Time-Step-Size-Independent Conditioning And Sensitivity To Perturbations in the Numerical Solution of Index-Three Differential Algebraic Equations*

Carlo L. Bottasso
Dipartimento di Ingegneria Aerospaziale,
Politecnico di Milano, Milano, Italy
Olivier A. Bauchau
D. Guggenheim School of Aerospace Engineering,
Georgia Institute of Technology, Atlanta, GA, USA
Alberto Cardona
Cimec-Intec, Universidad Nacional del Litoral – Conicet,
Güemes 3450, 3000 Santa Fe, Argentina

Abstract

We propose a simple preconditioning for the equations of motion of constrained mechanical systems in index three form. The scaling transformation is applied to the displacement-velocity-multiplier and to the reduced displacement-multiplier forms. The analysis of the transformed system shows that conditioning and sensitivity to perturbations become independent of the time step size, as in the case of well behaved ordinary differential equations. The new scaling transformation is simple to implement and does not require the re-writing of the system equations as other approaches. The theoretical analysis is confirmed by numerical examples.

1 Introduction

It is well known that the amplification of small errors and perturbations in the solution of high index differential algebraic equations (DAEs) causes severe numerical difficulties. For example, Petzold and Lötstedt have shown in [1] that index three DAEs for constrained mechanical (multibody) systems are severely ill conditioned for small time step sizes when discretized using BDF-type formulas. Their analysis indicates that, unless corrective actions are taken, the condition number of the iteration matrix is $\mathcal{O}(h^{-3})$. Furthermore, errors propagate in the displacement, velocity and multiplier fields at rates which are shown to be $\mathcal{O}(h^{-1})$, $\mathcal{O}(h^{-2})$ and $\mathcal{O}(h^{-3})$, respectively. A related perturbation analysis that deals with the propagation of the error in the satisfaction of the constraint equations is described in Arnold [2], which also considers systems with friction. These results indicate that errors will grow very rapidly as the time step size is reduced, preventing in practice the use of time refinement procedures, and imposing a very tight tolerance on the solution of the non-linear discrete equations. Furthermore, the ill conditioning of the linear system does not favor the use of iterative solvers.

SIAM Journal on Scientific Computing, **29**(1), pp 397–414, 2007

Several methods have been proposed in the literature to address these problems. For example, the system can be recast in index two form, by replacing the displacement level constraints with the velocity level ones. While the index two form is more robust to perturbations [2], this approach suffers from the well known drift effect, which calls for further ad-hoc corrective actions.

An alternative approach is to re-write the governing equations so as to include, together with the position level constraints, also their derivatives, as in the GGL [3] and Embedded Projection (EP) [4] methods. These methods however require additional multipliers which imply a somewhat increased problem size and, hence, a possibly higher computational cost.

Petzold and Lötstedt [1] discuss a simple scaling transformation of the index three governing equations which yields a condition number $\mathcal{O}(h^{-2})$ and an improvement of one order in the errors for all solution fields. Although the sensitivity to perturbations is reduced with respect to the unscaled problem, difficulties can still be expected in practice.

In this paper we propose a solution to the conditioning and error propagation problems that is based on a left and right preconditioning. The left preconditioning amounts to a scaling of the dynamic equilibrium and constraint equations, while the right preconditioning amounts to a scaling of the unknowns. The proposed procedure is trivially implemented in an existing code and does no require the re-writing of the equations of motion as in the GGL and EP methods.

The new scaling transformation is applied first to the equations in the standard displacement-velocity-multiplier form, using a scaling of the sole Lagrange multipliers. A similar scaling was previously considered in Cardona and Geradin [5] in the context of the Hilbert, Hughes and Taylor (HHT) scheme. The analysis of the present paper shows that errors propagate in the displacement, velocity and multiplier fields at rates equal to $\mathcal{O}(h^0)$, $\mathcal{O}(h^{-1})$ and $\mathcal{O}(h^0)$, respectively, while the condition number is $\mathcal{O}(h^{-1})$. Next, it is shown that if also the velocities are scaled together with the Lagrange multipliers, one can achieve perfect time step size independence ($\mathcal{O}(h^0)$) for error propagation and conditioning.

Finally, we analyze the displacement-multiplier form, obtained by static condensation of the velocities at each time step. This form of the equations is more computationally interesting than the full form for applications denoted by a large number of degrees of freedom, since it leads to the solution of smaller linear problems. Even in this case, the preconditioning of the reduced form leads to perfect time step size independence for both the conditioning and the error propagation in the displacement-multiplier solution variables.

These results indicate that index three DAEs are as easy to integrate as well behaved ordinary differential equations (ODEs), once they are recast in one of the $\mathcal{O}(h^0)$ forms.

The paper is organized as follows. In Section 2 we formulate the equations of motion of multibody systems in index three three-field form for BDF-type schemes, and we derive their sensitivity to perturbations and conditioning. Preconditioning by scaling of the discretized equations is presented in Section 3; the left preconditioning strategy is detailed in §3.1, while the left and right preconditioning is described in §3.2. The classical pendulum problem in §3.3 numerically confirms the results of the analysis and concludes this section. The two field form of the problem is discussed in Section 4; the left preconditioning is described in §4.2, the left and right preconditioning is given in §4.2, and the pendulum problem verifies the analysis in §4.3. Finally, a more complex contact-impact problem for a flexible multibody system is presented in Section 5 to show the importance of scaling in complex problems involving time step size refinement. The paper is closed by Section 6, which reports the main conclusions and findings of this work.

2 Sensitivity to perturbations and conditioning of multibody system equations

The governing DAEs for multibody systems are written as follows

$$\boldsymbol{u}' = \boldsymbol{v}, \tag{1a}$$
$$\boldsymbol{v}' = \boldsymbol{f} + \boldsymbol{G}\boldsymbol{\lambda}, \tag{1b}$$
$$\boldsymbol{\phi} = \boldsymbol{0}, \tag{1c}$$

where equations (1a) are the kinematic equations, equations (1b) represent the equations of dynamic equilibrium, $\boldsymbol{u}$ are generalized coordinates, $\boldsymbol{v}$ generalized velocities, $\boldsymbol{\lambda}$ are Lagrange multipliers that enforce the constraints (1c), and finally $\boldsymbol{G}^T = \boldsymbol{\phi}_{,\boldsymbol{u}}$ is the constraint Jacobian. Following [1], we have considered for simplicity an identity mass matrix in equation (1b) without loss of generality.

The solution of equation (1) is sought by means of a k-step BDF method [6], where temporal derivatives in equation (1a) and (1b) are replaced with the following difference approximation

$$(\bullet)'_n = \frac{1}{h}\sum_{i=0}^{k} \alpha_i (\bullet)_{n-i}. \tag{2}$$

Although the analysis is here restricted to BDF methods, similar results and conclusions can be reached for other integration schemes for DAEs. For example, the case of the Newmark family of integrators is analyzed in Bottasso, Dopico, and Trainelli [7].

