Multibody System Dynamics (2005) 14: 155–187

Nonminimal Kane's Equations of Motion for Multibody Dynamical Systems Subject to Nonlinear Nonholonomic Constraints

ABDULRAHMAN H. BAJODAH[1], DEWEY H. HODGES[2] and YE-HWA CHEN[3]
[1]*Aeronautical Engineering Department, King Abdulaziz University, Jeddah 21589, Saudi Arabia*
[2]*School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A.*
[3]*School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, U.S.A.*

(Received: 29 July 2003; accepted in revised form: 28 September 2004)

Abstract. Nonholonomic constraint equations that are nonlinear in velocities are incorporated with Kane's dynamical equations by utilizing the acceleration form of constraints, resulting in Kane's nonminimal equations of motion, i.e. the equations that involve the full set of generalized accelerations. Together with the kinematical differential equations, these equations form a state-space model that is full-order, separated in the derivatives of the states, and involves no Lagrange multipliers. The method is illustrated by using it to obtain nonminimal equations of motion for the classical Appell–Hamel problem when the constraints are modeled as nonlinear in the velocities. It is shown that this fictitious nonlinearity has a predominant effect on the numerical stability of the dynamical equations, and hence it is possible to use it for improving the accuracy of simulations. Another issue is the dynamics of constraint violations caused by integration errors due to enforcing a differentiated form of the constraint equations. To solve this problem, the acceleration form of the constraint equations is augmented with constraint stabilization terms before using it with the dynamical equations. The procedure is illustrated by stabilizing the constraint equations for a holonomically constrained particle in the gravitational field.

Keywords: nonminimal Kane's equations, nonlinear nonholonomic constraints, acceleration form of constraint equations, numerical stability of dynamical equations, constraint stabilization

1. Introduction

Dynamicists have noticed the absence of nonlinear nonholonomic constraints from daily life observations, and some have gone so far as to argue the nonexistence of such constraints in nature [1]. Among the few examples of mechanical systems with nonlinear nonholonomic constraints in the literature of analytical mechanics is the one due to Appell [2] and Hamel [3]. The constraints were modeled as nonlinear in the velocities by a limiting process on the nonholonomic constraint equations. However, the validity of the resulting equations of motion was questioned [4] because of the reduction in order associated with this limit condition, which yields a qualitative change in the system behavior and a huge difference in the results versus those associated with taking the limit after obtaining the equations of motion. The reinterpretation of nonholonomic constraints of the rolling type as nonlinear is originally due to Saletan and Cromer [5]. A further study of the Appell–Hamel problem for the purpose of analyzing nonlinear nonholonomic constraints in the context of Kane's method is found in [6]. An example of nonlinear nonholonomic constraints of the nonintegrable in accelerations type is due to Kitzka [7]. The fact that the system is inherently nonlinearly constrained was dismissed in [8], and an alternative derivation was presented with the constraint equations turning out to be linear and nonholonomic.

Despite the controversy regarding the feasibility of nonlinear, nonholonomic constraints of the passive type, active constraints (also called servo- or program constraints) certainly can be nonlinear and nonholonomic. Furthermore, they need not be ideal [9] (of the Chetaev type) or limited to second order in the generalized coordinates. The importance of understanding servoconstraints for control system analysis and synthesis can be considered the main reason for studying the various categories of constraints, including nonlinear, nonholonomic constraints. Such is the focus of the present paper.

The treatment of nonlinear, nonholonomic constraints in Kane's approach began with the extension of Passerello–Huston equation to include such constraints [10]. Later, Huston's method of undetermined multipliers [11, 12] was generalized [6] to include nonlinear nonholonomic constraints.

During the past three decades, Kane's equations of motion were successfully applied to numerical analysis and simulation of multibody systems. The original treatment of using the minimal (reduced) set of equations [13] becomes less useful when the multibody system is composed of a large number of bodies that are heavily constrained. In such cases the resulting equations of motion increase in complexity and, hence, become more difficult to analyze and less efficient for time simulations. To alleviate this problem, dynamical equations with orders exceeding the numbers of degrees of freedom were derived. Examples are [11, 14–19].

In [20], a nonminimal version of Kane's equations of motion for constrained multibody systems is derived with the aid of the acceleration form of constraints and the tangential properties of Kane's method that provide a relationship between the constrained and unconstrained generalized active and inertia forces. The resulting equations of motion are explicit in the generalized accelerations, and involve no Lagrange multipliers. The derivation is based on simple mathematical operations on the unconstrained equations of motion. This is particularly advantageous in the case where the equations are already derived and more constraints are to be added to the system for the purpose of improving its design or studying its performance, because the method does not require a totally new derivation.

The use of the acceleration form of nonlinear constraint equations with Kane's equations was first depicted in [6, 11, 21, 22], where the (nonunique) orthogonal complement of the constraint matrix is multiplied by the full-order, constrained form of Kane's equations to eliminate the contribution of the generalized constraint forces. It is shown in [20] that a particular choice of the orthogonal complement matrix is embedded in the minimal Kane's equations and is obtained by expanding these equations in terms of the unconstrained generalized active and inertia forces. This particular choice implies the consistency among the governing equations because it guarantees the nondeficiency of the augmented matrix which becomes a generalized "constrained" inertia matrix, regardless what the system constraint Jacobian might be. This provides a tolerance towards the dependency among constraint equations, as it waives the need to extract the largest independent set of constraints, a process that can be difficult for highly constrained multibody systems.

However, the above mentioned treatment of nonholonomic constraints was limited to the "simple nonholonomic" type [23]. It is shown in this paper that the method is also capable of handling nonlinear constraints. This is exploited from the fact that the acceleration form of constraint equations is linear in the generalized accelerations, even if the nonholonomic constraint equations are nonlinear in the generalized speeds. On the other hand, the relations between the holonomic and the nonholonomic partial velocities and partial angular velocities of the system are preserved in the case of nonlinear nonholonomic constraints, and hence the special structure of the resulting constraint matrices A_1 and A_2 is also preserved.

In spite of the advantages of modeling multibody systems using the acceleration form of constraint equations, the accuracy of numerical simulations may degenerate due to constraint violations caused by enforcing the constraint equations at the acceleration level. This is especially true for the case of holonomic constraints, as the equations must be numerically integrated twice to obtain the generalized coordinates. It is shown in this paper that nonlinear nonholonomic constraints substantially alter the constraint violation dynamics, and can reduce the deterioration in accuracy of the numerical simulations. Furthermore, the explicit appearance of the acceleration form of constraint equations facilitates the augmentation of Baumgarte type damping terms [24] before inverting the generalized inertia matrix, in case it is necessary to modify the dynamical equations in order to suppress this violation.

The main contributions in this paper are twofold. First, nonminimal Kane's equations are formulated for nonlinearly constrained multibody dynamical systems. Second, the nonminimal Kane's equations of motion are modified in a manner to provide accuracy to numerical integration schemes for the purpose of improving the robustness of time simulations.

2. Nonminimal Kane's Equations with Nonlinear Nonholonomic Constraints

The development in this introduction can be found in [23], and is provided here for convenience. Consider a nonholonomic dynamical system $\mathcal{S}$ with p degrees of freedom consisting of a set of ν particles and μ rigid bodies, and let $\mathcal{R}$ be an inertial frame of reference in which the configuration of the system is described by a set of *n generalized coordinates* $q_1, \dots, q_n$. The *generalized speeds* $u_1 \dots u_n$ are scalar variables satisfying the *kinematical differential equations* [23]

$$\dot{q} = C(q, t)u + D(q, t). \tag{1}$$

In the above equation, q denotes a column matrix containing the n generalized coordinates, u denotes a column matrix containing the generalized speeds, $C \in \mathbb{R}^{n\times n}$, $D \in \mathbb{R}^n$, C^{-1} exists for all $q \in \mathbb{R}^n$, and all $t \in \mathbb{R}$, and $(\dot{\ }) = \mathrm{d}()/\mathrm{d}t$. The velocity of a generic particle P of the system relative to $\mathcal{R}$ can be written as [23]

$$^{\mathcal{R}}\mathbf{v}^P = \sum_{r=1}^{p} {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^P(q, t)u_r + {}^{\mathcal{R}}\tilde{\mathbf{v}}_t^P(q, t), \quad r = 1, \dots, p. \tag{2}$$

The *nonholonomic partial velocities* $^{\mathcal{R}}\tilde{\mathbf{v}}_1^P \dots {}^{\mathcal{R}}\tilde{\mathbf{v}}_p^P$ in Equation (2) are vector entities that can be obtained by inspecting the velocity expression of the particle for the coefficients of the *independent* generalized speeds $u_1 \dots u_p$. Another way to express the velocity of the particle P is by using the full set of generalized speeds $u_1 \dots u_n$ as [23]

$$^{\mathcal{R}}\mathbf{v}^P = \sum_{r=1}^{n} {}^{\mathcal{R}}\mathbf{v}_r^P(q, t)u_r + {}^{\mathcal{R}}\mathbf{v}_t^P(q, t), \quad r = 1, \dots, n, \tag{3}$$

where $^{\mathcal{R}}\mathbf{v}_1^P \dots {}^{\mathcal{R}}\mathbf{v}_n^P$ are the *holonomic partial velocities* of the particle P in $\mathcal{R}$, and $u_1 \dots u_n$ are constrained according to some constraint relations, as discussed in the next section. In a similar manner, the angular velocity of a generic body B of the system relative to $\mathcal{R}$ may be written as [23]

$$^{\mathcal{R}}\boldsymbol{\omega}^B = \sum_{r=1}^{p} {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_r^B(q, t)u_r + {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_t^B(q, t), \quad r = 1, \dots, p. \tag{4}$$

The coefficients of the independent generalized speeds in Equation (4), $^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_1^B \dots {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_p^B$, are called the *nonholonomic partial angular velocities* of B relative to $\mathcal{R}$. Another expression for $^{\mathcal{R}}\boldsymbol{\omega}^B$ is in terms of the full constrained set of generalized speeds [23],

$$^{\mathcal{R}}\boldsymbol{\omega}^B = \sum_{r=1}^{n} {}^{\mathcal{R}}\boldsymbol{\omega}_r^B(q, t)u_r + {}^{\mathcal{R}}\boldsymbol{\omega}_t^B(q, t), \quad r = 1, \dots, n, \tag{5}$$

where $^{\mathcal{R}}\boldsymbol{\omega}_1^B \dots {}^{\mathcal{R}}\boldsymbol{\omega}_n^B$ are the *holonomic partial angular velocities* of the body B in $\mathcal{R}$.

