NUMERICALLY STABLE FORMULAS FOR A PARTICLE-BASED EXPLICIT EXPONENTIAL INTEGRATOR*

PRASHANTH NADUKANDI†

Abstract. Numerically stable formulas are presented for the closed-form analytical solution of the X-IVAS scheme in 3D. This scheme is a state-of-the-art particle-based explicit exponential integrator developed for the Particle Finite Element Method. Algebraically, this scheme involves two steps: 1) the solution of tangent curves for piecewise linear vector fields defined on simplicial meshes and 2) the solution of line integrals of piecewise linear vector-valued functions along these tangent curves. Hence, the stable formulas presented here have general applicability, e.g. exact integration of trajectories in particle-based (Lagrangian-type) methods, flow visualization and computer graphics. The Newton form of the polynomial interpolation definition is used to express exponential functions of matrices which appear in the analytical solution of the X-IVAS scheme. The divided difference coefficients in these expressions are defined in a piecewise manner, i.e. in a prescribed neighbourhood of removable singularities their series approximations are computed. An optimal series approximation of divided differences is presented which plays a critical role in this methodology. At least ten significant decimal digits in the formula computations are guaranteed to be exact using double-precision floating-point arithmetic. The worst case scenarios occur in the neighbourhood of removable singularities found in fourth-order divided differences of the exponential function.

Key words. X-IVAS scheme, particle finite element method, explicit exponential integrators, tangent curves, closed-form analytical solutions, finite arithmetic, loss of significance, numerically stable formulas

AMS subject classifications. Primary: 65-04; Secondary: 34A05, 39-04, 65G30, 70B05, 76M10, 76M28

1. Introduction.

The particle finite element method [11, 16] (PFEM) is a versatile numerical method in which each fluid particle is followed in a Lagrangian manner. It is shown to successfully simulate a wide variety of complex engineering problems, e.g. free-surface/multi-fluid flows with violent interface motions, polymer melting and burning simulations, multi-fluid mixing and buoyancy driven segregation problems, etc. A recent development within the framework of the PFEM is the X-IVAS (eXplicit Integration along the Velocity and Acceleration Streamlines) scheme [10]. Its development was motivated in the need for a faster and more accurate time integrator for incompressible flows.

The X-IVAS scheme targets the explicit time integration of the kinematics of the fluid particles using large time steps. The equations of motion are obtained from the incompressible Navier–Stokes equations subjected to a second-order pressure segregation method. In the Lagrangian formulation the segregated momentum balance equations define the acceleration of the fluid particles. The adopted hypothesis is that a streamline in the configuration at the start of the time step (reference configuration) is a good approximation to the pathline of the fluid particle for a relatively large time step. For a chosen time step each fluid particle is advected along the streamline passing through its position in the reference configuration. The particle velocity is updated for this time step by doing a line integral of the acceleration existing in the reference configuration along the streamline. It is expected that the explicit time integration of the particle position and velocity along the streamline yields a better and more stable approximation than doing so via standard finite difference time integrators.

The X-IVAS hypothesis has been tested [10, 9] by successfully simulating some benchmark CFD and FSI examples using very large time steps, e.g. 10–15 times the standard Courant–Friedrichs–Lewy stability limit. Further, the simulation results of multi-fluid flows

* This study was partially supported by the SAFECON project of the European Research Council (European Commission) and the WAM-V project funded under the Navy Grant N62909-12-1-7101 issued by Office of Naval Research Global. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein.

† Centre Internacional de Mètodos Numèrics en Enginyeria (CIMNE), Edifici C1, Gran Capitan s/n, 08034 Barcelona, Spain (npras@cimne.upc.edu).

were compared [9] with those obtained using OpenFOAM [5] and for similar accuracy the PFEM+X-IVAS method took nearly half the simulation time taken by OpenFOAM.

In the PFEM+X-IVAS method the data is stored in a discrete manner, i.e. all the essential variables are stored as intrinsic properties with the particles. To perform the explicit time integration we need a spatially continuous description of the velocity and acceleration. For this purpose a fixed auxiliary simplicial mesh is used to interpolate the variables of interest in a piecewise linear manner. Thus, the integration of the position and velocity of the fluid particles is actually driven by *approximate piecewise linear vector fields*. Nevertheless, the higher-order fine-scale details of the initial solution and of the integration results are retained with the particles. This is essential for the accuracy of the scheme and avoids the so-called erosion artefact associated to the Eulerian formulations.

Despite the possibility to compute analytically the positions and velocities of the particles, numerical sub-stepping methods based on simple finite-difference schemes (e.g. forward Euler) were used for this purpose in the original proposal and in the subsequent developments until now. On the one hand, this introduces an additional source of numerical error and the repercussions of the same are not well understood. One has to carefully choose an appropriate sub-time-step for stability. On the other hand, numerical sub-stepping might weaken the X-IVAS hypothesis. For instance, a situation originally used to motivate the X-IVAS hypothesis is that the streamlines of a flow never cross fixed impermeable domain boundaries (e.g. vortices near corners of wall bounded flows). The X-IVAS hypothesis guarantees that the material points never leave impermeable domain boundaries. Numerical sub-stepping procedures clearly compromise this guarantee.