Using a Newton-type method, at each time step one solves a linear system of equations in the form

$$\boldsymbol{A}\boldsymbol{z} = \boldsymbol{b}, \tag{3}$$

where the iteration matrix is

$$\boldsymbol{A} = h\boldsymbol{J}_n = \begin{bmatrix} \alpha_0\boldsymbol{I} & -h\boldsymbol{I} & \boldsymbol{0} \\ h\boldsymbol{X} & \alpha_0\boldsymbol{I} + h\boldsymbol{Y} & -h\boldsymbol{G} \\ h\boldsymbol{G}^T & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}, \tag{4}$$

with

$$\boldsymbol{X} = -\boldsymbol{f}_{,\boldsymbol{u}} - \boldsymbol{G}_{,\boldsymbol{u}}\boldsymbol{\lambda}, \tag{5a}$$
$$\boldsymbol{Y} = -\boldsymbol{f}_{,\boldsymbol{v}}. \tag{5b}$$

The three block rows of this matrix correspond to the kinematic, equilibrium and constraint equations of system (1), while the three block columns correspond to the $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{\lambda}$ variables, respectively.

Petzold and Lötstedt [1] show that, by using Gaussian elimination with partial pivoting, the accuracy in the i-th component of the solution can be estimated as

$$|\Delta z_i| \leq r\varepsilon \sum_j |(\boldsymbol{A}^{-1})_{ij}| \parallel \boldsymbol{A} \parallel_\infty \parallel \boldsymbol{z} + \Delta\boldsymbol{z} \parallel_\infty, \tag{6}$$

where r is an unknown coefficient, ε is the machine accuracy, and

$$(\boldsymbol{A} + \Delta\boldsymbol{A})(\boldsymbol{z} + \Delta\boldsymbol{z}) = \boldsymbol{b}, \tag{7}$$

with

$$\parallel \Delta\boldsymbol{A} \parallel_\infty \leq r\varepsilon \parallel \boldsymbol{A} \parallel_\infty . \tag{8}$$

The leading terms in the inverse of the iteration matrix (4) were derived in [1] as

$$\boldsymbol{A}^{-1} = (h\boldsymbol{J}_n)^{-1} = \frac{1}{\alpha_0}\begin{bmatrix} \boldsymbol{I}-\boldsymbol{T} & \gamma(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \gamma^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\gamma^{-1}\boldsymbol{T} & (\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \gamma^{-2}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\gamma^{-2}\boldsymbol{S}\boldsymbol{G}^T & -\gamma^{-1}\boldsymbol{S}\boldsymbol{G}^T\boldsymbol{R}^{-1} & \gamma^{-3}\boldsymbol{S} \end{bmatrix}, \tag{9}$$

where $\gamma = h/\alpha_0$ and

$$\boldsymbol{R} = \boldsymbol{I} + \gamma\boldsymbol{Y} + \gamma^2\boldsymbol{X}, \tag{10a}$$
$$\boldsymbol{S} = (\boldsymbol{G}^T\boldsymbol{R}^{-1}\boldsymbol{G})^{-1}, \tag{10b}$$
$$\boldsymbol{T} = \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S}\boldsymbol{G}^T. \tag{10c}$$

Therefore, using (6) and (9), we conclude that the roundoff errors in the solution are

$$\Delta u_i = \mathcal{O}(h^{-1}), \tag{11a}$$
$$\Delta v_i = \mathcal{O}(h^{-2}), \tag{11b}$$
$$\Delta \lambda_i = \mathcal{O}(h^{-3}). \tag{11c}$$

This shows that, as $h \to 0$, the accuracy in the Lagrange multipliers deteriorates quickly. Furthermore, it is readily verified that

$$\| \boldsymbol{A} \|_\infty = \mathcal{O}(h^0), \tag{12a}$$
$$\| \boldsymbol{A}^{-1} \|_\infty = \mathcal{O}(h^{-3}). \tag{12b}$$

Hence, the condition number $C = \| \boldsymbol{A} \|_\infty \| \boldsymbol{A}^{-1} \|_\infty$ becomes

$$C = \mathcal{O}(h^{-3}), \tag{13}$$

and the iteration matrix becomes severely ill conditioned for small h. The $\mathcal{O}(h^{-3})$ dependence of roundoff errors and conditioning on the time step size poses limitations to the practical use of variable step size solvers and time refinement procedures.

3 Preconditioning by scaling

3.1 Left preconditioning

As suggested in [1], this situation can be improved by considering a left diagonal scaling transformation of the equations. The scaled system can be written as

$$\widetilde{\boldsymbol{A}}\boldsymbol{z} = \widetilde{\boldsymbol{b}}, \tag{14}$$

where

$$\widetilde{\boldsymbol{A}} = \boldsymbol{D}\boldsymbol{A}, \tag{15a}$$
$$\widetilde{\boldsymbol{b}} = \boldsymbol{D}\boldsymbol{b}, \tag{15b}$$

and $\boldsymbol{D}$ is a diagonal scaling matrix defined as

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & h^{-1}\boldsymbol{I} \end{bmatrix}. \tag{16}$$

The inverse of the iteration matrix is in this case

$$\widetilde{\boldsymbol{A}}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} \boldsymbol{I}-\boldsymbol{T} & \gamma(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\gamma^{-1}\boldsymbol{T} & (\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0\gamma^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\gamma^{-2}\boldsymbol{S}\boldsymbol{G}^T & -\gamma^{-1}\boldsymbol{S}\boldsymbol{G}^T\boldsymbol{R}^{-1} & \alpha_0\gamma^{-2}\boldsymbol{S} \end{bmatrix}, \tag{17}$$

and, using again (6), we find

$$\Delta u_i = \mathcal{O}(h^0), \tag{18a}$$
$$\Delta v_i = \mathcal{O}(h^{-1}), \tag{18b}$$
$$\Delta \lambda_i = \mathcal{O}(h^{-2}). \tag{18c}$$

The situation is improved with respect to the original (unscaled) system, but the error in the Lagrange multipliers still grows rapidly as $h \to 0$ even for the left preconditioned problem. Similarly, we find $\| \widetilde{\boldsymbol{A}} \|_\infty = \mathcal{O}(h^0)$ and $\| \widetilde{\boldsymbol{A}}^{-1} \|_\infty = \mathcal{O}(h^{-2})$, which leads to the conclusion that the condition number is

$$C = \mathcal{O}(h^{-2}), \tag{19}$$

again not a favorable behavior.

3.2 Left and right preconditioning

A further improvement on the situation expressed by equation (18) can be obtained by considering a left and right preconditioning, i.e a scaling of the equations together with a scaling of the unknowns. In particular, since the Lagrange multipliers are affected by the largest roundoff errors according to (18), the governing equations are here re-written in the following form

$$\boldsymbol{u}' = \boldsymbol{v}, \tag{20a}$$
$$\boldsymbol{v}' = \boldsymbol{f} + s\boldsymbol{G}\hat{\boldsymbol{\lambda}}, \tag{20b}$$
$$\boldsymbol{\phi} = \boldsymbol{0}, \tag{20c}$$

where $\hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}/s$ are scaled Lagrange multipliers, and the scaling factor is $s = \mathcal{O}(h^{-2})$. The idea is now to solve the problem in terms of the scaled multipliers $\hat{\boldsymbol{\lambda}}$ and generalized coordinates $\boldsymbol{u}$ and velocities $\boldsymbol{v}$, and to recover the values of $\boldsymbol{\lambda}$ once at convergence.

