Engineering Analysis with Boundary Elements 73 (2016) 181–190

Contents lists available at ScienceDirect

Engineering Analysis with Boundary Elements

journal homepage: www.elsevier.com/locate/enganabound

A new direct time integration method for the semi-discrete parabolic equations

John T. Katsikadelis

School of Civil Engineering, National Technical University of Athens, Athens GR-15780, Greece

ARTICLE INFO

Keywords:
Semi-discrete parabolic equations
Analog equation method
Differential equations
Integral equation methods
Numerical solution

ABSTRACT

A direct time integration method is presented for the solution of systems of first order ordinary differential equations, which represent semi-discrete diffusion equations. The proposed method is based on the principle of the analog equation, which converts the N coupled equations into a set of N single term uncoupled first order ordinary differential equations under fictitious sources. The solution is obtained from the integral representation of the solution of the substitute single term equations. The stability and convergence of the numerical scheme is proved. The method is simple to implement. It is self-starting, unconditionally stable, accurate, while it does not exhibit numerical damping. The stability does not demand symmetrical and positive definite coefficient matrices. This is an important advantage, since the scheme can solve semi-discrete diffusion equations resulting from methods that do not produce symmetrical matrices, e.g. the boundary element method. The method applies also to equations with variable coefficients as well as to nonlinear ones. It performs well when long time durations are considered and it can be used as a practical method for integration of stiff parabolic equations in cases where widely used methods may fail. Numerical examples, including linear as well as non linear systems, are treated by the proposed method and its efficiency and accuracy are demonstrated.

1. Introduction

The initial value problem for the linear semi-discrete diffusion equation is stated as

$$\mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{p}(t), \quad t \in [0, T], \qquad T > 0 \tag{1}$$

$$\mathbf{u}(0) = \mathbf{u}_0 \tag{2}$$

where $\mathbf{u} = \mathbf{u}(t)$ represents the vector of the unknown functions, $\mathbf{p}(t)$ the external source vector, and $\mathbf{u}_0$ a given vector.

Eq. (1) represents the semi-discrete parabolic equations, that is, parabolic equations of which the space variables have been discretized and the time variable is left continuous. The matrices $\mathbf{C}$ and $\mathbf{K}$ may be symmetrical or not, depending on the method used for the spatial discretization.

The most well known and commonly used methods for solving Eq. (1) are members of the generalized trapezoidal family or $\alpha -$ family of methods, in which the time derivative is approximated by a weighted average of the field function at two consecutive time steps. Some well known members are identified from the value of the parameter α. Thus we have, the forward differences or forward Euler ($\alpha = 0$), the trapezoidal rule or Crank-Nicolson ($\alpha = 1/2$), and backward difference or backward Euler ($\alpha = 1$). The $\alpha -$ family of approximations is unconditionally stable for $\alpha \geq 1/2$, while for $\alpha < 1/2$ the methods are stable if the condition $\Delta t < \Delta t_{cr} \equiv 2/(1 - 2a)\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of the matrix $\mathbf{C}^{-1}\mathbf{K}$, is satisfied with $\mathbf{C}, \mathbf{K}$ being symmetrical and positive definite [1,2].

In this paper a new direct time integration method is presented for the numerical solution of the initial value problem (1) and (2). The proposed method is based on the principle of the analog equation [3], which converts the N coupled equations into a set of N single term uncoupled first order ordinary differential equations under fictitious sources, unknown in the first instance. The fictitious sources are established from the integral representation of the solution of the substitute single term equations. The solution procedure is analogous to that presented in [4]. But in this case the substitute equation is of the first order and the unknown fictitious source represents the first derivative. The stability is proved and the convergence is shown through well corroborated numerical results. The method is simple to implement. It is self starting, unconditionally stable and accurate. The stability does not demand symmetrical and positive definite coefficient matrices $\mathbf{C}, \mathbf{K}$ as the widely used methods, but can solve equations with non-symmetrical matrices provided that the eigenvalues of the matrix $\mathbf{C}^{-1}\mathbf{K}$ are nonnegative or have nonnegative real part for the complex eigenvalues. This is an important advantage, since the scheme can solve semi-discrete diffusion equations resulting from methods that do not produce symmetrical matrices, e.g. the boundary element method.

E-mail addresses: jkats@central.ntua.gr.

http://dx.doi.org/10.1016/j.enganabound.2016.09.009
Received 10 May 2016; Received in revised form 15 September 2016; Accepted 15 September 2016
0955-7997/

Moreover, the method performs well when long time durations are considered as it conserves energy and, thus, it can be used as a practical method for integration of the parabolic equations in cases where widely used methods do not apply or may fail. It applies also to the case of time dependent coefficient matrices, i.e., $\mathbf{C}(t)$, $\mathbf{K}(t)$, as well as for nonlinear equations. The method is illustrated by solving several equations, including linear as well as non linear systems. The obtained results are in excellent agreement with those obtained from exact solutions.

2. The linear system

2.1. The AEM solution

We illustrate the AEM (Analog Equation Method) with the linear one-degree-of-freedom system

$$c\dot{u} + ku = p(t), \quad c, k > 0 \tag{3}$$

$$u(0) = u_0 \tag{4}$$

Let $u = u(t)$ be the sought solution. Then, if the operator d/dt is applied to it, yields

$$\dot{u} = q(t) \tag{5}$$

where $q(t)$ is a fictitious source, unknown in the first instance. Eq. (5) is the analog of Eq. (3) [3]. It indicates that the solution of Eq. (3) can be obtained by solving Eq. (5) under the initial condition (4), if $q(t)$ is first established. This is implemented as following.

Taking the Laplace transform of Eq. (5) we obtain

$$sU(s) - u(0) = Q(s)$$

or

$$U(s) = \frac{1}{s}u(0) + \frac{1}{s}Q(s) \tag{6}$$

where $U(s)$, $Q(s)$ are the Laplace transforms of $u(t)$, $q(t)$. The inverse Laplace transform of expression (6) yields the solution of Eq. (5) in integral form

$$u(t) = u(0) + \int_0^t q(\tau)d\tau \tag{7}$$

Thus the initial value problem of Eqs. (3) and (4) is transformed into an equivalent Volterra integral equation for $q(t)$, Eq. (7).

Eq. (7) is solved numerically within a time interval $[0, T]$. Following a procedure analogous to that presented in [4], the interval $[0, T]$ is divided into N equal intervals $\Delta t = h$, $h = T/N$ (Fig. 1), in which $q(t)$ is assumed to vary according to a certain law, e.g. constant, linear etc.

Hence, Eq. (7) at instant $t = nh$ can be written as

$$\begin{aligned} u_n &= u_0 + \int_0^h q(\tau)d\tau + \int_h^{2h} q(\tau)d\tau + \ldots + \int_{(n-1)h}^{nh} q(\tau)d\tau \\ &= u_0 + \sum_{r=1}^{n-1}\int_{(r-1)h}^{rh} q(\tau)d\tau + \int_{(n-1)h}^{nh} q(\tau)d\tau \end{aligned} \tag{8}$$

or taking into account that

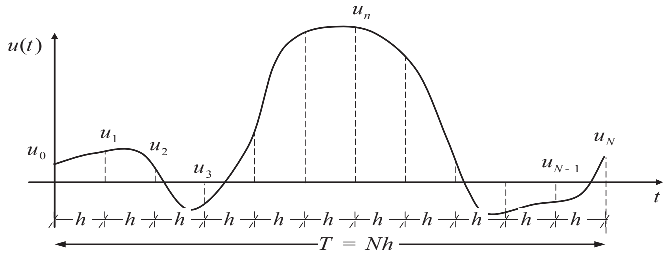

Fig. 1. Discretization of the interval $[0, T]$ into N equal intervals $h = T/N$.

$$u_{n-1} = u_0 + \sum_{r=1}^{n-1}\int_{(r-1)h}^{rh} q(\tau)d\tau \tag{9}$$

we may write Eq. (8) as

$$u_n = u_{n-1} + \int_{(n-1)h}^{nh} q(\tau)d\tau \tag{10}$$

2.2. Solution procedure for constant fictitious source $q(t)$

Without excluding higher order variation laws for $q(t)$, we assume that it is constant within the integration interval $[(r-1)h, rh]$ and equal to the mean value

$$q_r^m = \frac{q_{r-1} + q_r}{2} \tag{11}$$

Substituting Eq. (11) into Eq. (10) gives

$$u_n = u_{n-1} + \frac{h}{2}q_{n-1} + \frac{h}{2}q_n \tag{12}$$

Moreover, Eq. (3) at time $t = nh$ is written as

$$cq_n + ku_n = p_n \tag{13}$$

Eqs. (12) and (13) can be combined as

$$\begin{bmatrix} c & k \\ -\frac{h}{2} & 1 \end{bmatrix}\begin{Bmatrix} q_n \\ u_n \end{Bmatrix} = \begin{bmatrix} 0 & 0 \\ \frac{h}{2} & 1 \end{bmatrix}\begin{Bmatrix} q_{n-1} \\ u_{n-1} \end{Bmatrix} + p_n\begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \tag{14}$$

