Nonlinear Dyn (2021) 105:1191–1211
https://doi.org/10.1007/s11071-021-06500-5

ORIGINAL PAPER

Numerical integration of multibody dynamic systems involving nonholonomic equality constraints

P. Passas · S. Natsiavas · E. Paraskevopoulos

Received: 16 March 2021 / Accepted: 27 April 2021 / Published online: 8 July 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract This work considers a class of multibody dynamic systems involving bilateral nonholonomic constraints. An appropriate set of equations of motion is employed first. This set is derived by application of Newton's second law and appears as a coupled system of strongly nonlinear second-order ordinary differential equations in both the generalized coordinates and the Lagrange multipliers associated with the motion constraints. Next, these equations are manipulated properly and converted to a weak form. Furthermore, the position, velocity and momentum type quantities are subsequently treated as independent. This yields a three-field set of equations of motion, which is then used as a basis for performing a suitable temporal discretization, leading to a complete time integration scheme. In order to test and validate its accuracy and numerical efficiency, this scheme is applied next to challenging mechanical examples, exhibiting rich dynamics. In all cases, the emphasis is put on highlighting the advantages of the new method by direct comparison with existing analytical solutions as well as with results of current state-of-the-art numerical methods. Finally, a comparison is also performed with results available for a benchmark problem.

P. Passas · S. Natsiavas (✉) · E. Paraskevopoulos
Department of Mechanical Engineering, Aristotle University, 54 124 Thessaloniki, Greece
e-mail: natsiava@auth.gr

Keywords Multibody dynamics · Analytical dynamics · Nonholonomic constraints · Three-field weak formulation · Rolling bodies · Bicycle stability

1 Introduction

This work is concerned with dynamics of multibody mechanical systems subject to motion constraints. Such constraints arise frequently in modeling the dynamics of a plethora of engineering systems and need a special treatment [1–3]. In particular, the class of systems examined involves nonholonomic bilateral constraints. That is, nonintegrable equality constraints on their generalized velocities. In contrast to integrable (or holonomic) constraints, which limit both the velocities and the positions of a system, nonholonomic constraints impose restrictions on the velocity level only, without limiting the possible positions on their configuration space. In classical applications, such constraints arise in the presence of rolling or sliding motions. They lead to quite interesting dynamics, not exhibited by holonomic systems and represent a central subject of Nonholonomic Mechanics [4, 5].

The main emphasis of this work is placed on developing and presenting the essential steps of a new systematic methodology, leading to a robust, accurate and efficient numerical integration of the equations of motion of mechanical systems involving

nonholonomic constraints. Classical numerical methodologies, treating the equations of motion as a coupled system of differential–algebraic equations (DAEs), lead to severe difficulties and inaccuracies, related to response stability issues [6]. Specifically, small errors in the initial values as well as errors developed through numerical integration are amplified and lead to an erroneous response as time goes on, independently on the accuracy level imposed on the calculations. Most of the times, increasing these accuracy levels helps in just extending the acceptable accuracy of the solution for a little longer time duration but this is not sufficient to prevent an eventual drift from the real solution. This prompted the initiation of many research efforts, focusing toward eliminating or reducing these undesirable consequences. Besides consideration of dissipative time integration schemes [1, 3], this included efforts to properly scale the governing equations and constraints or to perform appropriate velocity projections [7–9].

In this work, the effort to eliminate the stability problems arising during integration of the equations of motion of constrained systems starts from their theoretical foundation and continues with the development and application of suitable numerical methodologies. More specifically, the essential reason for achieving an accurate and reliable numerical solution is the utilization of an appropriate set of equations, derived by a consistent application of Newton's law of motion [10]. Their derivation is based on classical concepts and tools of Analytical Dynamics and leads directly to a set of second-order ordinary differential equations (ODEs) in both the generalized constraints and the Lagrange multipliers, associated with the motion constraints [11]. This eliminates the well-known pathologies associated with the classical DAE formulations. Moreover, the robustness of the new numerical scheme is reinforced by first putting these equations in a suitable weak form and then treating the generalized coordinates, velocities and momenta as independent quantities, leading to a convenient three-field weak formulation [12]. This follows earlier mixed multi-field methods in computational mechanics, based on the de Veubeke–Washizu variational principle [13, 14]. Finally, the process is completed by developing and applying a suitable temporal discretization scheme.

A similar path was followed in a recent work of the authors [15]. In that work, the emphasis was placed on mechanical systems with holonomic constraints. Instead, the present work focuses on systems involving nonholonomic constraints. This leads to some important deviations in both the analytical formulation and the numerical scheme developed, which are noticed and explained in detail in Sects. 2–4. In addition, the set of unknowns arising in each time step of the present treatment includes both the generalized velocities and the time derivatives of the Lagrange multipliers. In contrast, the method applied in [15] was based on an augmented Lagrangian approach and treated the generalized velocities and the time derivatives of the Lagrange multipliers at different solution levels. Finally, all the numerical examples investigated in that work involved holonomic constraints, while the examples examined in this study involve nonholonomic constraints.

An outline of the present work is as follows. First, the set of equations of motion employed is presented briefly in Sect. 2, giving emphasis on explaining and modeling the effect of the nonholonomic constraints. Then, the essential details needed for putting these equations in a suitable three-field weak form are included in Sect. 3. This provides a solid foundation for developing an accurate and efficient methodology for the temporal discretization of the equations of motion, which is done in Sect. 4. Next, a selected set of numerical results is included in Sect. 5 for three typical mechanical examples. For the first two of them, analytical solutions are available, while the third example is a benchmark problem. In this way, direct comparison of the numerical results provides strong evidence and demonstrates the advantages of the new method, over classical DAE formulations. Finally, a synopsis of the study, together with possible future extensions, is presented in Sect. 6.

2 Equations of motion

The equations of motion employed for the class of systems examined in this work have been obtained and presented in a previous publication of the authors [11]. In this section, a brief summary of the essential steps performed and the main results derived in that work are presented in order to set up the notation and enhance the completeness of this study. Specifically, by adopting the classical viewpoint of Analytical Dynamics, the motion is described by a set of

generalized coordinates $q = (q^1 \ldots q^n)$, representing motion of a fictitious point p on a manifold M, as a function of time t [16–18]. Moreover, the tangent vector $\underline{v}$ to the motion curve at point p, known as generalized velocity, belongs to an n-dimensional vector space T_pM [4]. Therefore, if $\mathfrak{B}_e = \{\underline{e}_1 \quad \ldots \quad \underline{e}_n\}$ is a basis of T_pM, vector $\underline{v}$ can be put in the compact form $\underline{v} = v^i \underline{e}_i$ by using the summation convention on repeated indices [18]. Also, the elements of the corresponding cotangent (or dual) space T_p^*M, known as covectors, represent generalized momenta [17]. The relation between a vector $\underline{u}$ and the corresponding covector $\underset{\sim}{u}^*$ is established through the duality pairing

$$\underset{\sim}{u}^*(\underline{w}) \equiv \langle \underline{u}, \underline{w} \rangle, \quad \forall \underline{w} \in T_pM, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on T_pM [19]. Then, a dual basis $\mathfrak{B}_e^* = \{\underset{\sim}{e}^1 \quad \cdots \quad \underset{\sim}{e}^n\}$ to $\mathfrak{B}_e$ can be created for T_p^*M, through the condition $\underset{\sim}{e}^i(\underline{e}_j) = \delta_j^i$, where δ_j^i is a Kronecker's delta. In dynamics, the inner product is expressed in terms of the components g_{ij} of the metric tensor at point p, which are determined through consideration of the kinetic energy of the system.

When there are no motion constraints, a minimal and independent set of generalized coordinates can be selected. In general, the geometry of the corresponding configuration manifold M is non-Euclidean. For this reason, the exact solution path is determined by application of Newton's second law in the form

$$\nabla_{\underline{v}} \underset{\sim}{p}_M^* = \underset{\sim}{f}_M^*, \tag{2}$$

The term in the right-hand side is a covariant derivative along a path on M, with tangent vector $\underline{v}$ [10]. It is expressed in the component form

$$\nabla_{\underline{v}} \underset{\sim}{p}_M^*(t) = (\dot{p}_i - \Lambda_{ji}^k p_k v^j)\, \underset{\sim}{e}^i, \tag{3}$$

with $i, j, k = 1, \ldots, n$, where the quantities Λ_{ji}^k are known as affinities. They are components of the affine connection ∇ and dictate the transition from a tangent space to any neighboring tangent space of manifold M [19]. Moreover, covectors $\underset{\sim}{p}_M^* = p_I \underset{\sim}{e}^I$ and $\underset{\sim}{f}_M^* = f_I \underset{\sim}{e}^I$ represent the generalized momenta and the applied forces, respectively [10]. In fact, direct application of Eq. (1) leads to

$$p_i = g_{ij} v^j. \tag{4}$$

Then, by introducing the classical matrix notation. $\underline{q} = (q^1 \quad \cdots \quad q^n)^T$, $M = [g_{ij}]$ *and* $\underline{f} = (f_1 \quad \cdots \quad f_n)^T$, Equation (2) can be cast in the convenient form

$$(M(\underline{q})\dot{\underline{q}})^{\cdot} + \underline{h}(\underline{q}, \dot{\underline{q}}) = \underline{f}(\underline{q}, \dot{\underline{q}}, t), \tag{5}$$

where vector $\underline{h}(\underline{q}, \dot{\underline{q}})$ includes the inertia terms related to the affinities Λ_{ij}^k.

The picture changes drastically when the system is subjected to motion constraints. In particular, the class of mechanical systems examined is assumed to be subject to a set of k scleronomic nonholonomic and linearly independent constraints, with general form

$$\dot{\psi}^R \equiv a_i^R(q) v^i = 0 \text{ or } \underline{\dot{\psi}}(q, \underline{v}) \equiv A(q)\underline{v} = \underline{0}, \tag{6}$$

where $A = [a_i^R(q)]$ is a known $k \times n$ matrix. The manipulation of these constraints is quite critical for the derivation of the equations of motion. Essentially, they provide the tools to decompose the original configuration manifold M locally into an $(n-k)$-dimensional manifold M_A, described by an appropriate set of $n-k$ minimal coordinates, plus k one-dimensional manifolds M_R, one for each constraint. This is achieved by using Eq. (6) for defining a suitable set of linear operators, acting between the tangent and dual spaces of these manifolds. First, these operators help in selecting the geometric properties (i.e., metric and connection) of manifolds M_A and M_R, so that Newton's law of motion is transferred from M to M_A and to M_R (for $R = 1, \ldots, k$) in an invariant form, similar to that expressed by Eq. (2) [10]. In addition, they help in splitting each element of the cotangent space T_p^*M in a unique manner. More specifically, the elements of the special class of Newton covectors on M, defined by

$$\underset{\sim}{h}_M^* \equiv \nabla_{\underline{v}} \underset{\sim}{p}_M^* - \underset{\sim}{f}_M^*, \tag{7}$$

are expressed in the form

$$\underset{\sim}{h}_M^* = \underset{\sim}{h}_S^* + \underset{\sim}{h}_T^*, \tag{8}$$

with

$$\underset{\sim}{h}_S^* = E_{TD} \underset{\sim}{h}_A^* \text{ and } \underset{\sim}{h}_T^* = \sum_{R=1}^k T_{RD} \underset{\sim}{h}_R^* \tag{9}$$

The quantities E_{TD} and T_{RD} are specific linear operators defined by the constraints completely (for example, $T_{RD} \equiv a_i^R \underset{\sim}{e}^i \otimes \underset{\sim}{e}_R$). They transfer the corresponding Newton covectors $\underset{\sim}{h}_A^*$ from M_A and $\underset{\sim}{h}_R^*$ from each of the M_R, respectively, back to M. Covectors $\underset{\sim}{h}_A^*$ and $\underset{\sim}{h}_R^*$ are defined on M_A and M_R in accordance to Eq. (7).