Diachin and Herzog reported [3] that analytical solvers provide faster more accurate results for streamline calculations on a linear tetrahedra than the forward Euler and the fourth-order Runge–Kutta methods. The matrix functions were computed therein using a procedure based on matrix decomposition methods. The singular value decomposition was used to determine the matrix rank which in turn was used to classify the calculation procedure into four cases in 3D. Using a Schur decomposition the matrix functions were transformed to equivalent functions of upper triangular matrices. The calculation of the latter was done using a recursive relation proposed by Parlett [18]. Nielson and Jung presented [15] formulas in 2D and 3D to compute the closed-form analytical solution of tangent curves for linearly varying vector fields over tetrahedral domains. We infer from the algebraic structure of these formulas that the matrix functions were expressed therein using the Lagrange form of the polynomial interpolation definition. The analytical solution of the tangent curves were classified into five cases in 2D and nine cases in 3D which depend on the eigenvalues of the system matrix.

Idelsohn et al. presented [10] a procedure to compute the analytical solution of the X-IVAS scheme in 2D. The matrix functions are expressed therein using the Jordan canonical form definition. The solution in 3D was omitted pointing out that the extension to 3D is straightforward. Unlike in 2D where the procedure to express matrices in the Jordan canonical form is straightforward, in 3D (and for matrices of larger dimensions) this procedure is arduous as repeated eigenvalues with different Jordan blocks might exist. As the Jordan structure of a 3D matrix involves multiple cases, we beg to differ with Idelsohn et al. [10] that it is not trivial to derive and implement (code) their approach to compute the analytical solution of the X-IVAS scheme in 3D.

Moreover, using finite precision arithmetic the *as is* computation of the analytical solution procedure of Idelsohn et al. [10] and the formulas of Nielson and Jung [15] are conditionally stable. Errors creep into the computations in the neighbourhood of removable singularities where subtractive cancellations in finite precision arithmetic are brought to prominence. This leads to a *gradual* loss of significant digits (errors gradually build up) in the compu-

tations as we approach the points of removable singularities. In the absence of numerically stable formulas/computational techniques, the accuracy promise of analytical solvers is lost and what is worse, the loss of significance will go unnoticed or misdiagnosed. In the aforesaid papers [10, 15] we find expressions with removable singularities, e.g. $[\exp(\lambda t)-1]/\lambda$, which are typical examples used to demonstrate [12, 6] loss of significance in finite precision arithmetic.

The conditional stability issue also extends to the analytical solution procedure presented by Diachin and Herzog [3]. Although matrix decompositions are robust/stable with respect to rounding errors, the recursive algorithm used to compute the exponential function of block upper triangular matrices breaks down in certain situations. Moreover Parlett [18, p. 199] warned that implementations in finite precision arithmetic could be expected to give inaccurate results in particular situations. The message is clear: irrespective of the choice of the solution procedure, issues related to numerical instability exist and they need to be addressed.

Further, once such instabilities are identified, it is often not trivial to localize the terms in these formulas that participate to obtain a finite limit at removable singularities. Identifying such terms is crucial to control numerical instabilities and bound the loss of significant digits. We discuss these issues here and present algebraically equivalent yet numerically stable formulas for the X-IVAS scheme in 2D and 3D.

2. Preliminaries. Here we describe briefly the convention used in the description of the flow. The independent variables in Lagrangian kinematics are (χ,t), where χ represents a label to identify particles (material points) and t represents the time elapsed after labeling. The primary dependent variable is the fluid particle trajectory denoted as $\boldsymbol{X}(\chi,t)$. The initial particle positions denoted by $\boldsymbol{X}^0 := \boldsymbol{X}(\chi,0)$ are assumed to be given. A natural choice for the label χ is the ordered triple $\boldsymbol{X}^0$. The Lagrangian velocity and acceleration, denoted as $\dot{\boldsymbol{X}}(\chi,t)$ and $\ddot{\boldsymbol{X}}(\chi,t)$, respectively are defined as follows.

$$\dot{\boldsymbol{X}}(\chi,t) := \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}(\chi,t), \quad \ddot{\boldsymbol{X}}(\chi,t) := \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{X}(\chi,t) \tag{2.1}$$

On the other hand, the independent variables in Eulerian kinematics are $(\boldsymbol{x},t)$. Here $\boldsymbol{x}$ denotes the spatial coordinate. The primary dependent variable is the fluid velocity $\boldsymbol{u}(\boldsymbol{x},t)$. The so-called fundamental principle of kinematics [19] states that the velocity $\boldsymbol{u}(\boldsymbol{x},t)$ and acceleration $\boldsymbol{a}(\boldsymbol{x},t)$ at a given time t and fixed position $\boldsymbol{x}$ (Eulerian description) is equal to the velocity $\dot{\boldsymbol{X}}(\chi,t)$ and acceleration $\ddot{\boldsymbol{X}}(\chi,t)$ of a particle that is present at that position and at that instant (Lagrangian description). Thus,

$$\boldsymbol{u}(\boldsymbol{x},t) = \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}(\chi,t)\bigg|_{\boldsymbol{X}(\chi,t)=\boldsymbol{x}}, \quad \boldsymbol{a}(\boldsymbol{x},t) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{X}(\chi,t)\bigg|_{\boldsymbol{X}(\chi,t)=\boldsymbol{x}} \tag{2.2}$$

As a corollary we have the following exact but implicit equations of particle motion.

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}(\chi,t) = \boldsymbol{u}(\boldsymbol{X}(\chi,t),t), \quad \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{X}(\chi,t) = \frac{\mathrm{d}}{\mathrm{d}t}\dot{\boldsymbol{X}}(\chi,t) = \boldsymbol{a}(\boldsymbol{X}(\chi,t),t) \tag{2.3}$$

3. The X-IVAS scheme.

3.1. Introduction. The explicit time integration of the particle position and velocity along the streamline results in the following equations of motion.

