Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Sound and Vibration 284 (2005) 705–731
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

On practical integration of semi-discretized nonlinear equations of motion. Part 1: reasons for probable instability and improper convergence

Aram Soroushian[a,*], Peter Wriggers[b], Jamshid Farjoodi[a]

[a]*Department of Civil Engineering, Faculty of Engineering, University of Tehran, Tehran 11365, Iran*
[b]*Institute of Mechanics and Computational Mechanics, University of Hannover, Appelstrasse 9A, 30167 Hannover, Germany*

Received 8 September 2003; received in revised form 18 June 2004; accepted 7 July 2004
Available online 21 November 2004

Abstract

Time integration is the most versatile method for analyzing the general case of nonlinear semi-discretized equations of motion. However, the approximate responses of such analyses generally do not converge properly, and might even display numerical instability. This is a very significant shortcoming especially in practical time integration. Herein, after illustrating that this phenomenon is viable even for very simple nonlinear dynamic models, sources of the shortcoming are discussed in detail. The conclusion is that in time integration of nonlinear dynamic mathematical models of physically stable structural systems, responses may converge improperly for three major reasons. These reasons are: (1) inadequate number of iterations before terminating nonlinearity solutions; (2) deficiencies in the formulation of some time integration methods; and (3) the inherent behaviour of the models of some special dynamic systems. In addition, limitations on computational facilities and improper consideration of these limitations may impair the numerical stability and convergence of the computed responses. The differences between static and dynamic analyses are also discussed from the viewpoint of the numerical errors induced by nonlinearity.

*Corresponding author. Tehranpars, 115th Street, 136th alley, No. 35, Tehran 16517, Iran.
E-mail addresses: asoroush@ut.ac.ir, aramsoro@yahoo.com (A. Soroushian).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2004.07.008

1. Introduction

The main characteristic that numerical methods should possess to be successful is convergence [1]. Without this property, in order to be assured of accuracy, it is an accepted practice to analyze with limiting parameters. Time integration with very small time steps is an example. Nevertheless, in addition to its high computational cost, this approach is not reliable because of two reasons. First, the notion of limiting analysis, the corresponding parameters, and the numerical values that should be used for these parameters are vague and completely problem-dependent. Secondly, because of the limited computational facilities, the round-off error might considerably affect the implementation of the limiting parameters.

Concentrating on time integration of semi-discretized equations of motion, the following initial value problem:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{f}_{\text{int}} = \mathbf{f}(t),$$
$$\text{Initial conditions}: \left| \begin{array}{r} \mathbf{u}(t=0) = \mathbf{u}_0, \\ \dot{\mathbf{u}}(t=0) = \dot{\mathbf{u}}_0, \\ \mathbf{f}_{\text{int}}(t=0) = \mathbf{f}_{\text{int}_0}, \end{array} \right.$$
$$\text{Additional constraints}: \mathbf{Q}(\mathbf{u}, \dot{\mathbf{u}}) \tag{1}$$

should be studied [2]. In Eq. (1), $\mathbf{M}$ is the mass matrix; $\mathbf{u}, \dot{\mathbf{u}}$, and $\ddot{\mathbf{u}}$, respectively denote the vectors of displacement, velocity, and acceleration; $\mathbf{u}_0, \dot{\mathbf{u}}_0$, and $\mathbf{f}_{\text{int}_0}$ imply the initial conditions; $\mathbf{f}_{\text{int}}$ and $\mathbf{f}(t)$ stand for the vectors of the internal force and external excitation, respectively; and $\mathbf{Q}$ conceptually represents additional nonlinearity constraints, e.g., algebraic constraints in problems involved in impact and/or plasticity [3,4]. (The third initial condition in Eq. (1) should be considered in some special nonlinear cases, e.g., elastic–plastic dynamic models that are being studied after several loadings and unloadings.) The most practically accepted approach for analyzing Eq. (1) is time integration. There are many methods for integrating Eq. (1) in time. The development of such methods started from years ago [5,6], continued during decades [7–13], and is still in progress [14,15]. Returning to the concept of convergence, consider the analysis of Eq. (1) with an arbitrary time integration method. In view of the formal definition of error [16], Eq. (2), i.e.,

$$E(t_i) = \left\| \{\mathbf{u}_i \quad \dot{\mathbf{u}}_i\} - \left\{\mathbf{u}(t_i)\, \dot{\mathbf{u}}(t_i)\right\} \right\|, \tag{2}$$

is an appropriate representation for the error at time instant t_i. In Eq. (2), $\mathbf{u}_i$ and $\mathbf{u}(t_i)$, respectively, denote the computed and actual vectors for the unknown displacement at t_i. Correspondingly, $\dot{\mathbf{u}}_i$ and $\dot{\mathbf{u}}(t_i)$ stand for the computed and actual vectors of the unknown velocity at t_i and $\|\,\|$ implies an arbitrary norm [17]. After the first time integration analysis, if the analysis is repeated several times, consecutively, and each time, with time steps that are all similarly decreased in size, convergence of the computed responses implies [1,18–21]

$$\lim_{\Delta t \to 0} E(t_i) = 0. \tag{3}$$

In Eq. (3), Δt is a representative for the sizes of the time steps implemented in the analysis. From Eq. (3), it can be deduced, that the variation of $E(t_i)$ (and even the deviation of each displacement, velocity, internal force, etc., from the corresponding actual value) with respect to Δt can be illustrated by Fig. 1. This figure is valid regardless of the problem. Nevertheless, when the behaviour is very complicated and highly oscillatory, the intermediate linear part is postponed to smaller values of Δt. Fluctuations in the lower parts of the illustration in Fig. 1 (section a) are due to the negative effect of round off in the range of very small time steps. The second part in Fig. 1 (section b) can be expressed by

$$E_i = C_0 \Delta t^q + C_1 \Delta t^{q+1} + \cdots, \quad q \geqslant 1. \tag{4}$$

In Eq. (4), C_j, $j = 0, 1, 2, \ldots$ is a constant depending on the computed response, and q stands for a positive integer known as the convergence rate or rate of accuracy [2,19–21]. The main concern of convergence is the intermediate linear part in Fig. 1, where

$$E(t_i) \approx C_0\, \Delta t^q. \tag{5}$$

In presence of nonlinearity, changes of error, instead of Fig. 1, are as shown in Fig. 2 [22,23]. According to Fig. 2, in time integration of nonlinear equations of motion, the changes of error are generally unpredictable, and after an analysis, there is no reliable way to improve the accuracy. Moreover, numerical instability is viable even when the analysis is repeated with smaller time steps (x in Fig. 2). Though the shortcomings of convergence and specifically numerical instability are reported in the literature at times [22–33], it is instructive to illustrate these shortcomings in view of the following very simple (single degree of freedom) problem:

$$\begin{gathered} \ddot{u} + u = 0, \\ u(t=0) = 1, \\ \dot{u}(t=0) = 0, \\ \text{Elastic contact at } u = 0 \quad (\text{Coefficient of restitution} = 1.0). \end{gathered} \tag{6}$$

Denoting the number of collisions already occurred with $\bar{n}$, one can simply arrive at the exact response of Eq. (6), i.e.,

$$u(t) = \begin{cases} \cos(t), & 0 \leqslant t < \frac{\pi}{2}, \\ \cos(t - \bar{n}\pi), & (2\bar{n}-1)\frac{\pi}{2} \leqslant t < (2\bar{n}+1)\frac{\pi}{2}, \quad \bar{n} = 1, 2, \ldots. \end{cases} \tag{7}$$

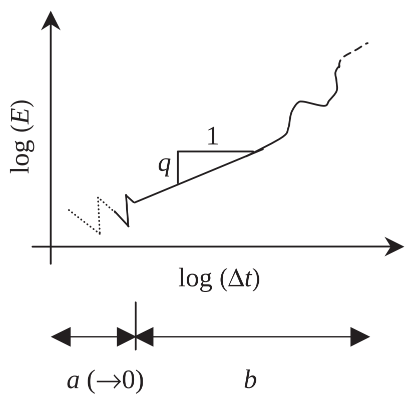

Fig. 1. Typical change of error for convergent responses.

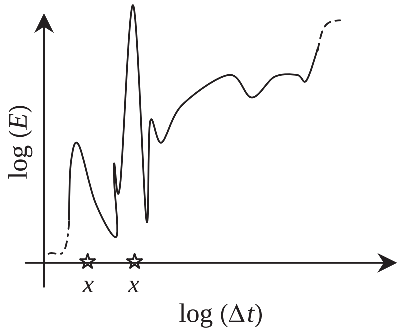

Fig. 2. Typical change of error in time integration of nonlinear equations of motion.

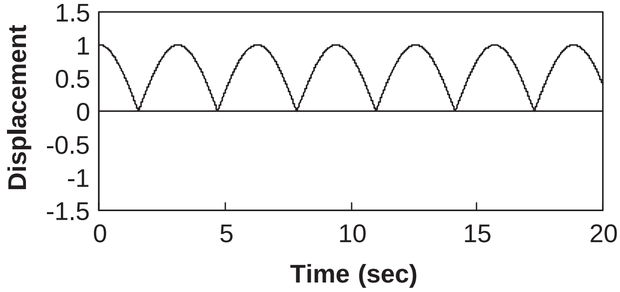

Fig. 3. The actual response of the system defined by Eq. (6).

This response is illustrated in Fig. 3 for the first 20 sec. Besides, the central difference time integration method [34] and the fractional-time-stepping nonlinearity solution method [35,36] are used in time integration of Eq. (6). (The fractional-time-stepping method is implemented due to its guaranteed convergence, and considering the small size of the problem.) The time step size is set equal to 1 sec (in absence of nonlinearity, the natural period of the model is 2π), and in order to consider the nonlinearities with sufficient accuracy, the fractional-time-stepping method will reduce the time step sizes by dividing them by 10, for 3 times, at most. The nonlinearity tolerance is defined in terms of the relative displacement and is set equal to 0.01. The solid line in Fig. 4 is obtained by consecutively repeating the analysis with time step sizes that are halved for each new analysis. This line illustrates the changes of the displacement error at $t = 20$ with respect to the implemented time step size. Comparing Fig. 4 with Figs. 1 and 2, or considering the dashed line in Fig. 4, illustrates the improper convergence of the computed responses. The result obtained from the above very simple example clearly demonstrates the probability of much worse shortcomings in time integration of complicated mathematical models of real structural systems. It is essential to note that even when the responses converge, the range of Δt for which the logarithm of the error in Fig. 1 reduces linearly, is problem-dependant. Therefore, maintaining convergence is especially significant from the viewpoint of numerical stability, and besides in order to arrive at meaningful responses.

Because of the importance of numerical stability and convergence, many researchers have tried to overcome the above-mentioned shortcomings [22–23,26,32,37–43]. However, yet, the problem

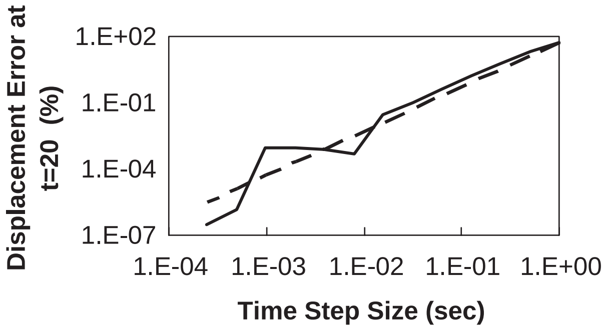

Fig. 4. Change of error for the response of the model defined by Eq. (6) when analyzed conventionally (nonzero residual errors and constant nonlinearity tolerance).

is not completely resolved [33,41,42]. A more thorough review of the existing literature shows that the reasons for improper convergence (in time integration of nonlinear dynamic models) require more study. The main cases already investigated are

- Inconvenient refinement of time steps when detecting nonlinearity [44,45];
- Insufficient precision of nonlinearity solutions [46,47];
- Divergence of nonlinearity solution methods [25];
- Numerical instability for linear analyses [2];
- Accumulation of the additional errors induced by nonlinearity [48].