Due to the presence of the motion constraints, the original set of coordinates $q = (q^1 \dots q^n)$ becomes redundant. Consequently, condition (2) is no longer true and is replaced by

$$\underset{\sim}{h}_A^* = \underset{\sim}{0}, \tag{10}$$

instead. Then, by taking Eqs. (9) and (10) into account, Eq. (8) becomes

$$\underset{\sim}{h}_M^* = \sum\nolimits_{R=1}^{k} T_{RD} \underset{\sim}{h}_R^*. \tag{11}$$

Next, substitution of Eqs. (3) and (4) in Eq. (7) yields

$$\underset{\sim}{h}_M^* = h_i \underset{\sim}{e}^i, \tag{12}$$

with components

$$h_i = (g_{ij} v^j)^{\cdot} - \Lambda_{\ell i}^m g_{mj} v^j v^\ell - f_i. \tag{13}$$

Likewise, each Newton covector $\underset{\sim}{h}_R^*$ on the constraint manifolds M_R (for $R = 1, \dots, k$) takes the form

$$\underset{\sim}{h}_R^* = h_R \underset{\sim}{e}^R, \tag{14}$$

with components

$$h_R = (\overline{m}_{RR} \dot{\lambda}^R)^{\cdot} + \overline{c}_{RR} \dot{\lambda}^R - \overline{f}_R. \tag{15}$$

In the last two equations and in the sequel, the convention on repeated indices does not apply to index R. Moreover, during the solution process, the terms

$$\overline{m}_{RR} = c_R^i g_{ij} c_R^j \quad \text{and} \quad \overline{f}_R = c_R^i f_i(q, \underline{v}, t) \tag{16}$$

are determined through a projection along special directions $\underline{c}_R$ on $T_p M$, for each of the constraints. In particular, the components of the n-vector $\underline{c}_R$ are chosen so that

$$a_i^R c_R^i = 1. \tag{17}$$

Finally, each coefficient $\overline{c}_{RR}$ is selected so that the corresponding term in Eq. (15) represents a corrective force applied on the figurative point, when a velocity violation tends to develop along direction $\underline{c}_R$. For instance, if the applied forces depend on the generalized velocity $\underline{v}$, a convenient choice is

$$\overline{c}_{RR} = -c_R^i \frac{\partial f_i}{\partial v^j} c_R^j. \tag{18}$$

Next, by combining Eqs. (11)-(15), it easily turns out that

$$h_i = \sum\nolimits_{R=1}^{k} a_i^R h_R, \tag{19}$$

or

$$(g_{ij} v^j)^{\cdot} - \Lambda_{\ell i}^m g_{mj} v^j v^\ell - f_i = \sum\nolimits_{R=1}^{k} a_i^R \left[(\overline{m}_{RR} \dot{\lambda}^R)^{\cdot} + \overline{c}_{RR} \dot{\lambda}^R - \overline{f}_R \right], \tag{20}$$

for $i = 1, \dots, n$. Therefore, after enhancing the notation introduced by Eq. (5) with

$$\underline{\lambda} = (\lambda^1 \quad \cdots \quad \lambda^k)^T, \ \overline{M} = diag(\overline{m}_{11} \quad \cdots \quad \overline{m}_{kk}), \\ \overline{C} = diag(\overline{c}_{11} \quad \cdots \quad \overline{c}_{kk}) \tag{21}$$

and the array $\underline{\overline{f}}$, as determined by Eqs. (16)-(18), Eq. (20) can be cast in a convenient matrix form

$$(M(\underline{q})\dot{\underline{q}})^{\cdot} + \underline{h}(\underline{q}, \dot{\underline{q}}) = \underline{f}(\underline{q}, \dot{\underline{q}}, t) + A^T(\underline{q})[(\overline{M}\dot{\underline{\lambda}})^{\cdot} + \overline{C}\dot{\underline{\lambda}} - \underline{\overline{f}}]. \tag{22}$$

Equation (20), or equivalently Eq. (22), corresponds to a set of n second-order ODEs, involving the $n + k$ unknowns q^i and λ^R. A complete mathematical formulation is obtained by including the k equations of the constraints. Originally, these are expressed by Eq. (6) but can eventually be put in the second-order ODE form

$$(\overline{m}_{RR} \dot{\psi}^R)^{\cdot} + \overline{c}_{RR} \dot{\psi}^R = 0, \tag{23}$$

through appropriate mappings from M to M_R, forcing $\dot{\psi}^R$ to become zero, for $R = 1, \dots, k$. In contrast to cases involving holonomic constraints, there is neither a ψ^R term in Eq. (23) nor a $\underline{\lambda}$ term in Eq. (22).

Complete details on the formulation presented in this section as well as an extensive comparison with previous formulations can be found in the original publication [11]. In brief, the present theoretical approach brings significant advantages when compared to other approaches applied so far in the field of

Multibody Dynamics. These advantages are related to the physically consistent and correct elimination of the singularities associated with the usual sets of high-index DAEs of motion [1–3]. More specifically, by treating the motion constraints as an integral part of the overall process of deriving the equations of motion led to a set of dynamic Lagrange multipliers, expressed by Eq. (15). In this way, each constraint introduces effective inertia terms in Eqs. (22) and (23). Consequently, both of these equations appear now in a pure second-order ODE form, from the onset and in a natural way. In addition, the extra terms appearing in these equations are evaluated in a systematic and analytical manner through exact application of Newton's second law on configuration manifolds possessing general geometric properties. As a result, there is no need for an ad hoc selection of their values for numerical stabilization or for scaling the constraints and the equations of motion. Some extra advantages are also realized during the development of the numerical integration process, which is started by first putting the equations of motion in a suitable multi-field weak form and is then integrated by performing an appropriate temporal discretization, as explained in Sects. 3 and 4, respectively. All these advantages are further supported and illuminated by the numerical results presented in Sect. 5.

3 Three-field weak form of the equations of motion

Starting from Eq. (11), it is straightforward to show that

$$\int_{t_1}^{t_2} (\underset{\sim}{h}_M^* - \sum\nolimits_{R=1}^{k} T_{RD}\underset{\sim}{h}_R^*)(\underline{w})dt = 0, \quad \forall \underline{w} \in T_pM, \tag{24}$$

for arbitrary time instances t_1 and t_2. Then, by performing lengthy manipulations, including a usual operation involving integration by parts (see ref. [12] for details), the above relation leads eventually to the following weak form of the equations of motion

$$\begin{aligned}&\left[(p_i - \sum\nolimits_{R=1}^{k} a_i^R \overline{m}_{RR}\dot{\lambda}^R)w^i\right]\Big|_{t_1}^{t_2}\\ &- \int_{t_1}^{t_2} \{p_i[\delta v^i + (\Lambda_{jk}^i - \tau_{jk}^i - \sigma_{jk}^i)v^k w^j] + f_i w^i\\ &+ \sum\nolimits_{R=1}^{k} [-\overline{m}_{RR}\dot{\lambda}^R\{\tfrac{Da_i^R}{Dt}w^i + a_i^R[\delta v^i\\ &+ (\Lambda_{jk}^i - \tau_{jk}^i - \sigma_{jk}^i)v^k w^j]\} + a_i^R(\overline{c}_{RR}\dot{\lambda}^R - \overline{f}_R)w^i]\}dt = 0.\end{aligned} \tag{25}$$

The quantities w^i and δv^i represent variations. They are evaluated by taking into account that the variation of any scalar function $f(q)$ on manifold M is defined as the derivative of f along an arbitrary vector $\underline{w}$ of T_pM, according to the rule

$$\delta f \equiv \underline{w}(f) = \frac{\partial f}{\partial q^i} w^i. \tag{26}$$

Therefore, for each holonomic coordinate this yields $w^i = \delta q^i$, while a little more involved relation is obtained in case of anholonomic coordinates [12, 19]. In this way, the variation of the velocity components is defined by $\delta v^i = \underline{w}(v^i)$. Moreover, the total time derivative Da_i^R/Dt, appearing in the second line of Eq. (25), is determined by

$$\frac{Da_i^R}{Dt} = (a_{i,j}^R - \Lambda_{ji}^{\ell} a_{\ell}^R)v^j, \tag{27}$$

with partial derivatives

$$a_{i,j}^R \equiv \partial a_i^R / \partial q^j.$$

Finally, the terms τ_{jk}^i represent components of the torsion of the connection selected on manifold M , while the quantities σ_{jk}^i are arbitrary constants, with $\sigma_{kj}^i = -\sigma_{jk}^i$ [12].

A similar treatment of each constraint equation, expressed by Eq. (23), leads first to

$$\int_{t_1}^{t_2} [(\overline{m}_{RR}\dot{\psi}^R)^{\cdot} + \overline{c}_{RR}\dot{\psi}^R]\,\delta\lambda^R dt = 0, \quad \forall \delta\lambda^R, \quad R = 1, \ldots, k, \tag{28}$$

which, after integration by parts, becomes

$$(\overline{m}_{RR}\dot{\psi}^R\delta\lambda^R)\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} (\overline{m}_{RR}\dot{\psi}^R\delta\dot{\lambda}^R - \overline{c}_{RR}\dot{\psi}^R\delta\lambda^R)dt = 0. \tag{29}$$

Next, in order to exploit more advantages of the weak formulation, the position, velocity and momentum variables are considered as independent quantities [13, 20, 21]. For this, a new velocity field $\underline{\upsilon}$ is introduced on manifold M, which is forced to become identical to the true velocity field $\underline{v} = d\underline{q}\big/dt$ in an average rather than in a pointwise sense through the relations

$$\int_{t_1}^{t_2} \delta\pi_i(\upsilon^i - v^i)\, dt = 0, i = 1, \ldots, n, \tag{30}$$

where $\delta\pi_i$ are arbitrary constants. Likewise, a new quantity μ^R is introduced for each constraint, which is forced to become identical to $\dot{\lambda}^R = d\lambda^R\big/dt$, by imposing the conditions

$$\int_{t_1}^{t_2} \delta\sigma_R(\mu^R - \dot{\lambda}^R)\, dt = 0,\ R = 1, \ldots, k, \tag{31}$$

where $\delta\sigma_R$ are arbitrary constants. In addition, these last sets of equations are enhanced by an accompanying set of conditions, consisting of

$$\int_{t_1}^{t_2} \pi_i\left(\delta\upsilon^i - \delta v^i\right) dt = 0 \tag{32}$$

and

$$\int_{t_1}^{t_2} \sigma_R(\delta\mu^R - \delta\dot{\lambda}^R)\, dt = 0, \tag{33}$$

where π_i and σ_R are appropriate Lagrange multipliers, leading to independent variations $\delta\upsilon^i$, δv^i and $\delta\mu^R$, $\delta\dot{\lambda}^R$, respectively. Therefore, by augmenting Eq. (25) with Eq. (32) and performing another round of manipulations, it can be shown (see [12] for details) that this leads to the following variational equation

$$\begin{aligned}
&\left[(p_i - \textstyle\sum_{R=1}^{k} a_i^R \overline{m}_{RR}\mu^R)w^i\right]\Big|_{t_1}^{t_2} \\
&+ \int_{t_1}^{t_2} \left[(\pi_i + \textstyle\sum_{R=1}^{k} a_i^R \overline{m}_{RR}\mu^R - p_i)\delta\upsilon^i - \pi_i\delta v^i\right]dt \\
&- \int_{t_1}^{t_2} \Big\{(\Lambda_{ij}^{\ell} - \tau_{ij}^{\ell} - \sigma_{ij}^{\ell})\, \upsilon^j (p_\ell - \textstyle\sum_{R=1}^{k} a_\ell^R \overline{m}_{RR}\mu^R) \\
&+ f_i + \textstyle\sum_{R=1}^{k} [-a_i^R \overline{m}_{RR}\mu^R + a_i^R(\overline{c}_{RR}\mu^R - \overline{f}_R)]\Big\}\, w^i dt = 0
\end{aligned} \tag{34}$$

A similar combination of Eqs. (29) and (33) yields

$$\begin{aligned}
&(\overline{m}_{RR}\dot{\psi}^R\delta\lambda^R)\Big|_{t_1}^{t_2} + \int_{t_1}^{t_2} \left[(\sigma_R - \overline{m}_{RR}\dot{\psi}^R)\delta\mu^R + \overline{c}_{RR}\dot{\psi}^R\delta\lambda^R - \sigma_R\delta\dot{\lambda}^R\right] dt = 0, \\
&R = 1, \ldots, k,
\end{aligned} \tag{35}$$

with

$$\dot{a}_i^R = (a_{i,j}^R - a_\ell^R \Lambda_{ji}^{\ell})\, \upsilon^j. \tag{36}$$

By construction, the variations w^i (representing δq^i or $\delta\vartheta^i$ for a true coordinate q^i or a pseudo-coordinate ϑ^i, respectively), $\delta\lambda^R$, $\delta\upsilon^i$, $\delta\mu^R$, $\delta\pi_i$ and $\delta\sigma_R$ are independent for all $i = 1, \ldots, n$ and $R = 1, \ldots, k$. Consequently, Eqs. (20)-(25) can be employed to construct a dynamically equivalent set of first-order ODEs. This was done in an earlier work, leading to the classical Hamilton's canonical equations [12]. Here, the weak form expressed by these equations is used, instead, as a foundation for performing an appropriate temporal discretization of the equations of motion, as explained in the following section.

