

A sensitized finite element solution approach to the time-dependent diffusion-advection-reaction equation

Eero-Matti Salonen and Jouni Freund

Summary. A new solution approach to the one-dimensional time-dependent diffusion-advection-reaction equation using the time-discontinuous Galerkin method is presented. The standard weak form is sensitized by least squares type extra terms consisting of the field equation residual and of the differentiated field equation residual. In addition, the initial condition residual and its differentiated form are included with three sensitizing parameters. The logic for the determination of the optimal values for these parameters is the main theme of the article. The approximation in each space-time slab is simplest possible: constant in time and with two node linear elements in space. Reference solution according to the von Neumann type analysis is used in a patch test. The parameters are determined by demanding the algorithmic and the exact amplification factors to coincide as well as possible. High accuracy formulations are found to emerge. The Mathematica program is used extensively in the analytical manipulations needed. The pure diffusion, the pure advection, and the pure reaction cases are dealt with in some detail to see the resulting expressions in a simple setting.

Key words: diffusion-advection-reaction equation, sensitized formulation, time-discontinuous Galerkin, space-time elements, von Neumann analysis, amplification factor

Introduction

There exists a vast literature on numerical solution of diffusion-advection-reaction problems in general. In particular, textbooks [1] and [2] and the references in them may be mentioned. Further, e.g. the finite element articles [3] to [8] are close to the theme of the present article but represent alternative approaches.

Diffusion-advection-reaction problems cover a large number of physical phenomena. Roughly, diffusion, advection and reaction are associated with second order, first order and zero order derivatives with respect to space coordinates, respectively. For instance, structural problems are mainly concerned with diffusive terms and in fluid mechanics advection is of main importance. To treat equations and equation system containing all the three ingredients — diffusion, advection, reaction — simultaneously produces a unifying point of view.

We will consider here just as an introductory limited case the one-dimensional linear time-dependent diffusion-advection-reaction equation

$$R(u) \equiv L(u) - s \equiv u_t - (du_x)_x + (au)_x + ru - s = 0 \quad (1)$$

valid in the space domain $\Omega = (0, L)$ and in the time domain $t > 0$ with some boundary and initial conditions. The unknown is $u = u(x, t)$. The data is $d \geq 0$ (diffusivity), a (advection velocity), $r \geq 0$ (reaction factor), and s (source term). The notation above

follows mainly that in Reference [1]. The terminology is not quite fixed in the literature. For example, in [1] d and a are called diffusion and advection coefficients, respectively. The reaction term is denoted there generally by the symbol f and $ru - s$ above is then just a linear reaction term. We will call ru as the reaction term. Further, quite commonly, instead of advection the term convection is employed as e.g. in Reference [2].

We will apply the time-discontinuous Galerkin method to solve the problem. In the time-discontinuous Galerkin method the approximation is taken on purpose to be discontinuous in the time direction between the time steps. At first sight this may seem unnatural when the actual solution is known to be continuous. However, this kind of approach has been found to have some useful properties. For example, the formulation contains the important flexible possibility to alter the mesh in space from slab to slab.

The finite element approximation used is the simplest possible: constant in time and with two node linear elements in space in a space-time slab. The — to our knowledge — main new feature in our approach consists of the use of totally three sensitizing parameters associated with the initial condition for a slab. The optimal parameter values are determined by a von Neumann type analysis followed by setting the algorithmic amplification factor in certain sense close to the exact amplification factor.

To proceed concisely, we could start by writing down directly the final discrete weak form to be used, see (34). However, we will continue here rather gently in an effort to convince the reader or the student (or ourselves) of the logic used to arrive at final expressions.

Steady case

Standard weak form

Two of the five optimal sensitizing parameter values associated with the unsteady field equation (1) are in fact found to be those obtained in the steady case. We therefore consider first the steady case.

In the steady case equation (1) simplifies to (superscript s from steady)

$$R^s(u) \equiv L^s(u) - s \equiv -(du_x)_x + (au)_x + ru - s = 0. \quad (2)$$

The well-known standard weak form corresponding to this is

$$\int_{\Omega} [w_x du_x + w(au)_x + wru - ws] d\Omega + bt^s = 0. \quad (3)$$

The notation bt^s refers to possible boundary terms from the boundary conditions. With Dirichlet boundary conditions bt^s will be zero. At a Dirichlet boundary the weighting function $w = w(x) = \delta u(x)$ is set to vanish.

Sensitized weak form

It is known that the Galerkin finite element method will not perform well with the standard weak form when the problem is advection dominated. The reaction term can also lead to unwanted oscillations. For a remedy, the standard formulation is stabilized,

or using the terminology of Reference [9], sensitized. We now present one way to explain the sensitizing similarly as was done in that reference.

We write down the following least squares functional

$$\Pi(u) = \frac{1}{2} \int_{\Omega} \tau^a \left(R^s(u) \right)^2 d\Omega + \frac{1}{2} \int_{\Omega} \tau^r \left(R^s(u)_x \right)^2 d\Omega, \quad (4)$$

where τ^a and τ^r are sensitizing parameters having correct dimensions so that the whole expression becomes dimensionally homogeneous. The first integral on the right-hand side comes from the field equation residual and the second from the differentiated field equation residual. The variation of (4) due to the variation δu becomes

$$\delta \Pi = \int_{\Omega} \tau^a R^s(u) L^s(\delta u) d\Omega + \int_{\Omega} \tau^r R^s(u)_x L^s(\delta u)_x d\Omega. \quad (5)$$

Demanding the variation (5) to vanish and making the interpretation $\delta u = w$ gives what might be called “a least squares weak form”:

$$\int_{\Omega} L^s(w) \tau^a R^s(u) d\Omega + \int_{\Omega} L^s(w)_x \tau^r R^s(u)_x d\Omega = 0. \quad (6)$$

The sensitized weak form is obtained as a linear combination of the two weak forms (3) and (6):

$$\int_{\Omega} \left[w_x du_x + w (au)_x + wru - ws \right] d\Omega + bt^s + \int_{\Omega} L^s(w) \tau^a R^s(u) d\Omega + \int_{\Omega} L^s(w)_x \tau^r R^s(u)_x d\Omega = 0. \quad (7)$$

Discrete sensitized weak form

The finite element approximation is taken to be the simplest possible:

$$\tilde{u}(x) = \sum_j N_j(x) u_j, \quad (8)$$

where u_j are the nodal values of u and N_j are the (global) shape functions corresponding to two node linear elements. The Galerkin method is used and the corresponding finite dimensional weighting function, denoted \tilde{w} , is thus of the same type as (8).