The authors of this paper believe that by a more detailed study on the reasons and especially the mechanisms of improper convergence (in presence of nonlinearity), better approaches to circumvent this undesired phenomenon are attainable. In this regard, some mechanisms for improper convergence are explained here in detail. First, a mathematical proposition is proved in Section 2. In view of this proposition, the main reason of the instability and improper convergence is explained in Section 3. Then, in Section 4, the sources of improper convergence originated in the formulation of some time integration methods are studied. Afterwards, the nonlinear dynamic models for which the concept of limiting analysis is not valid are discussed in Section 5. In Section 6, the effects of limited computational facilities and their inappropriate consideration are explained, and finally in Section 7, the conclusions are set out in brief.

2. A basic statement

Consider integrating an arbitrary nonlinear semi-discretized equation of motion subjected to specified initial conditions, and if needed some additional constraints. The approximate response will converge to the exact response of the limiting mathematical model (the model with characteristics obtainable in time integration with arbitrary small time steps), if

- The requirements of convergence in each linearized time step (including approximating the excitation with sufficient accuracy [20,49]), are fulfilled,

- The conditions at which the nonlinearity solutions result in zero residual errors are ideally provided.

Moreover, the convergence will occur with a rate equal to the convergence rate in time integration of linear initial value problems.

A very general nonlinear semi-discretized equation of motion can be expressed by Eq. (1). After carrying out the time integration analysis, the mathematical model is replaced with a piece-wisely linear model with characteristics that may differ at different time steps [13,48]. This replacement can be explained in detail when nonlinearities arise from changes of stiffness [46]. At each time step, after the iterations of nonlinearity solution are ended, the equations defining the integration method (generally three main equations) can be assumed satisfied [50,51]. From this point of view, consider the following definition:

$$\Delta\hat{\mathbf{f}}_i = \hat{\mathbf{K}}_{\mathrm{sec}}\, \Delta\mathbf{u}_i \tag{8}$$

for the equivalent dynamic secant stiffness matrix associated with the ith time step, $\hat{\mathbf{K}}_{\mathrm{sec}}$. In Eq. (8), $\Delta\hat{\mathbf{f}}_i$ denotes the equivalent dynamic force increment, e.g.,

$$\Delta\hat{\mathbf{f}}_i = \mathbf{f}_i - \mathbf{f}_{i-1} + \left(\frac{4}{\Delta t_i}\mathbf{M} + 2\mathbf{C}\right)\dot{\mathbf{u}}_{i-1} + 2\mathbf{M}\ddot{\mathbf{u}}_{i-1} \tag{9}$$

for the average acceleration method, where Δt_i and $\mathbf{C}$, respectively, imply the size of the time step ending at t_i (see Fig. 5) and the constant matrix of damping. For a specific $\Delta\hat{\mathbf{f}}_i$ and nonzero $\Delta\mathbf{u}_i$ the existence of $\hat{\mathbf{K}}_{\mathrm{sec}}$ is evident from linear algebra [17]. Hence, in view of the following relation:

$$\hat{\mathbf{K}}_{\mathrm{sec}} = \mathbf{K}_{\mathrm{sec}} + \frac{2}{\Delta t_i}\mathbf{C} + \frac{2}{\Delta t_i^2}\mathbf{M} \tag{10}$$

for the average acceleration method (see Ref. [34]), the result of nonlinearity iterations at the ith time step is identical to the one-step integration with $\mathbf{K}_{\mathrm{sec}}$ as the stiffness matrix. In other words, the iterations leading to $\Delta\mathbf{u}_{\mathrm{i}}$ can be replaced by one-step integration with the stiffness matrix $\mathbf{K}_{\mathrm{sec}}$ at the ith time step. Accordingly, the mathematical model will be precisely (with regard to the accuracies attained in nonlinearity solutions) replaced with a piece-wisely linear model. As a result, to study the phenomenon of convergence, the familiar error equation [2],

$$\mathbf{e}_i = \mathbf{A}_i\mathbf{e}_{i-1} + \boldsymbol{\tau}_i \tag{11}$$

can be implemented for an arbitrary oscillatory mode at the arbitrarily selected ith time step. (The assumption of appropriately modeling the excitation [20,49] is implicit in Eq. (11).) In Eq. (11), in view of the modal description at the ith time step, $\mathbf{A}_i$ denotes the amplification matrix [2], $\boldsymbol{\tau}_i$

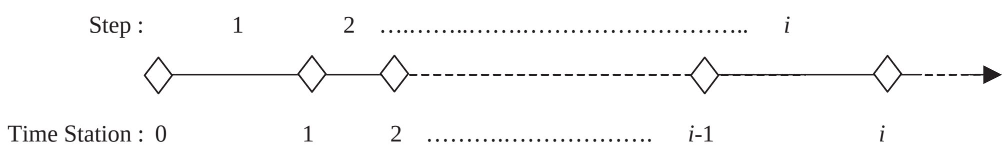

Fig. 5. Typical arrangement of time steps and time stations.

represents the vector of local truncation error (the error induced by time integration within the ith time step) [2,52], and correspondingly $\mathbf{e}_i$ stands for the deviation of the computed response from the exact response at the end of the step, i.e.,

$$\mathbf{e}_i = \begin{Bmatrix} v_i \\ \dot{v}_i \end{Bmatrix} - \begin{Bmatrix} v(t_i) \\ \dot{v}(t_i) \end{Bmatrix}. \tag{12}$$

In Eq. (12), v_i, $v(t_i)$, $\dot{v}_i$, and $\dot{v}(t_i)$ represent the computed and actual values of the modal displacements and velocities at the ith time station. Returning to Eq. (11), the definition and the proposed formulae for the local truncation error is based on the assumption of linear behaviour within the time step [52]. This assumption cannot be correct when the residuals of nonlinearity solutions are nonzero. In fact, nonzero residuals of nonlinearity solutions at the ith time step affect the corresponding local truncation error. Consequently, either $\boldsymbol{\tau}_i$ should be redefined, or Eq. (11) should be rearranged. With zero residual errors, neither of these considerations is essential. Accordingly, considering all oscillatory modes at the ith time step, one can expand Eq. (11) in

$$\begin{Bmatrix} {}_i\mathbf{e}_i^1 \\ {}_i\mathbf{e}_i^2 \\ \vdots \\ {}_i\mathbf{e}_i^n \end{Bmatrix} = \begin{bmatrix} \mathbf{A}_i^1 & & & \mathbf{0} \\ & \mathbf{A}_i^2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_i^n \end{bmatrix} \begin{Bmatrix} {}_{i-1}\mathbf{e}_i^1 \\ {}_{i-1}\mathbf{e}_i^2 \\ \vdots \\ {}_{i-1}\mathbf{e}_i^n \end{Bmatrix} + \begin{Bmatrix} \boldsymbol{\tau}_i^1 \\ \boldsymbol{\tau}_i^2 \\ \vdots \\ \boldsymbol{\tau}_i^n \end{Bmatrix}. \tag{13}$$

In Eq. (13), ${}_i\mathbf{e}_k^j$ implies the contribution of the jth oscillatory mode in the error of the response at the ith time station, according to the modal description at the kth time step. To say better, in view of the modal description at the ith time step, ${}_{i-1}\mathbf{e}_i^j$ and ${}_i\mathbf{e}_i^j$ represent the contribution of the jth mode in the response error, respectively, at the start and end of the ith time step (see Fig. 5). Correspondingly, $\mathbf{A}_i^j$ denotes the amplification matrix [2] for the jth oscillatory mode (that depends on both the structural characteristics and the size of the ith time step), $\boldsymbol{\tau}_i^j$ represents the local truncation error for the jth mode (all at the ith time step), and n stands for the total number of degrees of freedom. It is clear from Eq. (13) that at an arbitrary time step, e.g., the kth time step ($i = k$) the error term in the LHS of Eq. (13) implies the same physical concept as the error term in the RHS of Eq. (13) at the preceding time step ($i = k - 1$). For instance, $\begin{Bmatrix} {}_{k-1}\mathbf{e}_{k-1}^1 & {}_{k-1}\mathbf{e}_{k-1}^2 & \cdots & {}_{k-1}\mathbf{e}_{k-1}^n \end{Bmatrix}^{\mathrm{T}}$ and $\begin{Bmatrix} {}_{k-1}\mathbf{e}_k^1 & {}_{k-1}\mathbf{e}_k^2 & \cdots & {}_{k-1}\mathbf{e}_k^n \end{Bmatrix}^{\mathrm{T}}$ both imply the error at time instant t_{k-1} ('T' as the RHS superscript implies transposition), however, according to different modal descriptions, respectively, at time steps $k-1$ and k. Hence, there exists a rotational matrix such that [17]

$$\begin{Bmatrix} {}_{k-1}\mathbf{e}_k^1 \\ {}_{k-1}\mathbf{e}_k^2 \\ \vdots \\ {}_{k-1}\mathbf{e}_k^n \end{Bmatrix} = {}_{k-1}\boldsymbol{\Psi}_{k/k-1} \begin{Bmatrix} {}_{k-1}\mathbf{e}_{k-1}^1 \\ {}_{k-1}\mathbf{e}_{k-1}^2 \\ \vdots \\ {}_{k-1}\mathbf{e}_{k-1}^n \end{Bmatrix}, \tag{14.1}$$

$$\left\| {}_{k-1}\boldsymbol{\Psi}_{k/k-1} \right\| = 1. \tag{14.2}$$

In view of Eq. (14.1), considering Eq. (13) for all the time stations prior to the ith time station leads to

$$\left\{\begin{array}{c} {}_{i}\mathbf{e}_{i}^{1} \\ {}_{i}\mathbf{e}_{i}^{2} \\ \vdots \\ {}_{i}\mathbf{e}_{i}^{n} \end{array}\right\} = \left(\prod_{j=1}^{i} {}_{j}\boldsymbol{\psi}_{j+1/j} \begin{bmatrix} \mathbf{A}_{j}^{1} & & & \mathbf{0} \\ & \mathbf{A}_{j}^{2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_{j}^{n} \end{bmatrix} \right) \left\{\begin{array}{c} {}_{0}\mathbf{e}_{1}^{1} \\ {}_{0}\mathbf{e}_{1}^{2} \\ \vdots \\ {}_{0}\mathbf{e}_{1}^{n} \end{array}\right\} + \sum_{j=1}^{i} \left(\prod_{k=1}^{i-j} {}_{k+j}\boldsymbol{\psi}_{k+j+1/k+j} \begin{bmatrix} \mathbf{A}_{k+j}^{1} & & & \mathbf{0} \\ & \mathbf{A}_{k+j}^{2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_{k+j}^{n} \end{bmatrix} \right) \left\{\begin{array}{c} \boldsymbol{\tau}_{j}^{1} \\ \boldsymbol{\tau}_{j}^{2} \\ \vdots \\ \boldsymbol{\tau}_{j}^{n} \end{array}\right\}, \tag{15}$$

where for the sake of simplicity, ${}_{i}\boldsymbol{\psi}_{i+1/i}$ is defined as a unity matrix.

