tag{34}$$

If the numerical time integration were exact in both directions then the sum of these two increments would lead to the correct new configuration (Figure 1) that satisfies the constraints. Errors can be introduced in both linear increments to alter the solution. We will outline two possible ways to compensate for these errors.

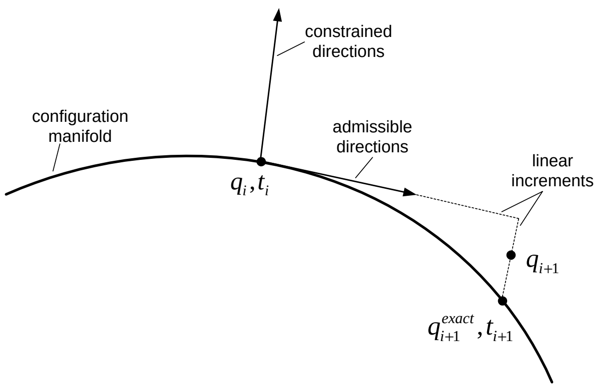

Figure 1. Relation between the system configurations in two adjacent time steps

We can consider that at the velocity level we have accurate values at time point t_i, which are admissible with the constraints. Based on this, we can introduce a simplified method by making the assumption that the linear increment along the admissible directions can be calculated by the integrator more accurately than the increment along the constrained directions. (It is emphasized that this assumption is not error free, but in comparison, these errors may be taken negligible.) Thus, the constraint violation $\mathbf{h}$ can actually be conceived as a violation along the constrained directions of the tangent space interpreted at t_i. This situation is depicted in Figure 1. This also means that the calculated increment along the constrained directions can be modified to apply correction to remedy the constraint violation. This correction formula can be written as

$$\mathbf{q}_{i+1}^{k+1} = \mathbf{q}_{i+1}^{k} - \mathbf{W}(\mathbf{q}_i)\mathbf{h}^k, \tag{35}$$

where k and $k+1$ represent two successive steps in the correction process. This formula can also be applied in a quasi-iterative way. It is called quasi-iterative because it is emphasized again that the generalized constrained directions during the correction process are always taken at $(\mathbf{q}_i, t_i)$ thus the constraint Jacobian is evaluated at t_i only. Therefore, it can improve the performance in an open-loop corrective fashion without closing the iterative loop.

The other possibility is to always modify the directions of corrections too based on the corrected configuration values to account for possible errors along the admissible directions as well, thus

$$\mathbf{q}_{i+1}^{k+1} = \mathbf{q}_{i+1}^{k} - \mathbf{W}(\mathbf{q}_{i+1}^{k})\mathbf{h}^k. \tag{36}$$

This formula can be applied in a fully iterative way or also as an "open-loop" corrective method. Our experience is that usually an open-loop corrective method can ensure substantial improvement in accuracy (one or two corrections), while still keeping the efficiency of the simulation.

Blajer [2] derived a similar methodology for the elimination of the constraint violations for the case of independent constraints based on the differential geometric representation of equation (9). In [2], corrections are applied for both the velocities and coordinates, instead of using a filtering algorithm for the velocities, and it is not defined how the constrained directions are maintained during the iterative process.

The methodology described in this section is simple to implement in any formulation for the velocity and configuration levels. It can be adopted to improve the performance of existing simulation environments. The user can easily try this method to test the performance.

It will require further extensive simulation studies to compare the two ways for the configuration correction as given by equations (35) and (36), and compare the potentials of this approach to other existing methods. It may turn out that equation (35) is giving more reliable results, because regardless of the assumption of being accurate along the admissible directions, we still have a good control over the constrained direction we are going in. Meanwhile, equation (36) seems attractive since the above assumption is released. However, we lose control over the definition of constrained direction, and it may happen that we actually find a constraint consistent solution which is physically meaningless. A general problem in constrained systems simulation is how to make sure that the the solution is accurate enough along the admissible directions, i.e., to check if we are finding the real solution not just something which satisfies the constraints. Currently, the only reliable, known method for this is to use accurate integration methods with appropriately small time steps. Another approach that has been believed to improve the accuracy is to apply so-called energy rate constraints. But, as was shown in [2], this could actually be quite misleading.