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}(\chi,t) = \boldsymbol{u}(\boldsymbol{X}(\chi,t),t^n), \quad \frac{\mathrm{d}}{\mathrm{d}t}\dot{\boldsymbol{X}}(\chi,t) = \boldsymbol{a}(\boldsymbol{X}(\chi,t),t^n) \tag{3.1}$$

Recall that the data corresponding to the dependent variables is stored with the particles which form a sufficiently large yet finite set. In other words, the data at any given time is available as discrete samples at the spatial locations occupied by the particles. Data interpolation is inevitable to have spatially continuous vector fields and to solve for the particle motion. Hence in the equations of motion $\boldsymbol{u}(\boldsymbol{x},t^n)$ and $\boldsymbol{a}(\boldsymbol{x},t^n)$, which are unknown for an arbitrary $\boldsymbol{x}$ are replaced by the interpolated counterparts $\boldsymbol{u}^h(\boldsymbol{x},t^n)$ and $\boldsymbol{a}^h(\boldsymbol{x},t^n)$, respectively. The superscript h represents the discretization size associated to the interpolation. It follows that the trajectory obtained from these interpolated vector fields needs to be represented as $\boldsymbol{X}^h(\chi,t)$.

In the following we describe the X-IVAS scheme to integrate the equations of particle motion from time t^n to t^{n+1} as a four step process.

Step 1: Projection. This step involves the projection of vector fields stored with the particles onto a simplicial mesh. Consider a simplicial mesh over the problem domain and a set of characteristic domains corresponding to every node of the mesh. Let $\mathscr{P}^i$ be an operator that projects data onto a mesh node with index i from a set of sample points in the corresponding characteristic domain. Using this projection operator we calculate the velocity $\overline{\boldsymbol{u}}^i(t^n)$ and acceleration $\overline{\boldsymbol{a}}^i(t^n)$ vector fields at the mesh nodes as follows.

$$\overline{\boldsymbol{u}}^i(t^n) := \mathscr{P}^i[\dot{\boldsymbol{X}}(\chi,t^n)], \quad \overline{\boldsymbol{a}}^i(t^n) := \mathscr{P}^i[\ddot{\boldsymbol{X}}(\chi,t^n)] \tag{3.2}$$

This projection step is unnecessary when $t^n = 0$ where we can obtain $\overline{\boldsymbol{u}}^i(0)$ and $\overline{\boldsymbol{a}}^i(0)$ directly from the prescribed initial conditions.

Step 2: Interpolation. In this step we do a piecewise linear interpolation of vector fields projected onto the mesh nodes. Using the velocity $\overline{\boldsymbol{u}}^i(t^n)$ and acceleration $\overline{\boldsymbol{a}}^i(t^n)$ vector fields at the mesh nodes we construct a piecewise linear interpolation of these vector fields as follows.

$$\boldsymbol{u}^h(\boldsymbol{x},t^n) := \mathrm{N}^i(\boldsymbol{x})\overline{\boldsymbol{u}}^i(t^n), \quad \boldsymbol{a}^h(\boldsymbol{x},t^n) := \mathrm{N}^i(\boldsymbol{x})\overline{\boldsymbol{a}}^i(t^n) \tag{3.3}$$

In the above equation $\mathrm{N}^i(\boldsymbol{x})$ represents the piecewise linear shape function corresponding to the node i. Let $\boldsymbol{x}^j$ denote the spatial coordinate of node j, $\langle\,\rangle_j$ denote the average operator over the index j and δ^{ij} denote the Kronecker delta. For a given simplex, we can express $\mathrm{N}^i(\boldsymbol{x})$ in terms of its gradient $\boldsymbol{\nabla}\mathrm{N}^i$ (which is constant within the simplex) and the spatial coordinate $\boldsymbol{x}$ as follows.

$$\mathrm{N}^i(\boldsymbol{x}) := \boldsymbol{\nabla}\mathrm{N}^i\cdot(\boldsymbol{x}-\langle\boldsymbol{x}^j\rangle_j) + \frac{1^i}{\delta^{kk}} \tag{3.4}$$

Using the above equation $\boldsymbol{u}^h(\boldsymbol{x},t^n)$ and $\boldsymbol{a}^h(\boldsymbol{x},t^n)$ can be expressed within each simplex as follows.

$$\boldsymbol{u}^h(\boldsymbol{x},t^n) = [\overline{\boldsymbol{u}}^i(t^n)\otimes\boldsymbol{\nabla}\mathrm{N}^i]\cdot(\boldsymbol{x}-\langle\boldsymbol{x}^j\rangle_j) + \langle\overline{\boldsymbol{u}}^j(t^n)\rangle_j = \mathbf{A}^n\cdot\boldsymbol{x}+\mathbf{b}^n \tag{3.5}$$

$$\boldsymbol{a}^h(\boldsymbol{x},t^n) = [\overline{\boldsymbol{a}}^i(t^n)\otimes\boldsymbol{\nabla}\mathrm{N}^i]\cdot(\boldsymbol{x}-\langle\boldsymbol{x}^j\rangle_j) + \langle\overline{\boldsymbol{a}}^j(t^n)\rangle_j = \mathbf{C}^n\cdot\boldsymbol{x}+\mathbf{d}^n \tag{3.6}$$

Here $\otimes$ denotes the tensor product. Further, $\mathbf{A}^n,\mathbf{b}^n,\mathbf{C}^n$ and $\mathbf{d}^n$ are constant tensors evaluated for each simplex at time t^n and are defined as follows.

$$\mathbf{A}^n := [\overline{\boldsymbol{u}}^i(t^n)\otimes\boldsymbol{\nabla}\mathrm{N}^i], \quad \mathbf{b}^n := \langle\overline{\boldsymbol{u}}^i(t^n)\rangle_i - \mathbf{A}^n\cdot\langle\boldsymbol{x}^i\rangle_i \tag{3.7}$$

$$\mathbf{C}^n := [\overline{\boldsymbol{a}}^i(t^n)\otimes\boldsymbol{\nabla}\mathrm{N}^i], \quad \mathbf{d}^n := \langle\overline{\boldsymbol{a}}^i(t^n)\rangle_i - \mathbf{C}^n\cdot\langle\boldsymbol{x}^i\rangle_i \tag{3.8}$$

Step 3: Integration. Here we describe the time integration of the approximate equations of particle motion. The approximate equations of motion for the particles in the X-IVAS scheme can be written as follows.