Considering the specified initial conditions,

$$\left\{\begin{array}{c} {}_{0}\mathbf{e}_{1}^{1} \\ {}_{0}\mathbf{e}_{1}^{2} \\ \vdots \\ {}_{0}\mathbf{e}_{1}^{n} \end{array}\right\} = \bar{\mathbf{O}}, \tag{16}$$

Eq. (15) can be rewritten as

$$\left\{\begin{array}{c} {}_{i}\mathbf{e}_{i}^{1} \\ {}_{i}\mathbf{e}_{i}^{2} \\ \vdots \\ {}_{i}\mathbf{e}_{i}^{n} \end{array}\right\} = \sum_{j=1}^{i} \left(\prod_{k=1}^{i-j} {}_{k+j}\boldsymbol{\psi}_{k+j+1/k+j} \begin{bmatrix} \mathbf{A}_{k+j}^{1} & & & \mathbf{0} \\ & \mathbf{A}_{k+j}^{2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_{k+j}^{n} \end{bmatrix} \right) \left\{\begin{array}{c} \boldsymbol{\tau}_{j}^{1} \\ \boldsymbol{\tau}_{j}^{2} \\ \vdots \\ \boldsymbol{\tau}_{j}^{n} \end{array}\right\}. \tag{17}$$

In order to evaluate the amount of error, an arbitrary norm is applied to Eq. (17), resulting in [17]

$$\left\| \left\{\begin{array}{c} {}_{i}\mathbf{e}_{i}^{1} \\ {}_{i}\mathbf{e}_{i}^{2} \\ \vdots \\ {}_{i}\mathbf{e}_{i}^{n} \end{array}\right\} \right\| \leqslant \sum_{j=1}^{i} \left(\prod_{k=1}^{i-j} \left\| \begin{bmatrix} \mathbf{A}_{k+j}^{1} & & & \mathbf{0} \\ & \mathbf{A}_{k+j}^{2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_{k+j}^{n} \end{bmatrix} \right\| \right) \left\| \left\{\begin{array}{c} \boldsymbol{\tau}_{j}^{1} \\ \boldsymbol{\tau}_{j}^{2} \\ \vdots \\ \boldsymbol{\tau}_{j}^{n} \end{array}\right\} \right\|. \tag{18}$$

When studying the convergence of responses in linear analyses, Eq. (18) is also being achieved constrained to the fact that

$$\forall r,\ k \in Z^+ \quad \mathbf{A}_r^k = \mathbf{A}^k. \tag{19}$$

Besides, since the requirements of convergence, including numerical stability [53], at each linearized time step are assumed satisfied [2],

$$\forall k+j \quad \forall p \in Z^+ \quad \left\| \begin{bmatrix} \mathbf{A}_{k+j}^1 & & & \mathbf{0} \\ & \mathbf{A}_{k+j}^2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{A}_{k+j}^n \end{bmatrix}^p \right\| \leqslant \text{const.} \tag{20}$$

Assuming $p = 1$, using the definition of norms [17], and applying mathematical induction, one can conclude that in presence of nonlinearity,

$$\forall j, i \quad \prod_{k=1}^{i-j} \left\| \begin{bmatrix} \mathbf{A}_{k+j}^1 & & & \\ & \mathbf{A}_{k+j}^2 & & \\ & & \ddots & \\ & & & \mathbf{A}_{k+j}^n \end{bmatrix} \right\| \leqslant \text{const.} \tag{21}$$

In view of Eq. (21), Eq. (18) results in

$$\left\| \begin{Bmatrix} {}_i\mathbf{e}_i^1 \\ {}_i\mathbf{e}_i^2 \\ \vdots \\ {}_i\mathbf{e}_i^n \end{Bmatrix} \right\| \leqslant \text{const.} \sum_{j=1}^{i} \left\| \begin{Bmatrix} \boldsymbol{\tau}_j^1 \\ \boldsymbol{\tau}_j^2 \\ \vdots \\ \boldsymbol{\tau}_j^n \end{Bmatrix} \right\| \tag{22}$$

leading to

$$\left\| \begin{Bmatrix} {}_i\mathbf{e}_i^1 \\ {}_i\mathbf{e}_i^2 \\ \vdots \\ {}_i\mathbf{e}_i^n \end{Bmatrix} \right\| \leqslant \text{const.}\ i \ \underset{j}{\text{Max}} \left\| \begin{Bmatrix} \boldsymbol{\tau}_j^1(t) \\ \boldsymbol{\tau}_j^2(t) \\ \vdots \\ \boldsymbol{\tau}_j^n(t) \end{Bmatrix} \right\|. \tag{23}$$

Furthermore, one of the requirements of convergence in linear analysis, which is here assumed fulfilled for each linearized time step, is consistency [53]. Hence, considering that the error vectors at the LHS and the local truncation error at the RHS of Eq. (23) have identical definitions in linear and nonlinear analyses,

$$\left\| \begin{Bmatrix} {}_i\mathbf{e}_i^1 \\ {}_i\mathbf{e}_i^2 \\ \vdots \\ {}_i\mathbf{e}_i^n \end{Bmatrix} \right\| \leqslant \text{const. } t_i \ \Delta t^q. \tag{24}$$

In Eq. (24), q ($q \geqslant 1$) denotes the rate of convergence in linear analysis. Since no restriction is considered for the norm utilized here, the convergence of ${}_i\mathbf{e}_i^i$ (to zero) with the rate of q can be deduced from Eq. (24) and the properties of norms [17]. Hence, the proof is complete.

Because of mathematical shortcomings, providing zero nonlinearity residual errors is generally impossible [48]. Nevertheless, for some simple sdof dynamic models analyzed by special time integration methods, the event-to-event nonlinearity solution method can result in zero residuals [48]. Therefore, to verify the stated proposition, the simple model of Section 1 is once again studied here, implementing the event-to-event nonlinearity solution method. The corresponding changes of errors for the displacement at $t = 20$, i.e., the dashed line in Fig. (4), clearly exhibit the proper convergence of the response.

3. The main reason for improper convergence

3.1. Preliminary notes

In order to arrive at the main reason and mechanism of probable improper convergence (in time integration of nonlinear dynamic models), an important point should be noted first. It is the fact that after analyzing a nonlinear dynamic mathematical model of a structural system by means of a time integration method and a nonlinearity solution method, the nonlinearity residual errors—though very small and negligible—are generally nonzero. Nevertheless, a different mathematical model can be defined such that by analyzing the new model with parameters (time integration method, time step sizes, nonlinearity solution method, and nonlinearity tolerances, etc.) similar to the parameters implemented in the analysis of the original model, a response identical to the response of the original model is obtained but with zero residual errors. To explain this, consider a specific time step, e.g., the ith step, at the end of which nonlinearity is detected. The iterations of the nonlinearity solution end, when at least one of the following equations:

$$|{}_k\delta_i| \leqslant \bar{\delta}, \tag{25.1}$$

$$k \geqslant N \tag{25.2}$$

is satisfied. In Eqs. (25), ${}_k\delta_i$ denotes the residual error after k iterations at the ith time step, $\bar{\delta}$ (>0) indicates the corresponding nonlinearity tolerance, and N represents the maximum acceptable number of the iterations. Nonlinearity might be simultaneously controlled by several equations similar to Eq. (25.1) when ${}_k\delta_i$ will be replaced with ${}_k\delta_i^j$. At the above-mentioned specific time instant, i.e., the time instant at which nonlinearity iterative solution is ended, some slight changes

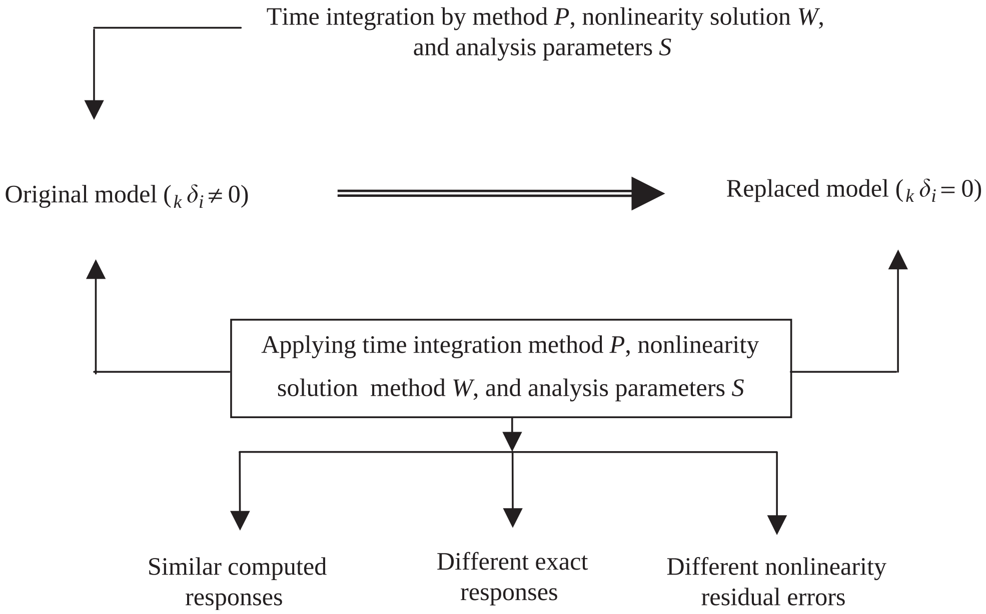

Fig. 6. Replacement of the nonlinear model with another model with the same response and different nonlinearity residual errors.

in the characteristics of the mathematical model (Eq. (1)), especially in the external excitation, would reduce the corresponding nonlinearity residuals to zero. Implementing these slight changes (e.g., taking the nonlinearity residuals to the RHS of the equation of motion, when nonlinearity iterations are with respect to unbalanced force), all through the time interval, one can define a new mathematical model. By its definition, the response obtained from time integration analysis of the new model with the parameters implemented in the original analysis is identical to the responses generated by the original analysis. Furthermore, once again according to the definition of the new model, in the analysis of the new model, the nonlinearity residual errors are zero. This is briefly illustrated in Fig. 6. In this stage, it is important to emphasize that the similarity of responses exists merely for the computed responses, and in general, the exact responses of the original and replaced models differ. Because of the importance of this dissimilarity, it is here explained in more detail for the case when Eq. (25.1) controls the unbalanced forces. (Other cases can be studied in a similar manner.) In this case, according to the definition of the replaced model, the original and replaced models differ in their excitations. For instance, consider the following two nonlinear mathematical models:

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{f}_{\text{int}} = \mathbf{f}^1, \tag{26.1}$$

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{f}_{\text{int}} = \mathbf{f}^2, \tag{26.2}$$

$$\mathbf{u}(t=0) = \mathbf{u}_0, \quad \dot{\mathbf{u}}(t=0) = \dot{\mathbf{u}}_0, \quad \mathbf{f}_{\text{int}}(t=0) = \mathbf{f}_{\text{int}_0}, \quad \mathbf{Q}(\mathbf{u}, \dot{\mathbf{u}}), \tag{26.3}$$

$$\mathbf{f}^1 \neq \mathbf{f}^2. \tag{26.4}$$

The two models defined by Eqs. (26), though, are identical in both initial conditions and additional constraints (Eq. (26.3)), are subjected to different excitations (Eq. (26.4)). The actual responses of the two models differ unless at least one of the models represents the behaviour of a physically unstable system. To explain this, consider the opposite case, i.e., the two models defined by Eqs. (26.3) and (26.4), and either of Eqs. (26.1) and (26.2) have identical responses. Since $\mathbf{f}^1$ and $\mathbf{f}^2$ differ, one of them should be nonzero. In view of the symmetry of Eqs. (26) with respect to $\mathbf{f}^1$ and $\mathbf{f}^2$, $\mathbf{f}^2$ is here assumed nonzero (a similar discussion is valid for $\mathbf{f}^1$). From applied linear algebra [17], a linear time-dependent transformation, e.g., $\mathbf{D}$, exists such that

$$\mathbf{D}\left(\mathbf{f}^2\right) = \mathbf{f}^1. \tag{27}$$

Applying such a transformation to Eq. (26.2), results in

$$\mathbf{D}(\mathbf{M}\ddot{\mathbf{u}}) + \mathbf{D}(\mathbf{f}_{\text{int}}) = \mathbf{f}^1. \tag{28}$$