Consider first for simplicity $s = 1/h^2$. The linear iteration problem for the left and right preconditioned system can be written as

$$\hat{\boldsymbol{A}}\hat{\boldsymbol{z}} = \hat{\boldsymbol{b}}, \tag{21}$$

where

$$\hat{\boldsymbol{A}} = \boldsymbol{D}_L\boldsymbol{A}\boldsymbol{D}_R, \tag{22a}$$
$$\hat{\boldsymbol{z}} = \boldsymbol{D}_R^{-1}\boldsymbol{z}, \tag{22b}$$
$$\hat{\boldsymbol{b}} = \boldsymbol{D}_L\boldsymbol{b}. \tag{22c}$$

The left diagonal scaling matrix $\boldsymbol{D}_L$ is defined as

$$\boldsymbol{D}_L = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & h\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & h^{-1}\boldsymbol{I} \end{bmatrix}. \tag{23}$$

The scaling of the third block rows is the same as in [1], while the scaling of the second block rows is needed to ensure $\| \hat{\boldsymbol{A}} \|_\infty = \mathcal{O}(h^0)$, which is beneficial for the condition number. The right diagonal transformation $\boldsymbol{D}_R$ which scales the Lagrange multipliers is defined as

$$\boldsymbol{D}_R = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & h^{-2}\boldsymbol{I} \end{bmatrix}. \tag{24}$$

This gives the following form for the inverse of the iteration matrix

$$\hat{\boldsymbol{A}}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} \boldsymbol{I} - \boldsymbol{T} & \alpha_0^{-1}(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0 \boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\gamma^{-1}\boldsymbol{T} & \alpha_0^{-1}\gamma^{-1}(\boldsymbol{I} - \boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0\gamma^{-1}\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\alpha_0^2 \boldsymbol{S}\boldsymbol{G}^T & -\alpha_0 \boldsymbol{S}\boldsymbol{G}^T\boldsymbol{R}^{-1} & \alpha_0^3 \boldsymbol{S} \end{bmatrix}. \tag{25}$$

Using (6), the roundoff errors for the three solution fields are in this case

$$\Delta u_i = \mathcal{O}(h^0), \tag{26a}$$

$$\Delta v_i = \mathcal{O}(h^{-1}), \tag{26b}$$

$$\Delta \hat{\lambda}_i = \mathcal{O}(h^0). \tag{26c}$$

Notice that the roundoff errors in the displacement field and in the scaled Lagrange multipliers are now insensitive to the time step size, while the condition number is

$$C = \mathcal{O}(h^{-1}). \tag{27}$$

The Lagrange multipliers are recovered at convergence of the Newton process from the scaled multipliers using $\boldsymbol{\lambda} = s\hat{\boldsymbol{\lambda}}$. The scaled multipliers will be affected by an error due to machine accuracy and to the Newton termination tolerance, error which is $\mathcal{O}(h^0)$ for the results above. Since s is $\mathcal{O}(h^{-2})$, then $\boldsymbol{\lambda}$ will be affected by an error $\mathcal{O}(h^{-2})$ due to the recovery process. However, the reduced accuracy of the multipliers can not affect the iterative Newton process, since this is formulated in terms of the scaled variables. Hence, the Newton iterations can be carried out until as tight a tolerance as necessary has been achieved.

A more general form for the multiplier scaling factor can be derived based on a dimensional argument. Recall that in general the dynamic equilibrium equations (1b) express the balance of inertial forces, elastic forces, constraint reactions and externally applied forces. To reflect the different nature of these forces, s is selected as

$$s = k_{\text{ave}} + \frac{m_{\text{ave}}}{h^2}, \tag{28}$$

where k_{ave} represents an average stiffness term for the system and m_{ave} an average mass term. To understand the dimensional effect of this scaling, let u_{ave} denote a representative displacement of the system, $F_{\text{ave}}^{\text{elas}}$ a representative value of the elastic forces and finally $F_{\text{ave}}^{\text{iner}}$ a representative value of the inertial forces. For large time step sizes, $s \approx k_{\text{ave}}$, and $\hat{\boldsymbol{\lambda}} \approx \boldsymbol{\lambda}/k_{\text{ave}}$. Then the ratio of the constraint reactions and of the elastic forces is approximately $s\boldsymbol{G}\hat{\boldsymbol{\lambda}}/F_{\text{ave}}^{\text{elas}} \approx k_{\text{ave}}\boldsymbol{G}\hat{\boldsymbol{\lambda}}/(k_{\text{ave}}u_{\text{ave}}) \approx \boldsymbol{G}\hat{\boldsymbol{\lambda}}/u_{\text{ave}}$; this means that the scaled Lagrange multipliers will be roughly of the same order of magnitude as the displacements, balancing the magnitude of all the unknowns of the problem. For very small time step sizes, $s \approx m_{\text{ave}}/h^2$, and $\hat{\boldsymbol{\lambda}} \approx \boldsymbol{\lambda}/(m_{\text{ave}}/h^2)$. It then follows that $s\boldsymbol{G}\hat{\boldsymbol{\lambda}}/F_{\text{ave}}^{\text{iner}} \approx (m_{\text{ave}}\boldsymbol{G}\hat{\boldsymbol{\lambda}}/h^2)/(m_{\text{ave}}u_{\text{ave}}/h^2) \approx \boldsymbol{G}\hat{\boldsymbol{\lambda}}/u_{\text{ave}}$; this means that even in this case the scaled Lagrange multipliers will be roughly of the same order of magnitude as the displacements, balancing again the magnitude of all the unknowns of the problem.

The roundoff errors and the conditioning of the left and right scaled system are much improved with respect to the unscaled or the left scaled ones. Yet, the velocity field still shows a mild

dependence on the time step size, which in turn induces the same behavior in the conditioning. This effect can be eliminated by considering a scaling of the unknowns that includes, together with the scaling of the Lagrange multipliers, also a scaling of the velocities, as suggested by Arnold [8]. This amounts to the following redefinition of the right diagonal transformation $\boldsymbol{D}_R$:

$$\boldsymbol{D}_R = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & h^{-1}\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & h^{-2}\boldsymbol{I} \end{bmatrix}. \tag{29}$$

The inverse of the iteration matrix is readily found as

$$\hat{\boldsymbol{A}}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} \boldsymbol{I}-\boldsymbol{T} & \alpha_0^{-1}(\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\alpha_0\boldsymbol{T} & (\boldsymbol{I}-\boldsymbol{T})\boldsymbol{R}^{-1} & \alpha_0^2\boldsymbol{R}^{-1}\boldsymbol{G}\boldsymbol{S} \\ -\alpha_0^2\boldsymbol{S}\boldsymbol{G}^T & -\alpha_0\boldsymbol{S}\boldsymbol{G}^T\boldsymbol{R}^{-1} & \alpha_0^3\boldsymbol{S} \end{bmatrix}, \tag{30}$$

which does not dependent on h. Hence, the roundoff errors in all solution variables and the conditioning are $\mathcal{O}(h^0)$.