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}^h(\chi,t) = \boldsymbol{u}^h(\boldsymbol{X}^h(\chi,t),t^n), \quad \frac{\mathrm{d}}{\mathrm{d}t}\dot{\boldsymbol{X}}^h(\chi,t) = \boldsymbol{a}^h(\boldsymbol{X}^h(\chi,t),t^n) \tag{3.9}$$

Recall that the above equations are expressed in a piecewise manner as both $\boldsymbol{u}^h$ and $\boldsymbol{a}^h$ are defined in this manner. To be precise, within each simplex the particle motion is driven by the following equations which vary from one simplex to another.

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}^h(\chi,t) = \mathbf{A}^n \cdot \boldsymbol{X}^h(\chi,t) + \mathbf{b}^n, \quad \frac{\mathrm{d}}{\mathrm{d}t}\dot{\boldsymbol{X}}^h(\chi,t) = \mathbf{C}^n \cdot \boldsymbol{X}^h(\chi,t) + \mathbf{d}^n \tag{3.10}$$

Likewise, the time integration of these equations should also be done in a piecewise manner. In other words, if a particle tends to exit the current simplex prior to the end of the time step, its subsequent motion is driven by the equations written for the simplex in which it tends to enter and so forth until the end of the time step.

Further, certain relationships that existed between the dependent variables in the exact equations of motions no longer hold for the corresponding variables in the approximate equations of motion. That is,

$$\dot{\boldsymbol{X}}^h(\chi,t) \neq \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{X}^h(\chi,t), \quad \ddot{\boldsymbol{X}}^h(\chi,t) := \frac{\mathrm{d}}{\mathrm{d}t}\dot{\boldsymbol{X}}^h(\chi,t) \neq \frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{X}(\chi,t), \tag{3.11}$$

Nevertheless, the X-IVAS scheme is consistent in the sense that these relations are recovered as the discretization size $h \to 0$. The analytical solution to the pair of equations given in Eq. (3.10) can be written within each simplex as follows.

$$\boldsymbol{X}^h(\chi,t) = \mathrm{e}^{(t-t^n)\mathbf{A}^n} \cdot \boldsymbol{X}^h(\chi,t^n) + \left[\int_{t^n}^{t} \mathrm{e}^{(t-\tau)\mathbf{A}^n}\,\mathrm{d}\tau\right] \cdot \mathbf{b}^n \tag{3.12}$$

$$\dot{\boldsymbol{X}}^h(\chi,t) = \dot{\boldsymbol{X}}^h(\chi,t^n) + \mathbf{C}^n \cdot \left[\int_{t^n}^{t} \boldsymbol{X}^h(\chi,\tau)\,\mathrm{d}\tau\right] + (t-t^n)\mathbf{d}^n \tag{3.13}$$

Note that the particle motion is restricted to the tangent curve of $\boldsymbol{u}^h(\boldsymbol{x},t^n)$ (i.e. the streamline) on which it was located at time t^n and is accelerated along this curve up to time t^{n+1}. Note that the solution for the particle velocity $\dot{\boldsymbol{X}}^h(\chi,t)$ is given as a line integral along the curve $\boldsymbol{X}^h(\chi,t)$. This integral is left here *as is* for compactness and its evaluated form will be given in the following section.

Equations 3.12 and 3.13 justify the classification of the X-IVAS scheme as a *particle-based explicit exponential integrator*. Unlike classical exponential integrators [8, 17, 2] which integrate a global system of equations, the X-IVAS scheme integrates analytically a small and fixed-size local system of equations for the particles. This is an innovative approach that combines concepts of exponential integrators with the particle percept. On the one hand, it inherits the stability properties of exponential integrators which allow us to choose larger time steps. On the other hand, the associated algebra is computationally intensive, i.e. it is not memory bound. This conceptual setting is ideal for parallel computation which is an important strategy for faster simulations.

Step 4: Update. In this step we update the dependent variables at time t^{n+1} and repeat the process. At the end of the time step we obtain $\boldsymbol{X}^h(\chi,t^{n+1})$ and $\dot{\boldsymbol{X}}^h(\chi,t^{n+1})$ which are governed by the kinematics of the flow. The state of $\ddot{\boldsymbol{X}}^h(\chi,t^{n+1})$ is governed by the dynamics of the internal and the external force terms that appear in the momentum balance equation of the flow.