Subtracting Eq. (28) from Eq. (26.1) and denoting the unity matrix with $\mathbf{I}$, leads to

$$(\mathbf{I} - \mathbf{D})(\mathbf{M}\ddot{\mathbf{u}}) + (\mathbf{I} - \mathbf{D})(\mathbf{f}_{\text{int}}) = \bar{\mathbf{O}}. \tag{29}$$

From an other point of view, considering the identity

$$\mathbf{I}\left(\mathbf{f}^2\right) = \mathbf{f}^2, \tag{30}$$

Eqs. (27) and (30) result in

$$(\mathbf{D} - \mathbf{I})\left(\mathbf{f}^2\right) = \mathbf{f}^1 - \mathbf{f}^2. \tag{31}$$

Considering the nature of time integration analysis, the only restriction on $\mathbf{f}^1 - \mathbf{f}^2$ is Eq. (26.4). In addition,

$$(\mathbf{D} - \mathbf{I})\bar{\mathbf{O}} = \bar{\mathbf{O}} \tag{32}$$

is an identity. Hence, considering $\mathbf{D}-\mathbf{I}$ as a linear transformation defined by Eqs. (31) and (32) in the space of excitations, the image of this space is the total space. Therefore, $\mathbf{D}-\mathbf{I}$ is invertible [17], simplifying Eq. (29) to

$$\mathbf{M}\ddot{\mathbf{u}} + \mathbf{f}_{\text{int}} = \bar{\mathbf{O}}. \tag{33}$$

Since $\mathbf{f}^2$ in the RHS of Eq. (26.2) is reasonably considered nonzero, from Eq. (33) it can be deduced that if the responses of the two models are identical, the response should also satisfy a free vibration equation of motion. Comparing Eq. (26.2) with Eq. (33), the response satisfying Eqs. (26.2), (26.3), and (26.4) is independent of $\mathbf{f}^2$. This independence cannot occur for the models of physically stable dynamic systems. Hence, in view of the assumption of physical stability, the exact responses of the models defined in Eqs. (26) cannot be identical. This implies that, as stated before, replacing the nonlinear mathematical model with another model such that the nonlinearity residual errors are forced to be zero, results in the similarity of the computed responses but not that of the actual responses. This is also evident in Fig. 6.

3.2. Convergence to non-unique responses

Consider a time integration analysis carried out with a specific arrangement of equal or unequal time steps (Fig. 5). Nonlinearity residual errors are in general nonzero. Yet, according to

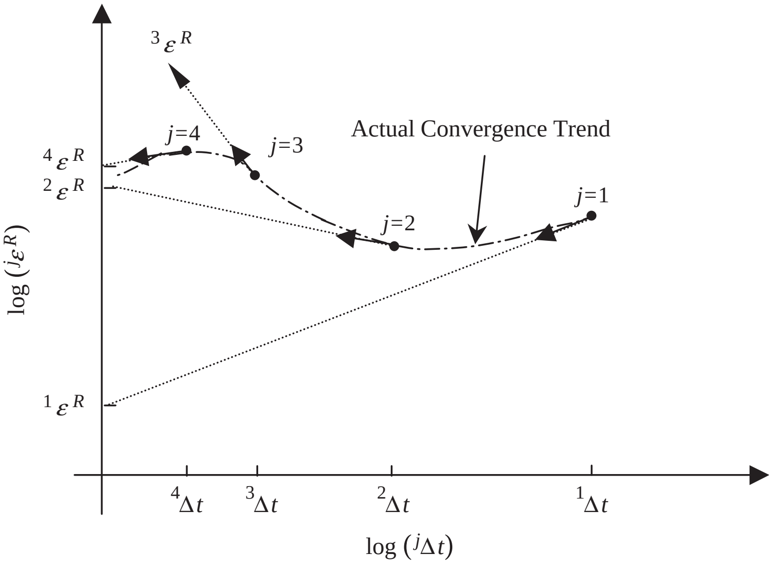

Fig. 7. The main mechanism of improper convergence in time integration of nonlinear equations of motion.

Section 3.1, the mathematical model can be replaced with a model that if analyzed with parameters identical to those in the analysis of the original model, a response identical to that of the original model will be obtained with zero residual errors. Consequently, according to Section 2, the response would converge. Nevertheless, since the zero residual errors are obtained in time integration of the replaced model, the convergence will be towards the exact response of the replaced model. Furthermore, according to the explanations in Section 3.1, the exact responses of the original and replaced models differ (Fig. 7). This consequence is independent of the time step size. Therefore, if the analysis is repeated with smaller time steps, a similar argument results in convergence towards a response different from the exact response (Fig. 7). Denoting this deviation with ${}^{j}\varepsilon^{R}$ ($j = 2$), and consecutively repeating the analysis with smaller time steps, results in

$$ {}^{j}\varepsilon^{R} = \left\| \left\{ {}^{j}\mathbf{u}^{\mathbf{R}}(t) \quad {}^{j}\dot{\mathbf{u}}^{\mathbf{R}}(t) \right\} - \left\{ \mathbf{u}(t) \quad \dot{\mathbf{u}}(t) \right\} \right\| \neq 0, \quad j = 1, 2, 3, \ldots \, . \tag{34} $$

In Eq. (34), ${}^{j}\mathbf{u}^{\mathbf{R}}(t)$,${}^{j}\dot{\mathbf{u}}^{\mathbf{R}}(t)$, $\mathbf{u}(t)$, and $\dot{\mathbf{u}}(t)$, respectively, denote the exact displacement and velocity vectors for the replaced and original models, and as before, $\| \|$ stands for an arbitrary norm. In view of the discussion in Section 3.1, the characteristics of the replaced mathematical model depend on both the characteristics of the original model and the nonzero residual errors obtained in time integration of the original model (${}_{k}\delta_{i}$). Therefore, for a specific original model analyzed by a specific time integration method, Eq. (34) leads to

$$ {}^{j}\varepsilon^{R} = R\left(\{ {}_{k}\delta_{i} \}, {}^{j}\Delta t \right). \tag{35} $$

In Eq. (35), R stands for a conceptual function, and ${}^{j}\Delta t$ denotes a representation for the size of time steps at the jth repetition of the time integration analysis. Since ${}_{k}\delta_{i}$ is the result of iterative computation terminated by Eqs. (25), ${}^{j}\varepsilon^{R}$ and ${}^{j+1}\varepsilon^{R}$ are generally unrelated. Therefore, consecutively repeating the analysis with smaller time steps does not yield any result better than the illustrations in Figs. 2 and 7 (two exceptional cases are discussed later in this sub-section).

In more detail, when repeating the time integration analysis, consecutively, with smaller time steps, responses may converge improperly, even when the errors induced by time integration, i.e., the truncation error, converge properly. The main reason is lack of control on the changes of ${}^{j}\varepsilon^{R}$ (or equivalently lack of sufficient control on the changes of nonlinearity residual errors) with respect to ${}^{j}\Delta t$. The consequence is the change of convergence direction towards the exact responses of other mathematical models. Depending on the analysis parameters, these changes of direction is considerable or negligible. Convergence towards different responses is indeed a type of inconsistency between time step sizes and nonlinearity residual errors. According to this inconsistency, while the sizes of time steps are being reduced in consistence with the desired error reduction, there is not such a control on nonlinearity residual errors. Particularly, because of the inequalities in Eqs. (25), the difference between ${}_{k}\delta_{i}$ in two consecutive time integration analyses might result in differences in order of instability between ${}^{j}\varepsilon^{R}$ and ${}^{j+1}\varepsilon^{R}$. Fig. 4 is a good example for convergence to different responses.

Since incremental computation and nonlinearity solutions should be implemented in both static and dynamic nonlinear analyses, it is instructive to briefly compare these analyses from the viewpoint of convergence. First, it is clear that in dynamic analysis, the increments imply time steps. Besides, in general, nonlinear static analysis of mathematical models associated with physically stable systems does not exhibit numerical shortcomings, specifically numerical instability. In view of the discussion already carried out for dynamic models, the reason is that in static nonlinear analysis there is no truncation error induced by incremental analysis. As a result, in static analysis, the inconsistency between the two sources of error (respectively brought about by time integration and nonlinearity iterative solution) loses meaning and the small errors that nonlinearity residual errors generate do not yield instability. From another point of view, in static analysis, the residuals of nonlinearity solutions are not only the sole source of error (neglecting the round-off error), but also unlike dynamic analysis, represent the true error at the end of nonlinearity iterations. Hence, the errors of incremental static nonlinear analysis are small in general, can be controlled by residual errors, and result in stable responses.

Returning to dynamic analysis and taking Eqs. (25) into consideration, when ${}^{j}\Delta t$ is sufficiently small the contribution of the nonlinearity residual errors in the RHS of Eq. (35) will depend on ${}^{j}\Delta t$. To explain the reason, when the time steps are very small, Eq. (25.1) is fulfilled without any iteration. Consequently, nonlinearity residual errors, depend only on ${}^{j}\Delta t$,

$$ {}^{j}\varepsilon^{R} = R\left(\left\{{}_{k}\delta_{i}\left({}^{j}\Delta t_{i}\right)\right\}, {}^{j}\Delta t\right) = g\left({}^{j}\Delta t\right), \tag{36} $$

where g stands for a conceptual function. In view of Eq. (36), ${}^{j}\Delta t$ and ${}^{j+1}\Delta t$ cause the relation between ${}^{j}\varepsilon^{R}$ and ${}^{j+1}\varepsilon^{R}$. In other words, in the range of sufficiently small time step sizes, regardless of the nonlinearity tolerances implemented in the RHS of Eq. (25.1), the ratio ${}^{j}\varepsilon^{R}/{}^{j+1}\varepsilon^{R}$ is indeed controllable by ${}^{j}\Delta t/{}^{j+1}\Delta t$. Hence, although in general, it is practically unattainable (because of round off), for many nonlinear mathematical models, i.e., Eq. (1), the responses eventually converge to exact responses. Including 'Improper Convergence' instead of 'Divergence' in the title of this paper is based on this argument. Another exceptional case is when the nonlinearity tolerance in Eq. (25.1) is very small, sufficient computational facilities are provided, and N in Eq. (25.2) is very large. In such a case, the nonlinearity residual errors can be assumed equal to zero, and once again, Eq. (36) holds. In fact, with sufficiently small nonlinearity residual errors, the exact response of the replaced model is generally very similar to that of the original model, and consequently ${}^{i}\varepsilon^{R}$ in Fig. 7 fluctuates in a narrow range about zero. Nevertheless, the notion of sufficiently small tolerances, sufficient computational facilities, and large enough N is not clear. Therefore, the responses of some nonlinear mathematical models converge properly, while the responses of some other models do not. This justifies the implementation of the word 'Probable' in the title of this paper.