Let $\mathbf{R}_i$ be the resultant active force on the ith particle, P_i. The resultant active forces on the ith rigid body B_i are equivalent to a force $\mathbf{Z}_i$ on a point Q_i on B_i, together with a torque $\mathbf{T}_i$. The *nonholonomic generalized active forces* are defined as [23]

$$\tilde{F}_r(q, u, t) = \sum_{i=1}^{\nu} {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^{P_i} \cdot \mathbf{R}_i + \sum_{i=1}^{\mu} {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^{Q_i} \cdot \mathbf{Z}_i + \sum_{i=1}^{\mu} {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_r^{B_i} \cdot \mathbf{T}_i, \quad r = 1, \ldots, p. \tag{6}$$

The inertia torque of B_i relative to $\mathcal{R}$ is [23]

$${}^{\mathcal{R}}\mathbf{T}^{\star}_{B_i} = -{}^{\mathcal{R}}\boldsymbol{\alpha}^{B_i} \cdot \underline{\mathbf{I}}^{B_i} - {}^{\mathcal{R}}\boldsymbol{\omega}^{B_i} \times \underline{\mathbf{I}}^{B_i} \cdot {}^{\mathcal{R}}\boldsymbol{\omega}^{B_i}, \tag{7}$$

where $\underline{\mathbf{I}}^{B_i}$ is the *central inertia dyadic of B_i relative to B_i*. The *nonholonomic generalized inertia forces* are defined as [23]

$$\tilde{F}^{\star}_r(q, u, \dot{u}, t) = -\sum_{i=1}^{\nu} m_{P_i} {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^{P_i} \cdot {}^{\mathcal{R}}\mathbf{a}^{P_i} - \sum_{i=1}^{\mu} m_{B_i} {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^{B_i} \cdot {}^{\mathcal{R}}\mathbf{a}^{b_i} + \sum_{i=1}^{\mu} {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_r^{B_i} \cdot {}^{\mathcal{R}}\mathbf{T}^{\star}_{B_i}, \quad r = 1, \ldots, p. \tag{8}$$

where b_i is the mass center of body B_i.

Kane's dynamical equations of motion are [23]

$$\tilde{F}_r(q, u, t) + \tilde{F}^{\star}_r(q, u, \dot{u}, t) = 0, \quad r = 1, \ldots, p. \tag{9}$$

The accelerations ${}^{\mathcal{R}}\mathbf{a}^{P_i}$ and ${}^{\mathcal{R}}\mathbf{a}^{b_i}$ used to form the expression (8) for $\tilde{F}^{\star}_r$ are found by differentiating Equation (2). To obtain the inertia torques ${}^{\mathcal{R}}\mathbf{T}^{\star}_{B_i}$, Equation (4) is differentiated, and the resulting angular accelerations ${}^{\mathcal{R}}\boldsymbol{\alpha}^{B_i}$ are used in expression (7). If the holonomic partial velocities are used instead to form Equations (6) and (8), then n holonomic generalized active forces and n holonomic generalized inertia force can be defined for the nonholonomic system, and are distinguished from their nonholonomic counterparts by removing the "$\sim$" from above the corresponding symbols. Defining holonomic quantities for nonholonomic systems is useful in constructing the nonminimal form, as will be shown later in the paper.

2.1. CONSTRAINTS INVOLVING NONLINEARITY IN GENERALIZED SPEEDS

The m nonholonomic constraint equations take the form

$$\phi(q, u, t) = 0, \tag{10}$$

where ϕ is the column matrix

$$\phi = \lfloor \phi_1 \ldots \phi_m \rfloor^T, \tag{11}$$

in which $\phi(q, u, t)$ is in general nonlinear in its arguments. Differentiating the constraint equations, Equation (10), with respect to time t, one obtains

$$\dot{\phi}(q, u, \dot{u}, t) = \frac{\partial \phi}{\partial q}\dot{q} + \frac{\partial \phi}{\partial u}\dot{u} + \frac{\partial \phi}{\partial t} = 0. \tag{12}$$

Substitution of the kinematical differential equations, Equation (1), in the above equation results in the acceleration form of the constraint equations

$$\dot{\phi}(q, u, \dot{u}, t) = \frac{\partial \phi}{\partial u}\dot{u} + B_1(q, u, t)u + B_2(q, u, t) = 0, \tag{13}$$

where

$$B_1(q, u, t) = \frac{\partial \phi}{\partial q}C(q, t) \tag{14}$$

$$B_2(q, u, t) = \frac{\partial \phi}{\partial q}D(q, t) + \frac{\partial \phi}{\partial t}. \tag{15}$$

Let

$$u = \lfloor u_I^T \quad u_D^T \rfloor^T, \tag{16}$$

where $u_I = \lfloor u_1 \ldots u_p \rfloor^T$ and $u_D = \lfloor u_{p+1} \ldots u_n \rfloor^T$. Define the $m \times p$ matrix

$$\begin{aligned} J_1(q, u, t) &:= \begin{bmatrix} \frac{\partial \phi_1}{\partial u_1} & \cdots & \frac{\partial \phi_1}{\partial u_p} \\ \vdots & \vdots & \vdots \\ \frac{\partial \phi_m}{\partial u_1} & \cdots & \frac{\partial \phi_m}{\partial u_p} \end{bmatrix} \\ &= \frac{\partial \phi}{\partial u_I} \end{aligned} \tag{17}$$

and the $m \times m$ matrix

$$\begin{aligned} J_2(q, u, t) &:= \begin{bmatrix} \frac{\partial \phi_1}{\partial u_{p+1}} & \cdots & \frac{\partial \phi_1}{\partial u_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial \phi_m}{\partial u_{p+1}} & \cdots & \frac{\partial \phi_m}{\partial u_n} \end{bmatrix} \\ &= \frac{\partial \phi}{\partial u_D}. \end{aligned} \tag{18}$$

We assume that $u_{p+1}, \ldots, u_n$ can be chosen such that the matrix J_2 is nonsingular for all q, u, and t that satisfy the constraint equations, Equation (10). This variable partitioning was first introduced in the seminal article [25]. To avoid singularities of J_2, variant formulations may be prepared in advance by redefining the sets of generalized speeds u_I and u_D in the vicinities of these singularities. A projective criterion for best variable partitioning is found in [26]. Thus, Equation (13) can be written as

$$\dot{\phi}(q, u, \dot{u}, t) = J_1 \dot{u}_I + J_2 \dot{u}_D + B_1(q, u, t)u + B_2(q, u, t) = 0. \tag{19}$$

Solving for $\dot{u}_D$ yields

$$\dot{u}_D = A(q, u, t)\dot{u}_I + B(q, u, t), \tag{20}$$

where

$$A(q, u, t) := -J_2^{-1} J_1$$
$$B(q, u, t) := -J_2^{-1}[B_1(q, u, t)u + B_2(q, u, t)].$$

Equation (20) can be written in matrix form as

$$A_1(q, u, t)\dot{u} = B(q, u, t), \tag{21}$$

where

$$A_1 = [-A \quad I]. \tag{22}$$

2.2. HOLONOMIC VERSUS NONHOLONOMIC PARTIAL VELOCITIES AND PARTIAL ANGULAR VELOCITIES

It is convenient to write Equation (2) in the matrix form

$$^{\mathcal{R}}\mathbf{v}^P = {}^{\mathcal{R}}\mathbf{V}^P(q, t)u + {}^{\mathcal{R}}\mathbf{v}_t^P(q, t) \tag{23}$$

where

$$^{\mathcal{R}}\mathbf{V}^P = \left\lfloor {}^{\mathcal{R}}\mathbf{v}_1^P \ldots {}^{\mathcal{R}}\mathbf{v}_n{}^P \right\rfloor. \tag{24}$$

Hence, the acceleration of the particle relative to $\mathcal{R}$ is

$$^{\mathcal{R}}\mathbf{a}^P = {}^{\mathcal{R}}\mathbf{V}^P(q, t)\dot{u} + {}^{\mathcal{R}}\mathbf{a}_t^P, \tag{25}$$

where

$$^{\mathcal{R}}\mathbf{a}_t^P = {}^{\mathcal{R}}\dot{\mathbf{V}}^P u + {}^{\mathcal{R}}\dot{\mathbf{v}}_t^P. \tag{26}$$

Defining

$$ {}^{\mathcal{R}}\mathbf{V}_I{}^P = \left\lfloor {}^{\mathcal{R}}\mathbf{v}_1{}^P \ldots {}^{\mathcal{R}}\mathbf{v}_p{}^P \right\rfloor \tag{27} $$

$$ {}^{\mathcal{R}}\mathbf{V}_D{}^P = \left\lfloor {}^{\mathcal{R}}\mathbf{v}_{p+1}{}^P \ldots {}^{\mathcal{R}}\mathbf{v}_n{}^P \right\rfloor, \tag{28} $$

Equation (25) can be written as

$$ {}^{\mathcal{R}}\mathbf{a}^P = {}^{\mathcal{R}}\mathbf{V}_I^P(q,t)\dot{u}_I + {}^{\mathcal{R}}\mathbf{V}_D^P(q,t)\dot{u}_D + {}^{\mathcal{R}}\mathbf{a}_t^P \tag{29} $$

Substituting Equation (20) for $\dot{u}_D$ in Equation (29) gives

$$ \begin{aligned} {}^{\mathcal{R}}\mathbf{a}^P = &\left[{}^{\mathcal{R}}\mathbf{V}_I^P(q,t) + {}^{\mathcal{R}}\mathbf{V}_D^P(q,t)A(q,u,t)\right]\dot{u}_I \\ &+ {}^{\mathcal{R}}\mathbf{V}_D^P(q,t)B(q,u,t) + {}^{\mathcal{R}}\mathbf{a}_t^P. \end{aligned} \tag{30} $$

Also, it is convenient to write Equation (4) in the matrix form

$$ {}^{\mathcal{R}}\mathbf{v}^P = {}^{\mathcal{R}}\tilde{\mathbf{V}}^P(q,t)u_I + {}^{\mathcal{R}}\tilde{\mathbf{v}}_t^P(q,t), \tag{31} $$

where ${}^{\mathcal{R}}\tilde{\mathbf{V}}^P$ is the row matrix containing the nonholonomic partial velocities

$$ {}^{\mathcal{R}}\tilde{\mathbf{V}}^P = \left\lfloor {}^{\mathcal{R}}\tilde{\mathbf{v}}_1^P \ldots {}^{\mathcal{R}}\tilde{\mathbf{v}}_p^P \right\rfloor. \tag{32} $$

Differentiating Equation (31) with respect to time in $\mathcal{R}$ gives

$$ {}^{\mathcal{R}}\mathbf{a}^P = {}^{\mathcal{R}}\tilde{\mathbf{V}}^P(q,t)\dot{u}_I + \frac{{}^{\mathcal{R}}\mathrm{d}[{}^{\mathcal{R}}\tilde{\mathbf{V}}^P(q,t)]}{\mathrm{d}t}u_I + \frac{{}^{\mathcal{R}}\mathrm{d}\left[{}^{\mathcal{R}}\tilde{\mathbf{v}}_t^P(q,t)\right]}{\mathrm{d}t}. \tag{33} $$