4 Temporal discretization of the equations of motion

First, the presence of the motion constraints causes a non-flatness of the configuration space of the class of systems examined. This, in turn, makes advantageous the utilization of a special type of curves in performing the temporal discretization. In this work, these curves are selected to be the autoparallels, corresponding to the "straightest" curves on the configuration space [19, 22]. Moreover, following earlier approaches [15, 23], all the variations in Eqs. (30)-(35) are assumed to remain constant within each time interval $[t_m, t_{m+1}]$, while the corresponding variables vary linearly in time. Then, it can be shown that these conditions convert Eqs. (30) and (31) into relations with the following general form

$$\underline{q}_{m+1} = \underline{q}_m + \underline{g}(\underline{\upsilon}_{m+1}, \underline{\upsilon}_m)\, \Delta t \tag{37}$$

and

$$\lambda_{m+1}^R = \lambda_m^R + \frac{1}{2}(\mu_{m+1}^R + \mu_m^R)\Delta t, R = 1, \ldots, k, \tag{38}$$

where $\Delta t = t_{m+1} - t_m$ is the time step. In particular, the function $\underline{g}(\underline{\upsilon}_{m+1}, \underline{\upsilon}_m)$ in Eq. (37) is determined in terms of the components of the connection selected.

For instance, in the special case of a Euclidean configuration space [23], where $\Lambda^k_{ij} = 0$, it turns out that

$$\underline{g}(\underline{\upsilon}_{m+1}, \underline{\upsilon}_m) = \frac{1}{2}(\underline{\upsilon}_{m+1} + \underline{\upsilon}_m).$$

Equations (37) and (38) represent update formulas for the generalized coordinates and the Lagrange multipliers, respectively. In addition, collecting the terms in Eqs. (34) and (35) multiplied by $\delta\upsilon^i$ and $\delta\mu^R$ yields

$$\int_{t_1}^{t_2} (\pi_i - p_i + \sum\nolimits_{R=1}^{k} a_i^R \overline{m}_{RR}\mu^R)\delta\upsilon^i \, dt = 0 \tag{39}$$

and

$$\int_{t_1}^{t_2} (\sigma_R - \overline{m}_{RR}\dot{\psi}^R)\delta\mu^R dt = 0, \tag{40}$$

respectively. These equations can be used in order to update the values of the generalized momenta π_i and σ_R, respectively, when necessary. Finally, collecting the remaining terms of Eqs. (34) and (35), multiplied by the variations w^i and $\delta\lambda^R$, yields

$$\begin{aligned}&(p_i - \sum\nolimits_{R=1}^{k} a_i^R \overline{m}_{RR}\mu^R)w^i|_{t_1}^{t_2} \\ &- \int_{t_1}^{t_2} (p_\ell - \sum\nolimits_{R=1}^{k} a_\ell^R \overline{m}_{RR}\mu^R)(\Lambda^\ell_{ij} - \tau^\ell_{ij} - \sigma^\ell_{ij})\,\upsilon^j + f_i \\ &+ \sum\nolimits_{R=1}^{k} [a_i^R(\overline{c}_{RR}\mu^R - \overline{f}_R) - \dot{a}_i^R \overline{m}_{RR}\mu^R]w^i dt = 0\end{aligned} \tag{41}$$

and

$$(\overline{m}_{RR}\dot{\psi}^R\delta\lambda^R)\big|_{t_m}^{t_{m+1}} + \int_{t_m}^{t_{m+1}} \overline{c}_{RR}\dot{\psi}^R\delta\lambda^R dt = 0, \quad R = 1, \ldots, k, \tag{42}$$

respectively.

In the sequel, a summary is provided for the essential information needed in understanding the basic features of the new numerical scheme. As a result of the three-field formulation performed in this work, the set of unknowns of the resulting mathematical problem is enhanced and includes the generalized coordinates q^i and λ^R (i.e., the Lagrange multipliers are also considered as generalized coordinates from hereon), together with the corresponding generalized velocities υ^i and μ^R as well as with the generalized momenta π_i and σ_R. In total, these quantities give rise to a set of $3(n+k)$ unknowns. Their determination is accomplished by employing the system of Eqs. (30), (31) and (39)-(42). Here, this is achieved by developing and applying an implicit numerical scheme in order to carry out the required temporal discretization. As usual, the outcome of this process is a system of nonlinear algebraic equations, whose solution provides the values of the unknown quantities at the end t_{m+1} of the time step considered in terms of their known values at earlier time steps. In fact, this set of equations is solved by applying a block-type iterative technique within each time step, as explained next.

First, it is assumed that the values of all the unknowns, but the components of the weak velocity vectors, defined by $\underline{\upsilon} = (\upsilon^1 \;\cdots\; \upsilon^n)^T$ and $\underline{\mu} = (\mu^1 \;\cdots\; \mu^k)^T$, are fixed. Then, an involved set of $n+k$ nonlinear algebraic equations is obtained through application of Eqs. (41) and (42), which can be put in the general form

$$\underline{\eta}(\underline{\upsilon}_{m+1}, \underline{\mu}_{m+1}; \underline{\upsilon}_m, \underline{\mu}_m, \underline{q}_m, \underline{\lambda}_m, \underline{\pi}_m, \underline{\sigma}_m) = \underline{0}. \tag{43}$$

In the last expression, $\underline{\upsilon}_{m+1}$, $\underline{\mu}_{m+1}$ and $\underline{\upsilon}_m$, $\underline{\mu}_m$ represent the values of $\underline{\upsilon}$, $\underline{\mu}$ at times t_{m+1} and t_m, respectively. Moreover, a similar meaning is also given to the vector quantities, $\underline{q} = (q^1 \;\cdots\; q^n)^T$, $\underline{\lambda} = (\lambda^1 \;\cdots\; \lambda^k)^T$, $\underline{\pi} = (\pi_1 \;\cdots\; \pi_n)^T$ and $\underline{\sigma} = (\sigma_1 \;\cdots\; \sigma_k)^T$. Following common practice, Eq. (43) is then solved by applying a Newton–Raphson approach, with respect to the unknown

$$\underline{x} \equiv (\underline{\upsilon}^T \quad \underline{\mu}^T)^T. \tag{44}$$

To achieve this, given an estimate $\underline{x}^\ell_{m+1}$, a corrected value $\underline{x}^{\ell+1}_{m+1}$ is obtained, according to

$$\underline{x}^{\ell+1}_{m+1} = \underline{x}^\ell_{m+1} + \Delta\underline{x}^\ell, \tag{45}$$

where the correction $\Delta\underline{x}^\ell$ is determined by solving the linearized problem

$$J^\ell_{m+1}\Delta\underline{x}^\ell = -\underline{\mathcal{R}}^\ell_{m+1}, \tag{46}$$

resulting by substituting Eq. (45) into Eq. (43), with Jacobian matrix

$$J^\ell_{m+1} = \frac{\partial\underline{\eta}}{\partial\underline{x}}(\underline{\upsilon}^\ell_{m+1}, \underline{\mu}^\ell_{m+1}; \underline{\upsilon}_m, \underline{\mu}_m, \underline{q}_m, \underline{\lambda}_m, \underline{\pi}_m, \underline{\sigma}_m) \tag{47}$$

and residual vector

$$\underline{\mathcal{R}}^{\ell}_{m+1} = \underline{\eta}(\underline{\upsilon}^{\ell}_{m+1}, \underline{\mu}^{\ell}_{m+1}; \underline{\upsilon}_m, \underline{\mu}_m, \underline{q}_m, \underline{\lambda}_m, \underline{\pi}_m, \underline{\sigma}_m). \quad (48)$$

The resulting system of equations involves only $n+k$ of the original unknowns. Each problem, as expressed by Eq. (46), is solved by employing a direct linear solver. The computations are stopped when the set of weak velocities $\underline{x}_{m+1}$ and the constraint equations are satisfied up to a prespecified accuracy or the iterations exceed a critical number. In the latter case, the time step Δt is reduced and the process is restarted. Next, the values of the generalized coordinates $\underline{q}_{m+1}$ and $\underline{\lambda}_{m+1}$ are determined through a direct update, based on Eqs. (37) and (38), respectively. The iteration process is completed when the residual in the right-hand side of Eq. (46) becomes also sufficiently small. Otherwise, the process is repeated after decreasing the time step. Finally, the new values of the momentum variables $\underline{\pi}_{m+1}$ and $\underline{\sigma}_{m+1}$ can also be obtained after the iterations are finished by using the subsystem of equations resulting from application of Eqs. (39) and (40), respectively.

For better clarity, the numerical implementation of the algorithm employed for solving the discretized set of equations of motion, given by Eqs. (41) and (42), is presented next. Here, the analysis is restricted to systems possessing rigid bodies only. In this case, through the choices $\sigma^{\ell}_{ij} = 0$ and $\tau^{\ell}_{ij} = \Lambda^{\ell}_{ij}$, with the affinities as determined in [10, 22], Eq. (41) is replaced by

$$\begin{aligned}&(p_i - \textstyle\sum_{R=1}^{k} a^R_i \overline{m}_{RR}\mu^R)w^i\big|_{t_1}^{t_2} \\ &- \int_{t_1}^{t_2} \{f_i + \textstyle\sum_{R=1}^{k} [a^R_i(\overline{c}_{RR}\mu^R - \overline{f}_R) - \dot{a}^R_i \overline{m}_{RR}\mu^R]\}w^i dt = 0.\end{aligned} \quad (49)$$

This equation can be written in a convenient matrix form

$$[\mathcal{A}(\underline{p} - A^T\overline{M}\underline{\mu})]\Big|_{t_m}^{t_{m+1}} - \int_{t_m}^{t_{m+1}} \mathcal{A}\underline{\ell} dt = \underline{0}, \quad (50)$$

with

$$\underline{\ell} \equiv \underline{f} + A^T(\overline{C}\underline{\mu} - \overline{\underline{f}}) - B^T\overline{M}\underline{\mu} \text{ and } B \equiv [\dot{a}^R_i], \quad (51)$$

where the terms $\dot{a}^R_i$ are defined by Eq. (36). Moreover, the action of operator $\mathcal{A}$ is equivalent to the identity or the corresponding rotation matrix of a rigid body when it applies to its translational or rotational degrees of freedom, respectively. In a similar manner, Eq. (42) is also put in a matrix form

$$[\overline{M}\,\underline{\dot{\psi}}]\Big|_{t_m}^{t_{m+1}} + \int_{t_m}^{t_{m+1}} \underline{\ell} dt = \underline{0}, \quad (52)$$

with

$$\underline{\ell} \equiv \overline{C}\,\underline{\dot{\psi}}. \quad (53)$$

Therefore, application of the classical trapezoidal rule to Eqs. (50) and (52) leads eventually to

$$[\mathcal{A}(\underline{p} - A^T\overline{M}\underline{\mu})]\Big|_{t_m}^{t_{m+1}} - \frac{1}{2}\Delta t\,(\mathcal{A}_m\underline{\ell}_m + \mathcal{A}_{m+1}\underline{\ell}_{m+1}) = \underline{0}, \quad (54)$$

and

$$[\overline{M}\,\underline{\dot{\psi}}]\Big|_{t_m}^{t_{m+1}} + \frac{1}{2}\Delta t(\underline{\ell}_{m+1} + \underline{\ell}_m) = \underline{0}, \quad (55)$$

respectively. Equations (54) and (55) comprise a set of nonlinear algebraic equations, having the structure of Eq. (43). Then, the residual covector of the last set of equations, defined by Eq. (48), takes the form

$$\underset{\sim}{\mathcal{R}}_T = (\underset{\sim}{\mathcal{R}}_v \quad \underset{\sim}{\mathcal{R}}_\mu), \quad (56)$$

where $\underset{\sim}{\mathcal{R}}_v$ and $\underset{\sim}{\mathcal{R}}_\mu$ is the numerical residual of Eq. (54) and (55), respectively. Then, the square of the norm of the total residual is evaluated in the form

$$\begin{aligned}||\underset{\sim}{\mathcal{R}}_T||^2 &= \begin{pmatrix}\underset{\sim}{\mathcal{R}}_v^T & \underset{\sim}{\mathcal{R}}_\mu^T\end{pmatrix}\begin{bmatrix}M^{-1} & 0\\ 0 & \bar{M}^{-1}\end{bmatrix}\begin{pmatrix}\underset{\sim}{\mathcal{R}}_v\\ \underset{\sim}{\mathcal{R}}_\mu\end{pmatrix} \\ &= \underset{\sim}{\mathcal{R}}_v^T M^{-1}\underset{\sim}{\mathcal{R}}_v + \underset{\sim}{\mathcal{R}}_\mu^T \bar{M}^{-1}\underset{\sim}{\mathcal{R}}_\mu\end{aligned} \quad (57)$$

Likewise, the square of the norm of the residual of the equation constraints, given by Eq. (6), is obtained by

$$||\underline{\mathcal{R}}_C^T||^2 = \underline{\dot{\psi}}^T\overline{M}\,\underline{\dot{\psi}}. \quad (58)$$

For a given tolerance level, the calculations are assumed to have converged when both $||\underset{\sim}{\mathcal{R}}_T||$ and $||\underline{\mathcal{R}}_C||$ are smaller than the selected tolerance value.