In the discrete weak form we obtain (see Remark 1)

$$R^s(\tilde{u}) \approx -d\tilde{u}_{xx} + a\tilde{u}_x + r\tilde{u} - s \approx a\tilde{u}_x + r\tilde{u} - s, \quad (9)$$

$$L^s(\tilde{w}) \approx -d\tilde{w}_{xx} + a\tilde{w}_x + r\tilde{w} \approx a\tilde{w}_x + r\tilde{w} \approx a\tilde{w}_x, \quad (10)$$

$$R^s(\tilde{u})_x \approx -d\tilde{u}_{xxx} + a\tilde{u}_{xx} + r\tilde{u}_x - s_x \approx r\tilde{u}_x - s_x, \quad (11)$$

$$L^s(\tilde{w})_x \approx -d\tilde{w}_{xxx} + a\tilde{w}_{xx} + r\tilde{w}_x \approx r\tilde{w}_x. \quad (12)$$

Remark 1. Several comments are in place. First, although not indicated directly in (7), the integrals are to be evaluated, as usual in the finite element method, just over the element interiors. Second, in the differentiations, we assume the data d , a , and r to have

some local constant values in an element. (Strictly, some new notations for the data in (9) to (12) would be needed. However, for simplicity, this is not indicated here.) Thus, no derivatives of d , a , and r appear. Third, with two node linear elements, the second or higher order derivatives of \tilde{u} and \tilde{w} vanish. Fourth, in (10) an additional simplification step has been taken by neglecting the term $r\tilde{w}$. The justification for all the simplifications performed above with respect to the terms in the sensitizing integrals is similar to that discussed in [9]. We assume that the standard part in (7) alone produces a correct formulation giving a converging solution when the mesh is refined infinitely. The sensitizing terms are used with finite meshes just in an effort to obtain good balance in stability and accuracy. If the sensitizing parameter values vanish with vanishing mesh sizes, the sensitizing terms vanish similarly and they cannot prevent convergence to the correct solution in the infinite limit.

The discrete analog of (7) is now

$$\begin{aligned} & \int_{\Omega} [\tilde{w}_x d \tilde{u}_x + \tilde{w} a \tilde{u}_x + \tilde{w} r \tilde{u} - \tilde{w} s] d\Omega + \tilde{\mathbf{b}} \mathbf{t}^s \\ & + \int_{\Omega} a \tilde{w}_x \tau^a (a \tilde{u}_x + r \tilde{u} - s) d\Omega + \int_{\Omega} r \tilde{w}_x \tau^r (r \tilde{u}_x - s_x) d\Omega = 0 \end{aligned} \quad (13)$$

It is seen that the standard part has been further simplified by taking a to have some constant value in an element. If wanted, a perhaps more accurate procedure could be to apply first the formula

$$(au)_x = au_x + a_x u \quad (14)$$

and to include the last term as a reaction term.

We define, what we call the advective damping diffusivity

$$d^a = a^2 \tau^a \quad (15)$$

and the reactive damping diffusivity

$$d^r = r^2 \tau^r. \quad (16)$$

These have the same physical dimension as d and are thus more illuminating than τ^a and τ^r as such. Using d^a and d^r , (13) obtains the form

$$\begin{aligned} & \int_{\Omega} \tilde{w}_x (d + d^a + d^r) \tilde{u}_x d\Omega + \int_{\Omega} \tilde{w} a \tilde{u}_x d\Omega + \int_{\Omega} \tilde{w} r \tilde{u} d\Omega + \int_{\Omega} \tilde{w}_x \frac{d^a r}{a} \tilde{u} d\Omega \\ & - \int_{\Omega} \tilde{w} s d\Omega - \int_{\Omega} \tilde{w}_x \frac{d^a}{a} s d\Omega - \int_{\Omega} \tilde{w}_x \frac{d^r}{r} s_x d\Omega + \tilde{\mathbf{b}} \mathbf{t}^s = 0. \end{aligned} \quad (17)$$

This is the final discrete sensitized weak form in the steady case.

Sensitizing patch test

As in Reference [9], the goal in sensitizing is also here to achieve as far as possible the nodally exact finite element solution. We define first suitable reference solutions to be used in a sensitizing patch test.

The governing field equation (2), with some constant representative data and with no source term is considered:

$$-du_{xx} + au_x + ru = 0. \quad (18)$$

The analytical solution of this is

$$u(x) = Ae^{s_1x} + Be^{s_2x}, \quad (19)$$

where

$$s_{1,2} = \frac{a \pm \sqrt{a^2 + 4dr}}{2d} \quad (20)$$

and where A and B are integration constants. Two reference solutions $A \exp s_1x$ and $B \exp s_2x$ are used.

An element patch is considered (Figure 1)

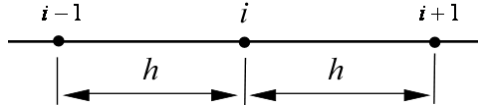


Figure 1. Uniform three-node, two-element patch.

The system equation for the middle node i is formed using the weak form (17). The nodal values are taken from the first reference solution. A linear equation in the two unknowns τ^a and τ^r is obtained. By proceeding similarly with the second reference solution a second equation is obtained. The unknowns are solved from these two equations (using the Mathematica program [10]). The expressions

$$d^a = pe \left(\frac{\frac{pe}{da} + \frac{\sinh \frac{pe}{2}}{2}}{\cosh \frac{pe}{2} - \cosh \sqrt{\frac{pe^2}{4} + da}} \right) d \quad (21)$$

and

$$d^r = \left[-\left(6pe^2 + 2da(3+da)\right) \cosh \frac{pe}{2} + \left(6pe^2 + da(6-da)\right) \cosh \sqrt{\frac{pe^2}{4} + da} - 6pe \cdot da \sinh \frac{pe}{2} \right] d / \left[da \left(\cosh \frac{pe}{2} - \cosh \sqrt{\frac{pe^2}{4} + da} \right) \right] \quad (22)$$

are obtained. The notations $pe = ah/d$ and $da = rh^2/d$ have been used. Quantity pe is (elementwise) *Péclet number* and da (elementwise) *Damköhler number*.