3.3 Numerical example

We consider the problem of the pendulum, with equations

$$u_x' = v_x, \tag{31a}$$
$$u_y' = v_y, \tag{31b}$$
$$v_x' = \lambda u_x, \tag{31c}$$
$$v_y' = \lambda u_y - 1, \tag{31d}$$
$$0 = \frac{1}{2}(u_x^2 + u_y^2 - 1), \tag{31e}$$

where u_x, u_y are the Cartesian coordinates of the point mass, and v_x, v_y the mass velocity components. Bar length, point mass and acceleration of gravity are all equal to 1. The point mass is initially at rest with $u_x = 1$, $u_y = 0$, and falls under the action of gravity. The integration of equation (31) is performed using the 2-step BDF formula until time $t = 1 \cdot 10^{-3}$, using the time step sizes $h = \{1 \cdot 10^{-4}, 1 \cdot 10^{-5}, 1 \cdot 10^{-6}, 1 \cdot 10^{-7}, 1 \cdot 10^{-8}\}$.

At each time step, the non-linear discrete equations are solved using Newton method. At the j-th Newton iteration of time step n, we solve the linear system of equations

$$\boldsymbol{A}_n^j \boldsymbol{z}_n^j = \boldsymbol{b}_n^j, \tag{32}$$

which yields the corrections $\boldsymbol{z}_n^j$. Typically, the norm of the corrections will decrease at each Newton step, until the accumulation of roundoff errors leads to a saturation value. If further iterations are carried out, one typically observes an oscillation of the correction norm around its saturation value. An example of this behavior is given in Figure 1, which shows $\| \boldsymbol{z}_n^j \|$ vs. the Newton iteration j for the scheme without preconditioning and $h = 1 \cdot 10^{-7}$. In this case, the saturation is reached at the fourth iteration, after which the correction norm has very small oscillations around a saturation value of approximatively $4 \cdot 10^{-3}$. To quantify this effect of roundoff errors, in this work iterations are arrested when the Newton corrections stop decreasing, i.e. when we detect the condition

$$\| \boldsymbol{z}_n^{j+1} \| \geq \| \boldsymbol{z}_n^j \|. \tag{33}$$

Hence, the magnitude of the last decreasing Newton correction gives an indication of the tightest achievable convergence of the Newton iterations, which can not be further improved no matter how many iterations one carries out.

At first, we consider the scaling of the sole Lagrange multipliers based on the simple choice $s = 1/h^2$, and we use the definition of $\boldsymbol{D}_R$ given in (24). Figure 2 shows the maximum throughout each simulation of the 2-norm of the last decreasing Newton corrections versus the time step size, i.e. $\max_n \parallel \boldsymbol{z}_n^j \parallel$ vs. h. The results obtained with the left preconditioned problem are shown with the $\square$ symbol, while the ones of the left and right preconditioning with the $\circ$ symbol. Notice that with left preconditioning alone it is in fact impossible to solve the non-linear iteration with a tight tolerance for small values of the time step.

Figure 3 shows the maximum throughout each simulation of the absolute value of the last decreasing Newton corrections by field type, i.e. $\max_n \max(|z_{u_x,n}^j|, |z_{u_y,n}^j|)$, $\max_n \max(|z_{v_x,n}^j|, |z_{v_y,n}^j|)$, $\max_n |z_{\lambda,n}^j|$. These results are in accordance with the analysis of the previous sections.

Finally, Figure 4 shows the condition number C vs. h. The condition number is computed by means of the singular value decomposition of the iteration matrix at convergence, i.e. at the last decreasing Newton iteration. Here again, a substantial improvement is observed for the left and right preconditioning with respect to the left preconditioning alone.

Next, we consider the previous scaling of the multipliers together with the scaling of the velocities, with $\boldsymbol{D}_R$ as given in (29). Figure 5 shows the maximum throughout each simulation of the 2-norm of the last decreasing Newton corrections versus the time step size, while Figure 6 shows the condition number C vs. h. In both cases, the solution of the left and right preconditioned problem does not depend on h, as predicted by the analysis.

4 Displacement based form

For large scale applications denoted by many degrees of freedom, the velocity-displacement-multiplier form discussed in the previous section leads to the solution of large linear systems. For example, this is the case when the governing DAEs (1) are obtained by spatial discretization using the finite element method of a flexible mechanical system. A substantial performance increase can be obtained by implementing the integration scheme in displacement-multiplier form, which is simply achieved by statically eliminating the velocity unknowns at each time step. In fact, considering again the k-step BDF method, velocities at time t_n are computed from the kinematic equations (1a) using the difference approximation (2) as

$$\boldsymbol{u}'_n = \boldsymbol{v}_n = \frac{1}{h} \sum_{i=0}^{k} \alpha_i \boldsymbol{u}_{n-i}. \tag{34}$$

Inserting this expression into the accelerations of the dynamic equilibrium equations (1b), we get

$$\boldsymbol{v}'_n = \frac{1}{h} \sum_{i=0}^{k} \alpha_i \boldsymbol{v}_{n-i} = \frac{1}{h} \left(\frac{\alpha_0}{h} \sum_{i=0}^{k} \alpha_i \boldsymbol{u}_{n-i} + \sum_{i=1}^{k} \alpha_i \boldsymbol{v}_{n-i} \right). \tag{35}$$

Using Newton method, one solves the reduced system of linear equations

$$\boldsymbol{B}\boldsymbol{y} = \boldsymbol{c}, \tag{36}$$

where the iteration matrix is in this case

$$\boldsymbol{B} = h\boldsymbol{J}_n = \begin{bmatrix} \alpha_0 \boldsymbol{U} & -h\boldsymbol{G} \\ h\boldsymbol{G}^T & \boldsymbol{0} \end{bmatrix}, \tag{37}$$

with

$$\boldsymbol{U} = \gamma^{-1}\boldsymbol{I} + \boldsymbol{Y} + \gamma \boldsymbol{X}. \tag{38}$$

The first block rows of this matrix correspond to the equilibrium equations in displacement-based form, while the second block rows correspond to the constraint conditions. The two block columns correspond to the $\boldsymbol{u}$ and $\boldsymbol{\lambda}$ variables, respectively. Once at convergence, velocities are recovered using the vector relationship (34).

The inverse of the iteration matrix is

$$\boldsymbol{B}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{U}^{-1} & \gamma^{-1}\boldsymbol{U}^{-1}\boldsymbol{G}\boldsymbol{V} \\ -\gamma^{-1}\boldsymbol{V}\boldsymbol{G}^T\boldsymbol{U}^{-1} & \gamma^{-2}\boldsymbol{V} \end{bmatrix}, \tag{39}$$

where

$$\boldsymbol{V} = (\boldsymbol{G}^T\boldsymbol{U}^{-1}\boldsymbol{G})^{-1}, \tag{40a}$$
$$\boldsymbol{W} = \boldsymbol{U}^{-1}\boldsymbol{G}\boldsymbol{V}\boldsymbol{G}^T. \tag{40b}$$

Using (6), and considering that

$$\lim_{h\to 0} \boldsymbol{U}^{-1} = \mathcal{O}(h), \tag{41a}$$
$$\lim_{h\to 0} \boldsymbol{V} = \mathcal{O}(h^{-1}), \tag{41b}$$

the roundoff errors in the primary solution variables are found to be

$$\Delta u_i = \mathcal{O}(h^{-1}), \tag{42a}$$
$$\Delta \lambda_i = \mathcal{O}(h^{-3}), \tag{42b}$$

while the condition number C becomes

$$C = \mathcal{O}(h^{-4}), \tag{43}$$

again very unfavorable behaviors with respect to the time step size.