5 Numerical results

In this section, a set of characteristic results is presented for three challenging mechanical examples.

Special emphasis is placed on highlighting the advantages of the integration scheme developed. For this, the attention is focused on comparing numerical results with analytical results as well as with similar results arising by employing standard DAE solvers for two classical nonholonomic systems. Finally, comparison with results for a benchmark problem involving nonholonomic constraints is also performed.

Specifically, the numerical solutions captured by the method presented in this work (indicated by the label NM-ODE) are compared with existing analytical solutions [5] and results of a well-known benchmark problem [24, 25]. Moreover, the results are compared with similar results, obtained by using two state-of-the-art commercial codes [26, 27]. More specifically, these codes set up the equations of motion as a system of high-index DAEs and solve them numerically, by employing classical integration schemes, based on backward differentiation formulas. In particular, the GSTIFF and DASPK method were selected in solving the equations of motion by ADAMS and Motion Solve. Finally, results obtained by applying the new method after setting

$$h_R = \lambda^R \quad \text{with} \quad \overline{m}_{RR} = 0, \ \overline{c}_{RR} = 1 \quad \text{and} \quad \overline{f}_R = 0, \tag{59}$$

in Eq. (20), for $R = 1, \ldots, k$, are also included. In this way, the set of equations employed is reduced to the following system of equations of motion

$$(M(\underline{q})\dot{\underline{q}})^{\cdot} + \underline{h}(\underline{q}, \dot{\underline{q}}) = \underline{f}(\underline{q}, \dot{\underline{q}}, t) + A^T(\underline{q})\underline{\lambda}, \tag{60}$$

which is identical to that employed by multibody dynamics DAE formulations [28–30]. Moreover, the value

$$\overline{m}_{RR} = 1 \tag{61}$$

is also selected in Eq. (23). In the sequel, this modified set of equations is referred to as MM-DAE.

5.1 Sphere rolling inside a cylinder

In the first example, motion of a sphere of unit mass and radius a, rolling on the rough inside wall of a fixed vertical cylinder with radius b, is investigated. The sphere and the cylinder, together with a fixed coordinate system $Oxyz$ and a rotating coordinate system $Ox_1y_1z_1$, are shown in Fig. 1. Moreover, the action of gravity is along the negative Oz axis. Then, if

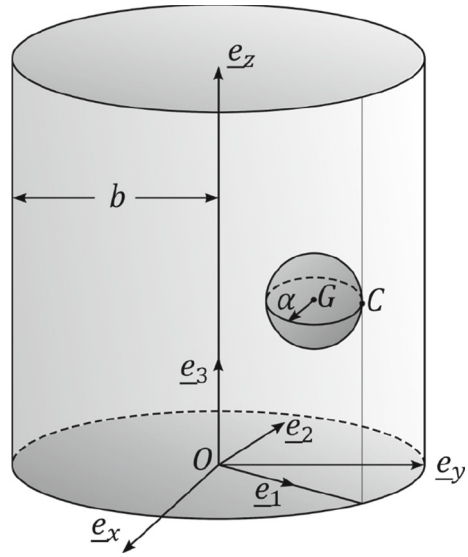

Fig. 1 A sphere rolling on a fixed vertical cylinder

$\{\underline{e}_x \quad \underline{e}_y \quad \underline{e}_z\}$ and $\{\underline{e}_1 \quad \underline{e}_2 \quad \underline{e}_3\}$ are appropriate orthonormal bases for the coordinate frames $Oxyz$ and $Ox_1y_1z_1$, respectively, it easily turns out that

$$\begin{aligned} \underline{e}_1 = \underline{e}_2 \times \underline{e}_3 = \frac{\underline{e}_z \times \underline{r}}{||\underline{e}_z \times \underline{r}||} \times \underline{e}_z \quad \Rightarrow \\ \underline{e}_1 = \frac{1}{||\underline{e}_z \times \underline{r}||}(I_3 - E)\underline{r}, \end{aligned} \tag{62}$$

where $\underline{r} = \underline{r}_{OG}$ is the position vector of the center of mass G of the sphere, I_3 is the 3×3 identity matrix and

$$E = \underline{e}_z \underline{e}_z^T.$$

In addition, if C is the contact point of the sphere with the cylinder, the condition of rolling without slipping of the sphere on the cylinder is expressed by

$$\underline{v}_C = \underline{0} \quad \Rightarrow \quad \underline{v}_G + \underline{\omega} \times \underline{r}_{GC} = \underline{0}, \tag{63}$$

where $\underline{\omega}$ is the angular velocity of the sphere and $\underline{r}_{GC} = a\underline{e}_1$. Therefore, after relating the components of the angular velocity of the sphere in the $Oxyz$ and a body frame (fixed in the sphere), through the corresponding rotation matrix R, i.e.,

$$\underline{\omega} = R\,\underline{\Omega},$$

the rolling condition is put eventually in the form

$$\underline{v}_G + \tilde{\omega}\,\underline{r}_{GC} = \underline{v}_G - \tilde{r}_{GC}(R\,\underline{\Omega}) = \underline{0}.$$

As usual, $\tilde{\omega}$ represents the 3×3 antisymmetric matrix

corresponding to the 3-vector $\underline{\omega}$ [1]. Equivalently, after performing some straightforward operations, this leads to

$$\left[I_3 - \frac{a}{||\underline{e}_z \times \underline{r}||}\widetilde{\{(I_3 - E)\underline{r}\}}R\right]\begin{pmatrix}\underline{v}_G \\ \underline{\Omega}\end{pmatrix} = \underline{0}. \tag{64}$$

This condition belongs to the class of constraints expressed by Eq. (3) with

$$A = \left[I_3 - \frac{a}{||\underline{e}_z \times \underline{r}||}\widetilde{\{(I_3 - E)\underline{r}\}}R\right] \quad \textit{and}$$
$$\underline{v} = \begin{pmatrix}\underline{v}_G \\ \underline{\Omega}\end{pmatrix}.$$

An exact solution for this problem can be found in pp. 95–98 of ref. [5].

Next, a selected set of numerical results is presented in Fig. 2. These results were obtained for the parameter values $a = 0.4$ m, $b = 1$ m, $z_0 = 1.224$ m, $\omega_{z0} =$

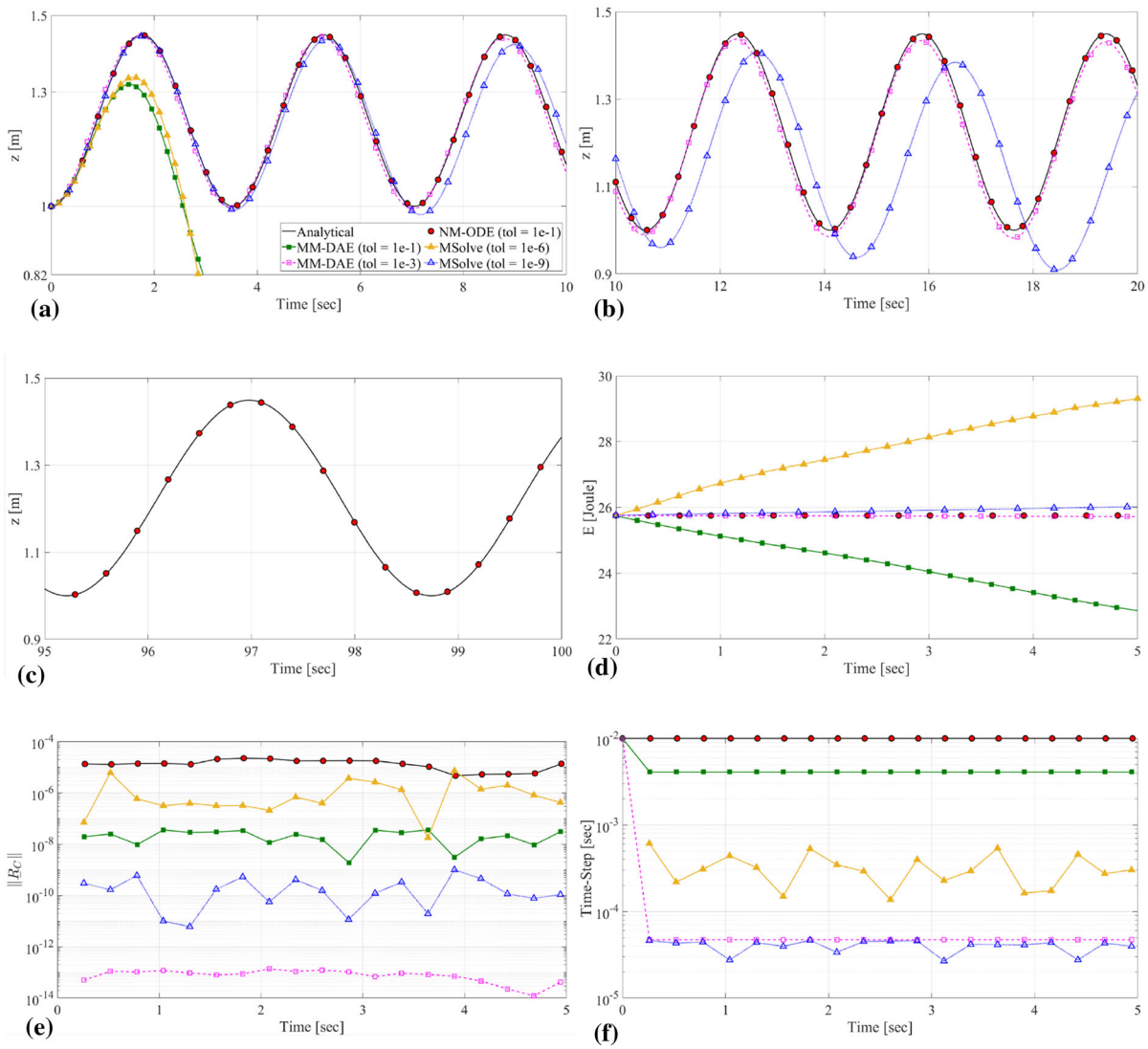

Fig. 2 Rolling sphere on a fixed vertical cylinder: **a** vertical displacement z of the sphere center for $0 < t < 10$ sec, **b** displacement z for 10 sec $< t < 20$ sec, **c** displacement z for 95 sec $< t < 100$ sec, **d** mechanical energy for $0 < t < 5$ sec, **e** norm $||\underline{\mathcal{R}}_C||$ of the residual of the constraint equations for $0 < t < 5$ sec, **f** time step for $0 < t < 5$ sec

$\omega_{30} = -5$ rad/s and a gravity acceleration $g = 9.81\ \mathrm{m/s^2}$. Also, selecting the amplitude $A = 1$ m and the phase $\mathscr{E} = 0$ in the exact expression of ω_2, presented by Eq. (2.24) in [5], leads to the initial angular velocity values $\omega_{20} = 0$ and $\omega_{10} = -20.265$ rad/s.

First, in Fig. 2a are presented results for the time history of the vertical displacement z of the sphere center. According to the analytical predictions, the sphere center executes a harmonic oscillation. This is obvious from the continuous curve in Fig. 2a, representing the exact solution. Also, identical results were obtained by the new numerical method (NM-ODE), represented by the full dots. However, the picture arising by employing the DAE solvers is quite different. First, only Motion Solve run and provided results for this example, while attempts to run it with ADAMS were not successful. In any case, the results obtained by Motion Solve, using the higher tolerance value, start deviating from the very beginning. In addition, decreasing the tolerance value causes a temporary improvement, which is quickly lost as time increases. This is verified by the results presented in Fig. 2b. Similar behavior is exhibited by the MM-DAE method, for two different tolerance values. In fact, such behavior is typical of all DAE formulations. Due to the missing terms in the equations of motion and the constraint equations, as presented by Eqs. (59) and (60), initial errors in the numerical values increase with time and explode eventually, independently on their initial magnitude and the tolerances imposed. On the other hand, including the extra terms, as presented in Sect. 2, the new method is capable of preserving numerical stability in the calculations. For instance, the results of Fig. 2c indicate this accurate and stable performance of the new method for a much later time interval. This provides a clear demonstration of the advantages associated with the new set of equations of motion employed and the numerical scheme developed in this work.