3.2. Remarks on the analytical solution. In this section we simplify the analytical solution given in Eq. (3.12) and Eq. (3.13) and identify relationships among the terms that appear therein, if any. Consider three matrices $\mathbf{P},\mathbf{Q}$ and $\mathbf{R}$ which in turn are defined as functions of a given matrix $\mathbf{A}$ and a scalar τ as follows.

$$\mathbf{P}(\tau,\mathbf{A}) := \mathrm{e}^{\tau\mathbf{A}}, \quad \mathbf{Q}(\tau,\mathbf{A}) := \int_0^\tau \mathrm{e}^{\xi\mathbf{A}}\,\mathrm{d}\xi, \quad \mathbf{R}(\tau,\mathbf{A}) := \int_0^\tau\int_0^\eta \mathrm{e}^{\xi\mathbf{A}}\,\mathrm{d}\xi\,\mathrm{d}\eta \tag{3.14}$$

As it can be seen from the above equation, the considered matrices are exponential functions of the given matrix $\mathbf{A}$. The matrix $\mathbf{P}$ is usually called the propagator [4]. The following relationships can be identified between the matrices $\mathbf{P}$ and $\mathbf{Q}$.

$$\mathbf{Q}(\tau,\mathbf{A}) = \int_0^\tau \mathbf{P}(\xi,\mathbf{A})\,\mathrm{d}\xi = \left[\mathrm{e}^{\tau\mathbf{A}} - \mathbf{I}\right]\cdot\mathrm{inv}(\mathbf{A}) = [\mathbf{P}(\tau,\mathbf{A}) - \mathbf{I}]\cdot\mathrm{inv}(\mathbf{A}) \tag{3.15}$$

$$\Rightarrow \quad \mathbf{P}(\tau,\mathbf{A}) = \mathbf{Q}(\tau,\mathbf{A})\cdot\mathbf{A} + \mathbf{I} \tag{3.16}$$

Likewise, the matrices $\mathbf{Q}$ and $\mathbf{R}$ satisfy the following relationships.

$$\mathbf{R}(\tau,\mathbf{A}) = \int_0^\tau \mathbf{Q}(\xi,\mathbf{A})\,\mathrm{d}\xi = \left[\left(\mathrm{e}^{\tau\mathbf{A}} - \mathbf{I}\right)\cdot\mathrm{inv}(\mathbf{A}) - \tau\mathbf{I}\right]\cdot\mathrm{inv}(\mathbf{A}) \tag{3.17}$$

$$\mathbf{R}(\tau,\mathbf{A}) = [\mathbf{Q}(\tau,\mathbf{A}) - \tau\mathbf{I}]\cdot\mathrm{inv}(\mathbf{A}), \quad \Rightarrow \quad \mathbf{Q}(\tau,\mathbf{A}) = \mathbf{R}(\tau,\mathbf{A})\cdot\mathbf{A} + \tau\mathbf{I} \tag{3.18}$$

In the above equations $\mathrm{inv}(\mathbf{A})$ denotes the matrix inverse of $\mathbf{A}$. Further, the products involving $\mathrm{inv}(\mathbf{A})$ and $\mathbf{A}$ in these equations are commutative, i.e. the order in which they appear are irrelevant. Using these definitions, we can express the analytical solution of the equations of motion in the X-IVAS scheme as follows.

$$\boldsymbol{\chi}^h(\chi,t) = \mathbf{P}(t-t^n,\mathbf{A}^n)\cdot\boldsymbol{\chi}^h(\chi,t^n) + \mathbf{Q}(t-t^n,\mathbf{A}^n)\cdot\mathbf{b}^n \tag{3.19}$$

$$\begin{aligned}\dot{\boldsymbol{\chi}}^h(\chi,t) = \dot{\boldsymbol{\chi}}^h(\chi,t^n) + (t-t^n)\mathbf{d}^n \\ + \mathbf{C}^n\cdot[\mathbf{Q}(t-t^n,\mathbf{A}^n)\cdot\boldsymbol{\chi}^h(\chi,t^n) + \mathbf{R}(t-t^n,\mathbf{A}^n)\cdot\mathbf{b}^n]\end{aligned} \tag{3.20}$$

Recall that a nodal projection of the data carried by the particles onto the background mesh is done after every time step and the tensors $\mathbf{A}^n,\mathbf{b}^n,\mathbf{C}^n$ and $\mathbf{d}^n$ have to be recalculated for each element using the projected data. It follows that the matrices $\mathbf{P},\mathbf{Q}$ and $\mathbf{R}$ also need to be recalculated for each element after every time step.

Note that we need the exit points of the particles on the simplex boundary to perform the piecewise integration of particle motions described earlier in the paragraph following Eq. (3.10). To find the exit points we need to solve the intersection of their trajectories with the simplex boundary. A procedure to solve for the exit points using Newton linearisation was presented by Kipfer et al. [13]. In this procedure the matrices $\mathbf{P}$ and $\mathbf{Q}$ have to be evaluated at every iteration as the time increments need not be uniform. Should one decide to use *analytical* sub-stepping procedures to arrive at the exit point and if a constant *sub* time step is used for all the particles throughout the sub-stepping procedure, then we need to compute these matrices for each element just once. The incremental method [15] for computing tangent curves and the analytical time stepping algorithm called ANTS [3] are based on this idea.

4. Functions of matrices.

4.1. Introduction. Functions of matrices can be defined in various yet equivalent ways, a comprehensive presentation of which is made by Higham [7]. Let the size of $\mathbf{A}$ be $n\times n$

and assume that a given scalar function $\mathrm{f}(\lambda)$ takes well-defined values (including values associated with derivatives where appropriate) at the eigenvalues of $\mathbf{A}$ denoted by the sequence $Z := \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. We will use the Newton form of the polynomial interpolation definition for $\mathrm{f}(\mathbf{A})$. The coefficients in the Newton form of the interpolating polynomial have the algebraic structure of divided differences.