The coefficient matrix in Eq. (14) is not ill-conditioned for sufficient small h and the system can be solved successively for $n = 1, 2, \ldots$ to yield the solution u_n and its derivative $\dot{u}_n = q_n$ at instant $t = nh \le T$. For $n = 1$, the value q_0 appears in the right hand side of Eq. (14). This quantity can be readily obtained from Eq. (3) for $t = 0$. This yields

$$q_0 = (p_0 - ku_0)/c \tag{15}$$

Eq. (14) can be also written in matrix form

$$\mathbf{U}_n = \mathbf{A}\mathbf{U}_{n-1} + \mathbf{b}p_n, \quad n = 1, 2, \ldots N \tag{16}$$

in which

$$\mathbf{U}_n = \begin{Bmatrix} q_n \\ u_n \end{Bmatrix}, \quad \mathbf{A} = \begin{bmatrix} c & k \\ -\frac{h}{2} & 1 \end{bmatrix}^{-1}\begin{bmatrix} 0 & 0 \\ \frac{h}{2} & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} c & k \\ -\frac{h}{2} & 1 \end{bmatrix}^{-1}\begin{Bmatrix} 1 \\ 0 \end{Bmatrix} \tag{17a,b,c}$$

We can readily extend the previous procedure to the problem (1) and (2), which represents a system of, say L, first order ordinary differential equations. Obviously, the analog equations corresponding to Eq. (5) constitute the set of the L uncoupled equations

$$\dot{\mathbf{u}} = \mathbf{q}(t) \tag{18}$$

where $\dot{\mathbf{u}}$, $\mathbf{q}(t)$ are $L \times 1$ vectors. Thus, the numerical scheme for the solution becomes

$$\mathbf{U}_n = \mathbf{A}\mathbf{U}_{n-1} + \mathbf{b}\mathbf{p}_n, \quad n = 1, 2, \ldots, N \tag{19}$$

where

$$\mathbf{U}_n = \begin{Bmatrix} \mathbf{q}_n \\ \mathbf{u}_n \end{Bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{K} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix} \tag{20a,b}$$

$$\mathbf{b} = \begin{bmatrix} \mathbf{C} & \mathbf{K} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1}\begin{Bmatrix} \mathbf{1} \\ \mathbf{0} \end{Bmatrix}, \quad \mathbf{1} = \{1 \;\; 1 \;\; \cdots \;\; 1\}^T \tag{21a,b}$$

$$\mathbf{q}_0 = \mathbf{C}^{-1}(\mathbf{p}_0 - \mathbf{K}\mathbf{u}_0) \tag{22}$$

The solution algorithm is shown in Table 1.

Table 1
Algorithm for the numerical solution of the semi-discrete linear parabolic equations.

```
A.    Data for $\mathbf{C}\dot{\mathbf{u}} + \mathbf{K}\mathbf{u} = \mathbf{p}(t)$
Read: $\mathbf{C}$, $\mathbf{K}$, $\mathbf{u}_0$, $\mathbf{p}(t)$, $T$
B.    Initial computations
  Choose:    $h = \Delta t$ and compute $n_{\max}$
  Compute:   $\mathbf{q}_0 = \mathbf{C}^{-1}(\mathbf{p}_0 - \mathbf{K}\mathbf{u}_0)$
  Formulate  $\mathbf{U}_0 = \{\dot{\mathbf{u}}_0 \ \ \mathbf{u}_0\}^T$
  Compute:   $\mathbf{A} = \begin{bmatrix} \mathbf{C} & \mathbf{K} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} \mathbf{C} & \mathbf{K} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$
C.    Compute the solution
  for  $n$: =1 to $n_{\max}$
  $\mathbf{U}_n = \mathbf{A}\mathbf{U}_{n-1} + \mathbf{b}\mathbf{p}_n$
  end
```

2.3. Stability of the numerical scheme

Applying Eq. (19) for $n = 1, 2, \ldots$ yields

$$\begin{aligned} &\mathbf{U}_1 = \mathbf{A}\mathbf{U}_o + \mathbf{b}p_1\mathbf{U}_2 = \mathbf{A}\mathbf{U}_1 + \mathbf{b}p_2 = \mathbf{A}(\mathbf{A}\mathbf{U}_o + \mathbf{b}p_1) + \mathbf{b}p_2 \\ &\quad = \mathbf{A}^2\mathbf{U}_o + \mathbf{A}\mathbf{b}p_1 + \mathbf{b}p_2 \cdots = \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\qquad \mathbf{U}_n = \mathbf{A}^n\mathbf{U}_o + (\mathbf{A}^{n-1}p_1 + \mathbf{A}^{n-2}p_2 + \cdots\mathbf{A}^0p_n)\mathbf{b} \end{aligned} \tag{23}$$

We observe that the last of Eq. (23) gives the solution vector $\mathbf{U}_n$ at instant $t_n = nh$ using only the known vector $\mathbf{U}_0$ at $t = 0$. The matrix $\mathbf{A}$ and the vector $\mathbf{b}$ are computed only once.

The matrix $\mathbf{A}$ is the amplification matrix. In order that the solution is stable, $\mathbf{A}^n$ must be bounded. This is true if the spectral radius $\rho(\mathbf{A})$ satisfies the condition

$$\rho(\mathbf{A}) = \ \max(\lambda_i) \le 1 \tag{24}$$

where λ_i are the eigenvalues of $\mathbf{A}$. If $\rho(\mathbf{A}) < 1$ the method is strongly stable. The condition (24) is satisfied, if all eigenvalues of the matrix $\hat{\mathbf{K}} = \mathbf{C}^{-1}\mathbf{K}$ are nonnegative or have nonnegative real part for complex eigenvalues. This is proved in what follows.

Proof.

First we write Eq.(1) as

$$\dot{\mathbf{u}} + \hat{\mathbf{K}}\mathbf{u} = \mathbf{C}^{-1}\mathbf{p}(t), \ \ \hat{\mathbf{K}} = \mathbf{C}^{-1}\mathbf{K} \tag{25}$$

Thus the matrix $\mathbf{A}$ defined by Eq. (20b) becomes

$$\mathbf{A} = \begin{bmatrix} \mathbf{I} & \hat{\mathbf{K}} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix} \tag{26}$$

Using the formula for the inverse of a block matrix [5], we find

$$\begin{bmatrix} \mathbf{I} & \hat{\mathbf{K}} \\ -\frac{h}{2}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1} & -(\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1}\hat{\mathbf{K}} \\ \frac{h}{2}(\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1} & (\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1} \end{bmatrix} \tag{27}$$

Hence

$$\mathbf{A} = \begin{bmatrix} -\frac{h}{2}(\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1}\hat{\mathbf{K}} & -(\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1}\hat{\mathbf{K}} \\ \frac{h}{2}(\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1} & (\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}})^{-1} \end{bmatrix} \tag{28}$$

Next we find the eigenvalues of $\mathbf{A}$. For this purpose we write the pertinent eigenvalue problem in the form

$$\begin{bmatrix} \mathbf{A}_{11} - \lambda\mathbf{I} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} - \lambda\mathbf{I} \end{bmatrix} \begin{Bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{0} \end{Bmatrix} \tag{29}$$

or

$$\left.\begin{aligned} (\mathbf{A}_{11} - \lambda\mathbf{I})\mathbf{x}_1 + \mathbf{A}_{12}\mathbf{x}_2 = \mathbf{0} \\ \mathbf{A}_{21}\mathbf{x}_1 + (\mathbf{A}_{22} - \lambda\mathbf{I})\mathbf{x}_2 = \mathbf{0} \end{aligned}\right\} \tag{30}$$

where $\mathbf{A}_{ij}$ are the $L \times L$ matrices

$$\left.\begin{aligned} &\mathbf{A}_{11} = -\tfrac{h}{2}[\mathbf{I} + \tfrac{h}{2}\hat{\mathbf{K}}]^{-1}\hat{\mathbf{K}}, \qquad &&\mathbf{A}_{12} = -[\mathbf{I} + \tfrac{h}{2}\hat{\mathbf{K}}]^{-1}\hat{\mathbf{K}} \\ &\mathbf{A}_{21} = \tfrac{h}{2}[\mathbf{I} + \tfrac{h}{2}\hat{\mathbf{K}}]^{-1}, &&\mathbf{A}_{22} = [\mathbf{I} + \tfrac{h}{2}\hat{\mathbf{K}}]^{-1} \end{aligned}\right\} \tag{31}$$