4. The formulation of time integration methods

As explained in Section 2, time integration replaces the nonlinear dynamic mathematical models of structural systems with piece-wisely linear numerical models [13,48]. Furthermore, if the requirements of convergence are met for each of the linearized time steps and the nonlinearity residual errors are ideally assumed zero, the responses would properly converge. Some time integration methods, instead of the equation of motion, implement replaced balance equations. Examples of these methods are HHT [9] and Generalized-α [15,54]. In order to arrive at the response at each time station, these methods may require the characteristics of the model at more than one time step. Therefore, in view of the different characteristics of nonlinear models at different time steps, the convergence requirements at each linearized time step are not necessarily fulfilled by satisfying convergence requirements in linear analysis. To illustrate how the responses computed by different time integration methods converge in presence of nonlinearity, consider the following simple sdof model [23]:

$$\begin{gathered} M\left(\ddot{u}+\ddot{u}_g\right)+f_{\text{int}}=0, \\ M=1\times10^5\ \text{kg},\ \ C=0\ \text{N s/m},\ \ f_{\text{int}}=C\dot{u}+f_s, \\ f_s=\text{Linear Elastic}-\text{Plastic with kinematic hardening}, \\ K=1\times10^5\ \text{N/m},\ \ K_{\text{plastic}}=1\times10^4\ \text{N/m}, u_y=1\times10^{-2}\ \text{m}, \\ \ddot{u}_g=N-S\ \text{acceleration component of the El Centro strong motion [50]}, \\ u(t=0)=\dot{u}(t=0)=f_{\text{int}}(t=0)=0. \end{gathered} \tag{37}$$

One way to make this discussion independent of the discussion in Section 3 is to omit the residual errors of nonlinearity solutions. This can be achieved by employing the event-to-event nonlinearity solution method (see Section 2). The problem defined by Eq. (37) is analyzed with different time integration methods including HHT and Generalized-α (as two broadly accepted methods with balance equations different from the equation of motion) and the event-to-event nonlinearity solution method. The analyses are then repeated, consecutively, halving the time steps (and correspondingly linearly interpolating the excitation), for each new analysis. In view of the proposition in Section 2, when implementing the average acceleration method together with the event-to-event method, the only source of error vanishes in an analysis with very small time steps. (Time integration with smaller time steps will only cause negligible changes in the response.) Considering the resulting response of such an analysis as exact, one can determine the errors of displacement and velocity at $t=20$ sec and for different analyses. The result is displayed in Fig. 8. According to this figure, the responses computed by HHT, and Generalized-α converge in a comparatively improper manner. Furthermore, in view of the proposition deduced in Section 2,

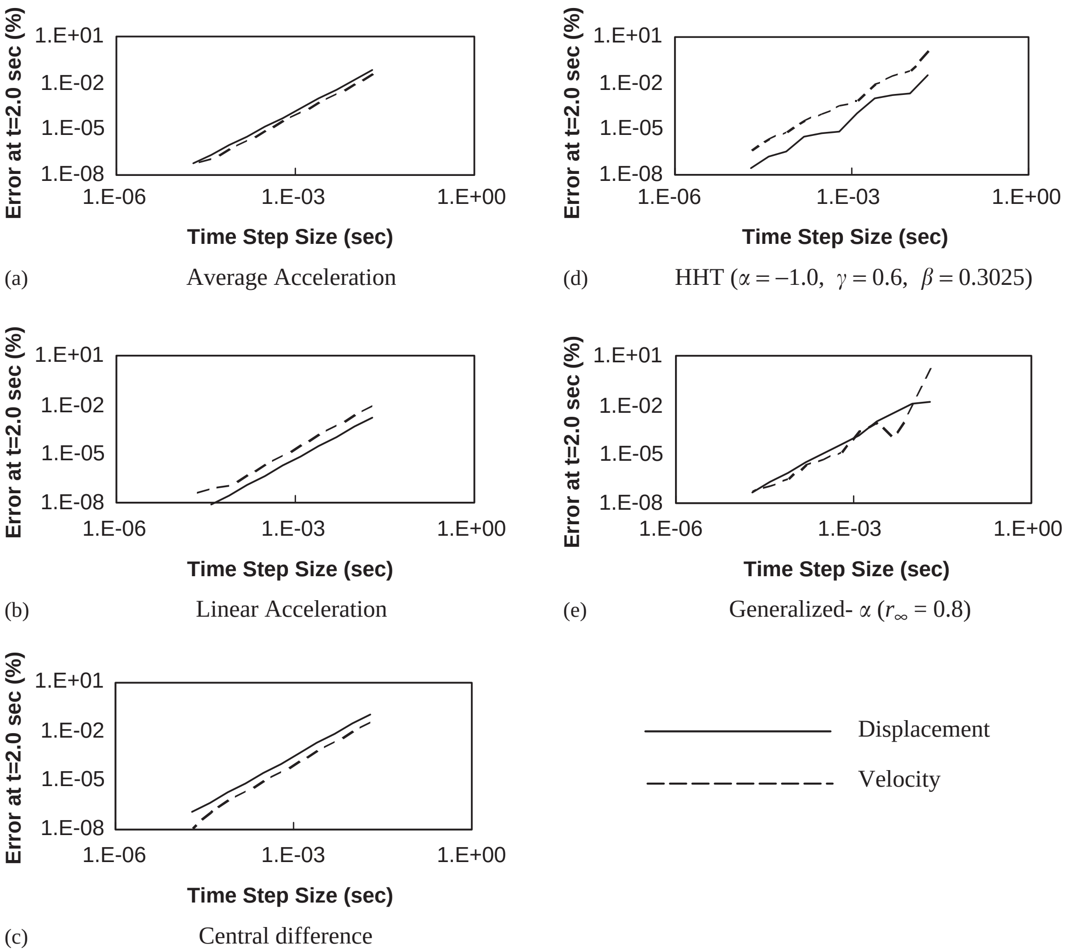

Fig. 8. Error at $t = 2.0$, when Eq. (37) is analyzed with different integration methods and the event-to-event nonlinearity solution method.

the reason should be related to the convergence requirements in linearized time steps. In order to explain this further and focus on the mechanism of improper convergence brought about by the formulations of integration methods, unless specified, in the rest of this section, the nonlinearity residual errors are ideally assumed zero and the approximation of the excitations are assumed sufficiently accurate. Now, consider the formulations of the HHT and the Generalized-α methods [20,46,54],

$$\mathbf{M}((1-\alpha_m)\ddot{\mathbf{u}}_{i+1}+\alpha_m\ddot{\mathbf{u}}_i)+\left((1-\alpha_f)\mathbf{f}_{\text{int}_{i+1}}+\alpha_f\mathbf{f}_{\text{int}_i}\right) = \left(1-\alpha_f\right)\mathbf{f}_{i+1}+\alpha_f\mathbf{f}_i, \tag{38.1}$$

$$\text{For HHT}: \alpha_m = 0, \;\; \alpha_f = -\alpha, \tag{38.2}$$

$$\dot{\mathbf{u}}_{i+1} = \dot{\mathbf{u}}_i + \Delta t((1-\gamma)\ddot{\mathbf{u}}_i + \gamma\ddot{\mathbf{u}}_{i+1}), \tag{38.3}$$

$$\mathbf{u}_{i+1} = \mathbf{u}_i + \Delta t\dot{\mathbf{u}}_i + \Delta t^2\left(\left(\tfrac{1}{2} - \beta\right)\ddot{\mathbf{u}}_i + \beta\ddot{\mathbf{u}}_{i+1}\right), \tag{38.4}$$

where α, γ, and β are the parameters of the HHT method, and α_m and α_f are the additional parameters of the Generalized-α method. Rewriting Eq. (38.1) as

$$\begin{aligned}&\left(1 - \alpha_f\right)\left(\mathbf{M}\ddot{\mathbf{u}}_{i+1} + \mathbf{f}_{\text{int}_{i+1}}\right) + \alpha_f\left(\mathbf{M}\ddot{\mathbf{u}}_i + \mathbf{f}_{\text{int}_i}\right)\\ &\quad = \left(1 - \alpha_f\right)\mathbf{f}_{i+1} + \alpha_f\mathbf{f}_i + \left(\alpha_m - \alpha_f\right)\mathbf{M}(\ddot{\mathbf{u}}_{i+1} - \ddot{\mathbf{u}}_i),\end{aligned} \tag{39}$$

and substituting $i = 0$, one can simply arrive at

$$\begin{aligned}&\left(1 - \alpha_f\right)\left(\mathbf{M}\ddot{\mathbf{u}}_1 + \mathbf{f}_{\text{int}_1}\right) + \alpha_f\left(\mathbf{M}\ddot{\mathbf{u}}_0 + \mathbf{f}_{\text{int}_0}\right)\\ &\quad = \left(1 - \alpha_f\right)\mathbf{f}_1 + \alpha_f\mathbf{f}_0 + \left(\alpha_m - \alpha_f\right)\mathbf{M}(\ddot{\mathbf{u}}_1 - \ddot{\mathbf{u}}_0).\end{aligned} \tag{40}$$

In addition, the equation of motion is satisfied at $t=0$, and for the sake of numerical stability in linear analyses $\alpha_f \leqslant 1/2$ [54]. Therefore, Eq. (40) can be simplified further to

$$\mathbf{M}\ddot{\mathbf{u}}_1 + \mathbf{f}_{\text{int}_1} = \mathbf{f}_1 + \frac{\alpha_m - \alpha_f}{1 - \alpha_f}\mathbf{M}(\ddot{\mathbf{u}}_1 - \ddot{\mathbf{u}}_0). \tag{41}$$

In view of Fig. (5), starting from Eq. (41) and implementing mathematical induction results in

$$\begin{aligned}&\mathbf{M}\ddot{\mathbf{u}}_{i+1} + \mathbf{f}_{\text{int}_{i+1}} = \mathbf{f}_{i+1} + \frac{\alpha_m - \alpha_f}{1 - \alpha_f}\mathbf{M}\sum_{j=0}^{i}\left(\frac{-\alpha_f}{1 - \alpha_f}\right)^{i-j}\left(\ddot{\mathbf{u}}_{j+1} - \ddot{\mathbf{u}}_j\right)\\ &\quad (\text{For HHT}: \alpha_m = 0,\ \alpha_f = -\alpha), \quad i = 0, 1, 2, 3, \ldots .\end{aligned} \tag{42}$$

In other words, to solve the initial value problem defined by Eq. (1), the Generalized-α method replaces the equation of motion with Eqs. (42), (38.3), and (38.4). Obviously, $\alpha_m = \alpha_f$ reduces the Generalized-α method to the Newmark method. This result is stronger than what mentioned in the literature. (According to the literature, only the case with $\alpha_m = \alpha_f = 0$ equates the Generalized-α method with the Newmark method [54].) As an evidence for this claim, the mathematical model defined by Eqs. (37) is studied here, again, with the optimized Generalized-α method [54] and $r_\infty = 1.0$ ($\alpha_m = \alpha_f = \gamma = 1/2$, $\beta = 1/4$). The resulting convergence plot is identical to what achieved in Fig. 8a. Hence, when $\alpha_m = \alpha_f$, the formulation of the Generalized-α method does not create any numerical shortcoming. In the other case, when $\alpha_m \neq \alpha_f$, Eq. (42) can be rewritten as

$$\begin{aligned}\frac{1 - \alpha_m}{1 - \alpha_f}\mathbf{M}\,\ddot{\mathbf{u}}_{i+1} + \mathbf{f}_{\text{int}_{i+1}} &= \mathbf{f}_{i+1} - \frac{\alpha_m - \alpha_f}{1 - \alpha_f}\mathbf{M}\ddot{\mathbf{u}}_i\\ &\quad + \frac{\alpha_m - \alpha_f}{1 - \alpha_f}\mathbf{M}\sum_{j=0}^{i-1}\left(\frac{-\alpha_f}{1 - \alpha_f}\right)^{i-j}\left(\ddot{\mathbf{u}}_{j+1} - \ddot{\mathbf{u}}_j\right)\\ &\quad (\text{For HHT}: \alpha_m = 0, \alpha_f = -\alpha), \quad i = 0, 1, 2, 3, \ldots .\end{aligned} \tag{43}$$

With regard to Eq. (43), implementing the definitions,

$$\tilde{\mathbf{M}} = \frac{1 - \alpha_m}{1 - \alpha_f}\mathbf{M}, \tag{44.1}$$

$$\tilde{\mathbf{f}}_{i+1} = \mathbf{f}_{i+1} - \frac{\alpha_m - \alpha_f}{1-\alpha_f}\left(\mathbf{M}\ddot{\mathbf{u}}_i - \mathbf{M}\sum_{j=0}^{i-1}\left(\frac{-\alpha_f}{1-\alpha_f}\right)^{i-j}(\ddot{\mathbf{u}}_{j+1} - \ddot{\mathbf{u}}_j)\right) \tag{44.2}$$

results in

$$\tilde{\mathbf{M}}\ddot{\mathbf{u}}_{i+1} + \mathbf{f}_{\text{int}_{i+1}} = \tilde{\mathbf{f}}_{i+1}. \tag{45}$$