We denote the kth divided difference of $\mathrm{f}(\lambda)$ on the subsequence $Z_i^k := \{\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+k}\}$ as $\mathrm{f}[\lambda_i; \lambda_{i+1}; \ldots; \lambda_{i+k}]$ and define it using the following recurrence equations.

$$\mathrm{f}[\lambda_i] := \mathrm{f}(\lambda_i) \tag{4.1}$$

$$\mathrm{f}[\lambda_i; \lambda_{i+1}; \ldots; \lambda_{i+k}] := \frac{\mathrm{f}[\lambda_{i+1}; \lambda_{i+2}; \ldots; \lambda_{i+k}] - \mathrm{f}[\lambda_i; \lambda_{i+1}; \ldots; \lambda_{i+k-1}]}{\lambda_{i+k} - \lambda_i} \tag{4.2}$$

$$\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+k} \Rightarrow \mathrm{f}[\lambda_i; \lambda_{i+1}; \ldots; \lambda_{i+k}] := \frac{1}{k!} \frac{\partial^k}{\partial \lambda^k} \mathrm{f}(\lambda) \bigg|_{\lambda=\lambda_i} \tag{4.3}$$

It is a well-known result that the value of $\mathrm{f}[\lambda_i; \lambda_{i+1}; \ldots; \lambda_{i+k}]$ does not depend on the order of $\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+k}$ in Z_i^k. The Newton form of the polynomial interpolation definition for $\mathrm{f}(\mathbf{A})$ is

$$\mathrm{f}(\mathbf{A}) = \mathrm{f}[\lambda_1]\mathbf{I} + \sum_{k=1}^{n-1} \mathrm{f}[\lambda_1; \lambda_2; \ldots; \lambda_{1+k}](\mathbf{A} - \lambda_1\mathbf{I})(\mathbf{A} - \lambda_2\mathbf{I}) \cdots (\mathbf{A} - \lambda_k\mathbf{I}) \tag{4.4}$$

The computation of *nearly confluent* divided differences are known to suffer from subtractive cancellations in floating point arithmetic. Thus, this definition gives *a priori* warning about the *gradual* loss of significance in the computation of $\mathrm{f}(\mathbf{A})$ in the neighbourhood of removable singularities. By grouping terms prone to loss of significance (as divided differences) it also paves way to systematically design procedures for the stable computation of $\mathrm{f}(\mathbf{A})$. Moreover this definition is independent of the Jordan structure of $\mathbf{A}$ [7, p. 6] which makes it convenient to implement in a computer program.

It is possible to write the kth divided difference $\mathrm{f}[\lambda_1; \lambda_2; \ldots; \lambda_{1+k}]$ as follows.

$$\mathrm{f}[\lambda_1; \lambda_2; \ldots; \lambda_{1+k}] = \sum_{i=1}^{1+k} \frac{\mathrm{f}(\lambda_i)}{\prod_{j \neq i}(\lambda_i - \lambda_j)}, \quad j \in \{1, 2, \ldots, 1+k\} \tag{4.5}$$

Using the above identity we can transform the formulas given in this article to compute tangent curves to the ones presented by Nielson and Jung [15]. Algebraic rearrangements using this identity do not avoid existing issues related to loss of significance and makes matters worse by obscuring them. The example in §5.1 drives this point home.

To express the closed-form solution of the X-IVAS scheme we need to consider the cases where $n = 2$ (2D) and $n = 3$ (3D). For $n = 3$, we can express $\mathrm{f}(\mathbf{A})$ as follows.

$$\mathrm{f}(\mathbf{A}) = \mathrm{f}(\lambda_1)\mathbf{I} + \mathrm{f}[\lambda_1; \lambda_2](\mathbf{A} - \lambda_1\mathbf{I}) + \mathrm{f}[\lambda_1; \lambda_2; \lambda_3](\mathbf{A} - \lambda_1\mathbf{I})(\mathbf{A} - \lambda_2\mathbf{I}) \tag{4.6}$$

Without loss of generality we assume that the eigenvalue λ_3 is a real number and the eigenvalues λ_1 and λ_2 might be complex numbers. Complex eigenvalues will always occur in conjugate pairs, i.e. $\{\lambda_1, \lambda_2\} = \{\lambda_\mathrm{c}, \lambda_\mathrm{c}^*\}$. The subscript c indicates that it is a complex number and the superscript $*$ indicates that it is a complex conjugate.

Although Eq. (4.6) holds for all eigenvalues, this form is convenient to implement in a computer program when the eigenvalues are real numbers. In the case of complex eigenvalues

Eq. (4.6) can be simplified to evaluate $\mathrm{f}(\mathbf{A})$ as follows.

$$\begin{aligned}\mathrm{f}(\mathbf{A}) = {} & \frac{\mathrm{Im}[f^*(\lambda_c)\lambda_c]}{\mathrm{Im}[\lambda_c]}\mathbf{I} + \frac{\mathrm{Im}[\mathrm{f}(\lambda_c)]}{\mathrm{Im}[\lambda_c]}\mathbf{A} \\ & + \left[\frac{\mathrm{f}(\lambda_3)\,\mathrm{Im}(\lambda_c) - \lambda_3\,\mathrm{Im}[\mathrm{f}(\lambda_c)] - \mathrm{Im}[f^*(\lambda_c)\lambda_c]}{\mathrm{Im}[\lambda_c]}\right]\left[\frac{\mathbf{A}^2 - 2\,\mathrm{Re}(\lambda_c)\mathbf{A} + |\lambda_c|^2\mathbf{I}}{\lambda_3^2 - 2\,\mathrm{Re}(\lambda_c)\lambda_3 + |\lambda_c|^2}\right]\end{aligned} \tag{4.7}$$

Here, the functions $\mathrm{Re}(\lambda_c)$ and $\mathrm{Im}(\lambda_c)$ return the real and imaginary parts of a complex argument λ_c, respectively.