4.1 Left preconditioning

Even in this case, the simple left preconditioning of Petzold and Lötstedt [1] gives a little benefit with respect to the unpreconditioned problem. In fact, defining the left scaling transformation as

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & h^{-1}\boldsymbol{I} \end{bmatrix}, \tag{44}$$

we find

$$\widetilde{\boldsymbol{B}}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} (\boldsymbol{I} - \boldsymbol{W})\boldsymbol{U}^{-1} & \alpha_0\boldsymbol{U}^{-1}\boldsymbol{G}\boldsymbol{V} \\ -\gamma^{-1}\boldsymbol{V}\boldsymbol{G}^T\boldsymbol{U}^{-1} & \alpha_0\gamma^{-1}\boldsymbol{V} \end{bmatrix}, \tag{45}$$

which gives for the propagation of errors

$$\Delta u_i = \mathcal{O}(h^0), \tag{46a}$$
$$\Delta \lambda_i = \mathcal{O}(h^{-2}), \tag{46b}$$

while the conditioning is

$$C = \mathcal{O}(h^{-3}). \tag{47}$$

4.2 Left and right preconditioning

The left and right scaling matrices for the reduced problem are simply obtained from (21–24), and are defined as

$$\boldsymbol{D}_L = \begin{bmatrix} h\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & h^{-1}\boldsymbol{I} \end{bmatrix}, \tag{48}$$

and

$$\boldsymbol{D}_R = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & h^{-2}\boldsymbol{I} \end{bmatrix}, \tag{49}$$

where we have considered again the simple choice $s = 1/h^2$. This gives

$$\hat{\boldsymbol{B}}^{-1} = \frac{1}{\alpha_0} \begin{bmatrix} \alpha_0^{-1}\gamma^{-1}(\boldsymbol{I} - \boldsymbol{W})\boldsymbol{U}^{-1} & \alpha_0 \boldsymbol{U}^{-1}\boldsymbol{G}\boldsymbol{V} \\ -\alpha_0 \boldsymbol{V}\boldsymbol{G}^T\boldsymbol{U}^{-1} & \alpha_0^3\gamma\boldsymbol{V} \end{bmatrix}, \tag{50}$$

which yields time-step-size-independent sensitivity to perturbations of the primary variables

$$\Delta u_i = \mathcal{O}(h^0), \tag{51a}$$

$$\Delta\hat{\lambda}_i = \mathcal{O}(h^0), \tag{51b}$$

and a time-step-size-independent conditioning, i.e.

$$C = \mathcal{O}(h^0), \tag{52}$$

the best possible behavior that can be expected of a numerical method.

4.3 Numerical example

We consider again the model problem (31), using the same data. At the j-th Newton iteration of time step n, we solve the linear system of equations

$$\boldsymbol{B}_n^j \boldsymbol{y}_n^j = \boldsymbol{c}_n^j, \tag{53}$$

which yields the corrections $\boldsymbol{y}_n^j$. As in the previous case, iterations are arrested only when the Newton corrections stop decreasing.

The results of the analysis are confirmed by Figure 7 and Figure 8. In particular, the former shows the maximum throughout each simulation of the 2-norm of the last decreasing Newton corrections versus the time step size, $\max_n \parallel \boldsymbol{z}_n^j \parallel$ vs. h, while the latter reports the condition number C vs. h. For the left and right preconditioning, exact time step size independence is achieved.

5 Numerical example: the need for scaling in contact-impact analysis of flexible multibody systems

Fig. 9a depicts a cam of length 0.4 m connected to the ground at point **A** by means of a revolute joint, which relative rotation is prescribed to be $\phi(t) = \phi_0(1 - cos\Omega t)$, where $\phi_0 = \pi/3$ rad, and $\Omega = 2\pi/3$ rad/sec. At the tip of the cam, a rigid impactor of total mass $M_I = 4$ kg is attached. A flexible beam of length 2.4 m is clamped to the ground at point **R**, and a point mass $M_T = 40$ kg is attached at its free end, point **T**. At the middle point **P** of the beam, a rigid body of total mass $M_P = 4$ kg is attached. As the impactor moves, it contacts the outer surface of this rigid body, as shown in fig. 9b. The physical properties of the flexible beam are: axial stiffness $EA = 44000$ kN, bending stiffnesses $I_{22} = 300$, $I_{33} = 23$ kN·m^2, torsional stiffness $GJ = 28$ kN·m^2, and mass per

unit span $m = 1.6$ kg/m. The bending stiffnesses of the cam are $I_{22} = 23$, $I_{33} = 300$ kN·m^2, and the axial stiffness, torsional stiffness and mass per unit span are the same as those of the beam.

When the two bodies come in contact, a contact model and friction model are used to define the normal and tangential contact forces, respectively. A linear spring with a stiffness constant $k = 2$ GN/m in parallel with a viscous damper of constant $\mu = 10^{-3}$ were used to evaluate the normal contact force. The friction force was computed with the help of the LuGre model, using the following parameters: $\sigma_0 = 10^5$ m^{-1}, $\sigma_1 = \sigma_2 = 0$ sec/m, $v_s = 10^{-3}$ m/sec, $\mu_k = 0.30$, $\mu_s = 0.30$, and $\gamma = 2$. More details about this modeling technique and the definition of these parameters can be found in Bauchau and Ju [9].

The simulation was conducted for a 1 sec duration. From $t = 0$ to 0.4 sec a constant time step size, $h = 10^{-2}$ sec, was used and at that point, an automated time step size adaptivity algorithm was activated [10]; the desired local error was set to $\varepsilon_{\text{loc}} = 10^{-7}$. In light of equation (28), the scaling factor was selected as $s = k_{\text{ave}} + m_{\text{ave}}/h^2$, where $k_{\text{ave}} = 10^3$ and $m_{\text{ave}} = 0$ or 1. When $m_{\text{ave}} = 0$, the scaling is time step size independent, whereas it becomes time step dependent when $m_{\text{ave}} = 1$.

For the problem at hand, large time step sizes were used before contact occurs, i.e. when $t < 0.494$ sec. This portion of the simulation is not challenging and both scaling factors performed equally well. At time $t \approx 0.494$ the time step size was drastically reduced by the adaptivity algorithm to decrease the large contact force gradients that would result from using a constant time step. Figure 10 depicts the time step size as a function of time during the simulation, for both $m_{\text{ave}} = 0$ and 1. As should be expected from the analysis developed in this paper, when $m_{\text{ave}} = 0$, the simulation fails to converge as soon as the time step size becomes small, whereas for $m_{\text{ave}} = 1$, convergence is achieved at each time step, despite the complex dynamic behavior of the system after impact. Figure 11 shows the corresponding history of the scaling factor. Note that scaling factors of the order of 10^{13} were needed at the instant of contact; after impact, the highly vibratory response of the system required small time steps and the scaling factor was in the range of $s \in [10^8,\ 10^{11}]$.