In the pure diffusion-advection case d^a simplifies to the more familiar form

$$d^a = \left(\frac{pe}{2} \coth \frac{pe}{2} - 1 \right) d \quad (23)$$

and in the pure diffusion-reaction case d^r simplifies to

$$d^r = \left(\frac{da \cosh \sqrt{da} + 2}{6 \cosh \sqrt{da} - 1} - 1 \right) d. \quad (24)$$

In practice the above expressions are often simplified further by suitable approximations depending on the ranges of pe and da .

The formulas for d^a and d^r were obtained in the sensitizing patch test using a regular mesh and constant data, but of course, in an analysis with an irregular mesh and varying data, local data values and element length for an element are applied in the relevant formulas.

Unsteady case

Time-discontinuous standard weak form

We repeat for convenience the governing field equation (1) in the time-dependent case:

$$R(u) \equiv L(u) - s \equiv u_t - (du_x)_x + (au)_x + ru - s = 0. \quad (25)$$

The most straightforward approach to extend the steady case into the unsteady case is the so-called method of lines or the semi-discretization. It would mean here that (8) would be written as $\tilde{u}(x, t) = \sum N_j(x) u_j(t)$. An ordinary differential system in time would be obtained. However, as remarked in Reference [1, p. 95]:", if we apply a standard ODE method to a semi-discrete system (6.1), information about the underlying PDE problem might be neglected". Numerical experience shows also that the optimal sensitizing parameters found in the steady case do not work any more well in the unsteady case with the semi-discrete approach.

In the finite element method literature the time-discontinuous Galerkin method is often advocated. We take this approach also here as the starting point.

Figure 2 shows the relevant solution domain and its division into space-time slabs and some additional notations.

We will describe the procedure in a typical slab $S_n = \Omega \times I_n$ where $I_n = (t_n, t_{n+1})$. However, for simplicity of presentation we drop in the following the index n referring to the slab; we consider a generic slab.

In the time-discontinuous Galerkin method the initial (or the continuity or the jump) condition for the present slab is

$$u^+(x) - u^-(x) = 0 \quad (26)$$

in Ω and it is satisfied only in a weak sense. The $+$ and $-$ superscript notations refer to the limiting values at the lower time level of the slab as indicated in Figure 2. We notice that u^- is a given function from the point of view of the present slab. Using similar notations as for the field equation residual, we could write (26) also more pedantically as (superscript i from initial)

$$R^i(u^+) \equiv L^i(u^+) - u^- \equiv u^+ - u^- = 0 \quad (27)$$

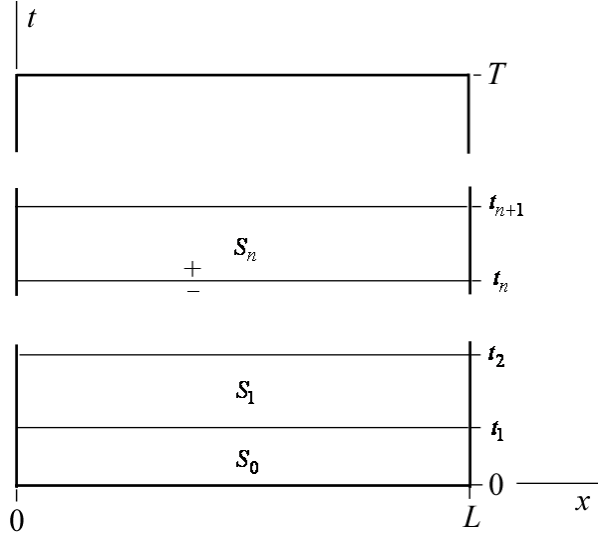


Figure 2. Solution domain and space-time slabs.

The standard time-discontinuous weak form for a generic slab is

$$\int_I \int_{\Omega} \left(w u_t + w_x du_x + w (au)_x + w ru - ws \right) d\Omega dt + \int_{\Omega} w^+ (u^+ - u^-) d\Omega + bt = 0, \quad (28)$$

where $w = w(x, t)$ is the weighting function.

Sensitized time-discontinuous weak form

The sensitized time-discontinuous weak form for a generic slab is now taken to be

$$\begin{aligned} & \int_I \int_{\Omega} \left(w u_t + w_x du_x + w (au)_x + w ru - ws \right) d\Omega dt + \int_{\Omega} w^+ (u^+ - u^-) d\Omega + bt \\ & + \int_I \int_{\Omega} L(w) \tau^a R(u) d\Omega dt + \int_I \int_{\Omega} L(w)_x \tau^r R(u)_x d\Omega dt \\ & + \int_{\Omega} w^+ \tau^i (u^+ - u^-) d\Omega + \int_{\Omega} w^+ \varepsilon^i (u_x^+ - u_x^-) d\Omega + \int_{\Omega} w_x^+ \sigma^i (u_x^+ - u_x^-) d\Omega = 0. \end{aligned} \quad (29)$$

The two last lines on the left-hand side represent the sensitizing terms. The logic for arriving at the sensitizing integrals with parameters τ^a and τ^r can be explained similarly as in section “Sensitized weak form” starting from a least squares functional. Now however, the integrals are over the slab area. In a same way, the sensitizing terms with parameters τ^i and σ^i can be arrived at by starting from the least squares functional

$$\Pi(u^+) = \frac{1}{2} \int_{\Omega} \tau^i \left(R^i(u^+) \right)^2 d\Omega + \frac{1}{2} \int_{\Omega} \sigma^i \left(R^i(u^+)_x \right)^2 d\Omega. \quad (30)$$

It is formed from the initial condition residual and of its differentiated form. The sensitizing term with parameter ε^i is not obtained via the least squares formulation. However, the least squares route is just one very useful way to approach sensitizing and nothing demands us to restrict just to it. The main thing is that the formulation is consistent in the sense that the exact solution satisfies the weak form (29).

The three parameters τ^i , ε^i and σ^i are associated with the initial condition (27). They have the roles of some kind of artificial reaction, advection, and diffusivity, respectively. It will be seen that these additional parameters give possibilities to tune the discrete solution behavior to advantage.