The gradual worsening in the quality of the solution obtained by the DAE solvers, as time goes on, can also be realized by looking at the mechanical energy of the system. Since there is no slipping event, the system examined is conservative. This means that its mechanical energy should be preserved. This is clearly the case for the results obtained by the new method, as illustrated in Fig. 2d. On the contrary, the results obtained by the DAE solvers exhibit a gradual and noticeable change in the mechanical energy. Similar behavior has also been observed and reported in other applications [8, 9, 15].

Finally, some more light is shed on the performance of the numerical methods employed by looking at the time variation of the norm $||\underline{\mathcal{R}}_C||$ of the residual of the constraint equations and the corresponding time steps. In all cases, an initial value of 0.01 s was selected for the time step. First, the results in Fig. 2e and f indicate that the new method converges to the correct solution by keeping the time step constant to the initial value of 0.01 s throughout the calculations, for a prespecified value of the tolerance $||\underset{\sim}{\mathcal{R}}_{\mathcal{T}}||$. On the other hand, the results in Fig. 2e demonstrate that in order for all the DAE numerical schemes applied to converge to a solution, independently of how good or bad their ultimate accuracy is, the norm $||\underline{\mathcal{R}}_C||$ should be reduced to extremely small levels eventually. As a consequence, the corresponding time step must also be reduced dramatically (i.e., more than two orders of magnitude), as verified by the results included in Fig. 2f. This demonstrates the numerical efficiency of the new method.

5.2 The falling rolling disk on a horizontal plane

In the second example, dynamics of a circular rigid disk, rolling without sliding on a horizontal plane, is examined. The disk has a unit mass and its radius is a. Moreover, the diametral and polar mass moments of inertia of the disk with respect to its center of mass G are known. The position of the disk is completely specified by five generalized coordinates, defined in Fig. 3. Namely, the coordinates x and y of the lowest point A of the disk, which is in contact with the horizontal plane, with respect to a frame $Oxyz$, together with three orientation angles θ, φ and ψ. The frame $Oxyz$ is fixed on the ground, so that the gravity force is along the negative Oz axis, with gravity acceleration g.

If $\{\underline{e}_x \quad \underline{e}_y \quad \underline{e}_z\}$ and $\{\underline{e}_\xi \quad \underline{e}_\eta \quad \underline{e}_\zeta\}$ are orthonormal bases for the fixed coordinate frame $Oxyz$ and the moving frame $O\xi\eta\zeta$, shown in Fig. 3, respectively, it turns out that

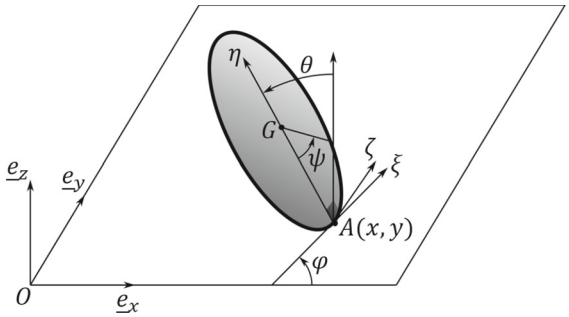

Fig. 3 A falling rolling disk on a horizontal plane

$$\underline{e}_\eta = \underline{e}_\zeta \times \underline{e}_\xi = \underline{e}_\zeta \times \left(\frac{\underline{e}_z \times \underline{e}_\zeta}{||\underline{e}_z \times \underline{e}_\zeta||}\right) \quad \Rightarrow \quad \underline{e}_\eta = \frac{(I_3 - RER^T)\underline{e}_z}{||(I_3 - RER^T)\underline{e}_z||}, \tag{65}$$

where R is the rotation matrix between the fixed frame $Oxyz$ and a frame fixed on the disk, while

$$\underline{e}_\zeta = R\underline{e}_2 \text{ and } E = \underline{e}_2\underline{e}_2^T = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Then, if the disk center of mass is located at point G

$$\underline{r}_{GA} = -a\underline{e}_\eta$$

and the condition of rolling of the disk without slipping on the ground is expressed by

$$\underline{v}_A = \underline{0} \quad \Rightarrow \quad \underline{v}_G + \underline{\omega} \times \underline{r}_{GA} = \underline{0}, \tag{66}$$

where $\underline{\omega}$ is the angular velocity of the disk. Therefore, after some manipulation, the rolling condition can eventually be written in the form

$$\left[I_3 - \frac{a}{||(I_3 - RER^T)\underline{e}_z||} \widetilde{\{(I_3 - RER^T)\underline{e}_z\}} R \right] \begin{pmatrix} \underline{v}_G \\ \underline{\Omega} \end{pmatrix} = \underline{0} \tag{67}$$

which belongs to the class of constraints expressed by Eq. (3).

In all the subsequent numerical examples, the disk radius was selected as $a = 1\ m$, the diametral and polar mass moment of inertia of the disk with respect to its center G was 0.25 and 0.5 $kg\,m^2$, respectively, while the gravity acceleration was chosen as $g = 9.81\ m/s^2$. Also, without a loss in generality, the initial values of the coordinates x, y, φ and ψ were selected to be zero. Finally, for compatibility in the notation with [5], the angular velocity $\underline{\omega}$ of the disk is expressed in the moving frame $O\xi\eta\zeta$ as

$$\underline{\omega} = -\dot{\theta}\underline{e}_\xi + \dot{\varphi}(\cos\theta\underline{e}_\eta + \sin\theta\underline{e}_\zeta) + \dot{\psi}\underline{e}_\zeta \equiv p\underline{e}_\xi + q\underline{e}_\eta + r\underline{e}_\zeta, \tag{68}$$

so that the components of the angular velocity in the moving frame are

$$p = -\dot{\theta}, \quad q = \dot{\varphi}\cos\theta, \quad = \dot{\psi} + \dot{\varphi}\sin\theta. \tag{69}$$

A quite complete analysis of the dynamics of the disk under the action of gravity, together with an investigation of certain stability issues, is presented in pp. 55–60 of ref. [5]. In general, an arbitrary set of initial conditions leads to complex motions, with a variable angle θ. However, certain combinations of the system initial parameters lead to special motions. One such category of motions are the cyclic motions, which are characterized by constant θ, q and r. Namely, for some special cases with $\theta = \theta_0 \neq 0$, the point of the disk in contact with the ground undergoes a circular orbit. Instead, when $\theta_0 = 0$, the disk can follow a uniform rectilinear rolling or a uniform spinning about a fixed vertical diameter [5].

First, in Fig. 4 are presented results obtained for $\theta_0 = 1^\circ$, $p_0 = q_0 = 0$ and three different initial values r_0. For this set of initial conditions, the results presented indicate that complex motions arise, with time varying θ, p, q and r. In all cases examined, the results of the new method are found to be virtually indistinguishable from those obtained by the analytical solution, verifying the accuracy of the new method. Moreover, in terms of dynamics, the results demonstrate that as the absolute value of r_0 increases, the magnitude of the variations of θ, p, q and r decreases gradually. Eventually, trajectories and dynamics with θ, q and r approaching constant values are reached, for sufficiently large absolute value of r_0, corresponding to cyclic motions.

The dynamics of the motions presented in Fig. 4 is illuminated in a better way by Fig. 5a, depicting the corresponding trajectories of the disk contact point on the horizontal plane. The orbits obtained exhibit a geometric complexity for a small absolute value of r_0. This complexity is diminished gradually as the absolute value of r_0 increases and the orbit tends to

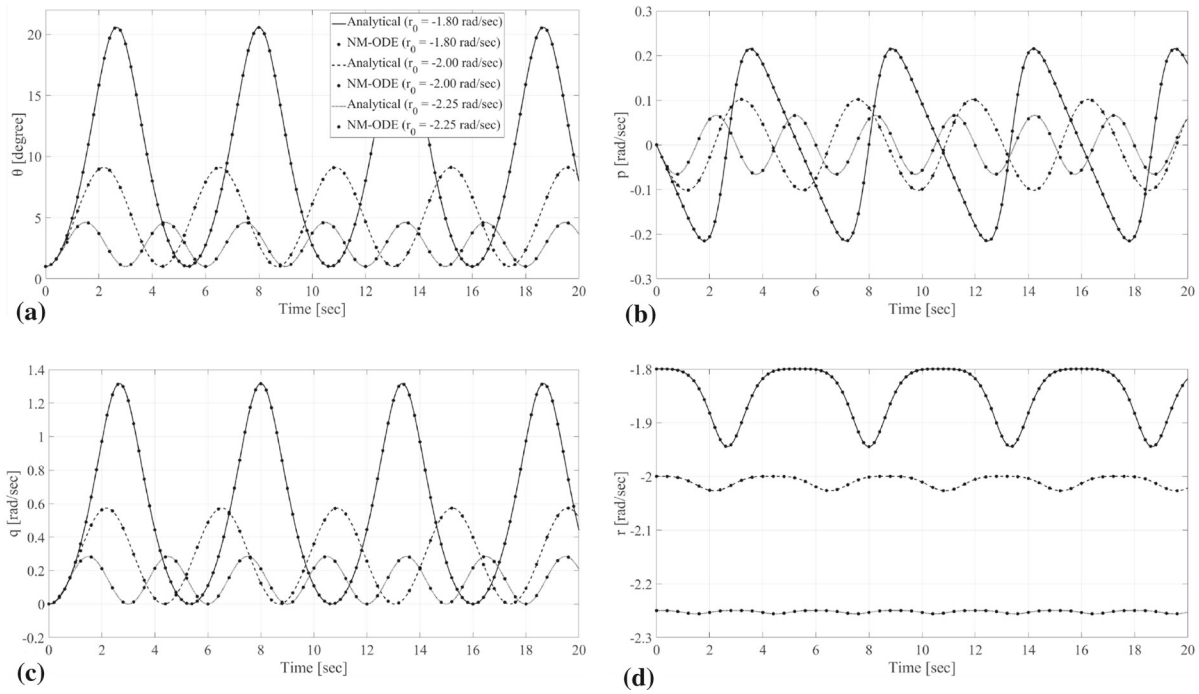

Fig. 4 Time histories for a rolling disk on a horizontal plane for $\theta_0 = 1^o$ and three different initial values r_0: **a** inclination angle θ, **b** p, **c** q and **d** r

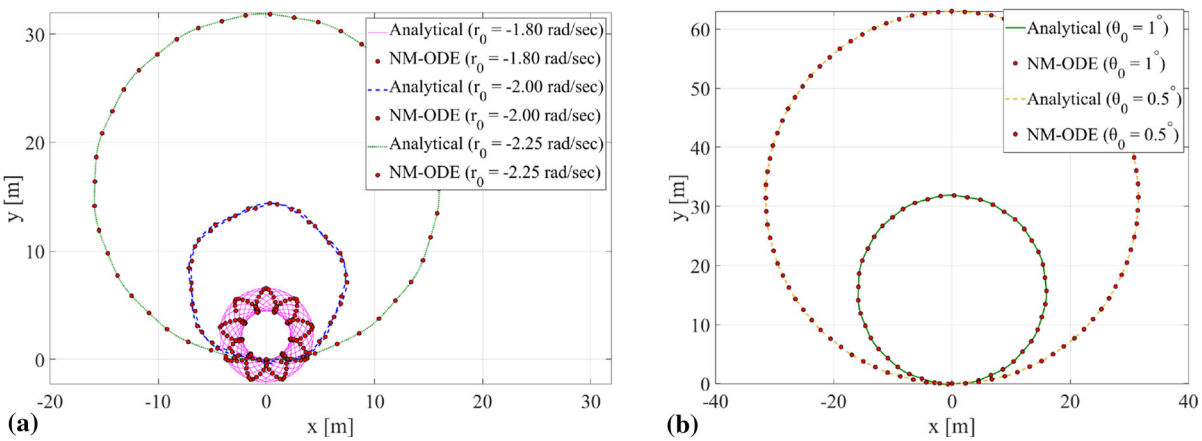

Fig. 5 Trajectories of the disk contact point on the horizontal plane: **a** $\theta_0 = 1^o$ and three different initial values r_0, **b** $\theta_0 = 0.5^o$ or $\theta_0 = 1^o$ and $r_0 = -2.25$ rad/s

reach a clean circular shape. In addition, the results of Fig. 5b, obtained for the largest absolute value of r_0, that is $r_0 = -2.25$ rad/s, with $\theta_0 = 0.5^o$, indicate that the radius of the circular trajectory increases as the value of θ_0 decreases. In this way, the motions examined reach another special type of motion, known as uniform rectilinear rolling, in the limit value $\theta_0 = 0$ [5].