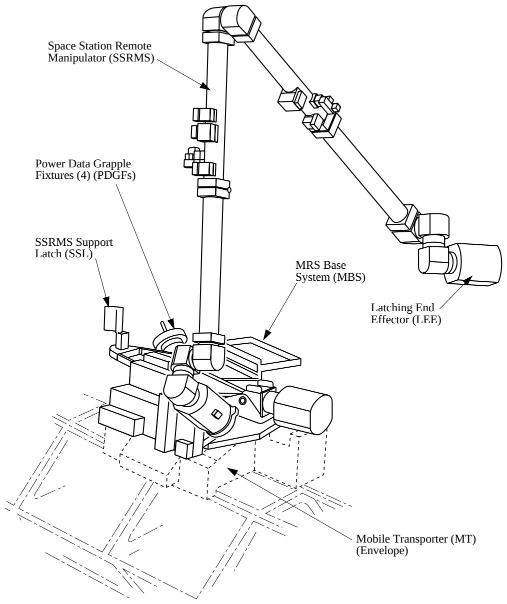

Figure 2. The Space Station Remote Manipulator System (SSRMS) – Canadarm2

As an example, we discuss the simulation of the seven jointed Space Station Remote Manipulator System (SSRMS) (Figure 2) when both of its ends (end effectors) are attached to interfaces on the International Space Station, and it forms a closed kinematic loop. The SSRMS was installed onto the space station in April 2001 and has since been used many times to assist in the assembly and operation of the space station. More detailed information on this robotic arm is available in [8], [12]. The scenario investigated here, where the closed-loop arm is subjected to loads similar to the ones observed during space shuttle docking maneuvers or space station attitude control. We consider that the base end effector is fully restrained, but the other end effector of the arm is attached to an interface that allows for rotations thus here we have three independent position constraints imposed on the arm. Thus, this closed loop system has ten degrees of freedom (six admissible motion possibilities for the space station and four for the SSRMS), and three constrained "degrees of freedom". The simulation we consider here is 80 second long run, where the inputs were considered to simulate the load case that can be observed during the above mentioned operations. The initial joint angles of the SSRMS in the closed loop configuration was $[-26.4°, -42.2°, -18.6°, -93.4°, -97.1°, 19.1°, 52.5°]$. The acceleration level calculations were performed using the classical Lagrange multiplier technique (simplest method). But, as was mentioned earlier, any other method could have been used to solve for the accelerations. Figures 3 and 4 show the results of the simulation for the constraint violations in velocities and configuration where no stabilization or correction technique was used at all. We can see that there are serious constraint violations. It has to be noted that in the context of space robotics, these violations are even more substantial than for some other applications, because of the lower operating speeds. However, we also have to notice, that these violations start to be considerable only after 25 seconds. Thus, for shorter simulations, the basic technique can also give reasonable results. (This, however, strongly depends on the excitation signal too.)

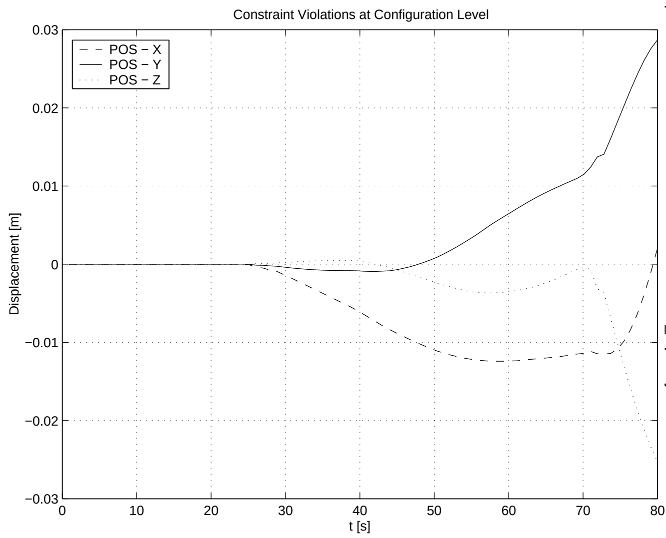

Figure 3. Constraint violations at the configuration level with no corrections and stabilization