Discrete sensitized time-discontinuous weak form

The constant in time approximation in the slab is an essential feature for obtaining a formulation simple enough for our purposes. So instead of an approximation of the type $\tilde{u} = \tilde{u}(x, t)$ we have just $\tilde{u} = \tilde{u}(x)$, or in more detail, still the form (8) now valid in the full slab and not only on a line. We are thus applying rectangular space-time elements. As we are going to use the Galerkin method, the finite dimensional weighting function \tilde{w} is also constant in time, so $\tilde{w} = \tilde{w}(x)$ and we see that

$$\tilde{u}^+ = \tilde{u}, \quad \tilde{w}^+ = \tilde{w} \quad (31)$$

and we obtain also

$$\tilde{u}_t = 0, \quad \tilde{w}_t = 0. \quad (32)$$

Further, it is seen that the simplifications applied to R^s and L^s in formulas (9) to (12) can be used equally well for R and L (see Remark 1) and the discrete analog of the weak form (29) becomes

$$\begin{aligned} & \int_I \int_{\Omega} (\tilde{w}_x d\tilde{u}_x + \tilde{w} a \tilde{u}_x + \tilde{w} r \tilde{u} - \tilde{w} s) d\Omega dt + \int_{\Omega} \tilde{w} (\tilde{u} - u^-) d\Omega + \tilde{b} t \\ & + \int_I \int_{\Omega} a \tilde{w}_x \tau^a (a \tilde{u}_x + r \tilde{u} - s) d\Omega dt + \int_I \int_{\Omega} r \tilde{w}_x \tau^r (r \tilde{u}_x - s_x) d\Omega dt \\ & + \int_{\Omega} \tilde{w} \tau^i (\tilde{u} - u^-) d\Omega + \int_{\Omega} \tilde{w} \varepsilon^i (\tilde{u}_x - u_x^-) d\Omega + \int_{\Omega} \tilde{w}_x \sigma^i (\tilde{u}_x - u_x^-) d\Omega = 0. \end{aligned} \quad (33)$$

As the dependence on t is missing (or assumed to be missing) in the space integrals, performing the integrations with respect to t produces values equal the integrands multiplied by $\Delta t = t_{n+1} - t_n$ (see Remarks 2 and 3). Thus dividing the resulting equation by Δt and introducing the notations (15) and (16) gives

$$\begin{aligned} & \int_{\Omega} \tilde{w}_x (d + d^a + d^r) \tilde{u}_x d\Omega + \int_{\Omega} \tilde{w} a \tilde{u}_x d\Omega + \int_{\Omega} \tilde{w} r \tilde{u} d\Omega + \int_{\Omega} \tilde{w}_x \frac{d^a r}{a} \tilde{u} d\Omega \\ & - \int_{\Omega} \tilde{w} s d\Omega - \int_{\Omega} \tilde{w}_x \frac{d^a}{a} s d\Omega - \int_{\Omega} \tilde{w}_x \frac{d^r}{r} s_x d\Omega \\ & + \frac{1}{\Delta t} \int_{\Omega} \tilde{w} (1 + \tau^i) (\tilde{u} - u^-) d\Omega + \frac{1}{\Delta t} \int_{\Omega} \tilde{w} \varepsilon^i (\tilde{u}_x - u_x^-) d\Omega \\ & + \frac{1}{\Delta t} \int_{\Omega} \tilde{w}_x \sigma^i (\tilde{u}_x - u_x^-) d\Omega + \frac{1}{\Delta t} \tilde{b} t = 0. \end{aligned} \quad (34)$$

This equation is useful due to its simple structure for further interpretations. We have a weak form only in space. The only difference with respect to the corresponding steady formulation (17) consists of the last three integrals on the left-hand side. Let us now consider such forcing terms in time that the steady solution is approached. Then \tilde{u} approaches u^- (u^- is given also by the finite element approximation) and the last three

integrals approach zero. Obviously then, the appropriate values for d^a and for d^r are those obtained in the steady case. Thus these values are to be used also in the time-dependent case when the present formulation is applied and the task now remains “just” to determine the optimal values for the parameters τ^i , ε^i , and σ^i .

Remark 2. If the data d , a , r , s depends on t , the integrand in the standard part on the first line in (33) is not constant in t . However, we can replace the possibly varying data with some local mean values in the time direction so that the integrand becomes constant in time. This produces no error in the limit as Δt approaches zero. Again, for notational convenience we do not care to use new symbols for these mean values.

Remark 3. Concerning the boundary conditions, let us consider an example case, say a setting with $u(0, t) = \bar{u}(t)$ and $-du_x(L, t) = \bar{q}(t)$. The overbar notation refers here to a given quantity. Then in the slab, $\tilde{u}(0)$ is set equal to $\bar{u}(t_n + \Delta t)$. The term $\text{bt} = \int_I w(L, t) \bar{q}(t) dt$ and thus $\tilde{\text{bt}} = \int_I \tilde{w}(L) \bar{q}(t) dt = \tilde{w}(L) \int_I \bar{q}(t) dt \approx \tilde{w}(L) \bar{q}_m \Delta t$, where \bar{q}_m is some mean value of \bar{q} on I . Thus finally, the last term on the left-hand side of (34) becomes $\tilde{w}(L) \bar{q}_m$, which corresponds to a similar term $\tilde{\text{bt}}^s$ with a given flux in the steady case.

Element contributions

As mentioned in the “Introduction” section, the time-discontinuous formulation contains the possibility to alter the mesh in space from slab to slab. In the applications to follow, we do not consider for simplicity of presentation this alternative. Thus here

$$u^-(x) = \sum_j N_j(x) u_j^-, \quad (35)$$

where the nodal values are those obtained in the previous slab (or at the beginning those obtained by approximation of the initial data).

For a generic element with length h and local nodes 1 and 2 with the first node having a lower x -coordinate value the element contributions to the system equations are expressed here as

$$[K] \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} - \begin{Bmatrix} b_1 \\ b_2 \end{Bmatrix}. \quad (36)$$

Performing the integrations, the element coefficient matrix becomes

$$\begin{aligned} [K] = & \frac{d + d^a + d^r}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{a}{2} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} + \frac{rh}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \frac{d^a r}{2a} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \\ & + \frac{(1 + \tau^i)h}{6\Delta t} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} + \frac{\varepsilon^i}{2\Delta t} \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} + \frac{\sigma^i}{h\Delta t} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \end{aligned} \quad (37)$$

and the column vector members become

$$b_1 = \int N_1 s dx - \frac{d^a}{ah} \int s dx - \frac{d^r}{rh} \int s_x dx$$

$$+\frac{(1+\tau^i)h}{6\Delta t}(2u_1^-+u_2^-)+\frac{\varepsilon^i}{2\Delta t}(-u_1^-+u_2^-)+\frac{\sigma^i}{h\Delta t}(u_1^- - u_2^-), \quad (38)$$

$$b_2 = \int N_2 s \, dx + \frac{d^a}{ah} \int s \, dx + \frac{d^r}{rh} \int s_x \, dx$$

$$+\frac{(1+\tau^i)h}{6\Delta t}(u_1^-+2u_2^-)+\frac{\varepsilon^i}{2\Delta t}(-u_1^-+u_2^-)+\frac{\sigma^i}{h\Delta t}(-u_1^-+u_2^-). \quad (39)$$

The meaning of the notations should be obvious. The data has been assumed generally as constant in the element. An exception is the source term for which the possibility for a more accurate presentation has been preserved.