6 Conclusions

We have proposed a simple scaling transformation for the index three DAEs describing constrained multibody dynamical systems. The approach amounts to a left and right preconditioning of the iteration matrix, which has the goal of making the solution less sensitive to the propagation of perturbations and of improving the condition number.

Using the proposed left preconditioning, the constraint equations are divided by the time step size as already proposed by [1], while the dynamic equilibrium equations are multiplied by the same quantity. The right preconditioning amounts to a definition of a set of modified unknowns: modified multipliers are obtained by scaling with the square of the time step size of the original Lagrange multipliers, while modified velocities are obtained by scaling with the time step size.

The new scaling was applied to the velocity-displacement-multiplier and to the reduced displacement-multiplier forms. In both cases, we obtain the remarkable results that both error propagations and conditioning are $\mathcal{O}(h^0)$, i.e. the numerical solution of index three DAEs behaves as in the case of regular ODEs.

The simple problem of the pendulum was used to numerically confirm the results of the analysis, while a more challenging problem dealing with the contact-impact analysis of a flexible multibody system was used to illustrate the effect of scaling when using time refinement procedures.

Although other approaches are available to deal with the problem of ill conditioning and error propagation for high index DAEs, the present approach has the possible advantages of being trivial to implement, of not requiring a re-writing of the equations of motion, and furthermore it does not introduce additional unknowns.

References

[1] L. Petzold and P. Lötstedt, *Numerical solution of nonlinear differential equations with algebraic constraints II: practical implications*, SIAM Journal on Scientific and Statistical Computing, 7 (1986), pp. 721–733.

[2] M. Arnold, *A perturbation analysis for the dynamical simulation of mechanical multibody systems*, Applied Numerical Mathematics, 18 (1995), pp. 37–56.

[3] C. Gear, B. Leimkuhler, and G. Gupta, *Automatic integration of Euler-Lagrange equations with constraints*, Journal of Computational and Applied Mathematics, 12–13 (1985), pp. 77–90.

[4] M. Borri, L. Trainelli, and A. Croce, *The embedded projection method: a general index reduction procedure for constrained system dynamics*, Computer Methods in Applied Mechanics and Engineering, (2005) accepted, to appear.

[5] A. Cardona, and M. Géradin, *Numerical integration of second order differential-algebraic systems in flexible mechanism dynamics*, in Computer-Aided Analysis of Rigid and Flexible Mechanical Systems, M.F.O. Seabra Pereira, and J.A.C. Ambrósio, Eds., Kluwer Academic Publishers, Dordrecht, Boston, 1994.

[6] C. Gear, *Simultaneous numerical solution of differential/algebraic equations*, IEEE Transactions on Circuits and Systems, 18 (1971), pp. 89–95.

[7] C.L. Bottasso, D. Dopico, and L. Trainelli, *On the optimal scaling of index three DAEs in multibody dynamics*, in Proceedings of the III European Conference on Computational Mechanics, Solids, Structures and Coupled Problems in Engineering, C.A. Mota Soares et al., Eds., Lisbon, Portugal, 5–8 June 2006.

[8] M. Arnold, *Private communication*, (2005).

[9] O.A. Bauchau and C.K. Ju, *Modeling Friction Phenomena in Flexible Multibody Dynamics.* Computer Methods in Applied Mechanics and Engineering, to appear, 2006.

[10] O.A. Bauchau, *Computational Schemes for Flexible, Nonlinear Multi-Body Systems*, Multibody System Dynamics, 2(2) (1998), pp. 169–225.

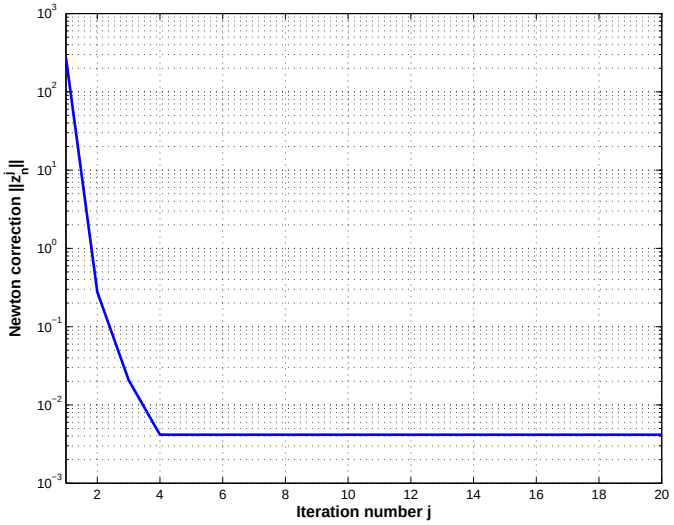

Figure 1: Convergence and saturation for the norm of the Newton corrections $\boldsymbol{z}_n^j$ plotted vs. the Newton iteration j for the scheme without preconditioning and $h = 1 \cdot 10^{-7}$.

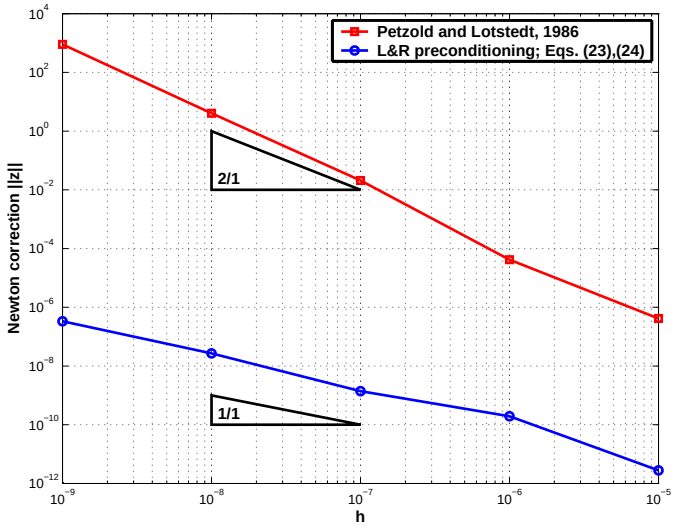

Figure 2: Last decreasing Newton corrections vs. time step size.