From an other point of view, for the mathematical model defined below:

$$\tilde{\mathbf{M}}\ddot{\mathbf{u}}(t) + \mathbf{f}_{\text{int}}(t) = \tilde{\mathbf{f}}(t), \tag{46}$$

where $\tilde{\mathbf{f}}(t)$ stands for the exact $\tilde{\mathbf{f}}$ at t, Eqs. (45), (38.3), and (38.4) define a time integration method identical to the Newmark method. Consequently, according to the proposition in Section 2, the responses computed by time integration will converge with the rate of two, if the approximation to the RHS of Eq. (45) has not a negative effect [49]. Eq. (44.2) defines the RHS of Eq. (45), i.e., $\tilde{\mathbf{f}}_{i+1}$, in terms of the responses up to the ith time station. Hence, in view of Eq. (46), if the time stations are taken into account one by one starting from the initial conditions, the appropriate convergence of $\tilde{\mathbf{f}}$ and, consequently, that of the response will be obvious. Furthermore, assuming the continuity of acceleration with respect to time, Eq. (44.2) results in

$$\tilde{\mathbf{f}}(t_{i+1}) = \lim_{\Delta t\to 0}\tilde{\mathbf{f}}_{i+1} = \mathbf{f}(t_{i+1}) - \frac{\alpha_m - \alpha_f}{1-\alpha_f}\mathbf{M}\ddot{\mathbf{u}}(t_i). \tag{47}$$

Therefore, in view of Eqs. (43)–(47), the result of integrating Eq. (46), starting from the initial conditions, converges towards the response of Eq. (48),

$$\frac{1-\alpha_m}{1-\alpha_f}\mathbf{M}\ddot{\mathbf{u}} + \mathbf{f}_{\text{int}} = \mathbf{f} - \frac{\alpha_m - \alpha_f}{1-\alpha_f}\mathbf{M}\ddot{\mathbf{u}}. \tag{48}$$

Eq. (48) is obviously identical to the equation of motion in Eq. (1). Regarding the assumption of continuous acceleration in Eqs. (47) and (48), it should be mentioned that in order to preserve unconditional stability in linear analyses, $\alpha_m \le \alpha_f \le 1/2$ [54], and hence,

$$\left|\frac{-\alpha_f}{1-\alpha_f}\right| \le 1. \tag{49}$$

Consequently, when parameters are set appropriately, e.g., $\alpha_m \le \alpha_f < 1/2$ (instead of $\alpha_m \le \alpha_f \le 1/2$ [54]) for the Generalized-α method, the negative effect of discontinuous acceleration will diminish after a few time steps. As a result, for such a selection of parameters, convergence is maintained. Nevertheless, even when convergence is preserved (Fig. 1), the errors change improperly (nonlinearly) for large enough time steps. Therefore, what Figs. 8d and e display is indeed error variations in the range of time steps larger than needed for the convergence trend illustrated in the intermediate part of Fig. 1. To explain the mechanism, Eq. (44.2) relates the convergence of the RHS of Eq. (46) (at an arbitrary time station) to the convergence of the acceleration increments at the preceding time stations. Hence, convergence of the response at a specific time station, i.e., t_{i+1}, depends on the convergence of acceleration increments at the preceding time stations, i.e., $\ddot{u}_j - \ddot{u}_{j-1}$, $j = 1, 2, \ldots, i$ [49]. Convergence of the acceleration increments at the preceding time stations can each be depicted in a log–log diagram similar to Fig. 1. Disregarding the effect of round off, these diagrams, each, have a linear and a nonlinear section (Fig. 1). Consequently, the error at a specific time station would converge improperly, if the increment of acceleration in one of the preceding time stations converges improperly. Besides, considering Eqs. (45), (44.2), (38.3), and (38.4), the responses computed by the Generalized-α method, at best, converge as the responses brought about by the corresponding Newmark method (the Newmark method with parameters equal to the corresponding parameters in the Generalized-α method). Returning to the convergence of acceleration increments, as the behaviour becomes more complicated, the linear part in Fig. 1 occurs for much smaller time steps [21]. For the problem under consideration, i.e., Eq. (37), abrupt changes of acceleration are the source of complexity (see Fig. 9). In view of the nonlinear behaviour, the time histories of stiffness and acceleration increment are depicted in Fig. 10. A comparison between Figs. 9 and 10 clearly demonstrates abrupt changes of stiffness as the reason of abrupt changes of acceleration. Therefore, in time integration with the Generalized-α method, the influence of responses at time stations prior to t_i on the response at t_i postpones the convergence. This delay depends on the complexity of the response (which is considerable in presence of the nonlinearities induced by abrupt changes of stiffness) and the values implemented for the parameters of the Generalized-α method. Accordingly, the responses in Figs. 8d and e converge, but satisfactorily only at time steps smaller than the time steps in Figs. 8a–c. In the limiting case that Eq. (49) is satisfied as an equality, i.e., $\alpha_f = 1/2$, and $\alpha_f \neq \alpha_m$, the convergence of acceleration increments at all time stations prior to t_{i+1} affects convergence of the response at t_{i+1}. Hence, in view of the increasing number of time steps in time integration with smaller time steps, convergence delay changes form to convergence with the reduced rate of one. Fig. 11 is a repetition of Fig. 8 with the Generalized-α

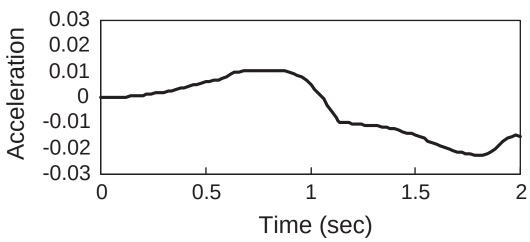

Fig. 9. Time history of acceleration, when Eq. (37) is analyzed with the average acceleration method ($\Delta t = 0.02/2^{10}$) and the event-to event nonlinearity solution method.

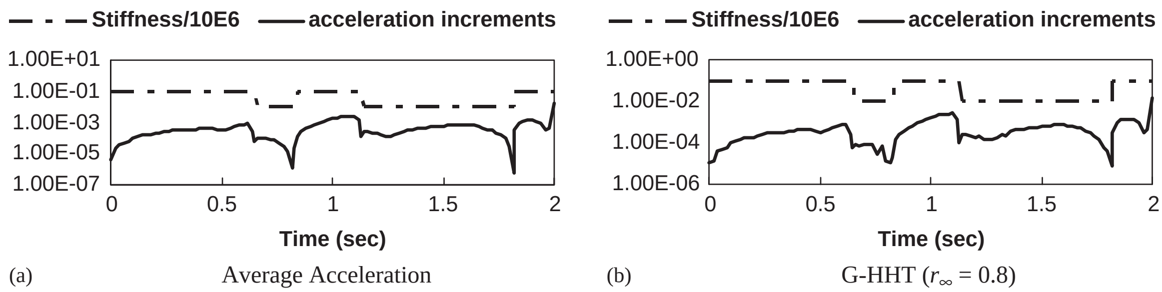

Fig. 10. Time histories of stiffness and acceleration increment, when Eq. (37) is analyzed with time integration ($\Delta t = 0.02/2^{10}$) and the event-to-event nonlinearity solution method.

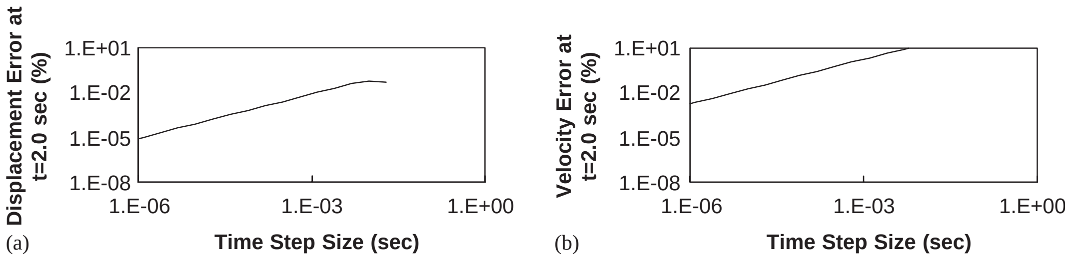

Fig. 11. Error at $t = 20$, when Eq. (37) is analyzed with the Generalized-α ($\alpha_m = 1/3$, $\alpha_f = 1/2$, $\beta = 1/3$, $\gamma = 2/3$) time integration method and the event-to-event nonlinearity solution method.

method using $\alpha_m = 1/3, \alpha_f = 1/2, \beta = 1/3, \gamma = 2/3$. Though, this selection fulfils the requirements of convergence with the rate of two [54], the rate of convergence in Fig. 11 is one.

Before closing this section, it is instructive to form an idea about the relative effects of the numerical shortcomings discussed here and in the previous section. In this regard, the simple problem defined in Eq. (37) is revisited after replacing the event-to-event method with the fractional-time-stepping nonlinearity solution method (with at most six iterations at each detection of nonlinearity, and at each iteration dividing the sizes of the time steps by 10) [35,36] and taking 0.001 as the relative displacement tolerance. Fig. 12 displays the outcome. In view of Fig. 1, the convergence shortcomings in Figs. 12a–e are almost similar and more severe than the shortcomings in Fig. 8. This comparison implies that, even for time integration methods with balance equations different from the equation of motion, improper control of nonlinearity residual errors is the major source of convergence shortcomings.

5. Nonzero residual errors with constant effects

Nonlinearity solution methods are in general iterative. These iterations end, when Eq. (25.1), Eq. (25.2), or both are satisfied. Furthermore, because of the iterative nature of nonlinearity solution methods, and lack of unlimited computational facilities,

$$_k\delta_i^j \neq 0. \tag{50}$$

In the limit of convergence, i.e., when time steps approach zero, $_k\delta_i^j$ approaches zero. Nevertheless, since time steps are never zero, Eq. (50) still holds. According to the discussion at the end of Section 3, in the limit of zero time steps, the effects of nonlinearity residual errors on convergence disappear. To put in more clearly, theoretically, for small enough time steps, the residuals of nonlinearity solutions decrease to the desired small values, and Fig. 2 reforms to Fig. 1. Yet, a shortcoming might occur. It is the fact that, for some special dynamic models, time integration cannot precisely consider the characteristics of the model even in the limit of zero time steps. Mathematical models defining the behaviour of dynamic systems involved in dry friction (Coulomb damping) can provide a good example. Because of Eq. (50), in time integration of these models, the instants at which velocities change sign are not determined exactly. This fact results in additional (residual) frictional forces, at least, at time instants at which velocities change sign. Regardless of the sizes of time steps and the precision achieved in Eq. (50), these additional forces