System equations for a patch

We consider a uniform three-node, two-element patch (Figure 1). More correctly, the points $i-1, i, i+1$ in the figure represent now nodal lines in the slab. In the parameter values determination we take a case with $s = 0$. The system equation for node i is found to become (the equation has been further multiplied by $\Delta t/h$ so that the coefficients of the nodal values become dimensionless)

$$\begin{aligned} & \left(-\frac{(d+d^a+d^r)\Delta t}{h^2} - \frac{a\Delta t}{2h} + \frac{r\Delta t}{6} + \frac{d^a r \Delta t}{2ah} \right) u_{i-1} + \left(\frac{1+\tau^i}{6} - \frac{\varepsilon^i}{2h} - \frac{\sigma^i}{h^2} \right) u_{i-1} \\ & + \left(2\frac{(d+d^a+d^r)\Delta t}{h^2} + \frac{4r\Delta t}{6} \right) u_i + \left(\frac{4(1+\tau^i)}{6} + \frac{2\sigma^i}{h^2} \right) u_i \\ & + \left(-\frac{(d+d^a+d^r)\Delta t}{h^2} + \frac{a\Delta t}{2h} + \frac{r\Delta t}{6} - \frac{d^a r \Delta t}{2ah} \right) u_{i+1} + \left(\frac{1+\tau^i}{6} + \frac{\varepsilon^i}{2h} - \frac{\sigma^i}{h^2} \right) u_{i+1} = \\ & \left(\frac{1+\tau^i}{6} - \frac{\varepsilon^i}{2h} - \frac{\sigma^i}{h^2} \right) u_{i-1}^- \\ & + \left(\frac{4(1+\tau^i)}{6} + \frac{2\sigma^i}{h^2} \right) u_i^- \\ & + \left(\frac{1+\tau^i}{6} + \frac{\varepsilon^i}{2h} - \frac{\sigma^i}{h^2} \right) u_{i+1}^-. \end{aligned} \quad (40)$$

Parameters τ^i , ε^i , and σ^i are determined by making use of equation (40) in connection with a reference solution of the kind used in the conventional von Neumann or Fourier analysis, e.g. [11]. (We will follow somewhat the notation of that reference.) This will be explained in detail below.

Reference solution

The governing differential equation with constant data and with no source term is

$$u_t - du_{xx} + au_x + ru = 0. \quad (41)$$

An analytical solution of the following separation of variables type $u(x, t) = A(t)\exp ikx$ is introduced. Here i is the imaginary unit and k a real quantity (wave number). Substituting this in (41) gives the solution

$$u(x, t) = A_0 e^{-(dk^2 + iak+r)t} e^{ikx}, \quad (42)$$

where A_0 is an integration constant.

The exact so-called *complex amplification factor* is defined as

$$G_e \equiv \frac{u(x, t + \Delta t)}{u(x, t)} = \frac{A_0 e^{-(dk^2 + iak+r)(t+\Delta t)} e^{ikx}}{A_0 e^{-(dk^2 + iak+r)t} e^{ikx}} = e^{-(dk^2 + iak+r)\Delta t}. \quad (43)$$

The expressions to follow in the discrete case and especially in the full diffusion-advection-reaction case become complicated and long. To simplify the formulas, certain non-dimensional short-hand notations are introduced in addition to pe and da used already in the steady case. These are defined in the Appendix. The double letter dimensionless quantities are written in antique. The meanings of these notations are difficult to remember but in fact there is no great need for that. The main thing is that dimensionless quantities are preferable when numerical manipulations in the end are used. When employing these new notations, (43) can be put in the form

$$G_e = e^{-fo \cdot \beta^2 - rs} e^{-ico \cdot \beta}. \quad (44)$$

The most important quantity to appear here and later is $\beta = kh$, which is a kind of dimensionless wave number. With a fixed h , the larger β , the more “wavy” the solution in space.

Further, for later purposes, we may represent the exact complex amplification factor alternatively as

$$G_e = |G_e| e^{i\phi_e} \quad (45)$$

with the magnitude

$$|G_e| = e^{-fo \cdot \beta^2 - rs} \quad (46)$$

and the phase angle

$$\phi_e = -co \cdot \beta. \quad (47)$$

In the discrete case using the von Neumann or Fourier approach, a form

$$\hat{u}(x, t) = \hat{A}_0 e^{\alpha t} e^{ikx} \quad (48)$$

similar to (42) is taken. It is to be noticed that the term in space, $\exp ikx$ is the same as in (42) but the term in time, $\exp \alpha t$ is not yet fixed. Expression (48) is evaluated at the

nodes of the patch and these values are used in the system equation (40). In this we associate the nodal values with superscript minus with the generic time level t and the nodal values without the superscript with the time level $t + \Delta t$. Similarly, node i is associated with the generic space coordinate value x . Performing the substitutions it is found that the term $\exp \alpha \Delta t$ can be solved from the resulting equation. In the literature, the ratio

$$G \equiv \frac{\hat{u}(x, t + \Delta t)}{\hat{u}(x, t)} = \frac{\hat{A}_0 e^{\alpha(t+\Delta t)} e^{ikx}}{\hat{A}_0 e^{\alpha t} e^{ikx}} = e^{\alpha \Delta t} \quad (49)$$

is called algorithmic amplification factor. We obtain here

$$G = \frac{1}{1 + \frac{6 pe(1 - \cos \beta) fo^t + pe \cdot rs (2 + \cos \beta) + 3i \left(pe \cdot co - (d^a / d) rs \right) \sin \beta}{(2 + \cos \beta)(1 + \tau^i) + 6(1 - \cos \beta)(\sigma^i / h^2) + 3i pe (\varepsilon^i / h) \cdot \sin \beta}}, \quad (50)$$

where $fo^t = fo + fo^a + fo^r$.