Eqs. (30) are solved using Gauss elimination. To avoid inversion of the singular matrix $(\mathbf{A}_{11} - \lambda\mathbf{I})$, we reorder these equation as

$$\left.\begin{aligned} \mathbf{A}_{12}\mathbf{x}_2 + (\mathbf{A}_{11} - \lambda\mathbf{I})\mathbf{x}_1 = \mathbf{0} \\ (\mathbf{A}_{22} - \lambda\mathbf{I})\mathbf{x}_2 + \mathbf{A}_{21}\mathbf{x}_1 = \mathbf{0} \end{aligned}\right\} \tag{32}$$

which after elimination of $\mathbf{x}_2$ give

$$\begin{bmatrix} \mathbf{A}_{12} & \mathbf{A}_{11} - \lambda\mathbf{I} \\ \mathbf{0} & -(\mathbf{A}_{22} - \lambda\mathbf{I})\mathbf{A}_{12}^{-1}(\mathbf{A}_{11} - \lambda\mathbf{I}) + \mathbf{A}_{21} \end{bmatrix} \begin{Bmatrix} \mathbf{x}_2 \\ \mathbf{x}_1 \end{Bmatrix} = \begin{Bmatrix} \mathbf{0} \\ \mathbf{0} \end{Bmatrix} \tag{33}$$

The characteristic equation of the matrix in Eq.(33) is

$$\Pi(\lambda) = \det\mathbf{A}_{12}\det[-(\mathbf{A}_{22} - \lambda\mathbf{I})\mathbf{A}_{12}^{-1}(\mathbf{A}_{11} - \lambda\mathbf{I}) + \mathbf{A}_{21}] = 0 \tag{34}$$

Taking into account that $\det\mathbf{A}_{12} = -\det([\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}}]^{-1}\hat{\mathbf{K}}) \neq 0$, we have

$$\begin{aligned} \Pi(\lambda) &= \det[-(\mathbf{A}_{22} - \lambda\mathbf{I})\mathbf{A}_{12}^{-1}(\mathbf{A}_{11} - \lambda\mathbf{I}) + \mathbf{A}_{21}] = 0 \\ &= \det(\mathbf{A}_{21} - \mathbf{A}_{22}\mathbf{A}_{12}^{-1}\mathbf{A}_{11} + (\mathbf{A}_{12}^{-1}\mathbf{A}_{11} + \mathbf{A}_{22}\mathbf{A}_{12}^{-1})\lambda - \mathbf{A}_{12}^{-1}\lambda^2) = 0 \end{aligned} \tag{35}$$

We can readily show that

$$\mathbf{A}_{21} - \mathbf{A}_{22}\mathbf{A}_{12}^{-1}\mathbf{A}_{11} = 0 \tag{36}$$

Hence, Eq. (35) becomes

$$\det(\mathbf{A}_{12}^{-1}\mathbf{A}_{11}\mathbf{A}_{12}^{1} + \mathbf{A}_{22} - \lambda\mathbf{I})\lambda = 0 \tag{37}$$

From Eq. (37) we conclude that the $2L \times 2L$ matrix $\mathbf{A}$ has L zero eigenvalues, while the other L nonzero eigenvalues are the eigenvalues of the matrix

$$\mathbf{A}^* = \mathbf{A}_{12}^{-1}\mathbf{A}_{11}\mathbf{A}_{12}^{1} + \mathbf{A}_{22} = \left[\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}}\right]^{-1}\left[\mathbf{I} - \frac{h}{2}\hat{\mathbf{K}}\right] \tag{38}$$

Let $\hat{\mathbf{\Lambda}}$ be the diagonal matrix of the eigenvalues of $\hat{\mathbf{K}}$ and $\hat{\mathbf{X}}$ the matrix of its eigenvectors normalized with respect to their measure (length). It holds $\det(\hat{\mathbf{X}}) \neq 0$, $\hat{\mathbf{X}}^{-1} = \hat{\mathbf{X}}^T$. Using the spectral decomposition of $\hat{\mathbf{K}}$, we may write

$$\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}} = \mathbf{I} + \frac{h}{2}\hat{\mathbf{X}}\hat{\mathbf{\Lambda}}\hat{\mathbf{X}}^{-1} = \hat{\mathbf{X}}\left(\mathbf{I} + \frac{h}{2}\hat{\mathbf{\Lambda}}\right)\hat{\mathbf{X}}^{-1} \tag{39a}$$

and

$$\left[\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}}\right]^{-1} = \hat{\mathbf{X}}\left[\mathbf{I} + \frac{h}{2}\hat{\mathbf{\Lambda}}\right]^{-1}\hat{\mathbf{X}}^{-1} \tag{39b}$$

Hence

$$\mathbf{A}^* = \left[\mathbf{I} + \frac{h}{2}\hat{\mathbf{K}}\right]^{-1}\left[\mathbf{I} - \frac{h}{2}\hat{\mathbf{K}}\right] = \hat{\mathbf{X}}\left[\mathbf{I} + \frac{h}{2}\hat{\mathbf{\Lambda}}\right]^{-1}\left[\mathbf{I} - \frac{h}{2}\hat{\mathbf{\Lambda}}\right]\hat{\mathbf{X}}^{-1} = \hat{\mathbf{X}}\mathbf{\Lambda}^*\hat{\mathbf{X}}^{-1} \tag{40}$$

or

$$\hat{\mathbf{X}}^{-1}\mathbf{A}^*\hat{\mathbf{X}} = \mathbf{\Lambda}^* \tag{41}$$

that is Λ^* is the diagonal matrix of the eigenvalues λ_i^* of $\mathbf{A}^*$. Hence

$$\lambda_i^* = \left(1 - \frac{h}{2}\hat{\lambda}_i\right)/\left(1 + \frac{h}{2}\hat{\lambda}_i\right) \tag{42}$$

The stability of the numerical scheme requires that

$$|\lambda_i^*| = \left|1 - \frac{h}{2}\hat{\lambda}_i\right|/\left|1 + \frac{h}{2}\hat{\lambda}_i\right| \le 1 \tag{43}$$

or

$$\left(1 - \frac{h}{2}\hat{\lambda}_i\right)^2 \le \left(1 + \frac{h}{2}\hat{\lambda}_i\right)^2 \tag{44}$$

which is satisfied

a) For real $\hat{\lambda}_i$, if $\hat{\lambda}_i \geq 0$
b) For complex $\hat{\lambda}_i = a_i + i\beta_i$, if $a_i \geq 0$

Hence the stability criterion can be stated as:

The proposed numerical scheme is stable, that is the stability condition (24) is satisfied, if the eigenvalues of the matrix $\hat{\mathbf{K}} = \mathbf{C}^{-1}\mathbf{K}$ are nonnegative or have nonnegative real part for complex eigenvalues.

2.4. Error analysis and convergence

The error is due to the approximation of the integrand in the integral of Eq. (8) in the r integration interval $[(r-1)h, rh]$

$$\int_{t_0}^{t_1} f(\tau)d\tau, \quad t_0 = (r-1)h, \quad t_1 = rh \tag{45}$$

$f(\tau)$ is approximated as

$$\tilde{f}(\tau) = \frac{q_{r-1} + q_r}{2} = q_r^m = q_0 = \text{constant} \tag{46}$$

Hence the error is

$$\int_{t_0}^{t_1} [f(\tau) - \tilde{f}(\tau)]d\tau \tag{47}$$

Expanding $f(\tau)$ in Taylor series at $\tau = 0$ and evaluating the integral of $f(\tau) - \tilde{f}(\tau)$ over the interval $[t_0, t_1]$ we find

$$\int_{t_0}^{t_1} [f(\tau) - \tilde{f}(\tau)]d\tau = \left(f_0 + hf'_0 + \frac{h^2}{2}f''_0 + \cdots\right) h - q_0 h = (f_0 - q_0)h + f'_0 h^2 + \frac{h^3}{2}f''_0 + \cdots \tag{48}$$

Taking into account that $f_0 = q_0$ we conclude that the convergence of the numerical scheme is $O(h^2)$ (see Fig. 4).

2.5. Numerical examples

Matlab codes, based on the developed numerical schemes have been written and various example problems have been solved. Note that the exact solutions, where no reference is made, have been obtained using the inverse method presented by Katsikadelis [4]. According to this method, a solution is assumed, which yields the corresponding source after inserting it in the equation.

Example 1. One-degree of freedom system.

Eq. (1) has been solved with data: $c = 5$, $k = 50$, $p(t) = -c\omega \sin \omega t + k \cos \omega t$, $u_0 = 1$, $\omega = 2$. Eq. (1) admits an exact solution $u_{ex} = \cos \omega t$. Fig. 2 shows the computed solution as compared with the exact solution for $h = 0.05$. Fig. 3 shows the error $\max|u(t_i) - u_{ex}(t_i)|$ and the mean square error MSE= $\sqrt{\frac{1}{n}\sum_{i=1}^{i=n}[u(t_i) - u_{ex}(t_i)]^2}$, $0 < t_i \leq 100$ versus the time step h, which validate the convergence of the numerical scheme. Moreover, Fig. 4 verifies that the convergence is of $O(h^2)$. Fig. 5 shows the solution for long duration ($0 \leq t \leq 50000$, $h = 0.01$). We observe that the scheme remains stable and the error within the same bounds. Finally, Fig. 6 shows that the quantity $u^T c\dot{u} + u^T ku = u^T p(t)$, that is the total "*work*" remains unchanged during the whole procedure, which means that the scheme does not exhibit numerical damping.