Cyclic motions, with θ, q and r constant, can be captured from the start, through an appropriate choice of the initial conditions. For instance, the results presented in Fig. 6 were obtained for $\theta_0 = 20^o$, $p_0 = 0$ and $q_0 = 1$ rad/s, while the value of r_0 was selected by using Eq. (2.9) of [5]. Once again, the new method (NM-ODE) gives identical results with the analytical solution. In addition, the pathologies of the DAE

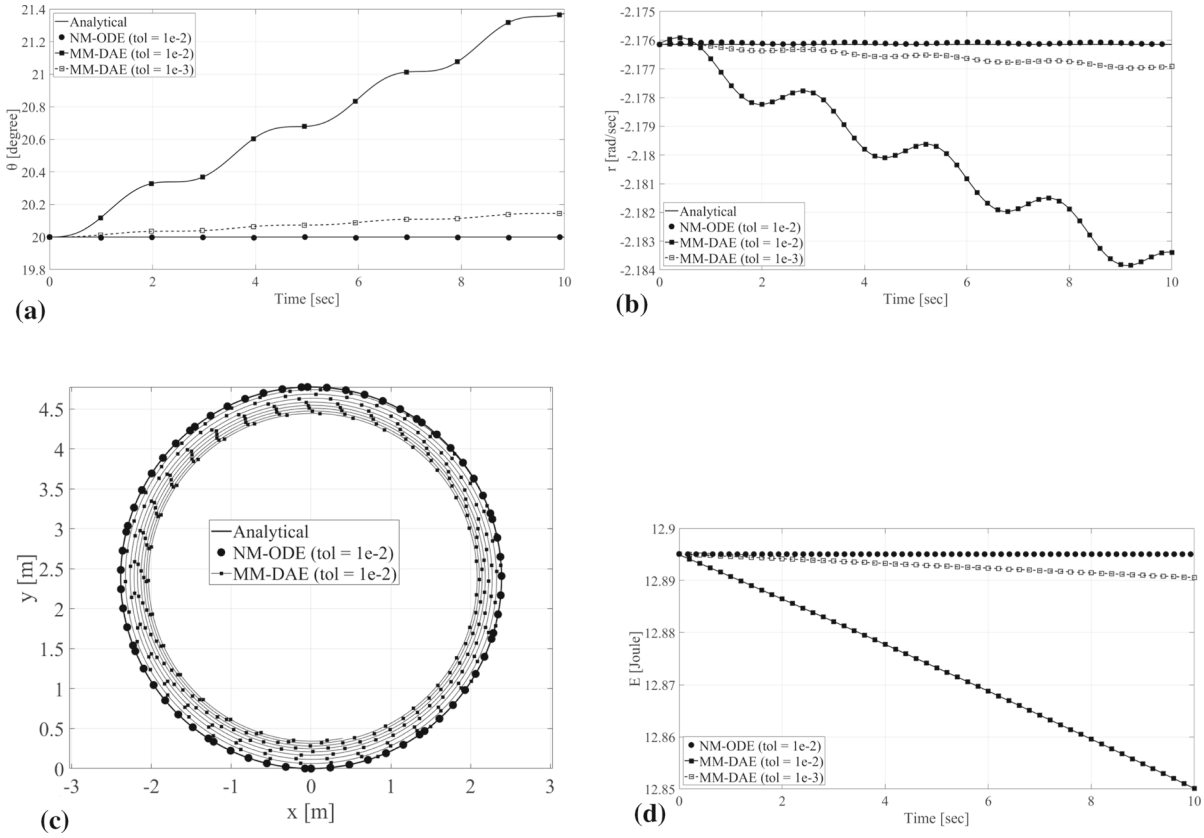

Fig. 6 Comparison of cyclic motions obtained by the analytical method, the new numerical method and its DAE counterpart: **a** history of the inclination angle θ, **b** history of angular velocity component r, **c** trajectory of the disk contact point on the horizontal plane and **d** mechanical energy of the disk

methods appear in this case as well. More specifically, results obtained by the MM-DAE method for the time histories of the inclination angle θ and angular velocity component r are first shown in Fig. 6a and b, respectively. These results demonstrate a gradual and monotonically increasing loss of accuracy in the computations, as time goes on. Furthermore, decreasing the tolerance value causes an improvement, which is only temporary. Similar behavior was also noticed for the other two angular velocity components p and q. As a consequence of these computational errors, the trajectory of the disk contact point on the horizontal plane starts deviating from the circular orbit, as shown in Fig. 6c, for the higher tolerance value. The result for the lower tolerance value is not included since it leads to similar behavior and just complicates the figure. In any case, the results indicate a gradual decrease in the radius of the orbit. This is justified by looking at the corresponding time histories of the disk mechanical energy, which are included in Fig. 6d. Obviously, the new method predicts a constant energy state, which is correct, while omission of the extra terms in the governing equations and the constraint equations leads to a gradual loss of energy.

Finally, the study focused on motions with $\theta = 0$, where the disk executes a uniform rectilinear rolling. In the first set of results, presented in Fig. 7, the initial values $\theta_0 = 0, p_0 = 0$ and $q_0 = 0$ were selected. At the same time, the initial value r_0 was chosen based on results of a stability analysis. Specifically, a critical value r_{cr} was selected, based on Eq. (2.14) of [5]. For the chosen combination of parameters, this numerical value was $r_{cr} = 1.808$ rad/s. Results were also obtained for a higher value ($r_0 = 2r_{cr}$) and a lower value ($r_0 = 0.9\, r_{cr}$), corresponding to a supercritical and a subcritical case, respectively. For each of these cases, the trajectory of the disk contact point on the horizontal plane and the corresponding history of the

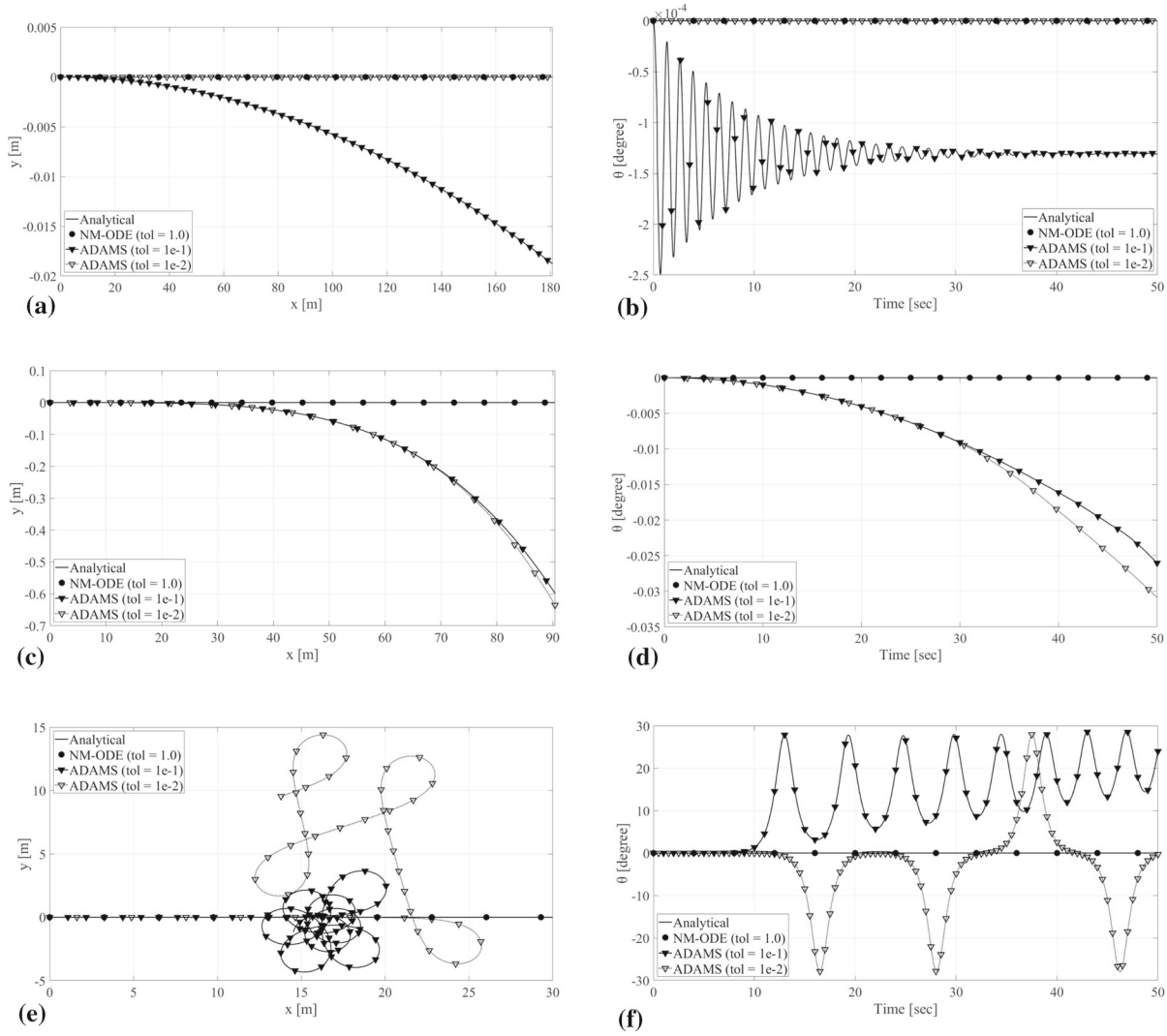

Fig. 7 Uniform rectilinear rolling motions of the disk ($p_0 = 0$). Trajectory of the disk contact point on the horizontal plane and history of angle θ for: **a** and **b** a supercritical case, with $r_0 = 2r_{cr}$, **c** and **d** the critical case ($r_0 = r_{cr}$), **e** and **f** a subcritical case, with $r_0 = 0.9\, r_{cr}$

inclination angle θ are shown in Fig. 7. In all cases, the results of the new method were found to coincide with those obtained by the analytical solution, once again. Specifically, the trajectory of the disk contact point on the horizontal plane was a straight path, while the angle θ remained zero for all times. In fact, this was possible even after selecting an arbitrarily large time step and tolerance value, since the specific motion (translation along a line plus rotation about a fixed axis) takes place along an autoparallel curve of the configuration space. However, the picture was quite different for the results obtained by using ADAMS. Namely, significant deviations were observed to occur, especially in the critical case and became even more pronounced in the subcritical case. Moreover, it is apparent that decreasing the tolerance level in the calculations does not prevent initial errors from increasing as time increases.