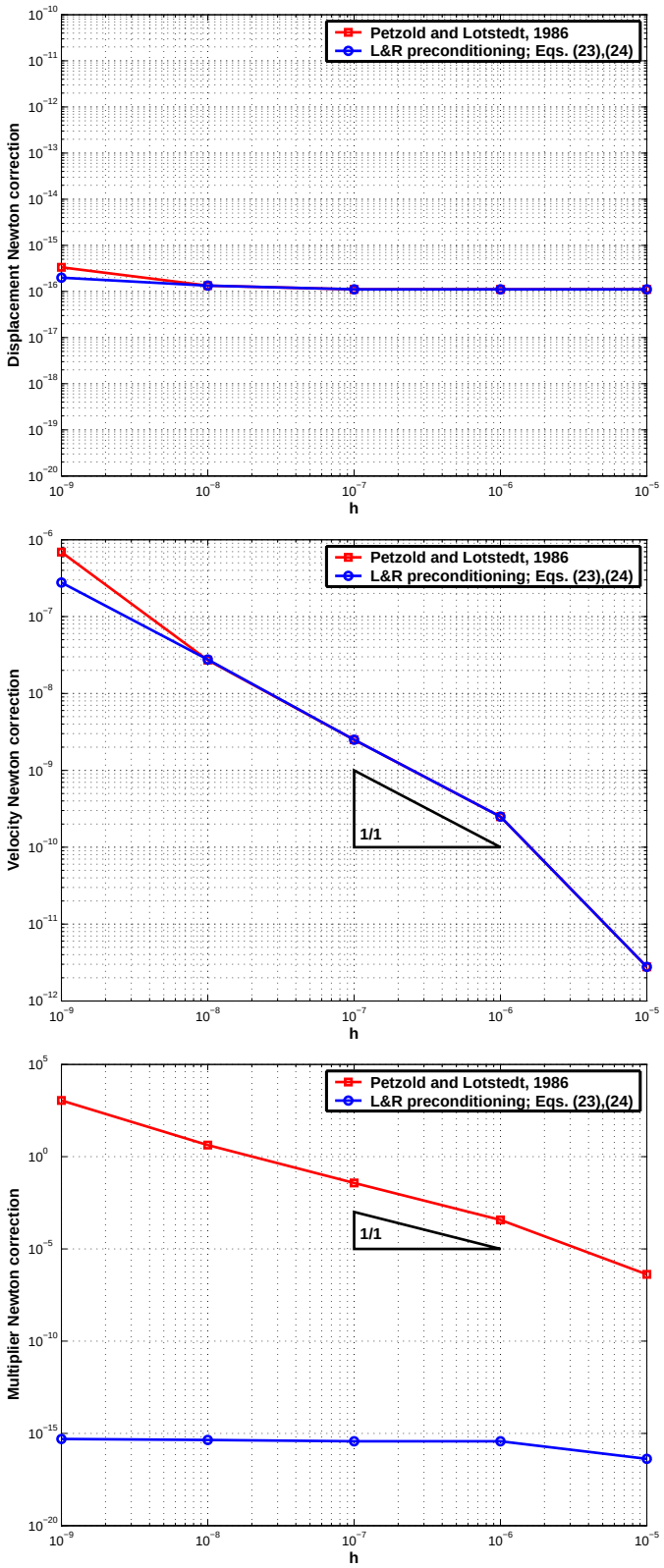

Figure 3: Last decreasing Newton corrections by field type vs. time step size. Top: displacements; center: velocities; bottom: Lagrange multipliers.

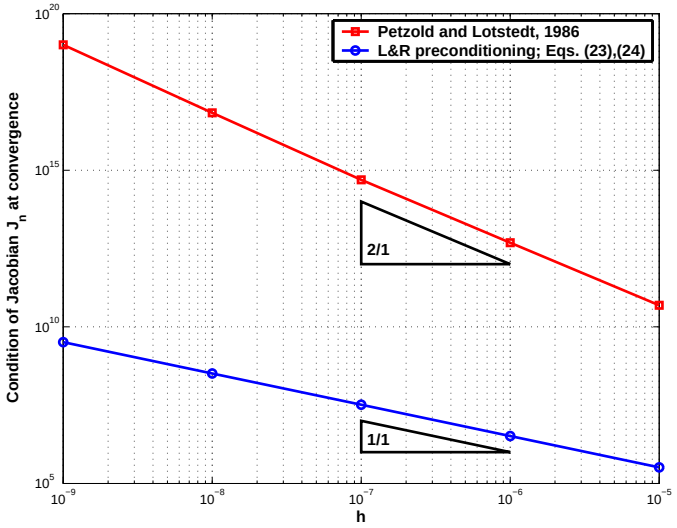

Figure 4: Condition number of Jacobian at convergence vs. time step size.

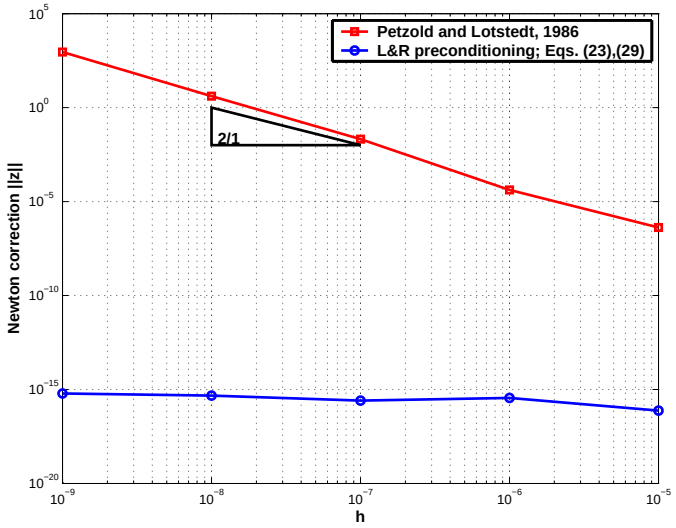

Figure 5: Last decreasing Newton corrections vs. time step size. Left and right preconditioning based on (23) and (29).

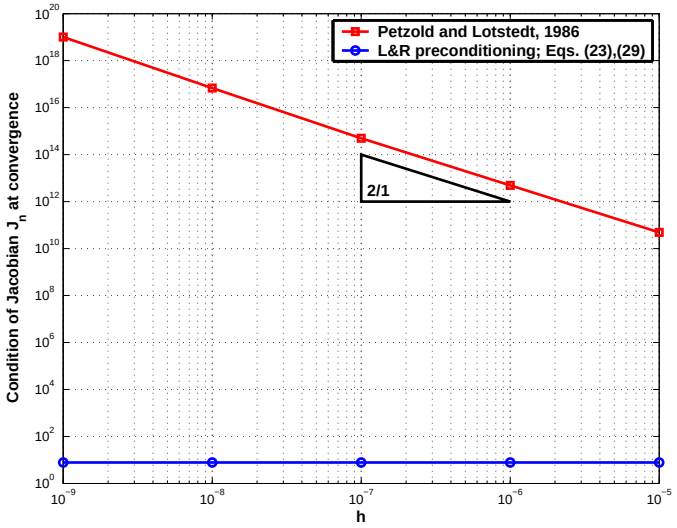

Figure 6: Condition number of Jacobian at convergence vs. time step size. Left and right preconditioning based on (23) and (29).

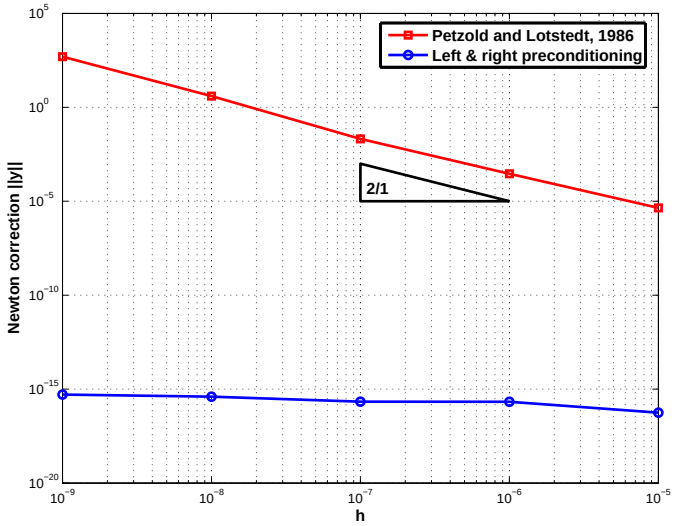

Figure 7: Last decreasing Newton corrections vs. time step size for the displacement-multiplier form.