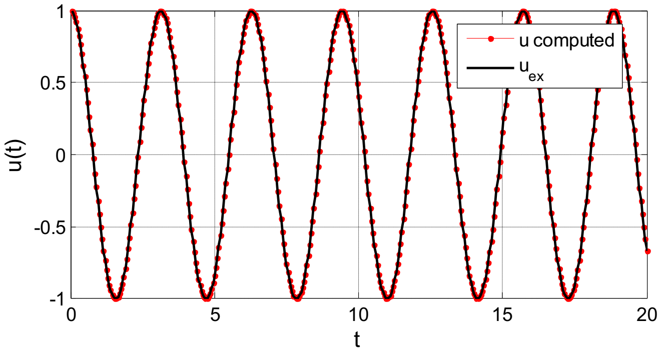

Fig. 2. Solution computed and exact in Example 1.

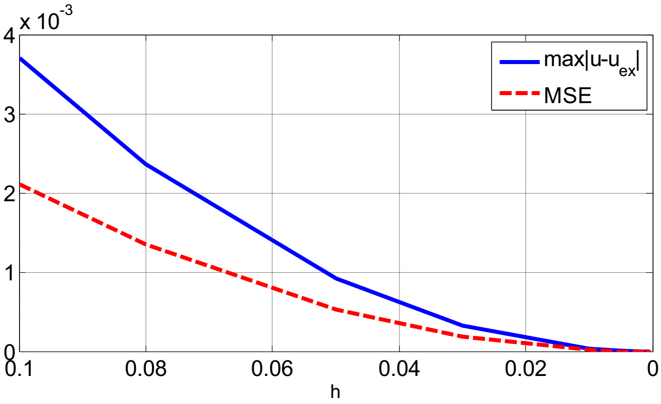

Fig. 3. Error $\max|u(t_i) - u_{ex}(t_i)|$ and MSE ($0 < t_i \leq 100$) in Example 1.

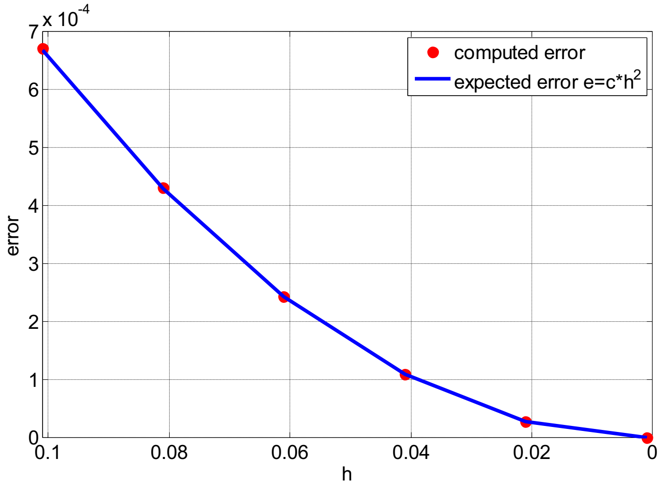

Fig. 4. Computed and expected error $e = e(h)$; $c = e(1)/h(1)$ in Example 1.

Example 2. System of equations. Symmetrical positive definite coefficient matrices…

In this example the system of equations

$$\begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}\begin{Bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{Bmatrix} + \begin{bmatrix} 25 & 20 \\ 20 & 20 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \exp(-0.1t)\begin{Bmatrix} 57/2 \cos t + 73/5 \sin t \\ 123/5 \cos t + 31/2 \sin t \end{Bmatrix} \tag{49}$$

with initial conditions $u_1 = 1$, $u_2 = 0$ is solved. The matrices $\mathbf{C}$, $\mathbf{K}$ are symmetrical and positive definite, $eig(\mathbf{C}) = \{1 \quad 9\}$, $eig(\mathbf{K})= \{2.3444 \quad 42.6556\}$. Eq. (49) admits an exact solution.

$$\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \exp(-0.1t)\begin{Bmatrix} \cos t) \\ \sin t \end{Bmatrix} \tag{50}$$

The computed solution for $T = 10$ and $h = 0.1$ is shown in Fig. 7 as compared with the exact one. Moreover Fig. 8 shows the error $\mathbf{u} - \mathbf{u}_{ex}$.

Example 3. System of equations. Nonsymmetrical non-positive definite coefficient matrices.

In this example the system of equations

$$\begin{bmatrix} 0.1493 & 0.8407 \\ 0.2575 & 0.2543 \end{bmatrix}\begin{Bmatrix} \dot{u}_1 \\ \dot{u}_2 \end{Bmatrix} + \begin{bmatrix} 0.8909 & 0.5472 \\ 0.9593 & 0.1386 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} p_1 \\ p_2 \end{Bmatrix} \tag{51}$$

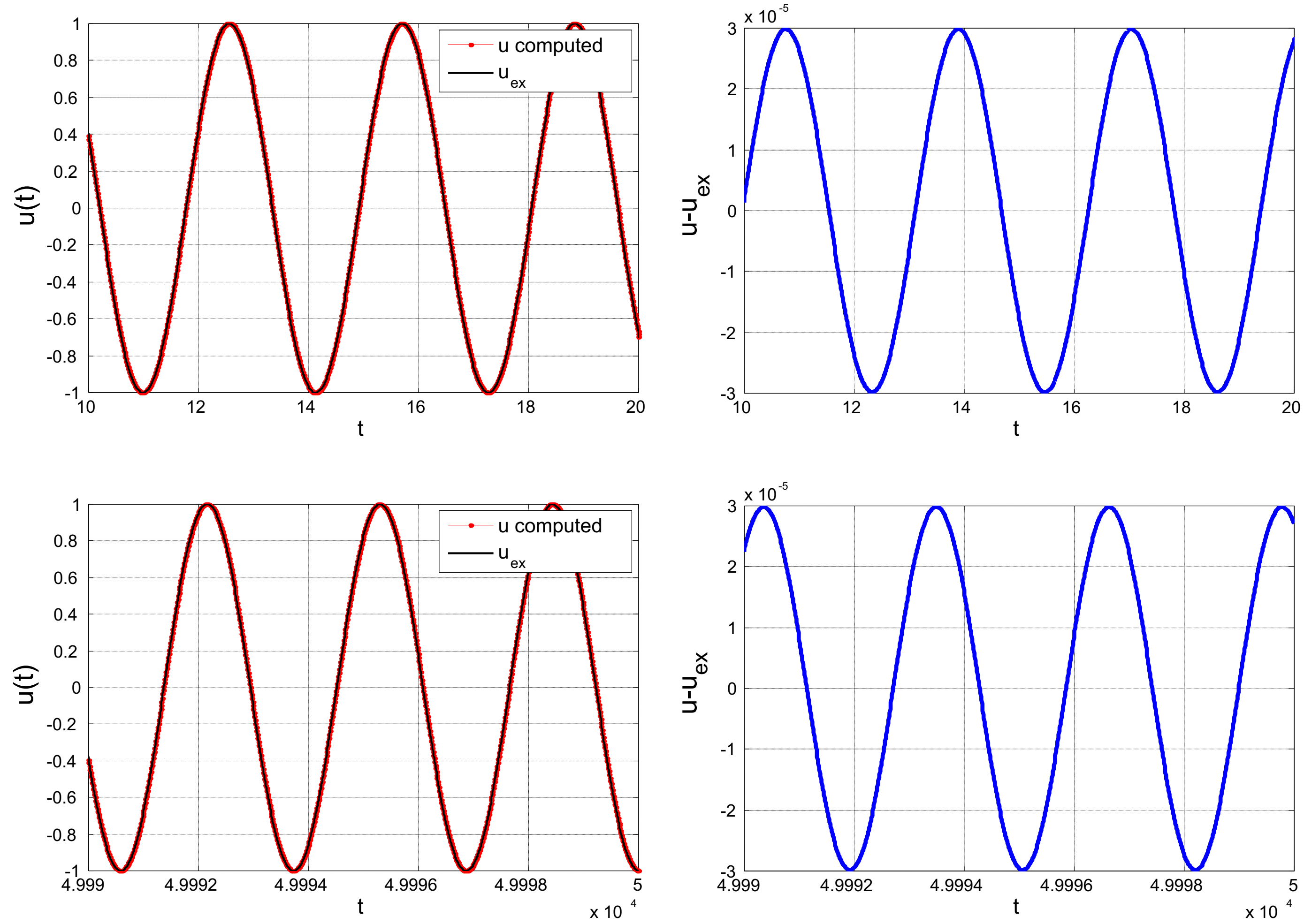

Fig. 5. Solution for long duration ($0 \le t \le 50000$) in Example 1.