Some important and rapid changes were observed to occur in the disk motion by imposing a small perturbation in the value of p_0. For instance, in Fig. 8 is presented a similar set of results, obtained for the same set of parameters as in the previous example, but with $p_0 = -0.1$ rad/s, instead. In this case, the trajectory exhibits some noticeable deviations from a straight line. At the same time, the inclination angle θ

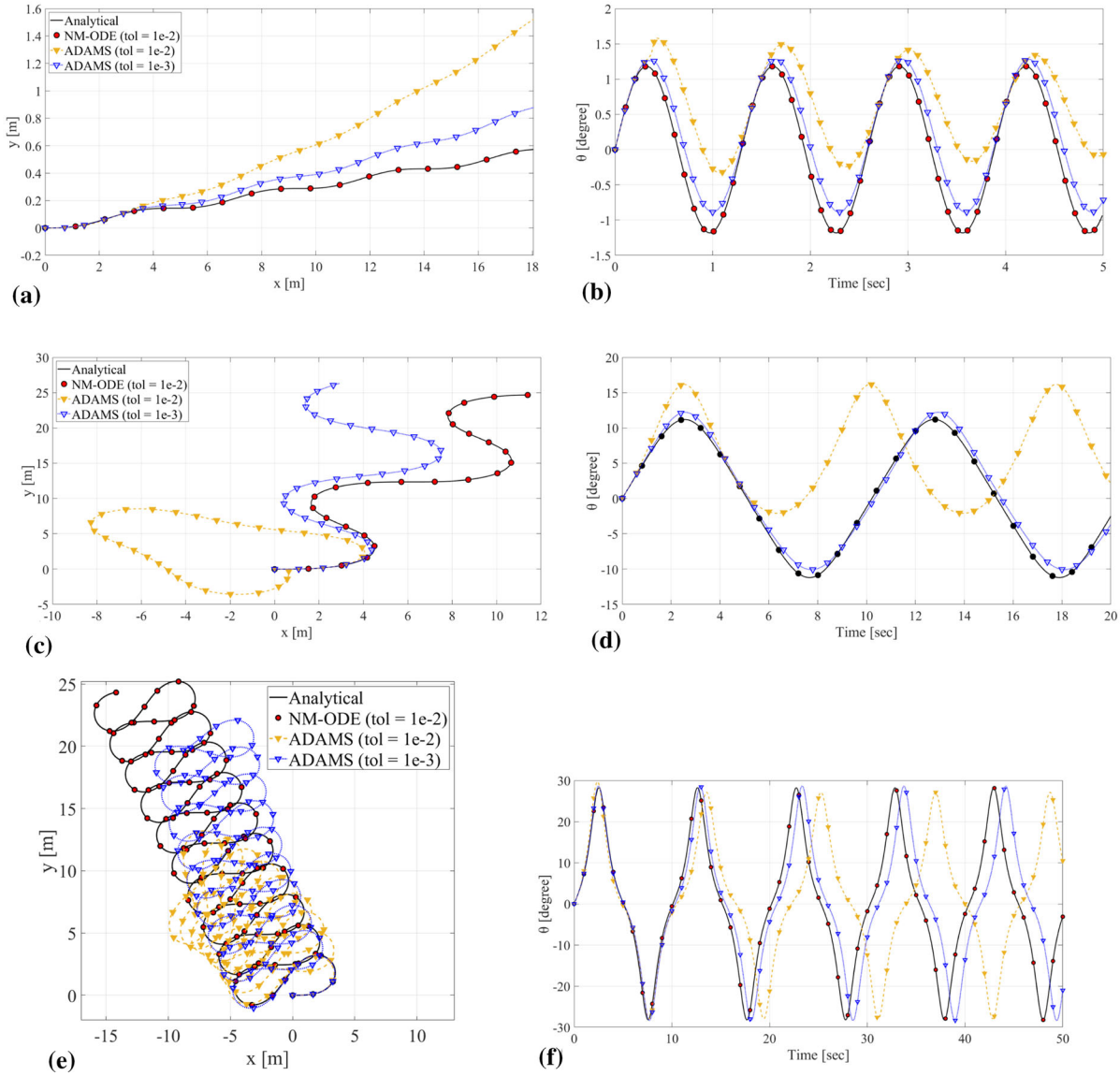

Fig. 8 Perturbed rectilinear rolling motions of the disk ($p_0 = -0.1$ rad/s). Trajectory of the disk contact point on the horizontal plane and history of inclination angle θ for: **a** and **b** a supercritical case, with $r_0 = 2\,r_{cr}$, **c** and **d** the critical case ($r_0 = r_{cr}$) and **e** and **f** a subcritical case, with $r_0 = 0.9\,r_{cr}$

also takes nonzero values. These deviations are relatively small in the supercritical range but are amplified quickly with a reduction in the value of parameter r_0, especially for values smaller than r_{cr}, where the changes are quite rapid and the motion deviates quickly and significantly from the rectilinear motion. Once again, the results of the new method coincide with those obtained by utilizing the analytical solution, in all cases examined. In contrast, all the predictions of the DAE method employed are quite unreliable. Moreover, these predictions are subject to qualitatively similar errors with the case $p_0 = 0$.

5.3 The uncontrolled bicycle

In the final example, results of the new method are compared with results available for a benchmark problem [31]. More specifically, response of an uncontrolled bicycle problem presented first in [24] is investigated. The corresponding mechanical model

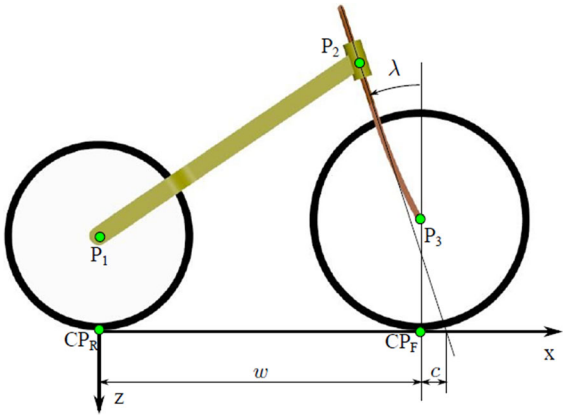

Fig. 9 The uncontrolled bicycle (taken from [31])

consists of four rigid members (i.e., the rear and front frame assembly, together with the rear and front wheel), as shown in Fig. 9. All the parameters of this model are taken from [31].

Assuming that both wheels roll on the ground without slipping, introduces nonholonomic constraints in the formulation. Eventually, the model examined possesses three degrees of freedom. Namely, the tilt or roll angle, the steering angle and the forward displacement of the bicycle. The bicycle is under gravity forcing, acting along the z axis. Also, at the starting position, the bicycle is in a vertical position and the steering is straight. Then, it is given an initial forward velocity, together with some roll velocity as a perturbation. According to a stability analysis of the reference position [24, 25], the subsequent motion of the bicycle is stable only when the forward speed lies within an interval of two critical values, say v_{c1} and v_{c2}.

In particular, when the forward speed of the bicycle is smaller than v_{c1}, the instability is static in nature [24, 25]. In accordance to the benchmark problem, selecting a forward speed of 4.0 m/s and a roll angle speed of 0.05 rad/s leads to Figs. 10a–c for the time histories of the forward velocity, the roll angle and the steer angle, respectively, over the first 20 s of the motion. Likewise, in Fig. 10d is shown the corresponding mechanical energy E_m of the system. In all cases, the benchmark results are represented by continuous curves, while the results of the new method are indicated by broken lines with full dots.

The virtual coincidence of the results illustrates the accuracy of the new method. Moreover, the corresponding percentage of variation of the mechanical energy E_m [31], defined by

$$p_E = 100\,[\max(E_m) - \min(E_m)]/E_{m0}, \tag{70}$$

was calculated to be $5.5 \cdot 10^{-4}$, which is much better than the value of $17.0 \cdot 10^{-4}$, obtained by the reported benchmark results.

Similarly, the results presented in Fig. 11 were obtained for an initial forward velocity value of the bicycle lying between the two critical values v_{c1} and v_{c2}, leading to an asymptotically stable solution [24, 25]. More specifically, these results were obtained for a forward speed of 4.6 m/s and a roll angle speed of 0.50 rad/s. In addition, the corresponding percentage of variation of the mechanical energy was calculated to be $5.2 \cdot 10^{-4}$, which is better than the value of $9.5 \cdot 10^{-4}$ determined by the reported benchmark results. Once again, the results illustrate the numerical accuracy and stability properties of the new method.

Finally, similar conclusions can be drawn by considering the set of results presented in Fig. 12. These results were obtained for an initial forward velocity value of the bicycle greater than v_{c2}, leading to an oscillatory unstable solution [24, 25]. More specifically, these results were obtained for a forward speed of 8.0 m/s and a roll angle speed of 0.05 rad/s. In this case, the corresponding percentage of variation of the mechanical energy was calculated to be $0.105 \cdot 10^{-4}$, which is about the same with the benchmark value of $0.106 \cdot 10^{-4}$.

In closing this section, it is mentioned that results obtained by application of all the DAE methodologies applied in the earlier examples were quite different than the benchmark results, in all three cases examined for the bicycle example. Also, qualitatively similar results were obtained for other typical mechanical systems subject to nonholonomic constraints, arising from a sliding rather than a rolling action (e.g., knife-edge, Chaplygin's sleigh [4, 5]).

6 Synopsis and extensions

The basic ingredients of a new method developed for the numerical integration of the equations of motion governing the behavior of multibody mechanical systems, involving nonholonomic motion constraints, were presented in the first part of this work. This

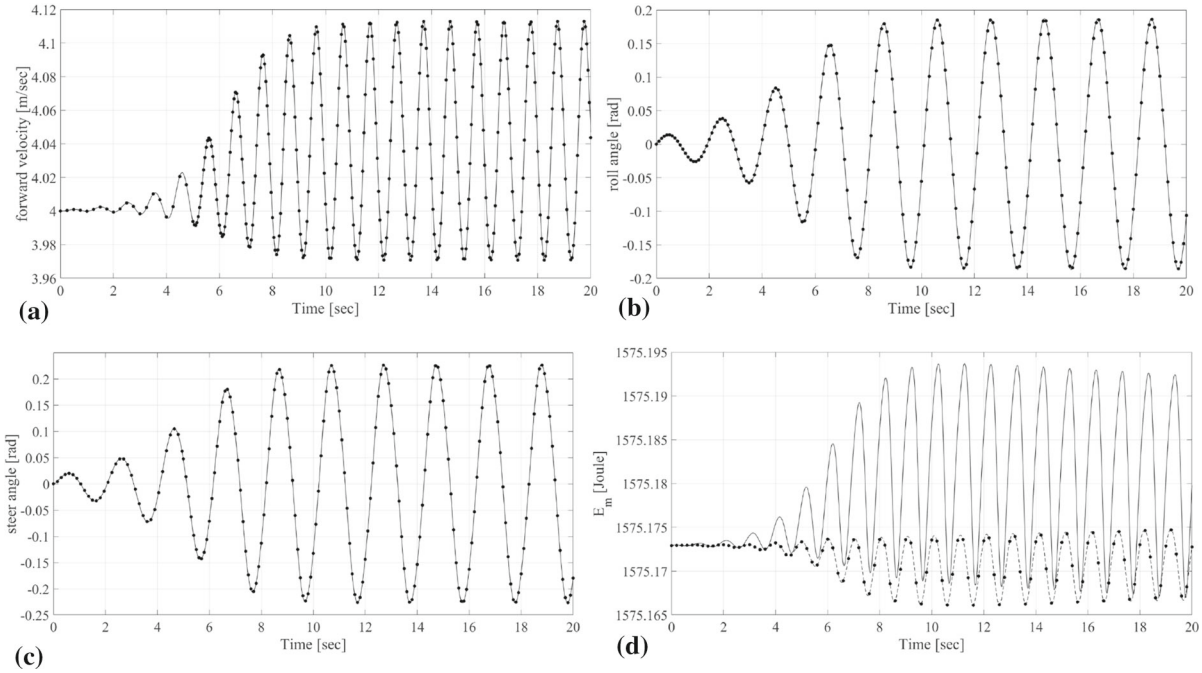

Fig. 10 Time histories for the: **a** forward velocity, **b** roll angle, **c** steer angle and **d** mechanical energy of the bicycle, for an initial forward velocity lower than v_{c1} (subcritical range)

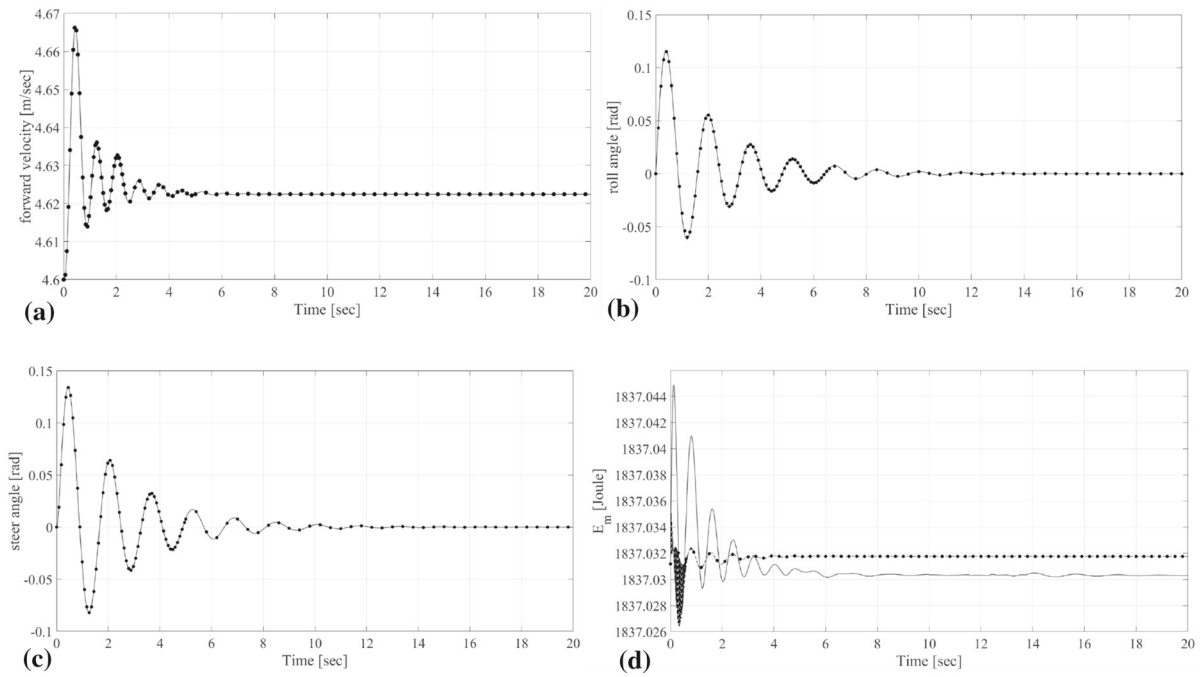

Fig. 11 Time histories for the: **a** forward velocity, **b** roll angle, **c** steer angle and **d** mechanical energy of the bicycle, for an initial forward velocity between v_{c1} and v_{c2} (stable range)