The above type of procedure to determine the algorithmic amplification factor is well documented in the literature; e.g. [11]. However, it is interesting to note that we have proceeded here quite in the same way as in the sensitizing patch test in the steady case and from this point of view expression (48) can well be called also as a reference solution.

G can be represented similarly to G_e as

$$G = |G| e^{i\phi}. \quad (51)$$

However, since the detailed expressions will be complicated, we do not present them here. It will be enough to evaluate them with numerical values.

The procedure proposed for the parameter values determination is finally now as follows. We would obviously like the algorithmic G to be close to the exact G_e . To this end we expand G and G_e to truncated Taylor series with respect to β at $\beta = 0$. We then demand the expansions to coincide as far as possible. It should be noted that the idea to set G in certain sense close to G_e in an effort to determine some parameter values is not new; see e.g. [12].

We do not continue here in the full diffusion-advection-reaction case as the expressions would become so long. (There is no problem to obtain the necessary formulas. In general, all the three parameters τ^i , ε^i , and σ^i are then non-zero.) It is perhaps more kind towards the reader to see how the idea of parameter values determination works in the simpler diffusion-advection and diffusion-reaction cases so we will deal only with them from this onwards.

Unsteady diffusion-advection

Parameter values determination

In the diffusion-advection case ($r = 0$) the exact amplification factor is

$$G_e = e^{-fo \cdot \beta^2} e^{-i co \cdot \beta} \quad (52)$$

and the algorithmic amplification factor is found to become

$$G = \frac{1}{1 + \frac{6(1 - \cos \beta)(fo + fo^a) + 3i co \cdot \sin \beta}{(2 + \cos \beta)(1 + \tau^i) + 6(1 - \cos \beta)(\sigma^i / h^2) + 3i(\varepsilon^i / h) \sin \beta}}. \quad (53)$$

The corresponding Taylor series are

$$G_e = 1 - i co \cdot \beta + \left(-\frac{co^2}{2} - fo \right) \beta^2 + i \left(\frac{co^3}{6} + co \cdot fo \right) \beta^3 + O[\beta^4] \quad (54)$$

and

$$G = 1 - i co \cdot \beta + \left(-co \left(co + \varepsilon^i / h \right) - \left(fo + fo^a \right) \right) \beta^2 + i \left(co \left(co + \varepsilon^i / h \right)^2 + co \left(2 \left(fo + fo^a \right) + \sigma^i / h^2 \right) + \left(\varepsilon^i / h \right) \left(fo + fo^a \right) \right) \beta^3 + O[\beta^4]. \quad (55)$$

In fact, the full series expression with τ^i included showed immediately that we have to put here $\tau^i = 0$ for the first powers of β to become the same in the two series. Therefore expression (55) is shown here for simplicity already without τ^i . The terms for zeroth and first powers of β in (54) and (55) are identical. Demanding the second and third powers (separately) to be equal gives two equations from which ε^i and σ^i can be determined. Taking further the notations in the Appendix into account, there is obtained

$$\varepsilon^i = - \left(\frac{co}{2} + \frac{d^a / d}{pe} \right) h, \quad (56)$$

$$\sigma^i = \left(\frac{d^a / d}{pe^2} - \frac{co^2}{12} - \frac{1 + d^a / d}{2} \frac{co}{pe} \right) h^2. \quad (57)$$

Unsteady pure diffusion

It is of some interest to see what is obtained in the simplified cases of pure diffusion and pure advection where it is easier to make comparison with traditional formulas in the literature.

The case of pure diffusion ($a = 0$) is considered first. In the *small advection* case the parameter d^a (see (23)) obtains first the simplified form

$$d^a \approx \frac{a^2 h^2}{12d}. \quad (58)$$

When this is used in (56) and (57) and when we let $a \rightarrow 0$, we obtain $\varepsilon^i = 0$ and

$$\sigma^i = \left(\frac{1}{12} - \frac{fo}{2} \right) h^2. \quad (59)$$

We form system equation (40). The result is

$$\begin{aligned} & \left(-\frac{fo}{2} + \frac{1}{12}\right)u_{i-1} + \left(fo + \frac{10}{12}\right)u_i + \left(-\frac{fo}{2} + \frac{1}{12}\right)u_{i+1} = \\ & \left(\frac{fo}{2} + \frac{1}{12}\right)u_{i-1}^- + \left(-fo + \frac{10}{12}\right)u_i^- + \left(\frac{fo}{2} + \frac{1}{12}\right)u_{i+1}^-. \end{aligned} \quad (60)$$

This becomes more transparent if we introduce the obvious notation used in the finite difference method, divide by Δt , make use of the relation $fo = d\Delta t / h^2$ and arrange:

$$\begin{aligned} & \frac{1}{12} \frac{\left(u_{i-1}^{n+1} + 10u_i^{n+1} + u_{i+1}^{n+1}\right) - \left(u_{i-1}^n + 10u_i^n + u_{i+1}^n\right)}{\Delta t} \\ & - d \frac{\frac{u_{i-1}^n - 2u_i^n + u_{i+1}^n}{h^2} + \frac{u_{i-1}^{n+1} - 2u_i^{n+1} + u_{i+1}^{n+1}}{h^2}}{2} = 0. \end{aligned} \quad (61)$$

This is almost the well-known Crank-Nicolson scheme applied to the diffusion equation (with $s = 0$), the only difference being that in the Crank-Nicolson scheme the upper line in (61) is replaced by the simpler expression

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} \quad (62)$$

In (61) we clearly have a weighted time derivative expression.

The amplification factors in pure diffusion are real. The exact factor is

$$G_e = e^{-fo \cdot \beta^2} \quad (63)$$

The algorithmic factor value of the present formulation is found to be

$$G = \frac{5 + \cos \beta - 6fo(1 - \cos \beta)}{5 + \cos \beta + 6fo(1 - \cos \beta)} \quad (64)$$

and the algorithmic value of the Crank-Nicolson scheme is (Reference [11, p. 112])

$$G = \frac{1 - fo(1 - \cos \beta)}{1 + fo(1 - \cos \beta)}. \quad (65)$$

These are shown as functions of β in Figure 3.

It is obvious that due to the way G is demanded to coincide with G_e , the accuracy of the present scheme is very good for small values of β .