The filtering and correction method suggested above can be easily added to the simulation algorithm. Here, we have used the method represented by equation (36). If we use this technique to improve the results in an open-loop corrective way, then using one velocity filtering step and two configuration correction steps,

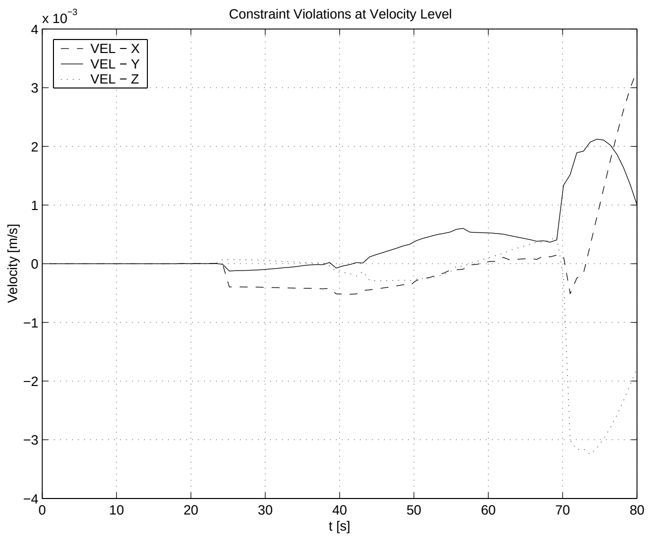

Figure 4. Constraint violations at the velocity level with no corrections and stabilization

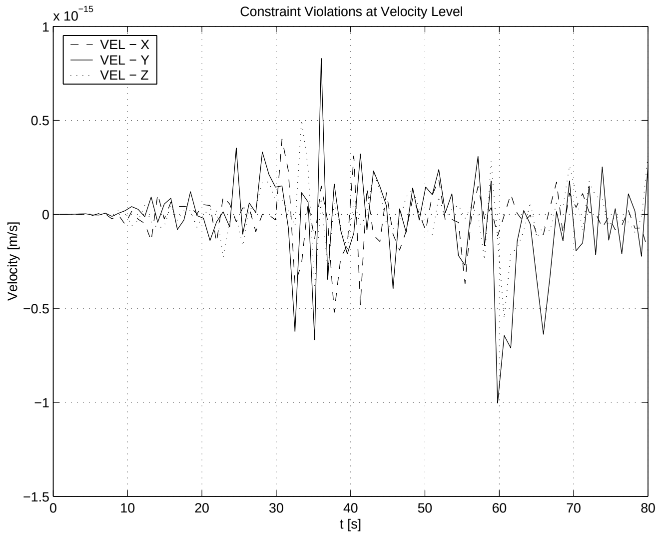

Figure 6. Constraint violations at the velocity level with velocity filtering

it is possible to virtually eliminate the constraint violations in this simulation, Figures 5 and 6 (the violations will approximately be in the range of $10^{-14} - 10^{-15}$, which can be considered machine precision in the Matlab/Simulink environment). This gain in accuracy and also keeping the method simple and efficient can be particularly useful in the longer simulations of complex systems where real-time aspects are also important.

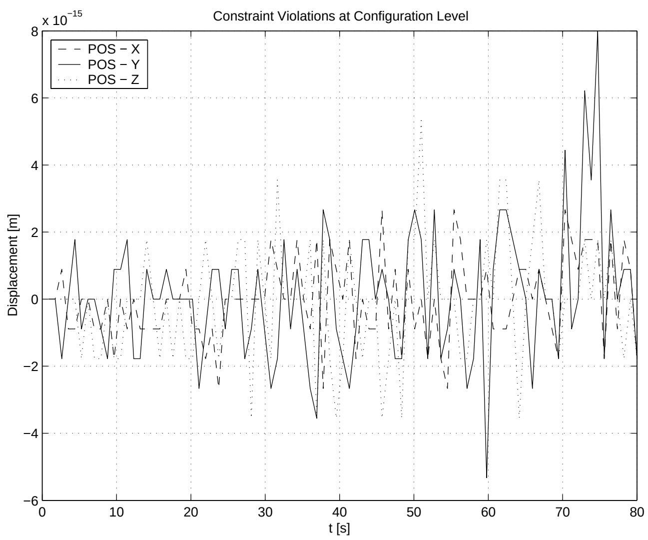

Figure 5. Constraint violations at the configuration level with two configuration corrections