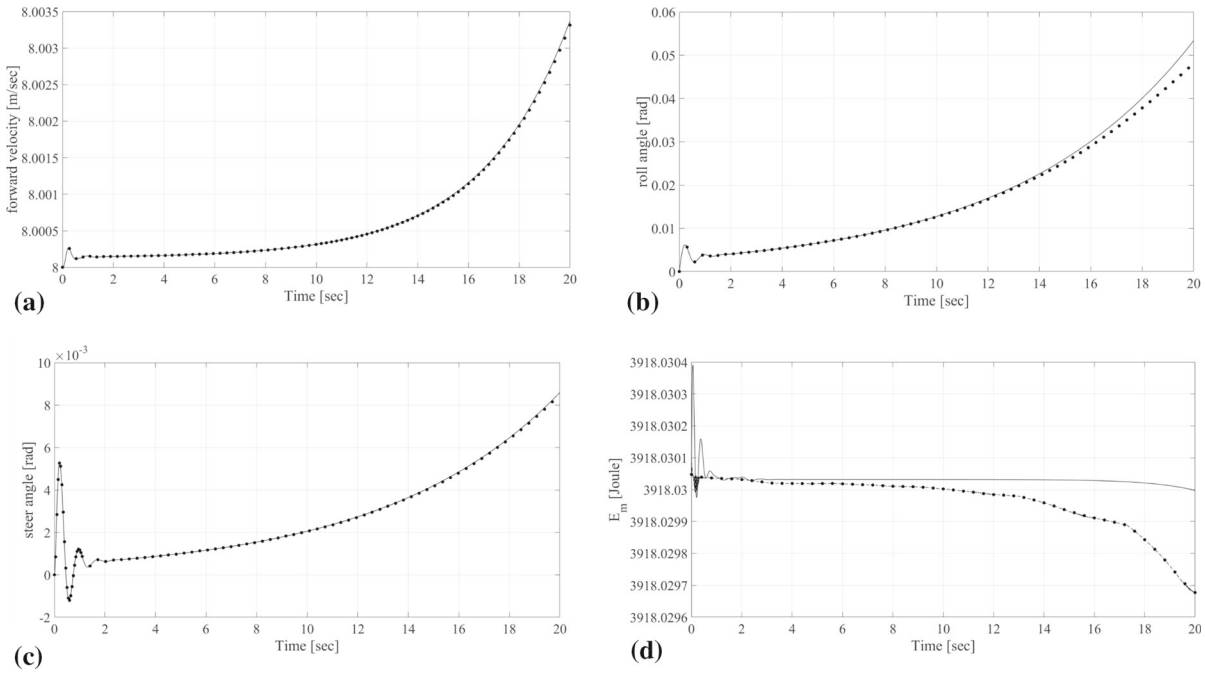

Fig. 12 Time histories for the: **a** forward velocity, **b** roll angle, **c** steer angle and **d** mechanical energy of the bicycle, for an initial forward velocity larger than v_{c2} (supercritical range)

method was founded on several new developments. Initially, an appropriate set of equations of motion was employed, consisting of a coupled and strongly nonlinear system of second-order ODEs for both the generalized coordinates and the Lagrange multipliers. These equations include suitable terms, causing an automatic stabilization and scaling of the governing equations and the equations of the constraints. These terms were obtained by a systematic and consistent application of Newton's law of motion, avoiding ad hoc and incorrect selections of parameters. Then, these equations were put eventually in a three-field weak form, by treating the generalized coordinates, velocities and momenta as independent quantities. Finally, a suitable temporal discretization scheme was developed for the class of systems examined, based on this solid theoretical foundation.

The robustness, accuracy and efficiency of the new numerical method were demonstrated in the second part of this study by applying it to typical mechanical examples of nonholonomic mechanics. First, dynamics of the classical example of a sphere rolling inside a cylinder was examined. Then, dynamics of a falling rolling disk on a horizontal plane was investigated in detail. For both of these examples, analytical solutions are available. In all cases examined, the results of the new method were found to be virtually coincident with those obtained by the analytical solution, validating the excellent accuracy and stability properties as well as the numerical efficiency of the new method in the best possible way. Moreover, special emphasis was placed on comparing the results of the new method with those obtained by application of classical DAE approaches. This brought up and illustrated the classical instability problems arising when using such methods. Similar conclusions were also drawn by considering a well-known benchmark problem, referring to response and stability of an uncontrolled bicycle model. Once again, comparison of the numerical results verified the accuracy and stability properties of the new method.

The new method can easily be applied to more complex and challenging mechanical systems, possessing a large number of degrees of freedom [32]. In such cases, it is quite possible to run into cases involving the presence of redundant constraints and/or the appearance of singular configurations during the motion of the system. Then, in order to be able to accommodate such numerically difficult situations, the new method should first be extended to an

augmented Lagrangian formulation. Finally, the present formulation can provide a solid basis to develop a numerical scheme for systems with unilateral constraints. In fact, it will be quite helpful in efforts to examine special situations arising in contact problems, where conditions of contact are fulfilled temporarily or change instantly [33–35]. In such cases, where the unilateral constraints can be treated as bilateral up to their violation, the constraints cannot be put in an integrable form and should therefore be treated as nonholonomic.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Geradin, M, Cardona, A: Flexible Multibody Dynamics. Wiley , NY (2001)
2. Shabana, A.A.: Dynamics of Multibody Systems, 3rd edn. Cambridge University Press, NY (2005)
3. Bauchau, O.A.: Flexible Multibody Dynamics. Springer, London (2011)
4. Bloch, A.M.: Nonholonomic Mechanics and Control. Springer, NY (2003)
5. Neimark, J.I., Fufaev, N.A.: Dynamics of Nonholonomic Systems. Translations of Mathematical Monographs, American Mathematical Society **33**, Providence, RI (1972)
6. Brenan, K.E., Campbell, S.L., Petzhold, L.R.: Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations. North-Holland, NY (1989)
7. Bauchau, OA, Epple, A, Bottasso, C.L.: Scaling of constraints and augmented Lagrangian formulations in multibody dynamics simulations. ASME J. Comput. Nonlinear Dyn. **4**, 021007 (2009)
8. García Orden, J.C.: Energy considerations for the stabilization of constrained mechanical systems with velocity projection. Nonlinear Dyn. **60**, 49–62 (2010)
9. García Orden, J.C., Conde, S.C.: Controllable velocity projection for constraint stabilization in multibody dynamics. Nonlinear Dyn. **68**, 245–257 (2012)
10. Paraskevopoulos, E., Natsiavas, S.: On application of Newton's law to mechanical systems with motion constraints. Nonlinear Dyn. **72**, 455–475 (2013)
11. Natsiavas, S., Paraskevopoulos, E.: A set of ordinary differential equations of motion for constrained mechanical systems. Nonlinear Dyn. **79**, 1911–1938 (2015)
12. Paraskevopoulos, E., Natsiavas, S.: Weak formulation and first order form of the equations of motion for a class of constrained mechanical systems. Int. J. Non-linear Mech. **77**, 208–222 (2015)
13. Washizu, K.: Variational Methods in Elasticity and Plasticity, 3rd edn. Pergamon Press, Oxford (1982)
14. Felippa, C.: On the original publication of the general canonical functional of linear elasticity. J. Appl. Mech. **67**, 217–219 (2000)
15. Potosakis, N., Paraskevopoulos, E., Natsiavas, S.: Application of an augmented Lagrangian approach to multibody systems with equality motion constraints. Nonlinear Dyn. **99**, 753–776 (2020)
16. Greenwood, D.T.: Principles of Dynamics. Prentice-Hall Inc., Englewood Cliffs, NJ (1988)
17. Murray, R.M., Li, Z, Sastry, S.S.: A Mathematical Introduction to Robot Manipulation. CRC Press, Boca Raton, FL (1994)
18. Papastavridis, J.G.: Tensor Calculus and Analytical Dynamics. CRC Press, Boca Raton (1999)
19. Frankel, T.: The Geometry of Physics: An Introduction. Cambridge University Press, NY (1997)
20. Rektorys, K.: Variational Methods in Mathematics Science and Engineering. D. Reidel Publishing Company, Dordrecht (1977)
21. Ženišek, A.: Nonlinear Elliptic and Evolution Problems and their Finite Element Approximations. Academic Press, London (1990)
22. Paraskevopoulos, E., Natsiavas, S.: A new look into the kinematics and dynamics of finite rigid body rotations using Lie group theory. Int. J. Solids Struct. **50**, 57–72 (2013)
23. Paraskevopoulos, E.A., Panagiotopoulos, C.G., Talaslidis, D.G.: Rational derivation of energy/momentum-preserving time integration algorithms. In: *Application to dynamic response under moving vehicles. ECCOMAS Thematic Conf. Comput. Meth. Struct. Dyn. Earthq. Eng*. Crete, Greece (2007)
24. Meijaard, J.P., Papadopoulos, J.M., Ruina, A., Schwab, A.L.: Linearized dynamics equations for the balance and steer of a bicycle: a benchmark and review. Proc. R. Soc. A Math. Phys. Eng. Sci. 463 (2084), 1955–1982 (2007)
25. García-Agúndez, A., García-Vallejo, D., Freire, E.: Linearization approaches for general multibody systems validated through stability analysis of a benchmark bicycle model. Nonlinear Dyn. **103**, 557–580 (2021)
26. MSC ADAMS 2018.1, User Guide, MSC Software Corporation, California, USA.
27. MotionSolve v19.0, User Guide, Altair Engineering Inc., Irvine, California, USA.
28. Bayo, E., Ledesma, R.: Augmented Lagrangian and mass-orthogonal projection methods for constrained multibody dynamics. Nonlinear Dyn. **9**, 113–130 (1996)
29. Dopico, D., Gonzalez, F., Cuadrado, J., Kövecses, J.: Determination of holonomic and nonholonomic constraint reactions in an index-3 augmented Lagrangian formulation with velocity and acceleration projections. ASME J. Comput. Nonlinear Dyn. **9**, p. 041006 (2014)
30. Gonzalez, F., Dopico, D., Pastorino, R., Cuadrado, J.: Behaviour of augmented Lagrangian and Hamiltonian methods for multibody dynamics in the proximity of singular configurations. Nonlinear Dyn. **85**, 1491–1508 (2016)
31. IFToMM T.C. for Multibody Dynamics, Library of Computational Benchmark Problems, http://www.iftomm-multibody.org/benchmark
32. Papalukopoulos, C., Natsiavas, S.: Dynamics of large scale mechanical models using multi-level substructuring. ASME J. Comput. Nonlinear Dyn. **2**, 40–51 (2007)

33. Lubarda, V.A.: Dynamics of a light hoop with an attached heavy disk: inside an interaction pulse. J. Mech. Mat. Struct. **4**, 1027–1040 (2009)
34. Bronars, A., O'Reilly, O.M.: Gliding motions of a rigid body: the curious dynamics of Littlewood's rolling hoop. Proc. R. Soc. A **475**, 20190440 (2019)
35. Antali, M., Stepan, G.: Nonsmooth analysis of three-dimensional slipping and rolling in the presence of dry friction. Nonlinear Dyn. **97**, 1799–1817 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.