Comparing the coefficients of $\dot{u}_I$ in Equation (30) and (33) gives the relations between the holonomic and the nonholonomic partial velocities of a particle in the system as

$$ {}^{\mathcal{R}}\tilde{\mathbf{v}}_r^P = {}^{\mathcal{R}}\mathbf{v}_r^P + \sum_{s=1}^{n-p} {}^{\mathcal{R}}\mathbf{v}_{p+s}^P A_{sr}(q,u,t), \quad r = 1,\ldots,p. \tag{34} $$

In a similar manner, the relations between the holonomic and the nonholonomic partial angular velocities of a body in the system is found by comparing the coefficients of $\dot{u}_I$ in the two expressions of the angular acceleration ${}^{\mathcal{R}}\alpha^B$ obtained by taking the time derivatives of ${}^{\mathcal{R}}\boldsymbol{\omega}^B$, as given in Equation (7) and (4).

$$ {}^{\mathcal{R}}\tilde{\boldsymbol{\omega}}_r^B = {}^{\mathcal{R}}\boldsymbol{\omega}_r^B + \sum_{s=1}^{n-p} {}^{\mathcal{R}}\boldsymbol{\omega}_{p+s}^B A_{sr}(q,u,t), \quad r = 1,\ldots,p. \tag{35} $$

Remark. The relation between $\dot{u}_I$ and $\dot{u}_D$ given by Equation (20) is similar to the relation between u_I and u_D in a simple nonholonomic system [23], except that the matrices A and B are functions of u. This yields relations between the holonomic and the nonholonomic partial velocities and partial angular velocities for nonlinearly constrained nonholonomic systems that are similar to their relations in a simple nonholonomic system, except that the matrix A is a function of u also, as given by Equation (34) and (35).

2.3. GENERALIZED ACTIVE AND INERTIA FORCES

Equations (34) and (35) can be used to represent the nonholonomic generalized active and inertia forces (6) and (8) in terms of the holonomic generalized active and inertia forces. Omitting the arguments for simplicity, these relations become

$$\tilde{F}_r = F_r + \sum_{s=1}^{n-p} F_{p+s} A_{sr} \tag{36}$$

$$\tilde{F}_r^\star = F_r^\star + \sum_{s=1}^{n-p} F_{p+s}^\star A_{sr}, \quad r = 1, \ldots, p. \tag{37}$$

Therefore, Equation (9) can be written as

$$F_r + F_r^\star + \sum_{s=1}^{n-p} (F_{p+s} + F_{p+s}^\star) A_{sr} = 0, , \quad r = 1, \ldots, p. \tag{38}$$

or in matrix form as

$$A_2 F^\star = -A_2 F, \tag{39}$$

where

$$A_2 := [I \quad A^T]. \tag{40}$$

The accelerations and angular accelerations are linear in $\dot{u}$; it follows that the generalized inertia forces are as well. Consequently, $F^\star$ can be written in the form

$$F^\star = -Q(q,t)\dot{u} - L(q,u,t), \tag{41}$$

where Q is a symmetric positive definite matrix. Then, Equation (39) becomes

$$A_2(q,u,t) Q(q,t) \dot{u} = A_2 P(q,u,t), \tag{42}$$

where

$$P(q, u, t) = -L(q, u, t) + F(q, u, t), \tag{43}$$

and Q is the generalized inertia matrix of the system.

Remark. Expanding the velocities and the angular velocities of the nonholonomic system $\mathcal{S}$ components in terms of the n generalized speeds allows one to define quantities that are related to the corresponding holonomic system, i.e. the system obtained by removing the nonholonomic constraints. This is crucial for the present development, as it permits construction of equations of motion for the nonholonomic system from those of its holonomic counterpart.

2.4. NONMINIMAL SYSTEM OF EQUATIONS

Equations (21) and (42) can be used to form the matrix system

$$T\dot{u} = V, \tag{44}$$

where $T := [A_1^T \; [A_2 Q]^T]^T$, and $V := \lfloor B^T \; [A_2 P]^T \rfloor^T$. The matrix T is a *constrained generalized inertia matrix* for the nonholonomic system $\mathcal{S}$. It is invertible for all choices of generalized coordinates and generalized speeds that render the elements of the constraint matrix A finite, i.e. render the matrix J_2 invertible. To show this, it is noticed that the row spaces of A_1 and A_2 are orthogonal complements. That is, both matrices are full row ranks, and $A_1 A_2^T = 0$. The row space of A_2 is unaltered if the rows of A_2 are scaled by scalars. Therefore, T is full rank if the holonomic system (41) is diagonal, i.e. the inertia matrix Q is diagonal. This diagonalization is possible by a proper choice of generalized speeds, and can be performed starting from an arbitrary choice of generalized speeds, by a Graham–Schmidt orthogonalization of the corresponding partial velocities [27]. The invertibility of T for this special choice of generalized speeds, denoted say by w, implies the invertibility of T for any other choice of generalized speeds. This can be seen by equating the right sides of the equations

$$\dot{q} = C_1(q, t)w + D_1(q, t) \tag{45}$$

$$\dot{q} = C_2(q, t)u + D_2(q, t) \tag{46}$$

which gives

$$w = \Phi_1(q, t)u + \Phi_2(q, t) \tag{47}$$

where

$$\Phi_1 = C_1^{-1} C_2 \tag{48}$$

$$\Phi_2 = C_1{}^{-1}(D_2 - D_1). \tag{49}$$

Equation (47) is a unique invertible transformation between the two sets of generalized speeds, which implies the equivalency of the existence of solution for one set and the existence of solution for the other. Therefore,

$$\dot{u} = T^{-1} V. \tag{50}$$

The above aggregation of the dynamical equations and the constraint equations can also be found in [18], where a nonminimal form is derived with the aid of differential geometry concepts.

Remark. The appearance of the constraint matrix A in the dynamical equations (39) as well as in the constraint equations (21) exploits the feature of deriving the nonminimal form of equations for a nonholonomic system by simple manipulations of the equations of motion for the corresponding holonomic system.

The two sets of ordinary differential Equations (1) and (50) form a complete separated-in-accelerations state space model for the constrained dynamical system, and involves no reduction in the dimension of the space of generalized speeds from the number of generalized coordinates to the number of degrees of freedom. Furthermore, this is obtained without employing Lagrange multipliers. Therefore, it enables the use of system analysis and control techniques that are related to state space model representation, in a unified treatment of holonomic and nonholonomic constraints. This complements the previous differential algebraic equations (DAE) approach.

The procedure of using the acceleration form of constraints in obtaining a consistent set of separated in accelerations equations of motion for nonlinear nonholonomic systems is summarized as follows:

1. A set of generalized speeds satisfying Equation (1) is chosen, and the nonlinear nonholonomic constraints, Equation (10) are differentiated with respect to time. The set of generalized speeds is partitioned according to Equation (16), and the dependency among the set is described at the acceleration level by Equation (20). If holonomic constraints are involved, the corresponding equations are twice differentiated in time to appear in the same acceleration form.
2. The matrix A is used to construct the matrices A_1 and A_2, Equations (22) and (40).

3. Expressions are obtained for holonomic partial velocities/angular velocities by inspecting the corresponding expressions for linear/angular velocities, as the coefficients of the generalized speeds.
4. Holonomic generalized active and inertia forces are found from the scalar (dot) product of the impressed and gravitational forces with the holonomic partial velocities/angular velocities, and used together with A_2 to form Equation (42).
5. Equations (21) and (42) are used to form the matrix equation, Equation (44), and T is inverted to yield the resulting equations of motion (50).

Example 1: The Appell–Hamel problem. The mechanism shown in Figure 1 consists of a frame with two legs that slide without friction on the x–y plane and supports two massless pulleys that are a distance ρ apart. A thread is passed around the pulleys, hanging a weight P that is idealized as a particle of mass m, and its movement is restricted to be along the vertical bar of the frame. The thread is wound around a drum of radius b, which is fixed to a wheel W of radius a, mass M, mass center W^*. The wheel rolls on the x–y plane, where ϕ is its angle of rotation in its

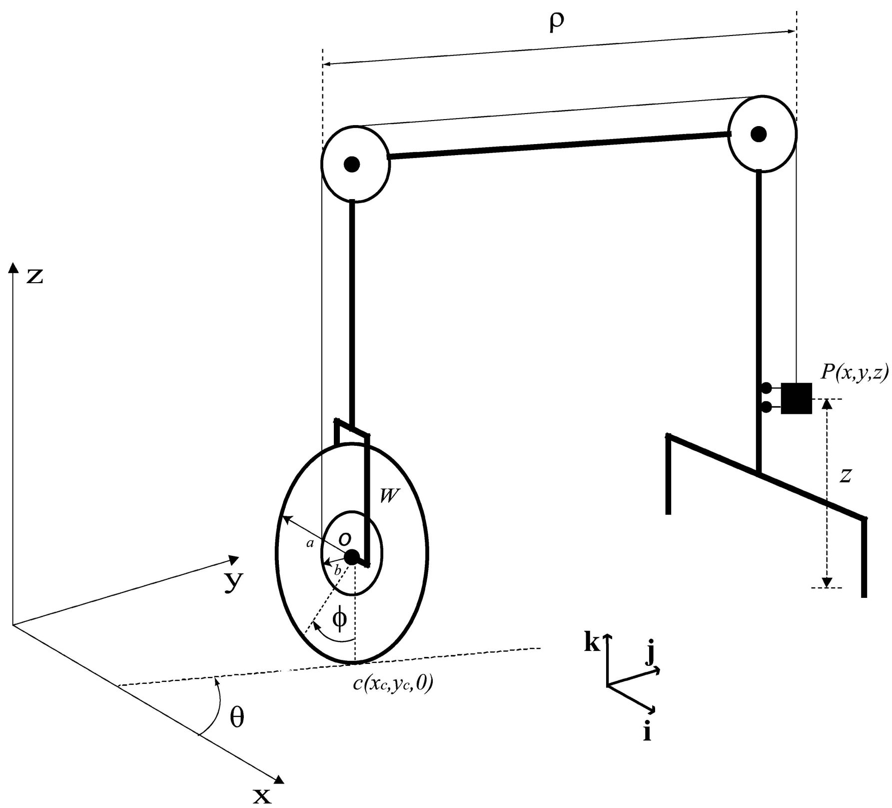

Figure 1. Schematic for the Appell–Hamel mechanism.