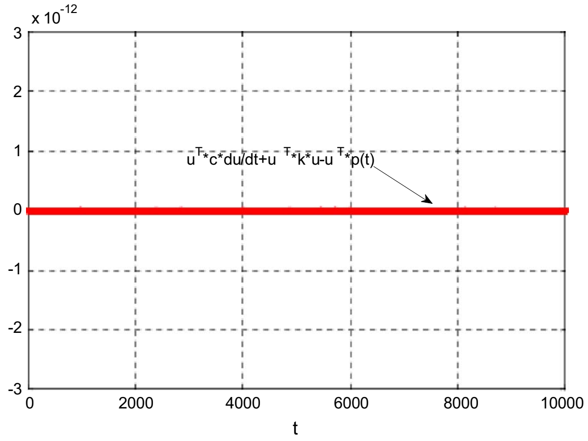

Fig. 6. Variation of the "*work*" for long duration procedures in Example 1.

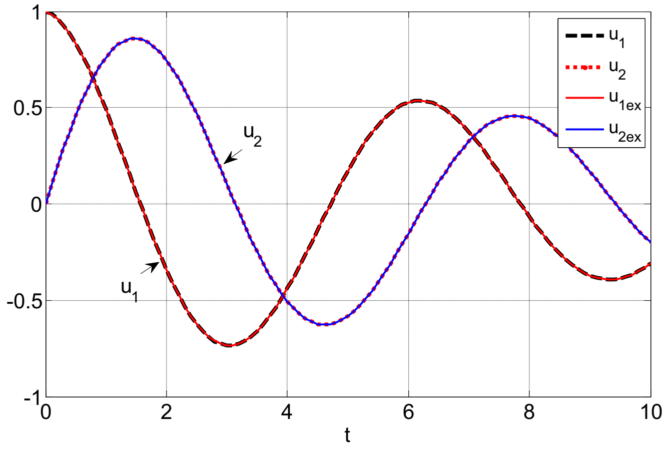

Fig. 7. Solution $\mathbf{u} = \{u_1 \quad u_2\}^T$ in Example 2.

with initial conditions $u_1 = 1, \quad u_2 = 0$ is solved. The matrices $\mathbf{C}$, $\mathbf{K}$ are nonsymmetrical and non-positive-definite, $eig(\mathbf{C}) = \{-0.2665 \quad 0.6770\}$ $eig(\mathbf{K}) = \{-1.1311 \quad -0.3016\}$. However, the nonsymmetrical matrix $\hat{\mathbf{K}} = \mathbf{C}^{-1}\mathbf{K}$ has positive eigenvalues, $eig(\hat{\mathbf{K}}) = \{3.2245 \quad 0.6973\}$. Therefore, the stability criterion is satisfied.

Eq. (51) for

$$\mathbf{p} = \exp(-0.1t)\begin{Bmatrix} -0.080543 \sin t + 1.07556 \cos t \\ 0.42495 \sin t + 0.48384 \cos t \end{Bmatrix} \tag{52}$$

admits an exact solution

$$\mathbf{u}_{ex} = \exp(-0.1t)\begin{Bmatrix} \sin t \\ 2 \cos t \end{Bmatrix} \tag{53}$$

The computed solution for $T = 10$ and $h = 0.01$ is shown in Fig. 9 as compared with the exact one. Moreover, Fig. 10 shows the error $\mathbf{u} - \mathbf{u}_{ex}$.

Example 4. Large linear systems of equations.

In this example the heat conduction equation which describes the

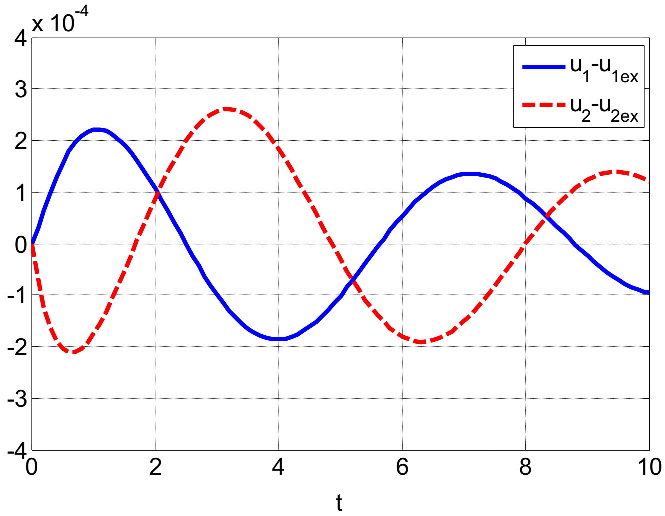

Fig. 8. Error $\mathbf{u}-\mathbf{u}_{ex}$ in Example 2.

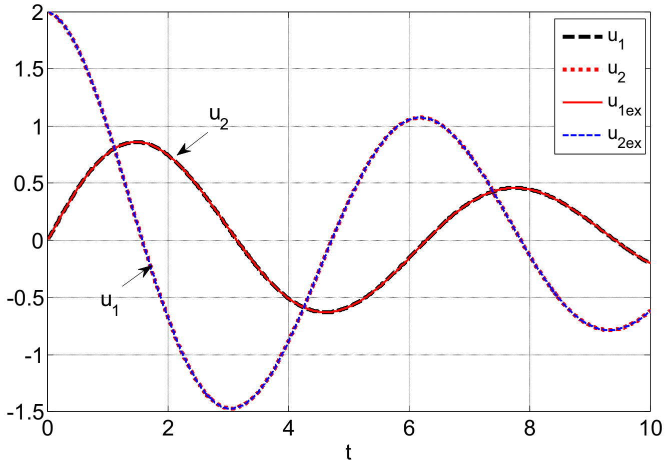

Fig. 9. Solution $\mathbf{u}=\{u_1 \quad u_2\}^T$ in Example 3.

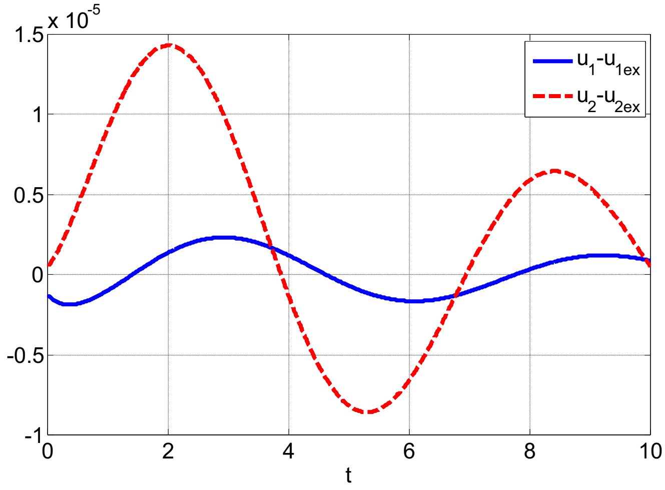

Fig. 10. Error $\mathbf{u}-\mathbf{u}_{ex}$ in Example 3.

transient temperature distribution $u(x,y,t)$ in a two-dimensional homogeneous orthotropic body occupying the rectangular domain Ω: $\{0\le x\le a,\ 0\le y\le b\}$ is studied. The temperature distribution is governed by the following initial value problem

$$\rho c u = k_x u,_{xx} + k_y u,_{yy} \quad \text{in } \Omega \tag{54}$$

$$u(0,y,t)=u(x,0,t)=u(a,y,t)=u(x,b,t)=0 \quad \text{on } \Gamma \tag{55a}$$

and

$$u(x,y,0)=u_0, \ \text{in } \Omega \tag{55b}$$

Eq. (54) admits an analytic solution [6]

$$u_{ex}(x,y,t)=\sum_{n=1}^{\infty}\sum_{j=1}^{\infty}A_n \sin\frac{n\pi x}{a}\sin\frac{j\pi y}{b}\exp\left[-\left(\frac{k_x n^2\pi^2}{a^2}+\frac{k_y j^2\pi^2}{b^2}\right)t\right] \tag{56a}$$

where

$$A_n=\frac{4u_0}{nj\pi^2}[(-1)^n-1][(-1)^j-1] \tag{56b}$$

The adopted numerical values are: $a=b=3$, $k_x=k_y=1.25$, $c=1$, $\rho=1$, and $u_0=30$.

The solution was computed using the AEM BEM [7] with $N=200$ constant boundary elements, $L=121$ domain nodal points distributed uniformly, and shape parameter of the multiquadrics $c=0.2$. This solution method produces, semi-discrete equations of the form (1) with nonsymmetrical and non-positive definite matrices $\mathbf{C}$, $\mathbf{K}$ with dimensions 121 × 121, whereas the matrix $\mathbf{C}^{-1}\mathbf{K}$ satisfies the stability criterion.

Fig. 11 shows the time history of the temperature at point (1.5,1.5) as compared with the exact solution, while Fig. 12 shows the relative error $[u(1.5,1.5;t)-u_{ex}(1.5,1.5;t)]/u_{ex}(1.5,1.5;t)$ with $\Delta t=0.005$.