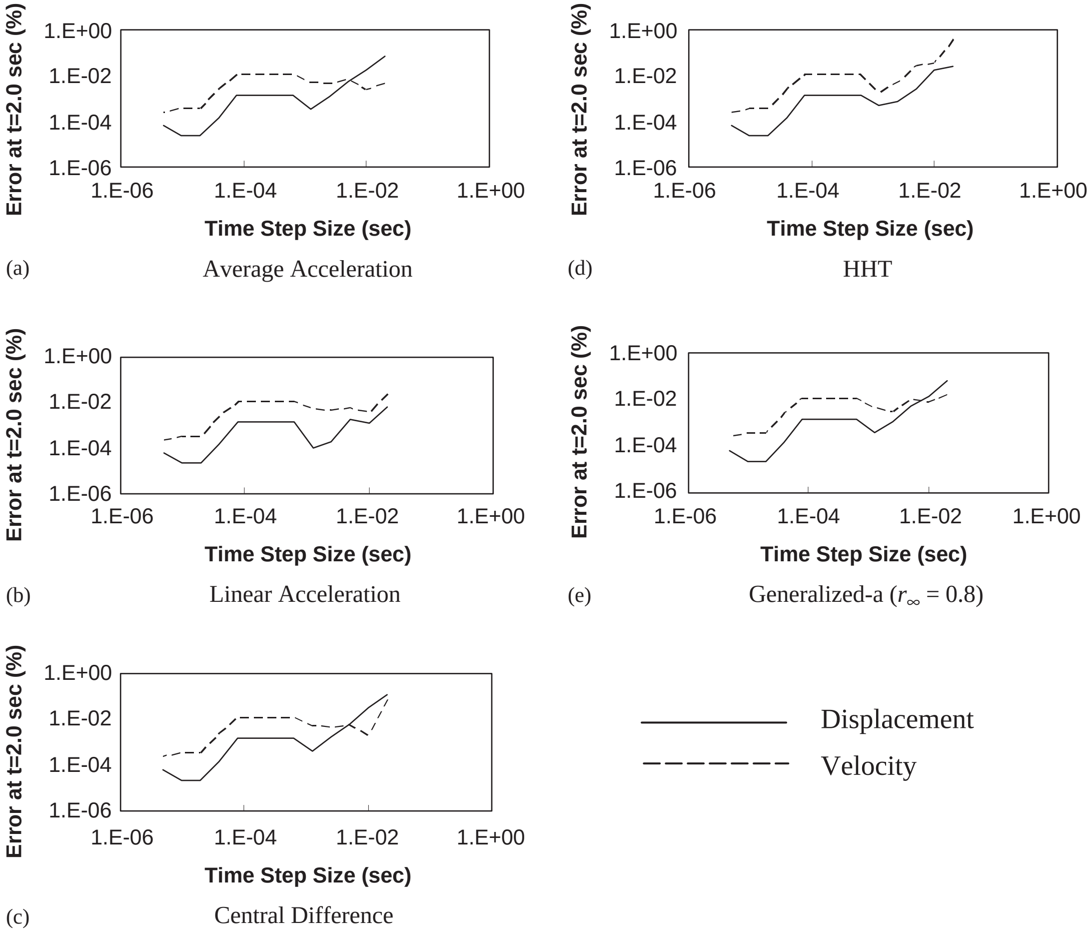

Fig. 12. Error at $t = 2.0$, when Eq. (37) is analyzed with different time integration methods and the fractional-time-stepping nonlinearity solution method.

are constant in size. This constancy, together with the fact that the residual forces at different time instants cannot cancel each other out, is the source of shortcoming. In brief, since at the end of iterative nonlinearity solutions, Eq. (50) generally holds, for models involved in friction, additional frictional forces exist even in the limit of zero time steps. These forces prevent convergence towards the response of the actual mathematical model. It is also worth noting that, artificially defining an appropriate relation between residual errors and time steps, e.g., replacing the function defining the friction with a regularized function, induces other error sources, and thus is not effective [55].

For more clarification, consider the following simple mathematical model [55]:

$$\begin{gathered} \ddot{u} + 2\xi\dot{u} + u + \bar{\alpha}sgn[\dot{u}] = \sin(\bar{\beta}t), \\ \xi = 0.05, \quad \bar{\alpha} = 0.5, \quad \bar{\beta} = 1.0, \\ u(t=0) = \dot{u}(t=0) = 0. \end{gathered} \tag{51}$$

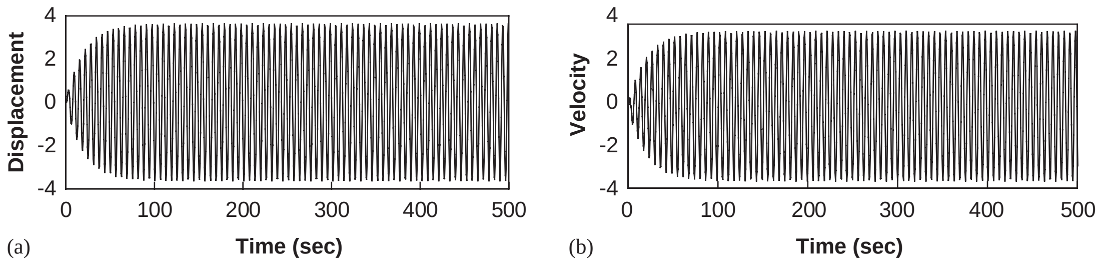

Fig. 13. The response of the model defined by Eq. (51).

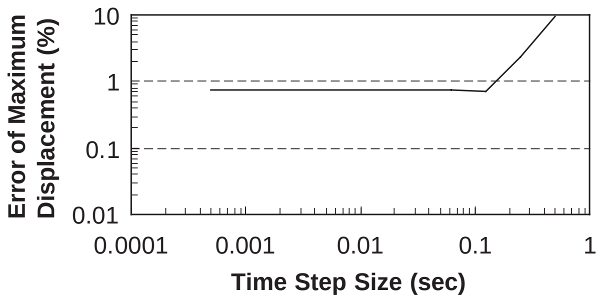

Fig. 14. Convergence trend for models defined by Eq. (51) and analyzed by time integration.

Eq. (51) is studied here in view of different time integration methods and different parameters of the fractional-time-stepping method [56]. The response is shown in Fig. 13 for the first 500 sec. According to the literature, the response that time integration generates for Eq. (51) converges only up to an especial accuracy, i.e., 1% [55]. After proper convergence to this limit, the response diverges gradually [55,56] (Fig. 14) mainly because of round off. A question that might arise here is why the response converges properly up to a certain limit, whereas according to Section 3, nonzero nonlinearity residual errors can induce shortcomings such as shown in Fig. 2. The response is simple. The occurrence of shortcomings explained in Section 3 is only probable. In addition, the shortcomings vanish for small enough time steps; and besides, the values of stiffness and mass in Eq. (51) dictate a resonance condition. Accordingly, in view of the type of nonlinearity, the linear behaviour dominates. Therefore, the response converges similar to linear analyses, but towards the response of the model definable in the limit of zero time steps. The effect of the shortcoming discussed in this section, seems insignificant. Nevertheless, it is a potential source of improper convergence, and even when nonlinearity residual errors are sufficiently small and the time integration formulation is appropriate, can lead to hidden shortcomings for large nonlinear dynamic models. This mechanism is in perfect agreement with the proposition in Section 2 and is independent from the sources discussed in Sections 3 and 4.

6. Limited computational facilities and their inappropriate consideration

Any limitation on computational facilities is by itself an obstacle to consistent reduction of the two major errors affecting time integration of nonlinear equations of motion (see Section 3).

Furthermore, after a special number of iterations, nonlinearity residual errors generally become so small that, even with more iteration, more accuracy is not attainable. With such small residual errors, if Eq. (25.1) is not satisfied, it will not be satisfied with more iteration. Unbiased round off is the likely result of more iteration. In view of the fact that, in many situations, computational facilities cannot be considered unlimited, in order to proceed to the next time step, it is essential to stop the iterations by considering the computational facilities besides controlling the residual errors. In this regard, it is an accepted practice to implement Eq. (25.2), and treat with N as an analysis parameter selected in advance [57]. From this point of view, if the nonlinearity solution method is considered in its convergence range and the rate of convergence is denoted with q', Fig. 15 will schematically exhibit the variation of the nonlinearity residual errors during the iterations of nonlinearity solution. Point A in Fig. 15 represents the residual error at the start of nonlinearity iterations. Carrying out the iterative computation, iterations will end at B in section I, II, or III, respectively when Eq. (25.1), Eq. (25.2), or both are satisfied. For a specific nonlinearity solution method, q' is a positive integer independent of the number of iterations, e.g., two for Newton Raphson. Therefore, the location of B in Fig. 15 depends on the starting point A. Nevertheless, regardless of the nonlinearity solution method and the corresponding parameters, point A might be located in a top region in Fig. 15. The dashed line in Fig. 15 displays the variation of residual errors (with respect to the number of iterations) starting from such a point. This line terminates in section II with a considerably large residual error, whereas, in view of the location of the line $\log\left(|{}_k\delta_i^j|\right) = \log\left(\bar{\delta}\right)$, more accuracy is attainable with more iteration. In general, N in Eq. (25.2) is being selected regardless of the residual error at the start of nonlinearity iterations. As a result, the iterative computation might end without satisfying Eq. (25.1), while the computational facilities can provide more accuracy. Therefore, even if the requirements of convergence in linear analysis are fulfilled, nonlinearity solution methods converge, the residual errors are controlled such that to avoid the problem explained in Section 3, appropriate time integration methods are implemented, and the models are not involved in problems explained in Section 5, limitations on computational facilities and their inappropriate consideration by parameter N may result in improper convergence for the responses that time integration brings

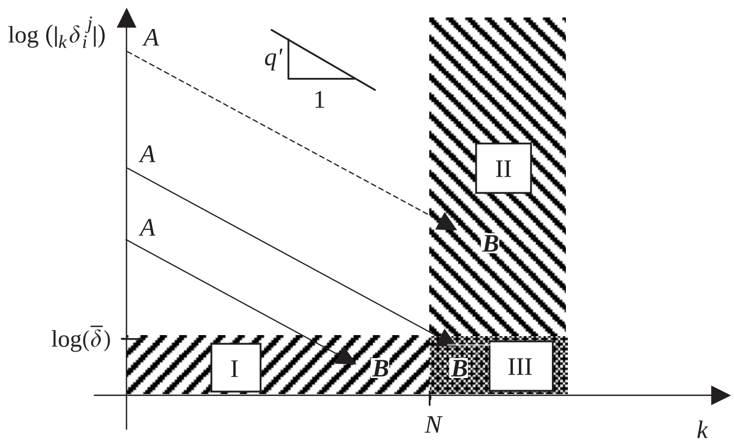

Fig. 15. Typical convergence for nonlinearity residual errors.

about for Eq. (1). Based on this discussion, underestimating computational facilities by selecting small values for N is in relation to the source of shortcoming discussed in Section 3. On the other hand, selecting large values for N adds to the effect of round off, and is therefore a source of shortcoming independent of the explanations in Sections 3–5.

7. Conclusions

In time integration of the general case of nonlinear dynamic models of structural systems, the responses do not necessarily converge in a proper manner. In this paper, the reasons are studied for nonlinear semi-discretized mathematical models of physically stable systems. The achievements are as noted below:

1. Improper convergence may occur even for small and simple nonlinear models.
2. The major reason for improper convergence is insufficient control on the residuals generated by iterative nonlinearity solutions. In this regard,
   (a) In time integration of nonlinear semi-discretized equations of motion, theoretically, the computed responses are affected by both the errors of time integration (truncation error) and the residuals of nonlinearity solutions.
   (b) Responses computed by time integration converge to responses that—depending on the implemented time steps—differ from the exact responses. Consequently, this source of improper convergence is indeed a continuous change of convergence direction (Fig. 7) that occurs because of the inconsistency between the inherent errors of time integration and the nonlinearity residual errors. Omitting either of these two errors, in static nonlinear or dynamic linear analyses, leaves out the possibility of improper convergence.
   (c) Improper convergence due to insufficient control of nonlinearity residual errors occurs regardless of the integration method, the iterative nonlinearity solution method, and the mathematical model.
   (d) Improper convergence due to insufficient control of nonlinearity residual errors may induce errors in order of instability.
   (e) The devastating effect of nonlinearity residual errors on convergence vanishes in the limit of very small time steps and/or very small nonlinearity residual errors.
3. Another source of improper convergence is the formulation of the time integration methods (e.g., Generalized-α) that replace the equation of motion with other balance equations. In time integration with these methods, convergence requirements in each linearized time step of a nonlinear analysis are not necessarily fulfilled by satisfying those requirements in linear analysis [2]. For the special case of the Generalized-α method (applied to linear or nonlinear problems),
   (a) The computed responses converge depending on both the time history of acceleration (the convergence of acceleration increments at the preceding time stations) and the values implemented for the parameters of the integration method.
   (b) The computed responses, at best, converge as the responses brought about by the Newmark method (with values of γ and β similar to those in the Generalized-α method).
   (c) When $\alpha_m = \alpha_f$, the Generalized-α method reduces to the Newmark method and the formulation does not induce improper convergence.