own plane. For simplicity, it is specified that the wheel has equal axial and polar moments of inertia, I. The plane of W makes the angle θ with the x axis, and the frame keeps it vertical relative to the x–y plane. Let x, y, z be the coordinates of the center of mass of P in the xyz coordinate system, which is fixed to an inertial frame of reference. The configuration parameters can be chosen as x, y, z, θ, and ϕ. Finally, let $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ be unit vectors parallel to the positive x, y, and z directions, respectively, and let $\mathbf{i}_w$, $\mathbf{j}_w$, and $\mathbf{k}_w$ be wheel-fixed unit vectors parallel to $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$, respectively when $\theta = 0$ and $\phi = 0$. The no slip condition of W on the plane xy gives rise to two relations that describe the velocity of the center of the wheel o,

$$\dot{x}_{\mathrm{o}} = a\dot{\phi}\cos\theta \tag{51}$$

$$\dot{y}_{\mathrm{o}} = a\dot{\phi}\sin\theta. \tag{52}$$

The velocity of o can also be described in terms of the velocity of P by the relations

$$\dot{x}_{\mathrm{o}} = \dot{x} + \rho\dot{\theta}\sin\theta \tag{53}$$

$$\dot{y}_{\mathrm{o}} = \dot{y} - \rho\dot{\theta}\cos\theta. \tag{54}$$

The relations (51) and (52) can be manipulated in order to create the nonlinear nonholonomic constraint Equation [5]

$$\dot{x}_{\mathrm{o}}^2 + \dot{y}_{\mathrm{o}}^2 = a^2\dot{\phi}^2, \tag{55}$$

and the linear nonholonomic constraint equation

$$\dot{x}_{\mathrm{o}}\sin\theta - \dot{y}_{\mathrm{o}}\cos\theta = 0. \tag{56}$$

Substituting Equations (53) and (54) into Equations (55) and (56) yields

$$(\dot{x} + \rho\dot{\theta}\sin\theta)^2 + (\dot{y} - \rho\dot{\theta}\cos\theta)^2 - a^2\dot{\phi}^2 = 0, \tag{57}$$

and

$$\dot{x}\sin\theta - \dot{y}\cos\theta + \rho\dot{\theta} = 0. \tag{58}$$

The inextensibility of the thread gives rise to the holonomic constraint equation

$$z = -b\phi + z_0, \tag{59}$$

where z_0 is a constant. Hence, the system has two degrees of freedom. Considering the generalized speeds $u_1 = \dot{\theta}$, $u_2 = \dot{\phi}$, $u_3 = \dot{x}$, $u_4 = \dot{y}$, $u_5 = \dot{z}$, and taking the time derivatives of the constraint Equations (57)–(59), the acceleration form of the constraint equations is

$$
\begin{aligned}
&(u_3 + \rho u_1 \sin\theta)\big(\dot{u}_3 + \rho\dot{u}_1 \sin\theta + \rho u_1^2 \cos\theta\big) \\
&\quad + (u_4 - \rho u_1 \cos\theta)\big(\dot{u}_4 - \rho\dot{u}_1 \cos\theta + \rho u_1^2 \sin\theta\big) = a^2 u_2 \dot{u}_2 &\quad (60)\\
&\dot{u}_3 \sin\theta + u_1 u_3 \cos\theta - \dot{u}_4 \cos\theta + u_1 u_4 \sin\theta + \rho\dot{u}_1 = 0 &\quad (61)\\
&\dot{u}_5 = -b\dot{u}_2. &\quad (62)
\end{aligned}
$$

Let

$$u_I = \lfloor u_1 \quad u_2 \rfloor \tag{63}$$

$$u_D = \lfloor u_3 \quad u_4 \quad u_5 \rfloor, \tag{64}$$

then the matrices J_1 and J_2 for the system are

$$J_1 = \begin{bmatrix} n_1 & n_2 \\ \rho & 0 \\ 0 & b \end{bmatrix} \tag{65}$$

$$J_2 = \begin{bmatrix} n_3 & n_4 & 0 \\ \sin\theta & -\cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{66}$$

where

$$n_1 = \rho^2 u_1 + \rho u_3 \sin\theta - \rho u_4 \cos\theta \tag{67}$$

$$n_2 = -a^2 u_2 \tag{68}$$

$$n_3 = u_3 + \rho u_1 \sin\theta \tag{69}$$

$$n_4 = u_4 - \rho u_1 \cos\theta. \tag{70}$$

The matrices A and B in Equation (20) for the system are

$$A(q, u, t) = \frac{1}{n_5} \begin{bmatrix} -\rho n_4 - n_1 \cos\theta & -n_2 \cos\theta \\ \rho n_3 - n_1 \sin\theta & -n_2 \sin\theta \\ 0 & -n_5 b \end{bmatrix} \tag{71}$$

and

$$B(q, u, t) = \frac{n_6}{n_5} \begin{bmatrix} -u_1(n_4 + \rho u_1 \cos\theta) \\ u_1(n_3 - \rho u_1 \sin\theta) \\ 0 \end{bmatrix}, \tag{72}$$

where

$$n_5 = n_3 \cos\theta + n_4 \sin\theta \tag{73}$$

$$n_6 = u_3 \cos\theta + u_4 \sin\theta. \tag{74}$$

Therefore, the matrices A_1 and A_2 are

$$A_1 = \begin{bmatrix} \frac{\rho n_4 + n_1 \cos\theta}{n_5} & \frac{n_2 \cos\theta}{n_5} & 1 & 0 & 0 \\ \frac{-\rho n_3 + n_1 \sin\theta}{n_5} & \frac{n_2 \sin\theta}{n_5} & 0 & 1 & 0 \\ 0 & b & 0 & 0 & 1 \end{bmatrix} \tag{75}$$

and

$$A_2 = \begin{bmatrix} 1 & 0 & \frac{-\rho n_4 - n_1 \cos\theta}{n_5} & \frac{\rho n_3 - n_1 \sin\theta}{n_5} & 0 \\ 0 & 1 & \frac{-n_2 \cos\theta}{n_5} & \frac{-n_2 \sin\theta}{n_5} & -b \end{bmatrix}. \tag{76}$$

The inertial velocity of P is

$$\mathbf{v}^P = u_3\mathbf{i} + u_4\mathbf{j} + u_5\mathbf{k}, \tag{77}$$

and its inertial acceleration is

$$\mathbf{a}^P = \dot{u}_3\mathbf{i} + \dot{u}_4\mathbf{j} + \dot{u}_5\mathbf{k}. \tag{78}$$

The applied force on P is

$$\mathbf{F}_P = -mg\mathbf{k}, \tag{79}$$

where g is the gravitational constant. Hence, the generalized active forces on P are contained in the column matrix

$$F_P = \lfloor 0 \quad 0 \quad 0 \quad 0 \quad -mg \rfloor^T. \tag{80}$$

Similarly, the generalized inertia forces are contained in the column matrix

$$F_P^\star = \lfloor 0 \quad 0 \quad -m\dot{u}_3 \quad -m\dot{u}_4 \quad -m\dot{u}_5 \rfloor^T. \tag{81}$$

Since the applied forces acting on W are all in the vertical direction, they do not contribute to the generalized active forces. The inertial angular velocity of W is

$$\boldsymbol{\omega}^W = -u_2 \sin\theta\mathbf{i} + u_2 \cos\theta\mathbf{j} + u_1\mathbf{k}, \tag{82}$$

and its angular acceleration is

$$\boldsymbol{\alpha}^W = (-\dot{u}_2 \sin\theta - u_1 u_2 \cos\theta)\,\mathbf{i} + (\dot{u}_2 \cos\theta - u_1 u_2 \sin\theta)\,\mathbf{j} + \dot{u}_1 \mathbf{k}. \tag{83}$$

The velocity of the center of mass of the wheel is

$$\mathbf{v}^{\mathrm{o}} = (u_3 + \rho u_1 \sin\theta)\mathbf{i} + (u_4 - \rho u_1 \cos\theta)\mathbf{j}, \tag{84}$$

and its acceleration is

$$\mathbf{a}^{\mathrm{o}} = \big(\dot{u}_3 + \rho \dot{u}_1 \sin\theta + \rho u_1^2 \cos\theta\big)\mathbf{i} + \big(\dot{u}_4 - \rho \dot{u}_1 \cos\theta + \rho u_1^2 \sin\theta\big)\mathbf{j}. \tag{85}$$

The generalized inertia forces of the wheel are given by

$$F^{\star}_{Wr} = \mathbf{F}^{\star}_{\mathrm{o}} \cdot \mathbf{v}^{\mathrm{o}}_r + \mathbf{T}^{\star}_W \cdot \boldsymbol{\omega}^W_r, \tag{86}$$

where the inertia force $\mathbf{F}^{\star}_{\mathrm{o}}$ is

$$\mathbf{F}^{\star}_{\mathrm{o}} = -M\mathbf{a}^{\mathrm{o}}, \tag{87}$$

and the inertia torque $\mathbf{T}^{\star}_W$ is

$$\mathbf{T}^{\star}_W = -\boldsymbol{\alpha}^W \cdot \underline{\mathrm{I}}^W - \boldsymbol{\omega}^W \times \underline{\mathrm{I}}^W \cdot \boldsymbol{\omega}^W. \tag{88}$$

Here, $\underline{\mathrm{I}}^W$ denotes the central inertia dyadic of W [23]. The relation between the wheel-fixed and the inertial frame-fixed unit vectors is given by:

$$\begin{Bmatrix} \mathbf{i}_w \\ \mathbf{j}_w \\ \mathbf{k}_w \end{Bmatrix} = \begin{bmatrix} \cos\phi\cos\theta & \cos\phi\sin\theta & -\sin\phi \\ -\sin\theta & \cos\theta & 0 \\ \sin\phi\cos\theta & \sin\phi\sin\theta & \cos\phi \end{bmatrix} \begin{Bmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{Bmatrix}. \tag{89}$$

Hence,

$$\underline{\mathrm{I}}^W = I(\mathbf{i}_w\mathbf{i}_w + \mathbf{j}_w\mathbf{j}_w + \mathbf{k}_w\mathbf{k}_w) \tag{90}$$

$$= I(\mathbf{ii} + \mathbf{jj} + \mathbf{kk}), \tag{91}$$

and the inertia torque is,

$$\mathbf{T}^{\star}_W = -I\dot{u}_1\mathbf{k} - I\dot{u}(\cos\theta\mathbf{j} - \sin\theta\mathbf{i}) - u_1u_2I(-\cos\theta\mathbf{i} - \sin\theta\mathbf{j}). \tag{92}$$