4.2. Formulas for exponential functions of 3×3 matrices. In this section we consider the case when $\mathrm{f}(\lambda) := \exp(\tau\lambda)$ and write the expressions for the matrices $\mathbf{P}(\tau,\mathbf{A})$, $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ which were defined earlier in Eq. (3.14). It is straightforward to verify the following results.

$$\int_0^\tau \mathrm{e}^{\xi\lambda}\,\mathrm{d}\xi = \frac{\mathrm{e}^{\tau\lambda}-1}{\lambda} = \tau\exp[0;\tau\lambda] \tag{4.8}$$

$$\int_0^\tau\int_0^\eta \mathrm{e}^{\xi\lambda}\,\mathrm{d}\xi\,\mathrm{d}\eta = \frac{\mathrm{e}^{\tau\lambda}-1-\tau\lambda}{\lambda^2} = \tau^2\exp[0;0;\tau\lambda] \tag{4.9}$$

Following this line we define two auxiliary functions $\mathrm{q}(x)$ and $\mathrm{r}(x)$ which are divided differences of the exponential function.

$$\mathrm{q}(x) := \exp[0;x] = \begin{cases}\dfrac{\mathrm{e}^x-1}{x} & \text{if } x\neq 0,\\ 1 & \text{if } x=0.\end{cases} \tag{4.10}$$

$$\mathrm{r}(x) := \exp[0;0;x] = \mathrm{q}[0;x] = \begin{cases}\dfrac{\mathrm{e}^x-1-x}{x^2} & \text{if } x\neq 0,\\ \dfrac{1}{2} & \text{if } x=0.\end{cases} \tag{4.11}$$

Using these auxiliary functions we can express $\mathbf{P}(\tau,\mathbf{A}),\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ as follows.

(4.12)
$$\mathbf{P}(\tau,\mathbf{A}) = \mathrm{e}^{\tau\lambda_1}\mathbf{I} + \tau\exp[\tau\lambda_1;\tau\lambda_2](\mathbf{A}-\lambda_1\mathbf{I}) + \tau^2\exp[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3](\mathbf{A}-\lambda_1\mathbf{I})(\mathbf{A}-\lambda_2\mathbf{I})$$

(4.13)
$$\mathbf{Q}(\tau,\mathbf{A}) = \tau\,\mathrm{q}(\tau\lambda_1)\mathbf{I} + \tau^2\,\mathrm{q}[\tau\lambda_1;\tau\lambda_2](\mathbf{A}-\lambda_1\mathbf{I}) + \tau^3\,\mathrm{q}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3](\mathbf{A}-\lambda_1\mathbf{I})(\mathbf{A}-\lambda_2\mathbf{I})$$

(4.14)
$$\mathbf{R}(\tau,\mathbf{A}) = \tau^2\,\mathrm{r}(\tau\lambda_1)\mathbf{I} + \tau^3\,\mathrm{r}[\tau\lambda_1;\tau\lambda_2](\mathbf{A}-\lambda_1\mathbf{I}) + \tau^4\,\mathrm{r}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3](\mathbf{A}-\lambda_1\mathbf{I})(\mathbf{A}-\lambda_2\mathbf{I})$$

The expression for $\mathbf{P}(\tau,\mathbf{A})$ is a trivial specialization of Eq. (4.6) for $\mathrm{f}(\lambda) := \exp(\tau\lambda)$. We obtain $\mathbf{Q}(\tau,\mathbf{A})$ by integrating the terms in $\mathbf{P}(\xi,\mathbf{A})$ with respect to ξ; cf. Eq. (3.15). Likewise, $\mathbf{R}(\tau,\mathbf{A})$ is obtained by integrating the terms in $\mathbf{Q}(\xi,\mathbf{A})$ with respect to ξ; cf. Eq. (3.17). We

have used the following results to arrive at these equations.

$$\int_0^\tau \xi \exp[\xi\lambda_1;\xi\lambda_2]\,\mathrm{d}\xi = \tau\frac{\exp[0;\tau\lambda_2]-\exp[0;\tau\lambda_1]}{\lambda_2-\lambda_1} = \tau^2\,\mathrm{q}[\tau\lambda_1;\tau\lambda_2] \tag{4.15}$$

$$\begin{aligned}\int_0^\tau \xi^2 \exp[\xi\lambda_1;\xi\lambda_2;\xi\lambda_3]\,\mathrm{d}\xi &= \int_0^\tau \xi\frac{\exp[\xi\lambda_2;\xi\lambda_3]-\exp[\xi\lambda_1;\xi\lambda_2]}{\lambda_3-\lambda_1}\,\mathrm{d}\xi\\ &= \tau^3\,\mathrm{q}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3]\end{aligned} \tag{4.16}$$

$$\int_0^\tau\int_0^\eta \xi \exp[\xi\lambda_1;\xi\lambda_2]\,\mathrm{d}\xi\,\mathrm{d}\eta = \tau^2\frac{\exp[0;0;\tau\lambda_2]-\exp[0;0;\tau\lambda_1]}{\lambda_2-\lambda_1} = \tau^3\,\mathrm{r}[\tau\lambda_1;\tau\lambda_2] \tag{4.17}$$