(d) In the other case, when $\alpha_m \neq \alpha_f$ and $\alpha_f \neq 1/2$, the complexities in the history of acceleration postpone convergence to the range of smaller time steps generally affected by round off. In presence of nonlinearity, abrupt changes of stiffness are a source of complexity.
(e) In the special case when $\alpha_m \neq \alpha_f$ and $\alpha_f = 1/2$, the rate of convergence may be reduced to one.

4. In time integration of some mathematical models, iterative nonlinearity solutions cannot omit all residuals, even asymptotically, and in the limit of small time steps and small nonlinearity tolerances. In these cases, convergence can at most occur towards the exact response of the limiting model, i.e., the model defined in time integration with sufficiently small time steps and sufficiently small nonlinearity tolerances. Problems involved in dry friction introduce an example.
5. Limitations on computational facilities and inadequate consideration of these limitations can impair convergence either by preventing the reduction of nonlinearity residual errors or by adding to the effects of round off.

Acknowledgements

The authors gratefully acknowledge the helpful discussions they had on different parts of this paper with Prof. Arthur R. Robinson, Dr. K. Ghayour, and Mr. K. Arzhangi. The first and third authors also express their sincere appreciation to the financial assistance of the Research Council of the University of Tehran for the research report in correspondence to this paper. In addition, the last but by no means the least thanks are given to Prof. A. Der Kiureghian and Prof. J. Retief who caused the need to such a study come into the view of the first author.

References

[1] P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1962.
[2] T. Belytschko, T.J.R. Hughes, *Computational Methods for Transient Analysis*, Elsevier, Amsterdam, 1983.
[3] P. Wriggers, *Computational Contact Mechanics*, Wiley, New York, 2002.
[4] T.J.R. Hughes, K.S. Pister, R.L. Taylor, Explicit-explicit finite elements in nonlinear transient analysis, *Computer Methods in Applied Mechanics and Engineering* 17/18 (1979) 159–182.
[5] J.C. Houbolt, A recurrence matrix solution for the dynamic response of elastic aircraft, *Journal of the Aeronautical Sciences* 17 (1950) 540–550.
[6] N.M. Newmark, A method of computation for structural dynamics, *Journal of Engineering Mechanics* 85 (1959) 67–94.
[7] C.W. Gear, The automatic integration of stiff ordinary differential equations, in: A.J.H. Morell (Ed.), *Information Processing*, North-Holland, Amsterdam, 1969, pp. 187–193.
[8] K.J. Bathe, E.L. Wilson, Stability and accuracy analysis of direct integration methods, *Earthquake Engineering and Structural Dynamics* 1 (1973) 283–291.
[9] H.M. Hilber, T.J.R. Hughes, R.L. Taylor, Improved numerical dissipation for time integration algorithms in structural dynamics, *Earthquake Engineering and Structural Dynamics* 5 (1977) 283–292.
[10] G. Bazzi, E. Anderheggen, The ρ-family of algorithms for time step integration with improved numerical dissipation, *Earthquake Engineering and Structural Dynamics* 10 (1982) 537–550.

[11] C. Hoff, P.J. Pahl, Development of an implicit method with numerical dissipation for time integration algorithms in structural dynamics, *Computer Methods in Applied mechanics and Engineering* 67 (1987) 367–385.
[12] G.M. Hulbert, J. Chung, Explicit time integration algorithms for structural dynamics with optimal numerical dissipation, *Computer Methods in Applied Mechanics and Engineering* 17 (1996) 175–188.
[13] C.C. Chen, A.R. Robinson, Improved time-history analysis for structural dynamics. I: Treatment of rapid variation of excitation and material nonlinearity, *Journal of Engineering Mechanics* 119 (1993) 2496–2513.
[14] J.Y. Cho, S.J. Kim, An explicit discontinuous time integration method for dynamic contact/impact problems, *Computer Modeling in Engineering and Science* 3 (2002) 687–698.
[15] S. Erlicher, L. Bonaventura, O.S. Bursi, The analysis of the Generalized-α method for nonlinear dynamic problems, *Computational Mechanics* 28 (2002) 83–104.
[16] A. Ralston, P. Rabinowitz, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1983.
[17] B. Noble, J.W. Daniel, *Applied Linear Algebra*, Prentice-Hall, Englewood Cliffs, NJ, 1977.
[18] C.W. Gear, *Numerical Initial Value Problem in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, 1971.
[19] A. Soroushian, J. Farjoodi, More reliable responses for time integration analyses, *Structural Engineering and Mechanics* 16 (2003) 219–240.
[20] W.L. Wood, *Practical Time Stepping Schemes*, Clarendon Press, Oxford, 1990.
[21] A. Soroushian, New Methods to Maintain Responses' Convergence and Control Responses' Errors in the Analysis of Nonlinear Dynamic Models of Structural Systems, PhD Thesis, University of Tehran, 2003.
[22] A. Soroushian, J. Farjoodi, Convergence of the responses that time integration generates for contact problems, *Proceedings of the Fourth International Conference on Seismology and Earthquake Engineering*, Tehran, May 2003 (in Persian)
[23] A. Soroushian, J. Farjoodi, Responses convergence in time integration of nonlinear semi-discrete equations of motion, *Proceedings of the Eighth International Conference on Recent Advances in Structural Dynamics*, Southampton, July 2003.
[24] Y.M. Xie, G.P. Steven, Instability, chaos, and growth and decay of energy of time-stepping schemes for nonlinear dynamic equations, *Communications in Numerical Methods in Engineering* 10 (1994) 393–401.
[25] K.C. Park, An improved stiffly stable method for direct integration of nonlinear structural dynamics equations, *Journal of Applied Mechanics* 42 (1975) 464–470.
[26] T. Belytschko, D.F. Schoeberle, On the unconditional stability of an implicit algorithm for structural dynamics, *Journal of Applied Mechanics* 42 (E) (1975) 865–869.
[27] W.L. Wood, M.E. Oduor, Stability properties of some algorithms for the solution of nonlinear dynamic vibration equation, *Communications in Applied Numerical Methods* 4 (1988) 205–212.
[28] P.G. Reinhall, T.K. Caughey, T.K. Storti, D.W. Storti, Order and chaos in a discrete Duffing oscillator: implications in numerical integration, *Journal of Applied Mechanics* 56 (1989) 162–167.
[29] A. Cardona, M. Geradin, Time integration of the equations of motion in mechanism analysis, *International Journal Computers and Structures* 33 (1989) 801–820.
[30] Y.M. Xie, An assessment of time integration schemes for nonlinear dynamic equations, *Journal of Sound and Vibration* 192 (1996) 321–331.
[31] S. Rashidi, M.A. Saadeghvaziri, Seismic modeling of multi-span simply supported bridges using Adina, *Computers and Structures* 64 (1997) 1025–1039.
[32] M.A. Crisfield, G. Jelenic, Y. Mi, H.J. Zhong, Z. Fan, Some aspects of the non-linear finite element method, *Finite Elements in Analysis and Design* 27 (1997) 19–40.
[33] T. Belytschko, W.K. Liu, B. Moran, *Nonlinear Finite Elements for Continua and Structures*, Wiley, New York, 2001.
[34] R.W. Clough, J. Penzien, *Dynamics of Structures*, McGraw-Hill, Singapore, 1993.
[35] J.M. Nau, Computation of inelastic spectra, *Journal of Engineering Mechanics* 109 (1983) 279–288.
[36] S. A. Mahin, J. Lin, Construction of inelastic response spectra for single degree-of-freedom systems, University of California, Berkeley Report No. UCB/EERC-83/17 1983.
[37] J. Chung, J.M. Lee, A new family of explicit time integration methods for linear and non-linear structural dynamics, *International Journal for Numerical Methods in Engineering* 37 (1994) 3961–3976.

[38] O.A. Bauchau, G. Damilano, N.J. Theron, Numerical integration of nonlinear elastic multibody systems, *International Journal for Numerical Methods in Engineering* 38 (1995) 2727–2751.
[39] X. Chen, K.K. Tamma, D. Sha, Virtual-pulse time integral methodology: a new approach for computational dynamics, *Finite Elements in Analysis and Design* 20 (1995) 195–204.
[40] T.A. Laursen, V. Chawla, Design of energy conserving algorithms for frictionless dynamic contact problems, *International Journal for Numerical Methods in Engineering* 40 (1997) 863–886.
[41] D. Kuhl, M.A. Crisfield, Energy-conserving and decaying algorithms in nonlinear structural dynamics, *International Journal for Numerical Methods in Engineering* 40 (1999) 569–599.
[42] A. Hernandez, C.H. Pinto, E. Amezua, H. Fernandez, Analysis of the components of discretization error in nonlinear structural problems, *Finite Elements in Analysis and Design* 39 (2003) 835–864.
[43] P. Betsch, P. Steinmann, Conservation properties of a time FE method-part IIII: time-stepping schemes for non-linear elastodynamics, *International Journal for Numerical Methods in Engineering* 8 (2001) 1931–1955.
[44] H. Kardestuncer, *Finite Element Handbook*, McGraw-Hill, New York, 1987.
[45] T.J.R. Hughes, *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
[46] T.J.R. Hughes, Stability, convergence and growth and decay of energy of the average acceleration method in nonlinear structural dynamics, *Computers and Structures* 6 (1976) 313–324.
[47] K.J. Bathe, A.P. Cimento, Some practical procedures for the solution of nonlinear finite element equations, *Computer Methods in Applied Mechanics and Engineering* 22 (1980) 59–85.
[48] D. Bernal, Locating events in step-by-step integration of equations of motion, *Journal of Structural Engineering* 117 (1991) 530–545.
[49] S.N. Penry, W.L. Wood, Comparison of some single-step methods for the numerical solution of the structural dynamic equation, *International Journal for Numerical Methods in Engineering* 21 (1985) 1941–1955.
[50] A.K. Chopra, *Dynamics of Structures: Theory and Application to Earthquake Engineering*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
[51] J.L. Humar, *Dynamics of Structures*, Balkema, Rotterdam, 2002.
[52] G.B. Warburton, Formulae for errors for initial displacement and velocity problems using the Newmark method, *Earthquake Engineering and Structural Dynamics* 18 (1989) 565–573.
[53] R.D. Richtmyer, K.W. Morton, *Difference Methods for Initial Value Problems*, Wiley, New York, 1967.
[54] J. Chung, G.M. Hulbert, A time integration algorithm for structural dynamics with improved numerical dissipation: the Generalized-α method, *Journal of Applied Mechanics* 60 (1993) 371–375.
[55] N. Mostaghel, T. David, Representations of coulomb friction for dynamic analysis, *Earthquake Engineering and Structural Dynamics* 26 (1997) 541–548.
[56] A. Soroushian, J. Farjoodi, Lack of convergence in time integration of dynamic systems involving friction, *Proceedings of the International Conference on Computational and Experimental Engineering and Sciences*, Korfu, July 2003.
[57] A. Soroushian, J. Farjoodi, An improvement in nonlinear analysis, *Proceedings of the 15th Engineering Mechanics Division Conference, ASCE*, Washington, June 2002.