Therefore, the contribution of the wheel to the generalized inertia forces is given by

$$F^{\star}_{W1} = -M(\rho^2\dot{u}_1 + \rho\dot{u}_3\sin\theta - \rho\dot{u}_4\cos\theta) - I\dot{u}_1 \tag{93}$$
$$F^{\star}_{W2} = -I\dot{u}_2 \tag{94}$$
$$F^{\star}_{W3} = -M\big(\dot{u}_3 + \rho\dot{u}_1\sin\theta + \rho u_1^2\cos\theta\big) \tag{95}$$
$$F^{\star}_{W4} = -M\big(\dot{u}_4 - \rho\dot{u}_1\cos\theta + \rho u_1^2\sin\theta\big) \tag{96}$$
$$F^{\star}_{W5} = 0. \tag{97}$$

The generalized inertia forces for the system are given by

$$F^{\star} = F^{\star}_P + F^{\star}_W. \tag{98}$$

Therefore, with the above mentioned choice of generalized speeds,

$$Q = \begin{bmatrix} M\rho^2 + I & 0 & M\rho\sin\theta & -M\rho\cos\theta & 0 \\ 0 & I & 0 & 0 & 0 \\ M\rho\sin\theta & 0 & M+m & 0 & 0 \\ -M\rho\cos\theta & 0 & 0 & M+m & 0 \\ 0 & 0 & 0 & 0 & m \end{bmatrix} \tag{99}$$

$$P = \begin{Bmatrix} 0 \\ 0 \\ -M\rho u_1^2\cos\theta \\ -M\rho u_1^2\sin\theta \\ -mg \end{Bmatrix}. \tag{100}$$

Forming Equations (21) and (42) for this system and augmenting the two equations yields Equation (44). $\dot{u}$ can be obtained by inverting the coefficient matrix $T := [A_1^T[A_2Q]^T]^T$. This can be done for all values of generalized coordinates and generalized speeds that give nonzero values of n_5.

The inversion of the matrix T can be done either numerically or symbolically. The symbolic inversion results in lengthy expressions that are not needed for our purpose. The time simulations must run with initial conditions that satisfy the constraint equations. These are chosen to be $\dot{\theta} = \dot{\phi} = 1.0$ rad/s, $\dot{x} = 1.0$ m/s, $\dot{y} = 5.0$ m/s, $\dot{z} = -0.5$ m/s, $z = 30$ m, and zero initial conditions for the remaining generalized coordinates. Time simulations are performed with $a = 1.0$ m, $b = 0.5$ m, $\rho = 5.0$ m, $m = 1.0$ kg, $M = 5.0$ kg, $z_0 = 30$ m. Figures 2 and 3 show the responses of θ and ϕ, respectively. The responses of the time rates of change of these angles, $\dot{\theta}$ and $\dot{\phi}$ are shown in Figures 4 and 5, respectively. The angle θ tends

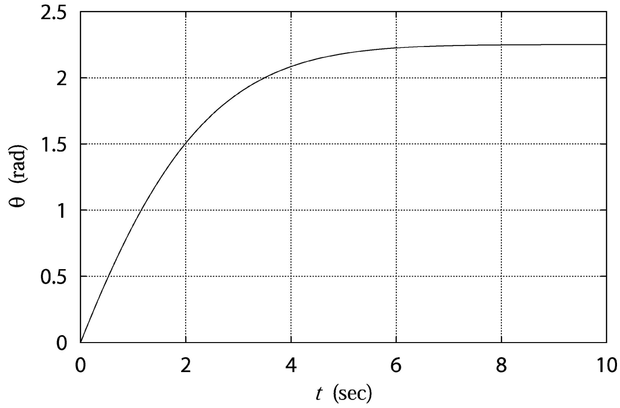

Figure 2. Example 1: θ vs. t.

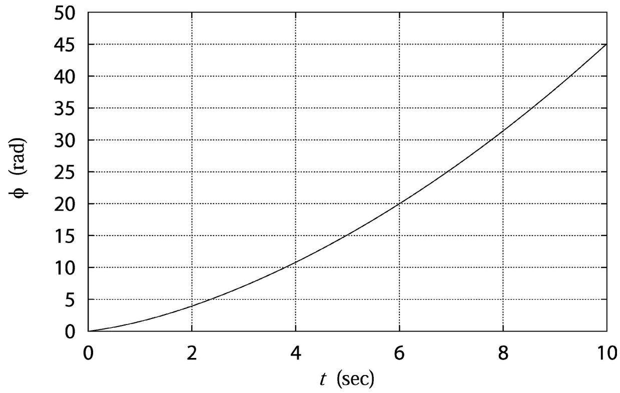

Figure 3. Example 1: ϕ vs. t.

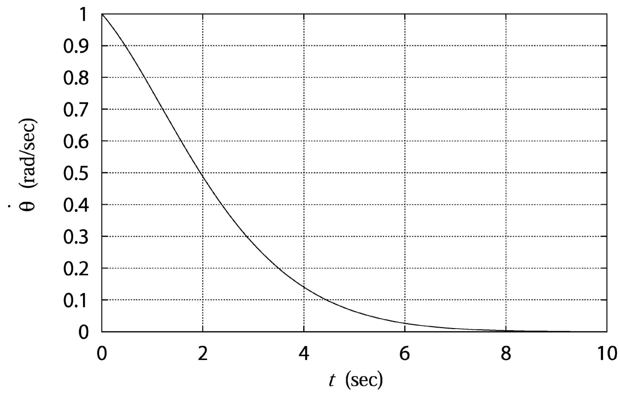

Figure 4. Example 1: $\dot{\theta}$ vs. t.

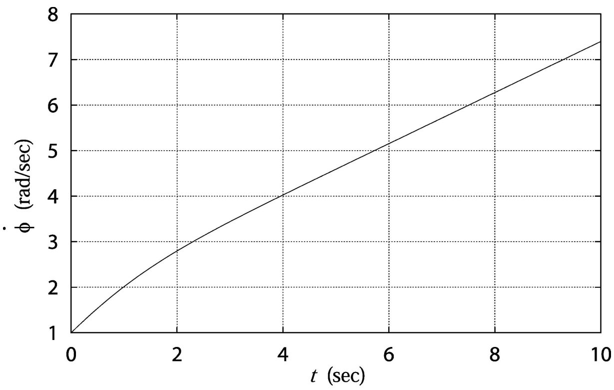

Figure 5. Example 1: $\dot{\phi}$ vs. t.

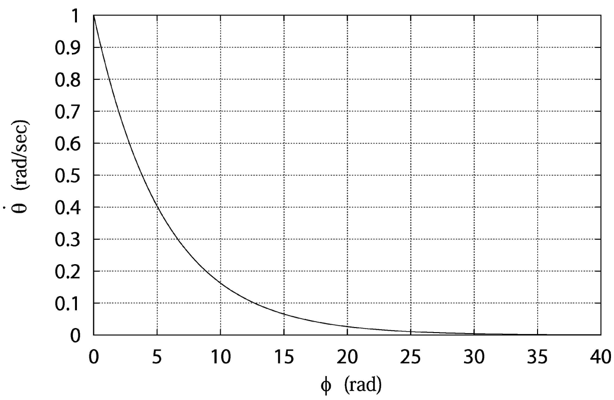

Figure 6. Example 1: $\dot{\theta}$ vs. ϕ.

to reach a constant steady state value as the wheel continues to roll over the x–y plane, as shown in Figure 6. Also, the load P intercept on the x–y plane is shown in Figure 7, and the time history of its height z is shown in Figure 8.

2.5. NONLINEAR NONHOLONOMIC CONSTRAINTS AND NUMERICAL STABILITY OF THE EQUATIONS OF MOTION

The problem of numerical drift of constraints and integrals of motion is well known in the solutions of differential equations subjected to constraints. Several methods have been introduced to remedy this problem. Every method has its advantages and disadvantages, but all these methods involve modifications to the dynamical equations in order to suppress the numerical violation. Stabilizing the constraint equations and the dynamical equations are not independent matters, and one should be careful when implementing a constraint stabilizing scheme, as the modification can alter the dynamics of the whole system in addition to its effect on the constraint

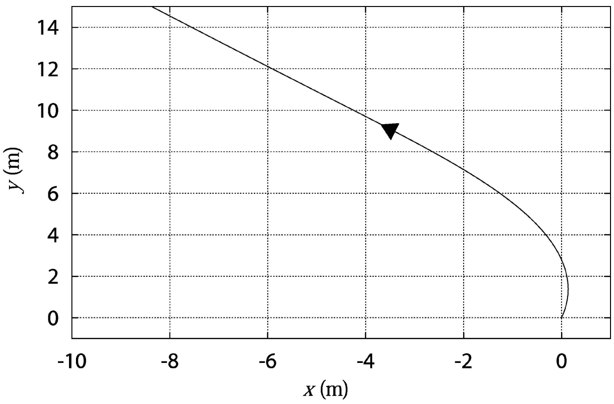

Figure 7. Example 1: y vs. x.

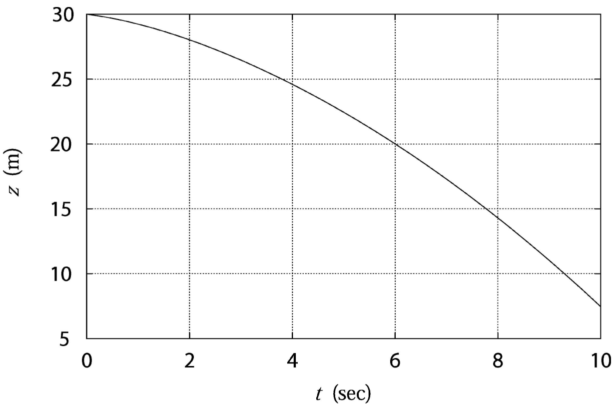

Figure 8. Example 3.1: z vs. t.

dynamics. Alternatively, it is possible to consider the issue of numerical stability during the modeling phase, to avoid the need to correct the motion by modifying the already formulated equations of motion.