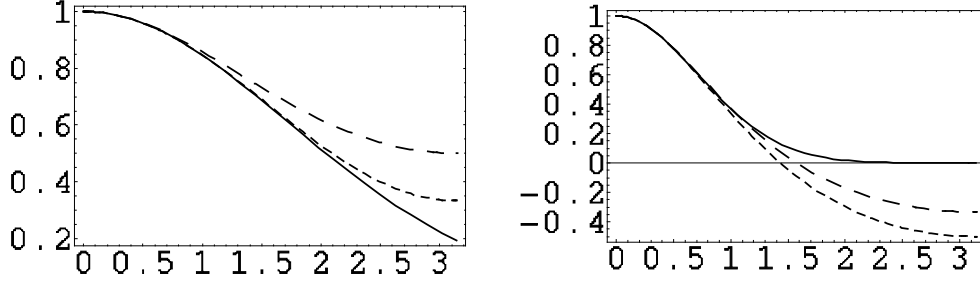


Figure 3. Amplification factors in pure diffusion as functions of β . On the left, Fourier number $fo = 1/6$. On the right, Fourier number $fo = 1$. Solid line, exact factor. Dense dashing, the present scheme. Coarse dashing, the Crank-Nicolson scheme.

Unsteady pure advection

In the case of very *large advection* or pure advection ($d \rightarrow 0$), the parameter d^a obtains the form

$$d^a \approx \frac{|a|h}{2}. \quad (66)$$

We find correspondingly from (56) and (57)

$$\varepsilon^i = -\left(\frac{co}{2} + \frac{\text{sgn } co}{2}\right)h, \quad (67)$$

$$\sigma^i = -\left(\frac{co^2}{12} + \frac{|co|}{4}\right)h^2. \quad (68)$$

System equation (40) becomes (we use the same notation as in (61))

$$\begin{aligned} & \left(\frac{1}{6} + \frac{\text{sgn } co}{4} - \frac{co}{4} - \frac{|co|}{4} + \frac{co^2}{12}\right)u_{i-1}^{n+1} \\ & + \left(\frac{4}{6} + \frac{|co|}{2} - \frac{co^2}{6}\right)u_i^{n+1} + \left(\frac{1}{6} - \frac{\text{sgn } co}{4} + \frac{co}{4} - \frac{|co|}{4} + \frac{co^2}{12}\right)u_{i+1}^{n+1} = \\ & \left(\frac{1}{6} + \frac{\text{sgn } co}{4} + \frac{co}{4} + \frac{|co|}{4} + \frac{co^2}{12}\right)u_{i-1}^n \\ & + \left(\frac{4}{6} - \frac{|co|}{2} - \frac{co^2}{6}\right)u_i^n + \left(\frac{1}{6} - \frac{\text{sgn } co}{4} - \frac{co}{4} + \frac{|co|}{4} + \frac{co^2}{12}\right)u_{i+1}^n \end{aligned} \quad (69)$$

and the amplification factor is found to become

$$G = \frac{4 - \text{co}^2 + (2 + \text{co}^2) \cos \beta - 3(1 - \cos \beta) |\text{co}| - 3i(\text{co} + \text{sgn co}) \sin \beta}{4 - \text{co}^2 + 3|\text{co}| + (2 + \text{co}^2 - 3|\text{co}|) \cos \beta + 3i(\text{co} - \text{sgn co}) \sin \beta}. \quad (70)$$

It is difficult to see a direct resemblance with some familiar finite difference formulations. We take here as a comparison case the Lax-Wendroff scheme, which has the amplification factor (Reference [11, p. 101])

$$G = 1 - \text{co}^2 (1 - \cos \beta) - i \text{co} \cdot \sin \beta. \quad (71)$$

The quantities $|G|/|G_e|$ (here $|G_e|=1$) and ϕ/ϕ_e are represented similarly as in Reference [4] in Figure 4 as radii in polar coordinates as functions of the polar angle β .

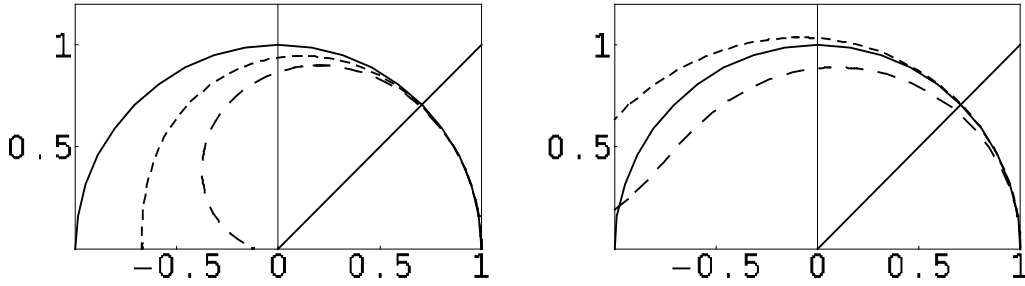


Figure 4. On the left, $|G|/|G_e|$ as function of the polar angle β . On the right, ϕ/ϕ_e as function of the polar angle β . The Courant number $\text{co} = 0.75$. Dense dashing, the present scheme. Coarse dashing, the Lax-Wendroff scheme.

The accuracy of the present scheme appears again to be very good for small values of β . When $\text{co} = 1$, both schemes are found to be without errors. However, when $|\text{co}| > 1$ also the present scheme gives a $|G|$ larger than one. As the proposed approach leads to an implicit scheme, the appearance of a stability limit is not welcome. This obviously is here the price paid due to the high accuracy. Ways to increase the stability limit at the expense of accuracy are under study.

Unsteady diffusion-reaction

Parameter values determination

In the diffusion-reaction case ($a = 0$) the exact amplification factor

$$G_e = e^{-\text{fo} \cdot \beta^2 - \text{rs}} \quad (72)$$

and the algorithmic amplification factor

$$G = \frac{1}{1 + \frac{6(1 - \cos \beta)(\text{fo} + \text{fo}^r) + (2 + \cos \beta) \text{rs}}{(2 + \cos \beta)(1 + \tau^i) + 6(1 - \cos \beta)(\sigma^i / h^2)}}. \quad (73)$$

G_e is here real and to have G also real we have to put $\varepsilon^i = 0$. This has already taken into account in expression (73).