3. Linear equation with variable coefficients

So far we have developed the method for the solution of Eq. (1) with constant coefficients. Obviously, if the coefficients c and k in Eq.(3) are functions of the independent variable t, i.e., $c(t)$, $k(t)$, the previously described solution procedure remains the same except that the elements c, k in the first row of the coefficient matrix in the left hand side of Eq. (14) depend on time. Therefore, this coefficient matrix in the respective solution algorithm must be reevaluated in each step. In the following, the efficiency of the method is demonstrated by solving an equation with variable coefficients.

Example 5. Variable coefficients.

We consider the initial value problem

$$(5+t)\dot{u}+(1+t^2)u=p(t), \ \ u_0=1 \tag{57}$$

Eq. (57) for $p(t)=[(0.5-0.1t+t^2)\cos t-(5+t)\sin t]\exp(-0.1t)$ admits an exact solution $u_{ex}(t)=e^{-0.1t}\cos t$. The computed solution for $T=30$ and $h=0.01$ is shown in Fig. 13 as compared with the exact one.

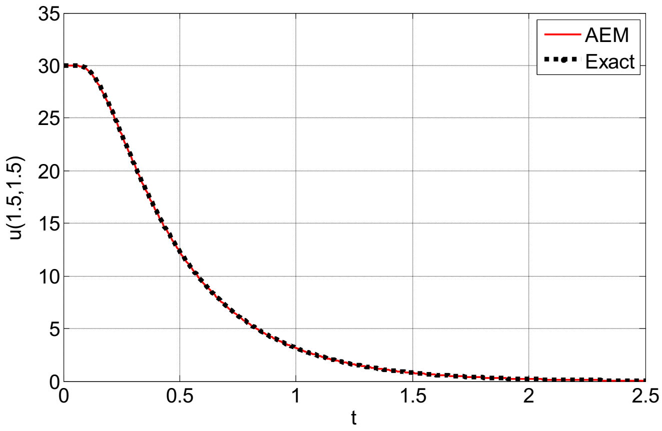

Fig. 11. Time history of the temperature at point (1.5, 1.5) in Example 4.

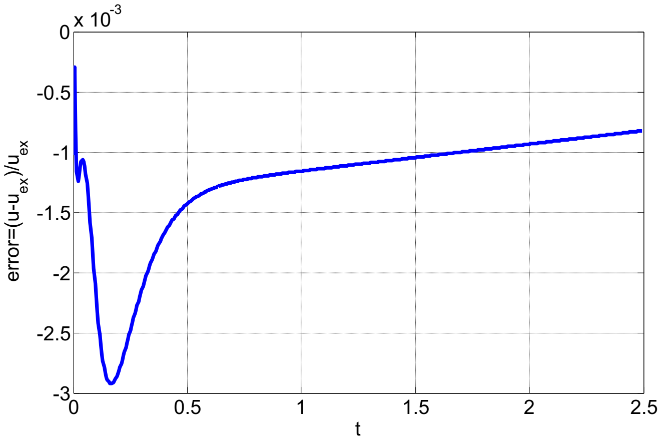

Fig. 12. Relative error at point (1.5, 1.5) in Example 4.

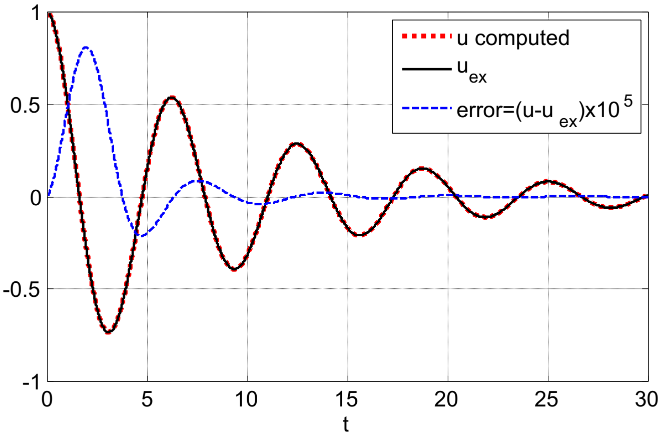

Fig. 13. Solution u and error $(u - u_{ex}) \times 10^5$ in Example 5.

4. Nonlinear equations

The solution procedure developed previously for the linear equations can be straightforwardly extended to nonlinear equations.

The nonlinear initial value problem for multi-degree of freedom systems is described as

$$\mathbf{C}\dot{\mathbf{u}} + \mathbf{F}(\mathbf{u}) = \mathbf{p}(t) \tag{58}$$

$$\mathbf{u}(0) = \mathbf{u}_0 \tag{59}$$

where $\mathbf{C}$ is $L \times L$ known coefficient matrix with $\det(\mathbf{C}) \neq 0$; $\mathbf{F}(\mathbf{u})$ is an $L \times 1$ vector, whose elements are nonlinear functions of the components of $\mathbf{u}$; $\mathbf{p}(t)$ is the external source vector and $\mathbf{u}_0$ a given constant vector.

The solution procedure is similar to that for the linear systems. Thus, Eq. (58) for $t = 0$ gives the vector

$$\mathbf{q}_0 = \mathbf{C}^{-1}[\mathbf{p}_0 - \mathbf{F}(\mathbf{u}_0)], \qquad \mathbf{q}_0 = \dot{\mathbf{u}} \tag{60}$$

Subsequently, we apply Eq. (58) for $t = t_n$

$$\mathbf{C}\mathbf{q}_n + \mathbf{F}(\mathbf{u}_n) = \mathbf{p}_n \tag{61}$$

Apparently, Eq. (12) is valid in this case, too. Thus we may write

$$\mathbf{u}_n = \mathbf{u}_{n-1} + \frac{h}{2}\mathbf{q}_n + \frac{h}{2}\mathbf{q}_{n-1} \tag{62}$$

Eqs. (61) and (62) are combined and solved for $\mathbf{q}_n$, $\mathbf{u}_n$ with $n = 1, 2, \ldots$. The solution can be obtained using an iterative procedure in each step. A simple procedure is to substitute Eq. (62) into Eq. (61). This yields a nonlinear equation for $\mathbf{q}_n$, which is solved by employing any ready-to-use subroutine for nonlinear algebraic equations. In our examples the functions *fsolve* of Matlab or the subroutine NEQNF of the IMSL have been employed to obtain the numerical results.

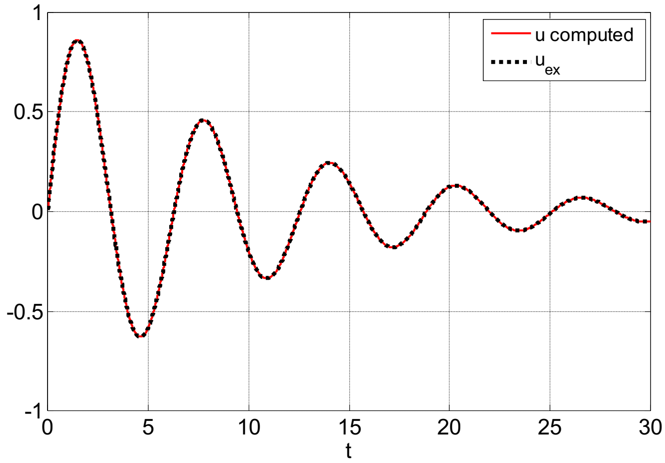

Fig. 14. Solution u in Example 6.

Example 6. Nonlinear one-degree of freedom system.

The numerical scheme is employed to solve the initial value problem

$$0.2\dot{u} + u + u^3 = p(t), \;\; u(0) = 0 \tag{63a,b}$$

For

$$p(t) = e^{-0.1t}[(0.01\,\sin t - 0.2\,\cos t - \;\sin t) - 0.2(0.1\,\sin t - \;\cos t), \qquad \text{Eq}$$
$$+ \;\sin t + e^{-0.2t}(\sin t)^3]$$

(63a) admits an exact solution $u_{exact}(t) = e^{-0.1t}\sin t$. Fig. 14 shows the solution with $\Delta t = 0.01$ as compared with the exact one and Fig. 15 presents the error $u - u_{ex}$.

Example 7. Large nonlinear systems of equations.