Modeling the constraints to be nonlinear in the velocities results in equations of motion that are different in appearance from the equations of motion resulting from modeling the constraints as linear in the velocities. However, the solutions of the resulting equations of motion and the time simulations should not be different, irrespective of the way the constraint equations are manipulated in order to be augmented with the dynamical equations. Nevertheless, the numerical stability of the solution is certainly affected by the constraint modeling. In that regard, it can be beneficial to use the nonlinearity of the constraint equations as a passive tool to suppress the numerical errors. To illustrate that, we create the linear nonholonomic constraint equations by equating Equations (51) and (52) with Equations (53) and

(54). The resulting equations are

$$a\dot{\phi}\cos\theta - \dot{x} - \rho\dot{\theta}\sin\theta = 0 \tag{101}$$

$$a\dot{\phi}\sin\theta - \dot{y} + \rho\dot{\theta}\cos\theta = 0. \tag{102}$$

Taking the time derivatives of Equations (101) and (102) , the same procedure can be used to obtain the equations of motion for the Appell–Hamel mechanism with the constraints modeled as linear nonholonomic. By running the time simulations for both systems of equations, a common numerical violation measure can be tested, that is the total energy E of the mechanism. Considering the x–y plane as the datum for computing the potential energy, E is given as

$$E = \frac{I}{2}\left(u_1^2 + u_2^2\right) + \frac{M}{2}[(u_3 + \rho u_1 \sin\theta)^2 + (u_4 - \rho u_1 \cos\theta)^2] + \frac{M}{2}ga + \frac{m}{2}\left(u_3^2 + u_4^2 + u_5^2\right) + mgz. \tag{103}$$

Figure 9 shows the plots of E by using the state variables obtained from integrating the equations of motion that correspond to the two types of constraint modeling, where ΔE is the difference between the computed value of the energy and its initial value. It is noticed that the nonlinearity in the constraint equations subdues the growing deviation in the total energy of the mechanism.

Nevertheless, the error dynamics for nonlinear systems depends on the initial errors in the state variables, and on the input forces on the system. These can vary substantially during the simulation process, and manipulating the constraint equations provides no guarantee of error convergence, which frequently necessitates an implementation of a corrective scheme. It is noticed also that when the constraints are modeled as nonlinear nonholonomic then the inversion of the resulting

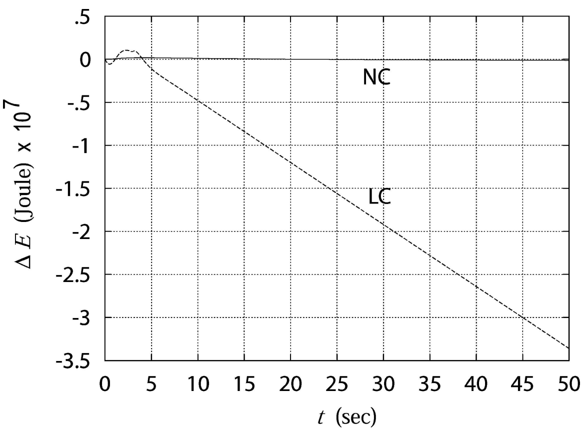

Figure 9. Example 1: Energy integral, LC: linear constraints, NC: nonlinear constraints.

constrained generalized inertia matrix substantially increases the complexity of the equations of motion. For that reason, this artificial nonlinearity of constraints should be used only when the purpose is to increase robustness of numerical simulations. The equivalent equations of motion that are obtained by modeling the constraints as linear nonholonomic are more suitable for analysis.

In the next section, we derive a version of the nonminimal equations that is free of the constraint drift phenomenon.

3. Elimination of Constraint Drift

The problem of constraint drift is generally unavoidable when a differential form of the constraint equations is used to formulate the equations of motion. Such a problem becomes more serious for the case in which the acceleration form of holonomic constraints is enforced, since two integrations are needed at each time step to obtain the generalized coordinates. The purpose of this section is to modify the derived equations of motion to suppress the errors resulting from this integration process. The explicit use of the acceleration form of constraint equations suggests the employment of the classical numerical stabilization method by Baumgarte [24]. Let ϕ be the set of m nonholonomic constraint equations. Instead of using the acceleration form of the constraint equations, Equation (13), the equations of Baumgarte type

$$\dot{\phi}(q, u, \dot{u}, t) - \Gamma\phi(q, u, t) = 0, \tag{104}$$

are considered. Here $\Gamma \in \mathbb{R}^{m\times m}$ is a matrix that has eigenvalues with strictly negative real parts. Using Equation (19) in Equation (104), one obtains

$$J_1\dot{u}_I + J_2\dot{u}_D - \Gamma\phi(q, u, t) + B_1(q, t)u + B_2(q, t) = 0. \tag{105}$$

Solving for $\dot{u}_D$ yields

$$\dot{u}_D = A(q, t)\dot{u}_I + \hat{B}(q, u, t), \tag{106}$$

where

$$A(q, u, t) = -J_2^{-1}J_1 \tag{107}$$

$$\hat{B}(q, u, t) = B(q, u, t) + J_2^{-1}\Gamma\phi(q, u, t). \tag{108}$$

Equation (106) can be written as

$$A_1(q, u, t)\dot{u} = \hat{B}(q, u, t), \tag{109}$$

which, together with Equation (42), form the matrix system

$$T\dot{u} = \hat{V}, \tag{110}$$

where $\hat{V} := \lfloor \hat{B}^T \ [A_2 P]^T \rfloor^T$. Therefore,

$$\dot{u} = T^{-1}\hat{V}. \tag{111}$$

A similar treatment can be developed for holonomic constraint equations. Instead of the acceleration form of the constraint equations, the following constraint equations are used:

$$\ddot{\phi}(q, u, \dot{u}, t) - \Gamma_1\dot{\phi}(q, u, t) - \Gamma_2\phi(q, t) = 0, \tag{112}$$

where Γ_1 and Γ_2 are chosen such that the dynamics of Equation (112) is stable. In this case, $\hat{B}$ becomes

$$\hat{B}(q, u, t) = B(q, u, t) + J_2^{-1}[\Gamma_1\dot{\phi}(q, u, t) + \Gamma_2\phi(q, t)]. \tag{113}$$

Remark. The matrices Γ, Γ_1, Γ_2 can be thought of as feedback gains of a control system that is aimed to regulate the constraint functions ϕ at the zero value. In order to obtain from the numerical integration scheme a true and accurate state of the dynamical system, these gains must also keep the entire system of Equation (111) stable. The choice of the gain matrices can affect the stability of Equation (111), beside its effect on the convergence rate of ϕ. For adaptive choices of gain matrices for Baumgarte type of constraint violation stabilization, the reader is referred to [28].

The procedure for deriving nonminimal form of Kane's equations of motion that is free of constraint drift for dynamical systems subjected to nonlinear nonholonomic constraints is summarized as follows:

1. Stable constraint dynamics equations are constructed by augmenting the constraint functions ϕ with their differentiated forms by means of the matrix Γ in case ϕ is nonholonomic, resulting in Equation (104). In case ϕ is holonomic, both $\dot{\phi}$ and $\ddot{\phi}$ are augmented with ϕ by means of Γ_1 and Γ_2, resulting in Equation (112). In both cases, these matrices are chosen to damp out any nonzero values of ϕ.
2. A generalized speeds partitioning according to Equation (16) is used to put the equations in the form (106), which results in the upper subsystem of the unreduced form, Equation (109).
3. Equation (109) is used with Equation (42) to form the system of Equation (110), which can be solved for $\dot{u}$ by inverting the matrix T to obtain Equation (111).

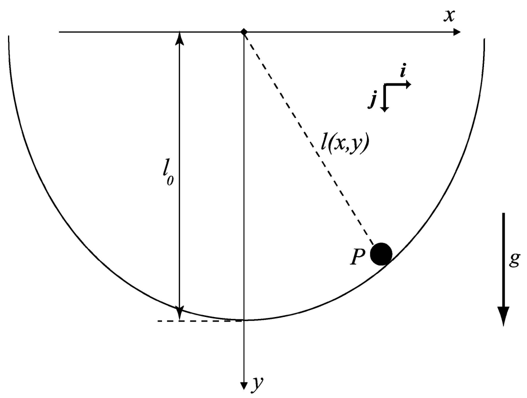

Figure 10. Schematic for Example 2.

The following example illustrates the procedure for numerically stabilizing a holonomic constraint equation.

Example 2: Sliding Particle. A particle is sliding over a surface is shown in Figure 10. The surface and the x-y coordinate system are fixed in an inertial reference frame $\mathcal{R}$. The surface defines the holonomic constraint equation

$$\phi(x, y) = y + \frac{x^2}{l_0} - l_0 = 0, \tag{114}$$

where l_0 is a positive constant equal to $l(x, y) = \sqrt{(x^2 + y^2)}$ when either $x = 0$ or $y = 0$. Consider x and y as the generalized coordinates, and let the generalized speeds be $u_1 = \dot{x}$ and $u_2 = \dot{y}$. Also, let $\mathbf{i}$ and $\mathbf{j}$ be unit vectors in the x and y directions, respectively. The position vector of the particle P is

$$\mathbf{p} = x\mathbf{i} + y\mathbf{j}. \tag{115}$$

The inertial velocity of the particle is

$$^{\mathcal{R}}\mathbf{v}^P = u_1\mathbf{i} + u_2\mathbf{j}.$$

The velocity form of the constraint equation above is thus

$$\dot{\phi}(x, u) = u_2 + \frac{2}{l_0}xu_1 = 0,$$

and the acceleration form is

$$\ddot{\phi}(x, u, \dot{u}) = \dot{u}_2 + \frac{2}{l_0}\left(u_1^2 + x\dot{u}_1\right) = 0.$$

Choosing $u_I = u_1$,

$$\begin{aligned} A_1 &= \left\lfloor \tfrac{2}{l_0}x \quad 1 \right\rfloor, \\ A_2 &= \left\lfloor 1 \quad -\tfrac{2}{l_0}x \right\rfloor, \end{aligned} \tag{116}$$

and

$$B = -\frac{2}{l_0}u_1^2. \tag{117}$$

The unconstrained equations of motion are

$$\begin{aligned} \dot{u}_1 &= 0 \\ \dot{u}_2 &= g. \end{aligned} \tag{118}$$

Equation (44) for this system is

$$\frac{2}{l_0}x\dot{u}_1 + \dot{u}_2 = -\frac{2}{l_0}u_1^2 \tag{119}$$

$$\dot{u}_1 - \frac{2}{l_0}x\dot{u}_2 = -\frac{2}{l_0}gx. \tag{120}$$

Given the parameter $l_0 = 1.0$ m, solving for $\dot{u}_1$ and $\dot{u}_2$ yields,

$$\dot{u}_1 = \frac{-2x\left(g + 2u_1^2\right)}{1 + 4x^2} \tag{121}$$

$$\dot{u}_2 = \frac{4gx^2 - 2u_1^2}{1 + 4x^2}. \tag{122}$$

Setting the initial condition $x(0) = 1.0$ m, and using the Kutta-Merson numerical integration scheme, the above system of equations is solved for x, y, u_1, and u_2, with the constraint violation ϕ evaluated at each time step. The time history of ϕ after 500 s of simulation time is plotted for two small values of the time integration step Δt, as shown in Figures 11 and 12. It is clearly seen that ϕ grows with time. Both the pattern and the magnitude of ϕ are affected by the choice of Δt. Reducing Δt reduces the growth of ϕ, at the cost of increasing the required time to perform the simulation. The same is concluded for a constant of motion of this conservative system, namely the energy integral E. This is simply the sum of kinetic and potential energies of the system,

$$E = K + V = \frac{1}{2}m\left(u_1^2 + u_2^2\right) + mg(l_0 - y), \tag{123}$$

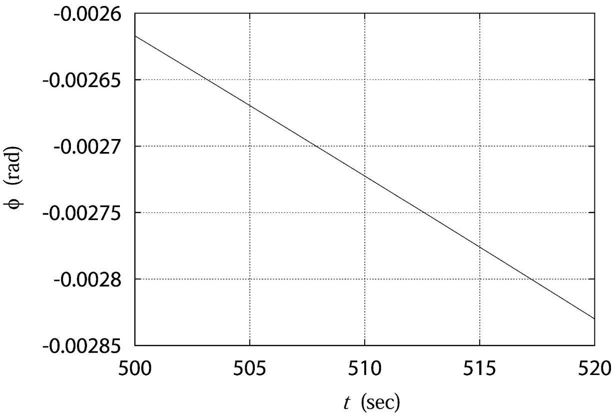

Figure 11. Example 2: Constraint violation, ϕ: $\Gamma_1 = \Gamma_2 = 0$, $\Delta t = 0.01$ s.