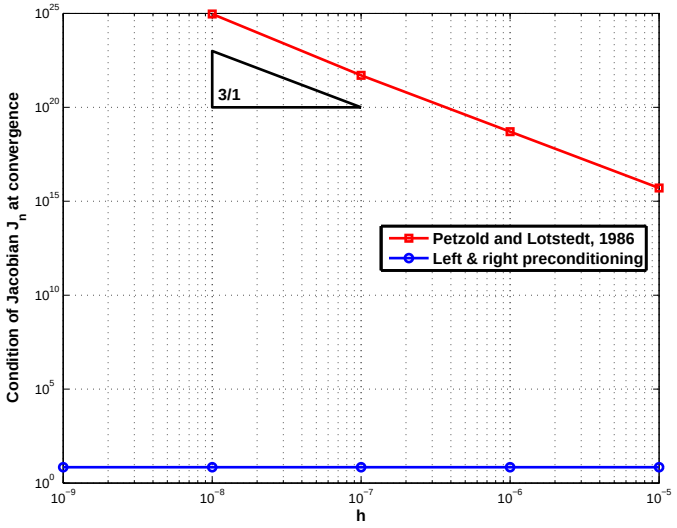

Figure 8: Condition number of Jacobian at convergence vs. time step size for the displacement-multiplier form.

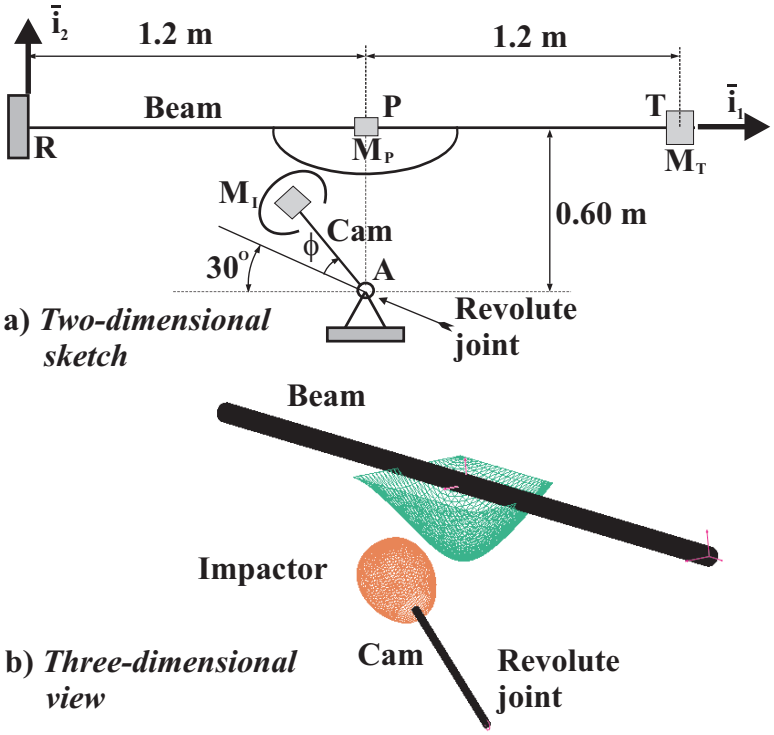

Figure 9: Configuration of the contact problem.

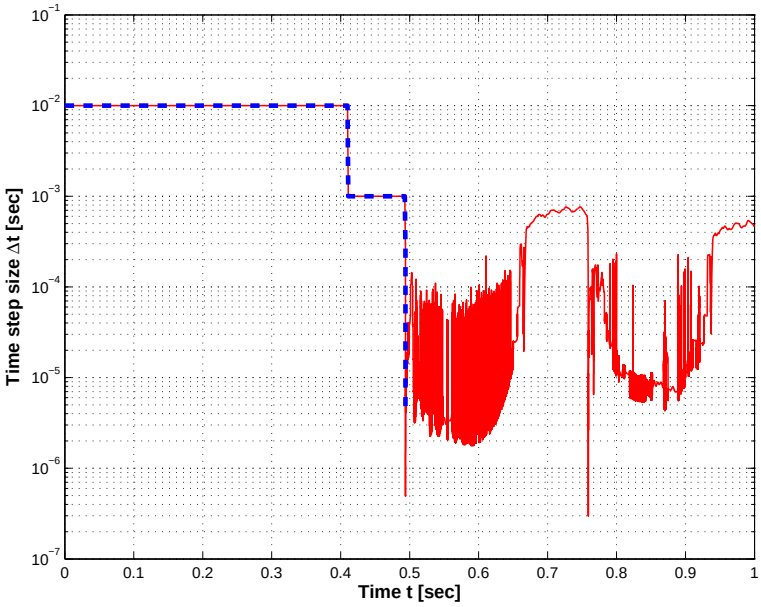

Figure 10: Time step size as a function of time. $m_{\text{ave}} = 0$: thick dashed line; $m_{\text{ave}} = 1$: thin solid line.

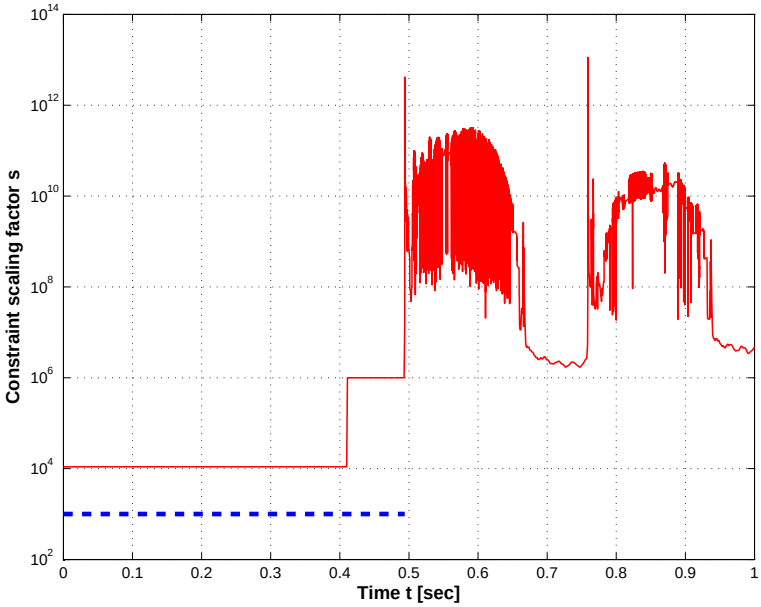

Figure 11: Constraint scaling factor s. $m_{\text{ave}} = 0$: thick dashed line; $m = 1$: thin solid line.