In this example we consider the transient heat conduction in a plane body with temperature dependent conductivity $k = k_0(1 + \beta u)$. The temperature distribution $u(\mathbf{x}, t)$ at time t, when the temperature on the boundary is kept zero, is described by the following initial boundary value problem

$$\eta\dot{u} = k\nabla^2 u + k_0\beta(u,{}_{x}^{2} + u,{}_{y}^{2}) + f(\mathbf{x}, t) \text{ in } \Omega \tag{64a}$$

$$u = 0 \text{ on } \Gamma \tag{64b}$$

$$u(\mathbf{x}, 0) = [1 - (x^2 + y^2)] \text{ in } \Omega \tag{65}$$

where $f(\mathbf{x}, t)$ is the internal heat source, and k_0, β are material

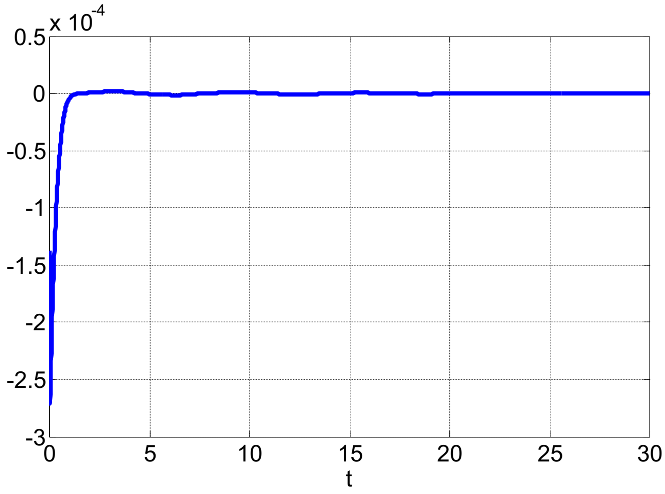

Fig. 15. Error $u - u_{ex}$ with $\Delta t = 0.01$ in Example 6.

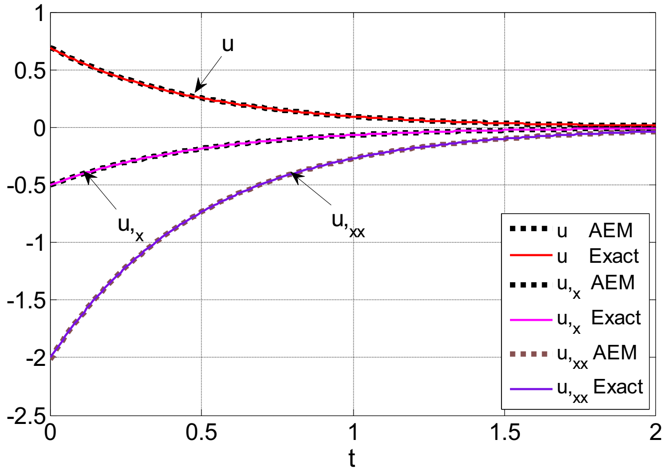

Fig. 16. Time variation of u and its derivatives $u,_x$, $u,_{xx}$ at point (0, 0) in Example 7.

constants.

The problem is solved for a circular domain Ω of unit circle with $\eta = 1/2$, $k_0 = 1$, $\beta = 3$, $f(\mathbf{x}, t) = (3 + x^2 + y^2)\exp(-2t) + 12[1-2(x^2 + y^2)]\exp(-4t)$. The results were obtained using the AEM/BEM [7] with $N = 200$ constant boundary elements, $L = 129$ domain nodal points uniformly distributed, and shape parameter of the multiquadrics $c = 0.1$. This solution method produces, a system of 129 semi-discrete equations of the form (58), which are solved using the procedure described in Section 4. Fig. 16 shows the time variation of the computed temperature u and its derivatives $u,_x$, $u,_{xx}$ at point (0, 0) as compared with the exact ones, while Fig. 17 shows the respective relative errors for $\Delta t = 0.001$.

Example 8. The Rober problem.

In this example the problem describing the kinetics of an autocatalytic reaction given by Robertson [8] is solved. It is governed by the nonlinear system of equations

$$\begin{Bmatrix}\dot{u}_1\\ \dot{u}_2\\ \dot{u}_3\end{Bmatrix} = \begin{Bmatrix} -k_1u_1 + k_3u_2u_3 \\ k_1u_1 - k_2u_2^2 - k_3u_2u_3 \\ k_2u_2^2\end{Bmatrix} \tag{66}$$

with initial conditions $\mathbf{u}_0 = \{1 \ \ 0 \ \ 0\}^T$. The variables u_1, u_2, u_3 denote the concentrations of the three involved chemical species and k_1, k_2, k_3 are the rate constants.

This problem, known as ROBER problem, is very popular in numerical studies and it is often used as a benchmark problem to test the efficiency of stiff numerical integrators. The numerical values of the rate constants used in the test problem are $k_1 = 0, 04$, $k_2 = 3 \times 10^7$, $k_3 = 10^4$. The large difference among the reaction rate constants is the reason for stiffness. It was observed that many integration codes, though for small intervals ($0 \le t \le 40$) perform well, fail if t becomes very large. In this case, u_2 may accidentally become negative, and then tends to $-\infty$, causing overflow [9]. A Matlab code has been written

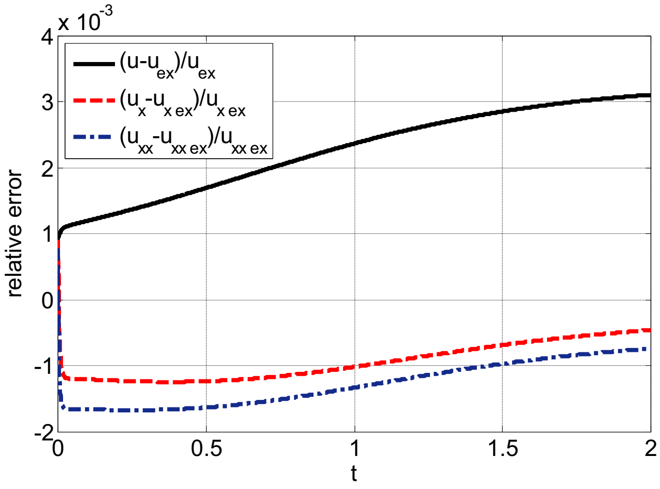

Fig. 17. Relative error at point (0, 0) with $\Delta t = 0.001$ in Example.

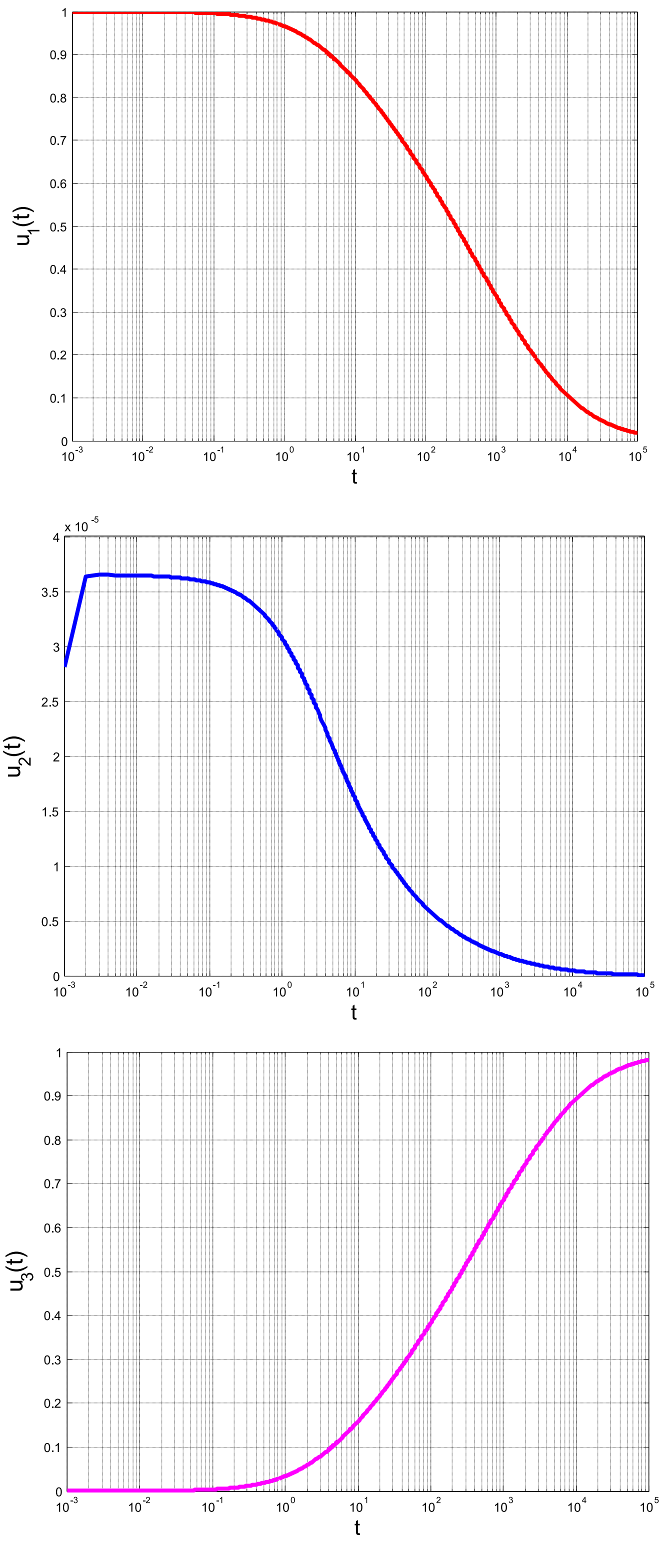

Fig. 18. Behavior of the solution of the Rober problem over the integration interval in Example 8.