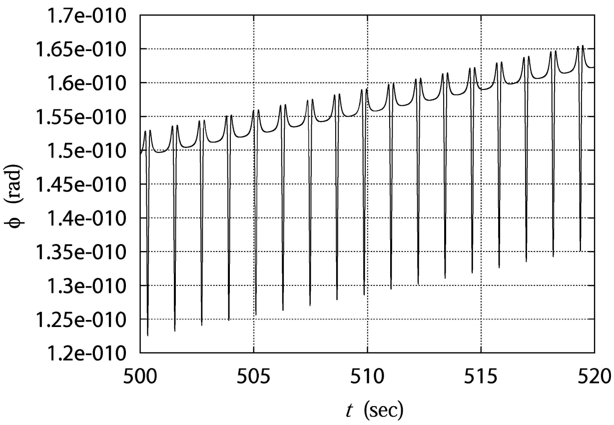

Figure 12. Example 2: Constraint violation, ϕ: $\Gamma_1 = \Gamma_2 = 0$, $\Delta t = 0.001$ s.

where the datum for computing the potential energy is chosen as $y_d = l_0$. For $m = 1.0$ kg and $l_0 = 1$. Figures 13 and 14 show the deviations in E for the two choices of Δt after 500 s of simulation time.

Next, Equation (112) for this system is used. The resulting Equation (110) with $l_0 = 1.0$ m becomes

$$
\begin{aligned}
2x\dot{u}_1 + \dot{u}_2 &= -2u_1^2 + \Gamma_1\dot{\phi}(x, u) + \Gamma_2\phi(x, y, u) \\
\dot{u}_1 - 2x\dot{u}_2 &= -2gx.
\end{aligned} \tag{124}
$$

Solving for $\dot{u}$ yields

$$
\dot{u}_1 = \frac{-2x\left(g + 2u_1^2\right)}{1 + 4x^2} + \frac{2x(u_2 + 2xu_1)}{1 + 4x^2}\Gamma_1 - \frac{2x(1 - y - x^2)}{1 + 4x^2}\Gamma_2 \tag{125}
$$

$$
\dot{u}_2 = \frac{4gx^2 - 2u_1^2}{1 + 4x^2} + \frac{u_2 + 2xu_1}{1 + 4x^2}\Gamma_1 - \frac{1 - y - x^2}{1 + 4x^2}\Gamma_2. \tag{126}
$$

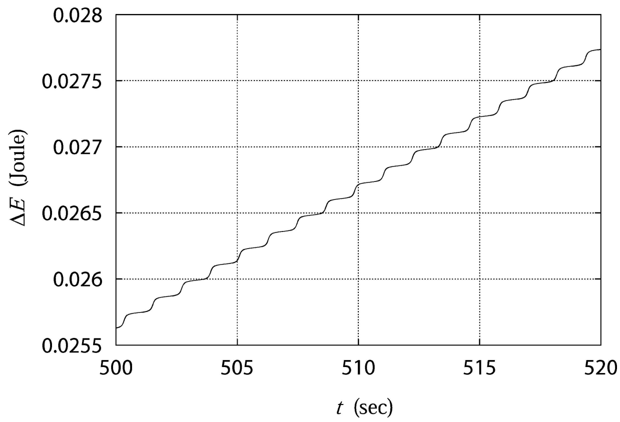

Figure 13. Example 2: Energy integral: $\Gamma_1 = \Gamma_2 = 0$, $\Delta t = 0.01$ s.

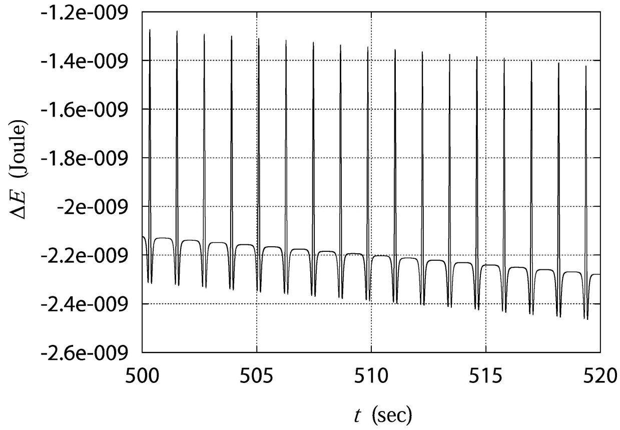

Figure 14. Example 2: Energy integral: $\Gamma_1 = \Gamma_2 = 0$, $\Delta t = 0.001$ s.

The time history of ϕ after 500 s is shown in Figure 15. It is noticed that the constraint violation becomes bounded during the time simulation period. Arbitrarily small bounds can be obtained by increasing the values of Γ_1 and Γ_2. However, this also increases the relative magnitudes of the damping terms, which results in an increase in the stiffness of the differential equations, and requires smaller time steps. For $\Gamma_1 = -20\,\mathrm{s}^{-1}$ and $\Gamma_2 = -100\,\mathrm{s}^{-2}$, a bound of $|\phi| < 2.0 \times 10^{-11}$ m is obtained for $\Delta t = 0.001$ s. The choice of Γ_1 and Γ_2 affects the numerical stability of the whole nonminimal system of equations, as discussed below. For the purpose of comparison, Kane's minimal equation may be derived. The solution of this equation is free from the constraint drift, because the constraint equation is used in its algebraic form. Substituting y from Equation (114) into Equation (115)

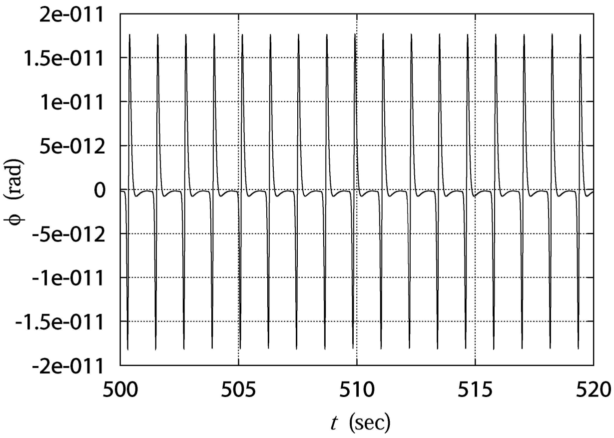

Figure 15. Example 2: Constraint violation, ϕ: $\Gamma_1 = -20$, $\Gamma_2 = -100$, $\Delta t = 0.001$ s.

yields

$$\mathbf{p} = x\mathbf{i} + (1 - x^2)\mathbf{j}. \tag{127}$$

The first and second time derivatives of the above equation relative to $\mathcal{R}$ are the inertia velocity and acceleration of P, represented in terms of u_1 and $\dot{u}_1$,

$$\begin{aligned} {}^{\mathcal{R}}\mathbf{v}^P &= u_1\mathbf{i} + (1 - 2xu_1)\mathbf{j} && (128)\\ &= u_1(\mathbf{i} - 2x\mathbf{j}), && (129) \end{aligned}$$

and

$$^{\mathcal{R}}\mathbf{a}^P = \dot{u}_1\mathbf{i} - \left(2u_1^2 + 2x\dot{u}_1\right)\mathbf{j}. \tag{130}$$

The coefficient of u_1 in Equation (129) is the holonomic partial velocity of P,

$$^{\mathcal{R}}\mathbf{v}_1^P = \mathbf{i} - 2x\mathbf{j}. \tag{131}$$

The holonomic generalized active force F on the particle is the contribution of gravity, given by

$$\begin{aligned} F &= mg\mathbf{j} \cdot {}^{\mathcal{R}}\mathbf{v}_1^P && (132)\\ &= -2mgx, && (133) \end{aligned}$$

and the generalized inertia force is

$$F^\star = -m^{\mathcal{R}}\mathbf{a}^P \cdot {}^{\mathcal{R}}\mathbf{v}_1^P \tag{134}$$
$$= -m\big[\dot{u}_1 + 2x\big(2u_1^2 + 2x\dot{u}_1\big)\big]. \tag{135}$$

Kane's dynamical equation of motion is

$$F + F^\star = 0, \tag{136}$$

yielding to

$$2gx + \dot{u}_1 + 2x(2u_1^2 + 2x\dot{u}_1) = 0. \tag{137}$$

To illustrate the improvement in the numerical solution of the nonminimal equations resulting from the augmentation of the damping terms, the solutions for y after 1 h of simulation time obtained from the integrations of Equations (121)–(122) and Equations (125)–(126) are compared with the most accurate one obtained from the integration of Equation (137), as shown in Figure 16. To reduce the computer memory and time costs, a bigger integration step Δt of 0.1 s is chosen. Clearly, the damping of constraint violations is crucial for accurate fast long-term simulations of such systems. It should be noted, however, that despite the correction in the computed holonomic constraint violation, this does not imply necessarily an improvement in the accuracy of the individual states. The energy integral provides a check on the stability of the whole system of nonminimal equations of motion, and is independent of the constraint violation measure ϕ. Figure 17 shows the deviation in the energy for the constraint-stabilized system. This deviation represents

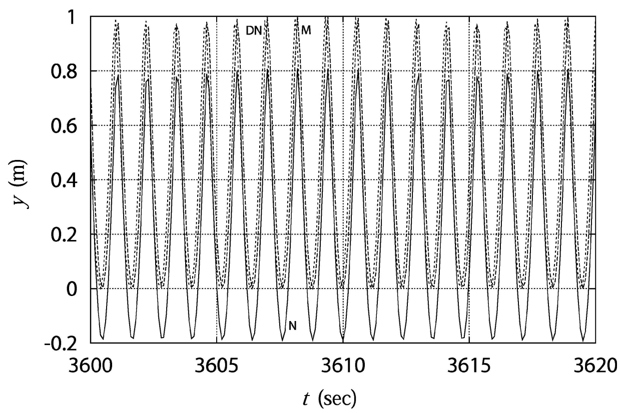

Figure 16. Example 2: y solution after 1 h, $\Delta t = 0.1$ s: M minimal, N nonminimal ($\Gamma_1 = \Gamma_2 = 0$), DN damped nonminimal ($\Gamma_1 = -20$, $\Gamma_2 = -100$).