The Taylor expansions are found to be

$$G_e = e^{-rs} - fo \cdot e^{-rs} \beta^2 + O[\beta^3] \quad (74)$$

and

$$G = \frac{1 + \tau^i}{1 + rs + \tau^i} + \frac{(rs(\sigma^i / h^2) - (fo + fo^r))(1 + \tau^i)}{(1 + rs + \tau^i)^2} \beta^2 + O[\beta^3]. \quad (75)$$

Demanding the zeroth and second powers to be equal gives two equations from which τ^i and σ^i can be determined. The solutions are

$$\tau^i = \frac{1 + rs - e^{rs}}{-1 + e^{rs}}, \quad (76)$$

$$\sigma^i = \frac{-1 + (1 - rs)e^{rs} + (-1 + e^{rs})d^r / d}{(-1 + e^{rs})^2} d\Delta t. \quad (77)$$

Pure reaction

In the case of *large reaction* or pure reaction ($d \rightarrow 0$) we obtain from (24) by a limiting process the value

$$d^r \approx \frac{rh^2}{6}. \quad (78)$$

A careful study shows that now

$$\tau^i = -1 + \frac{r\Delta t}{-1 + e^{r\Delta t}} = -1 + \frac{6\sigma^i}{h^2} \quad (79)$$

and the system equation (40) simplifies finally just to

$$u_i^{n+1} = e^{-r\Delta t} u_i^n. \quad (80)$$

This result has an obvious interpretation when the solution of the field equation $u_t + ru = 0$ is considered.

Conclusions

A to our knowledge new approach to solve one-dimensional time-dependent diffusion-advection-reaction problems has been presented. The main new feature consists of the use of three different sensitizing parameters associated with the initial condition from slab to slab. Optimizing the parameter values make it possible to tune the discrete solution behavior to advantage. The rather nice feature in the effort to determine all the sensitizing parameter values is that it can be split into two parts: First certain parameters

are determined in the steady case. These need no updating when the additional parameters taking care of the time dimension are found.

The approach demands involved analytical manipulations, which would be by hand calculations in practice nearly impossible. However, modern tools, such as the Mathematica program, make the task rather easy.

In certain simple special cases (pure diffusion, pure advection, pure reaction) equations have been specialized far enough to see that very accurate formulations emerge.

The approach leads normally to implicit schemes. It seems that the great accuracy achieved has the price that there are stability limits. The use Taylor series in setting G close to G_e is obviously not the only possibility. We are presently working on ways to extend the stability limits and on ways to extend the approach to two and three space dimensions.

The concepts used in the approach are not mathematically complicated. If one has assimilated the ideas in sensitizing (stabilization) first in the steady case, in the step into the unsteady case quite similar tools are applied.

Appendix

The following notations are defined:

$$co = a\Delta t / h \quad \underline{\text{C}}ourant \text{ number,}$$

$$da = rh^2 / d \quad \underline{\text{D}}amk\ddot{o}hler \text{ number,}$$

$$fo = d\Delta t / h^2 \quad \underline{\text{F}}ourier \text{ number,}$$

$$fo^a = d^a \Delta t / h^2 \quad \underline{\text{a}}dvective \text{ damping diffusivity associated } \underline{\text{F}}ourier \text{ number,}$$

$$fo^r = d^r \Delta t / h^2 \quad \underline{\text{r}}eactive \text{ damping diffusivity associated } \underline{\text{F}}ourier \text{ number,}$$

$$pe = ah / d \quad \underline{\text{P}}\acute{e}clet \text{ number,}$$

$$rs = r\Delta t \quad \text{time } \underline{\text{s}}tep \text{ associated } \underline{\text{r}}eaction \text{ number,}$$

$$\beta = kh \quad \text{dimensionless wave number.}$$

All of these dimensionless quantities are not independent. For example, $fo = co/pe$ and $rs = da \cdot co/pe$.

References

- [1] W. Hundsdorfer, J.G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*, Corrected 2nd printing, Springer, 2007.
- [2] J. Donea, A. Huerta, *Finite Element Methods for Flow Problems*, Wiley, 2003.
- [3] F. Shakib, T.J.R. Hughes, A new finite element formulation for computational fluid dynamics: IX. Fourier analysis of space-time Galerkin/least-squares algorithms, *Comput. Methods Appl. Mech. Engrg.*, 87 (1991) 35-58.
- [4] A. Huerta, J. Donea, Time-accurate solution of stabilized convection-diffusion-reaction equations: I—time and space discretization, II—accuracy analysis and

- examples, *Communications in Numerical Methods in Engineering*, 18 (2002) 565-573, 575-584.
- [5] G. Hauke, M.H. Doweidar, Fourier analysis of semi-discrete and space-time stabilized methods for the advective-diffusive-reactive equation: I. SUPG, *Comput. Methods Appl. Mech. Engrg.*, 194 (2005) 45-81.
 - [6] E. Oñate, J. Miquel, G. Hauke, Stabilized formulation for the advection-diffusion-absorption equations using finite calculus and linear finite elements, *Comput. Methods Appl. Mech. Engrg.*, 195 (2006) 3926-3946.
 - [7] G. Hauke, M.H. Doweidar, Fourier analysis of semi-discrete and space-time stabilized methods for the advective-diffusive-reactive equation: III. SGS/GSGS, *Comput. Methods Appl. Mech. Engrg.*, 195 (2006) 6158-6176.
 - [8] P. Nadukandi, E. Oñate, J. Carcia, Analysis of a consistency recovery method for the 1D convection-diffusion equation using linear finite elements, *Int. J. Numer. Methods Fluids*, 57 (2008) 1291-1320.
 - [9] J. Freund, E-M. Salonen, Sensitizing according to Courant the Timoshenko beam finite element solution, *Int. J. Numer. Methods Engrg.*, 47 (2000) 1621-1631.
 - [10] www.wolfram.com/products/mathematica.
 - [11] D.A. Anderson, J.C. Tannehill, R.H. Pletcher, *Computational Fluid Mechanics and Heat Transfer*, McGraw-Hill, 1984.
 - [12] A.J. Baker, J.W. Kim, A Taylor weak-statement algorithm for hyperbolic conservation laws, *Int. J. Numer. Methods Fluids*, 7 (1987) 489-520.

Eero-Matti Salonen
 Aalto University, Department of Civil and Structural Engineering
 P.O.Box 12100, 00076, Aalto
 Finland
 eero-matti.salonen@aalto.fi

Jouni Freund
 Aalto University, Department of Applied Mechanics
 P.O.Box 14300, 00076 Aalto
 Finland
 jouni.freund@aalto.fi