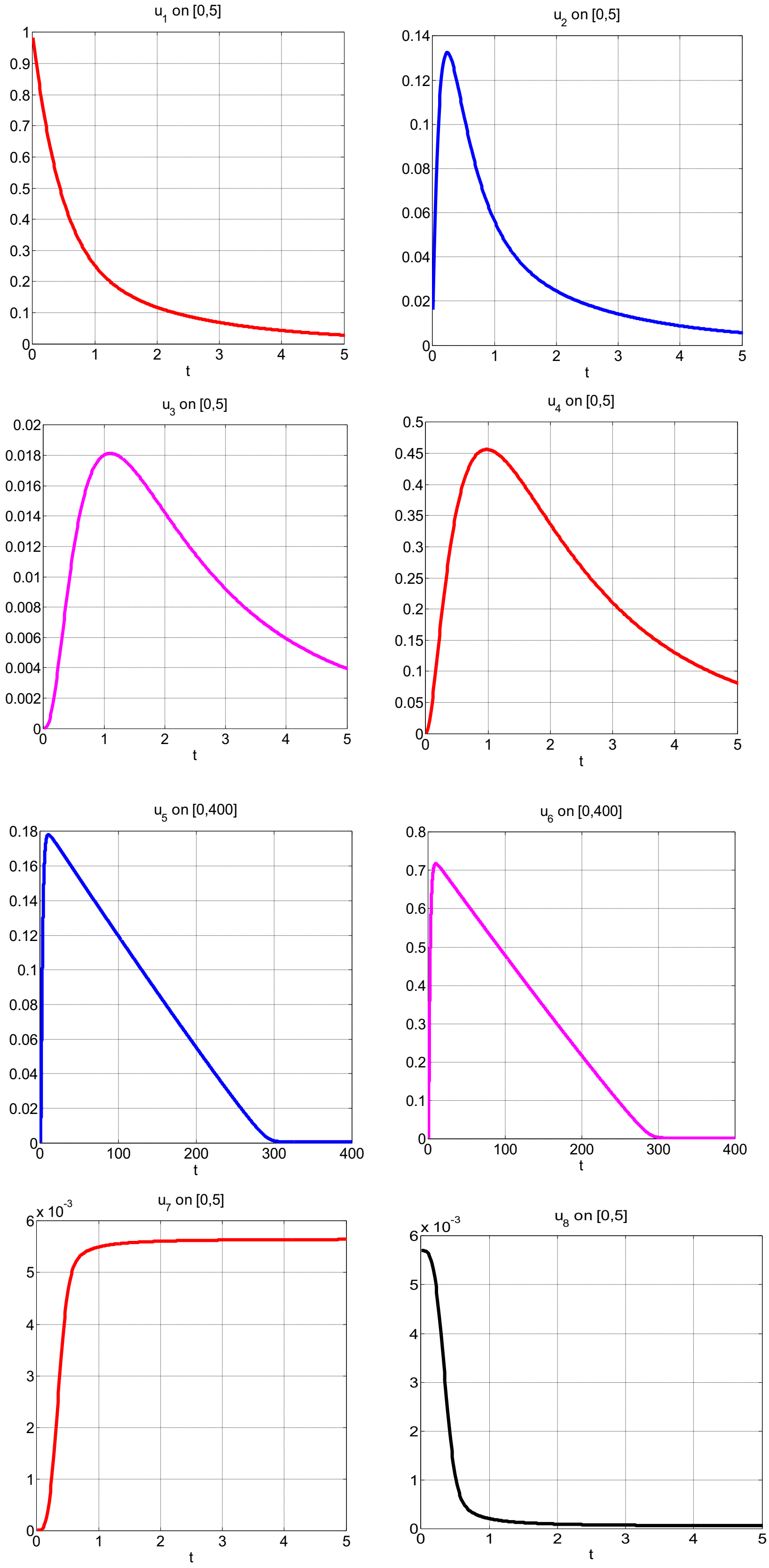

Fig. 19. Behavior of the solution of the HIRES problem over the integration interval in Example 9.

based on the procedure described in Section 4. The code was run on a Toshiba KIRAbook I7 computer. The solution was obtained with $\Delta t = 0.001$ in the interval $0 \le t \le 3$, and $\Delta t = 0.1$ in the interval $3 < t \le 10^5$. Fig. 18 shows the behavior of the solution, which coincides with the respective reference figure shown in: http://www.dm.uniba.it/~testset/testsetivpsolvers/?page_id=26#ODE.

Example 9. The HIRES problem..

This initial value problem is another stiff system of 8 non-linear ordinary differential equations. It was proposed by Schäfer in 1975 [10]. It refers to 'High Irradiance RESponse', which is described by this ODE. It is used also as a benchmark problem to test the efficiency of stiff numerical integrators (http://www.dm.uniba.it/~testset/testsetivpsolvers/?page_id=26#ODE).

The problem is described by the set of equations.

$$\begin{Bmatrix} \dot{u}_1 \\ \dot{u}_2 \\ \dot{u}_3 \\ \dot{u}_4 \\ \dot{u}_5 \\ \dot{u}_6 \\ \dot{u}_7 \\ \dot{u}_8 \end{Bmatrix} = \begin{Bmatrix} -1.71u_1 + 0.43u_2 + 8.32u_3 + 0.0007 \\ 1.71u_1 - 8.75u_2 \\ -10.03u_3 + 0.43u_4 + 0.035u_5 \\ 8.32u_2 + 1.71u_3 - 1.12u_4 \\ -1.745u_5 + 0.43u_6 + 0.43u_7 \\ -280u_6u_8 + 0.69u_4 + 1.71u_5 - 0.43u_6 + 0.69u_7 \\ 280u_6u_8 - 1.81u_7 \\ -280u_6u_8 + 1.81u_7 \end{Bmatrix} \tag{67}$$

with initial conditions $\mathbf{u}_0 = \{1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0\ \ 0.0057\}^T$. The solution obtained with $\Delta t = 0.01$ is shown in Fig. 19. It coincides with the respective reference figure shown in: http://www.dm.uniba.it/~testset/testsetivpsolvers/?page_id=26#ODE..

5. Conclusions

A direct time integration method has been developed for the numerical solution of first order linear and nonlinear parabolic differential equations. The developed numerical scheme is applied to the solution of the semi-discrete equations arising in diffusion problems after spatial discretization using modern computational methods. Contrary to the widely used methods, the stability of the scheme does not demand symmetrical and positive definite coefficient matrices $\mathbf{C}$, $\mathbf{K}$. Thus, it can solve equations with nonsymmetrical and non-positive definite matrices provided that the eigenvalues of the matrix $\mathbf{C}^{-1}\mathbf{K}$ are nonnegative or have nonnegative real part for the complex eigenvalues. This is an important advantage, since the scheme can solve semi-discrete diffusion equations resulting from methods that do not produce symmetrical matrices, e.g. the boundary element method. It applies also to equations with time dependent coefficient matrices, i.e. variable coefficients. The method is simple to implement. It is self-starting, unconditionally stable, second order accurate and does not exhibit numerical damping. It performs well when long time durations are considered. It can be used as a practical method for integration of stiff diffusion equations in cases where widely used time integration procedures fail. Besides, the present paper highlights the capability of the AEM to solve differential equations of any order. The efficiency and accuracy of the method is validated through well corroborated examples and benchmark problems.

References

[1] Hughes TJR. The finite element method. Englewood Cliffs, NJ, USA: Prentice Hall Inc; 1987.

[2] Reddy JN. An introduction to nonlinear finite element analysis. New York: Oxford University Press Inc.; 2004.

[3] Katsikadelis JT. The boundary element method for plate analysis. Elsevier, U.K: Academic Press; 2014.

[4] Katsikadelis JT. A new direct time integration method for the equations of motion in structural dynamics. ZAMM Z Angew Math Mech 2014;94(9):757–74. http://dx.doi.org/10.1002/zamm.20120024.

[5] Bierens HJ. The Inverse of a Partitioned Matrix. 2014. ⟨http://grizzly.la.psu.edu.~hbierens⟩.

[6] Bruch JC, jr., Zyvoloski G. Transient two-dimensional heat conduction problems solved by the finite element method. Int J Numer Methods Eng 1974;8:481–94.

[7] Katsikadelis JT. The boundary element method for engineers and scientists. Theory and applications, 2nd ed. Elsevier, U.K: Academic Press; 2016.

[8] Robertson HH. The solution of a set of reaction rate equations. In: Walsh J, editor. Numerical Analysis, An Introduction. London: Academic Press; 1966. p. 178–82.

[9] Hairer E, Wanner G. Solving ordinary differential equations II: stiff and differential-algebraic problems, second revised edition. Berlin Heidelberg GmbH: Springer-Verlag; 1996.

[10] Schäfer E. A new approach to explain the "high irradiance responses" of photomorphogenesis on the basis of phytochrome. J Math Biol 1975;2:41–56.