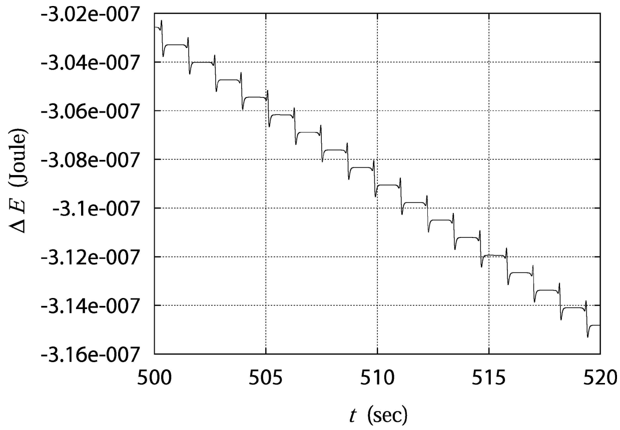

Figure 17. Example 2: Energy integral deviation ΔE: $\Gamma_1 = -20$, $\Gamma_2 = -100$, $\Delta t = 0.001$ s.

a deterioration in accuracy compared to the constraint-unstabilized system for the same Δt, as noticed by comparing with Figure 14. Careful choice of Γ_1 and Γ_2 is therefore important to preserve the stability of all the states. Nevertheless, the only tangible effect of the constraint stabilization in this example is favorable on y, as illustrated in Figure 16.

The achieved constraint numerical violation dynamics is compared in Figure 18 with those corresponding to two constraint numerical violation geometric elimination methods, given in [29] and [30]. Both methods are based on appropriate corrections of the state variables, without modifying the equations of motion. Clearly the augmentation of Baumgarte numerical stabilization terms yields better constraint violation dynamics, provided that the coefficients of the stabilizing terms are chosen properly.

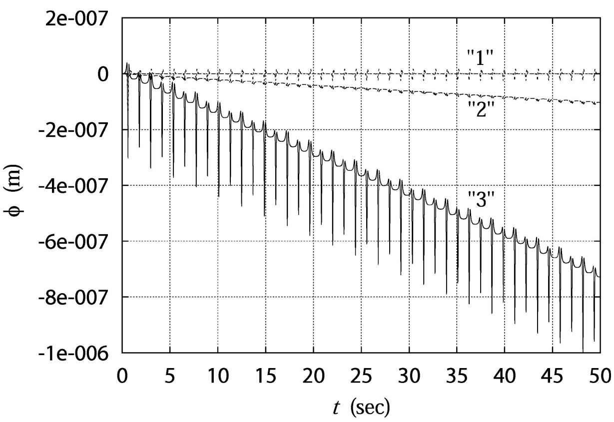

Figure 18. Example 2: Constraint violation ϕ: $\Delta t = 0.01$ s; “1” Baumgarte, “2” Blajer, “3” Yoon et al.

4. Conclusions and Future Work

By taking advantage of the mathematical conformity of the acceleration form of the constraint equations and the dynamical equations, a procedure for deriving nonminimal Kane's equations of motion is presented, where the constraint equations might be nonlinear in the generalized speeds. This is possible because of the special structure that the constraint matrices A_1 and A_2 retain as the variables of the dynamical system evolve in time, which implies the invertibility of the generalized constrained inertia matrix T. The resulting equations of motion are used to solve the Appell–Hamel problem, when the constraints are modeled as nonlinear in the generalized speeds. This nonlinearity is shown advantageous in stabilizing the solution of the equations of motion when numerically integrating the equations of motion. Furthermore, a systematic procedure is provided to modify the equations of motion for the purpose of suppressing the constraint violation due to numerical integration errors, by augmenting the constraint equations with damping terms of the Baumgarte type. The associated coefficients can be chosen to obtain any desired constraint dynamics without affecting the invertibility of the generalized constrained inertia matrix, as the coefficients of the acceleration terms in both the dynamical and the constraint equations remain unaltered. An illustrative example shows a significant reduction in a holonomic constraint violation, although the effect on the whole system of equations is slightly destabilizing. Therefore, the coefficients of the stabilizing terms must be chosen such that the improvement in the numerical stability of the constraint equations does not deteriorate the numerical stability of the resulting nonminimal system of equations. Employing optimal control theory to obtain the best coefficients implies minimizing an integral that involves the constraint violations, the stabilizing coefficients, and possibly the energy of the dynamical system. Solving the resulting optimality conditions is an on-going research by the authors.

Acknowledgement

The authors thank the reviewers for their insightful comments and suggestions for revision of this article.

References

1. Roberson, R.E. and Schwertassek, R., *Dynamics of Multibody Systems*, Springer-Verlag, Berlin, 1988.
2. Appell, P., 'Exemple de mouvement d'un point assujetti à une liaison exprimée par une relation non linéaire entre les composantes de la vitesse', *Rend. Circular Mat. Palermo* **32**, 1911, 48–50.
3. Hamel, G., *Theoretische Mechanik*, Spring-Verlag, Berlin, 1949.
4. Neimark, J.I. and Fufaev, N.A., *Dynamics of Nonholonomic Systems*, American Mathematical Society, Providence, RI, 1972.

5. Saletan, E. and Cromer, A., 'A variational principle for non-holonomic systems', *American Journal of Physics* **38**(7), 1970, 892–897.
6. Xu, M., Liu, C. and Huston, R.L., 'Analysis of nonlinearly constrained non-holonomic multibody systems', *International Journal of Non-Linear Mechanics* **25**(5), 1990, 511–519.
7. Kitzka, F., 'An example for the application of nonholonomic constraint of second order in particle mechanics', *Zeitschrift Fur Angewandte Mathematik und Mechanik* **66**, 1986, 312–314.
8. Fufaev, N.S., 'About an example of a system with a non-holonomic constraint of a second order', *Zeitschrift Fur Angewandte Mathematik und Mechanik* **70**, 1990, 593–594.
9. Blajer, W., 'Dynamics and control of mechanical systems in partly specified motion', *Journal of the Franklin Institute-Engineering and Applied Mechanics* **334B**(3), 1997, 407–426.
10. Huston, R.L. and Passerello, C.E., 'Nonholonomic systems with non-linear constraint equations', *International Journal of Nonlinear Mechanics* **11**, 1976, 331–336.
11. Wang, J.T. and Huston, R.L., 'Kane's equations with undetermined multipiers-application to constrained multibody systems', *ASME Journal of Applied Mechanics* **54**, 1987, 424–429.
12. Wang, J.T. and Huston, R.L., 'Computational methods in constrained multibody dynamics: Matrix formalisms', *Computers and Structures* **29**, 1988, 331–338.
13. Kane, T.R. and Levinson, D.A., 'Multibody dynamics', *ASME Journal of Applied Mechanics* **50**, 1983, 1071–1078.
14. Anderson, K., *Recursive Derivation of Explicit Equations of Motion for the Efficient Dynamic/Control Simulation of Large Multibody Systems*, Ph.D. Thesis, Stanford University, Department of Mechanical Engineering, Stanford, CA, 1990.
15. Rosenthal, D., 'An order *n* formulation for robotic systems', *Journal of Astronautical Sciences* **38**(4), 1990, 511–529.
16. Banerjee, A., 'Block-diagonal equations for multibody elastodynamics with geometric stiffness and constraints', *Journal of Guidance, Control, and Dynamics* **16**(6), 1993, 1092–1100.
17. Blajer, W., 'A geometric unification of constrained system dynamics', *Multibody System Dynamics* **1**, 1997, 3–21.
18. Blajer, W., 'A geometrical interpretation and uniform matrix formulation of multibody system dynamics', *Zeitschrift Für Angewandte Mathematik und Mechanik* **81**, 2001, 247–259.
19. Anderson, K. and Critchley, J., 'Improved 'order-N' performance algorithm for the simulation of constrained multi-rigid-body dynamic systems', *Multibody System Dynamics* **9**(2), 2003, 185–212.
20. Bajodah, A.H., Hodges, D.H. and Chen, Y.H., 'New form of Kane's equations of motion for constrained systems', *Journal of Guidance, Control, and Dynamics* **26**(1), 2003, 79–88.
21. Wampler, C., Buffinton, K. and Shu-hui, J., 'Formulation of equations of motion for systems subject to constraints', *Journal of Applied Mechanics* **52**, 1985, 465–470.
22. Huston, R.L., 'Constraint forces and undetermined multipliers in constrained multibody systems', *Multibody System Dynamics* **3**, 1999, 381–389.
23. Kane, T.R. and Levinson, D.A., *Dynamics: Theory and Applications*, 1st edn., McGraw-Hill Series in Mechanical Engineering, 1985.
24. Baumgarte, J., 'Stabilization of constraints and integrals of motion in dynamical systems', *Computer Methods in Applied Mechanics and Engineering* **1**, 1972, 1–16.
25. Wehage, R.A. and Haug, E.J., 'Generalized coordinate partitioning for dimension reduction in analysis of constrained dynamic systems', *Journal of Mechanical Design* **104**, 1982, 247–255.
26. Blajer, W., Schiehlen, W. and Schirm, W., 'A projective criterion to the coordinate partitioning method for multibody dynamics', *Archive of Applied Mechanics* **64**, 1994, 86–98.
27. Lesser, M., 'A geometrical interpretation of Kane's equations', *Proceedings of the Royal Society of London. Series A, Mathematical Physical and Engineering Sciences* **436**, 1992, 69–87.

28. Chang, C.O. and E., N.P., 'An adaptive constraint violation stabilization method for dynamic analysis of mechanical systems', *Journal of Mechanisms, Transmissions, and Automation in Design* **107**, December 1985, 488–492.
29. Blajer, W., 'Elimination of constraint violation and accuracy aspects in numerical simulation of multibody systems', *Multibody Sysems Dynamics* **7**, 2002, 265–284.
30. Yoon, S., Howe, R. and Greenwood, D., 'Geometric elimination of constraint violations in numerical simulation of Lagrangean equations', *Journal of Mechanical Design* **116**, 1994, 1058–1064.