CONCLUSION

In this paper, problems in the dynamic modeling and simulation of constrained systems were considered. A novel dynamic formulation was introduced based on the D'Alembert-Lagrange principle for the case when a non-minimum set of variables is used based on the physically meaningful decomposition of the generalized virtual displacements. This formulation can be useful in inverse dynamics, forward dynamics and control analysis. The general formulation is valid for the cases of both independent and dependent constraint equations. Therefore, it can also offer advantages in the analysis of over-constrained and singular systems. Based on the formulation introduced, a simple and efficient approach was summarized for the simulation problem of constrained dynamic systems based on velocity filtering and configuration corrections. It can be incorporated into existing simulation environment and can be used along with the various dynamic formulations. Our experience is that with this methodology it is possible to reduce the constraint errors at the acceleration, velocity and configuration levels to machine precision. More analysis is currently underway to further evaluate the potentials of this approach and compare it to the other methods used in computational dynamics of constrained systems.

REFERENCES

[1] Blajer, W., A Geometric Unification of Constrained System Dynamics, *Multibody System Dynamics*, Vol. 1, pp. 3-21, 1997.

[2] Blajer, W., Elimination of Constraint Violation and Accuracy Aspects in Numerical Simulation of Multibody Systems, *Multibody System Dynamics*, Vol. 7, pp. 265-284, 2002.

[3] Garcia de Jalón J. and Bayo E., *Kinematic and Dynamic Simulation of Multibody Systems - The Real-Time Challenge* Springer-Verlag, New York, 1994.

[4] Golub, G.H. and van Loan C.F., *Matrix Computations* The John Hopkins University Press, third edition, 1996.

[5] Klema, V.C. and Laub, A.J., The Singular Value Decomposition: Its Computation and Some Applications, *IEEE Trans. Automatic Control*, Vol. 25, No. 2, pp. 164-176, 1980.

[6] Kövecses, J. Piedbœuf, J.-C. and Lange, C., Dynamics Modeling and Simulation of Constrained Robotic Systems, *IEEE/ASME Trans. Mechantronics*, Accepted for publication, 2003.

[7] Meirovitch, L.: *Principles and Techniques of Vibrations* Prentice-Hall, Upper Saddle River, New Jersey, 1997, pp. 280-283.

[8] Nguyen, P.K. and Hughes, P.C., Teleoperation: From the Space Shuttle to the Space Station, *AIAA - Teleoperation and Robotics in Space* edited by Skaar, S.B. and Ruoff, C.F., Vol 161 of Progress in Astronautics and Aeronautics AIAA, Washington, DC, 1994.

[9] Rosenberg, R.M., *Analytical Dynamics of Discrete Systems,* Plenum Press, New York, 1977.

[10] Shabana, A.A., *Computational Dynamics* John Wiley, New York, 2001.

[11] Spivak, M., *Calculus on Manifolds* W. A. Benjamin, New York, 1965.

[12] www.mdrobotics.ca/galleryframe.html

[13] Synge J.L. and Schild, A., *Tensor Calculus*, University of Toronto Press, Toronto, 1969.

[14] Udwadia, F.E. and Kalaba, R.E., A New Perspective on Constrained Motion, *Proc. Royal Society Ser. A*, Vol. 439, No. 1906, pp. 407-410, 1992.

[15] Udwadia, F.E. and Kalaba, R.E., The Geometry of Constrained Motion, *Z. Angew. Math. Mech. (ZAMM)*, Vol. 75, No. 8, pp. 637-640, 1995.