(4.18)

$$\begin{aligned}\int_0^\tau\int_0^\eta \xi^2 \exp[\xi\lambda_1;\xi\lambda_2;\xi\lambda_3]\,\mathrm{d}\xi\,\mathrm{d}\eta &= \int_0^\tau\int_0^\eta \xi\frac{\exp[\xi\lambda_2;\xi\lambda_3]-\exp[\xi\lambda_1;\xi\lambda_2]}{\lambda_3-\lambda_1}\,\mathrm{d}\xi\,\mathrm{d}\eta\\ &= \tau^4\,\mathrm{r}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3]\end{aligned}$$

Let α,β be real numbers and consider a complex number λ_c as defined below.

$$\mathrm{i} := \sqrt{-1}, \quad \lambda_\mathrm{c} := \alpha+\mathrm{i}\beta, \quad \Rightarrow \lambda_\mathrm{c}^* = \alpha-\mathrm{i}\beta \tag{4.19}$$

The cardinal sine function $\mathrm{sinc}(x)$ is defined as follows.

$$\mathrm{sinc}(x) := \begin{cases}\dfrac{\sin(x)}{x} & \text{if } x\neq 0,\\ 1 & \text{if } x=0.\end{cases} \tag{4.20}$$

Further, define two auxiliary functions $\Psi(x,y)$ and $\Phi(x,y)$ as follows.

$$\Psi(x,y) := \cos(y) - x\,\mathrm{sinc}(y) \tag{4.21}$$

$$\Phi(x,y) := \exp[-\mathrm{i}y;\mathrm{i}y;x] = \begin{cases}\dfrac{\mathrm{e}^x-\Psi(-x,y)}{x^2+y^2} & \text{if } (x,y)\neq(0,0),\\ \dfrac{1}{2} & \text{if } (x,y)=(0,0).\end{cases} \tag{4.22}$$

In the case of complex eigenvalues, i.e. $\{\lambda_1,\lambda_2\}=\{\lambda_\mathrm{c},\lambda_\mathrm{c}^*\}$ we can evaluate $\mathbf{P}(\tau,\mathbf{A}),\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ as follows.

$$\mathbf{P}(\tau,\mathbf{A}) = \mathrm{e}^{\tau\alpha}\big[\Psi(\tau\alpha,\tau\beta)\mathbf{I}+\tau\,\mathrm{sinc}(\tau\beta)\mathbf{A}+\tau^2\Phi(\tau\lambda_3-\tau\alpha,\tau\beta)[(\mathbf{A}-\alpha\mathbf{I})^2+\beta^2\mathbf{I}]\big] \tag{4.23}$$

$$\begin{aligned}\mathbf{Q}(\tau,\mathbf{A}) = \tau\mathrm{e}^{\tau\alpha}\big[&\mathrm{sinc}(\tau\beta)\mathbf{I}+\tau\Phi(-\tau\alpha,\tau\beta)(\mathbf{A}-2\alpha\mathbf{I})\\ &+\tau^2\Phi(\star,\tau\beta)[-\tau\alpha;\tau\lambda_3-\tau\alpha][(\mathbf{A}-\alpha\mathbf{I})^2+\beta^2\mathbf{I}]\big]\end{aligned} \tag{4.24}$$

(4.25)

$$\begin{aligned}\mathbf{R}(\tau,\mathbf{A}) = \tau^2\mathrm{e}^{\tau\alpha}\big[&\Phi(-\tau\alpha,\tau\beta)\mathbf{I}+\tau\Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha](\mathbf{A}-2\alpha\mathbf{I})\\ &+\tau^2\Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha;\tau\lambda_3-\tau\alpha][(\mathbf{A}-\alpha\mathbf{I})^2+\beta^2\mathbf{I}]\big]\end{aligned}$$

The expression for $\mathbf{P}(\tau,\mathbf{A})$ is a trivial specialization of Eq. (4.7) for $\mathrm{f}(\lambda) := \exp(\tau\lambda)$ and the choice of the auxiliary functions $\Psi(x,y)$ and $\Phi(x,y)$ is motivated by the structure of the same. The notation $\Phi(\star,y)[x_1;x_2]$ means that the divided differences are to be taken with respect to the variable in whose place the symbol $\star$ appears. The rest of this section describes some results which were used to arrive at the expressions for $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ from the expression for $\mathbf{P}(\tau,\mathbf{A})$.

The following integrals are straightforward.

(4.26)
$$\int_0^\tau \mathrm{e}^{\xi\alpha}\cos(\xi\beta)\,\mathrm{d}\xi = \frac{\mathrm{e}^{\tau\alpha}[\alpha\cos(\tau\beta)+\beta\sin(\tau\beta)]-\alpha}{\alpha^2+\beta^2} = \tau\mathrm{e}^{\tau\alpha}[\mathrm{sinc}(\tau\beta)-\tau\alpha\,\Phi(-\tau\alpha,\tau\beta)]$$

$$\int_0^\tau \mathrm{e}^{\xi\alpha}\sin(\xi\beta)\,\mathrm{d}\xi = \frac{\mathrm{e}^{\tau\alpha}[\alpha\sin(\tau\beta)-\beta\cos(\tau\beta)]+\beta}{\alpha^2+\beta^2} = \tau^2\beta\mathrm{e}^{\tau\beta}\,\Phi(-\tau\alpha,\tau\beta) \tag{4.27}$$

$$\Rightarrow \int_0^\tau \mathrm{e}^{\xi\alpha}\,\Psi(\xi\alpha,\xi\beta)\,\mathrm{d}\xi = \tau\mathrm{e}^{\tau\alpha}[\mathrm{sinc}(\tau\beta)-2\tau\alpha\,\Phi(-\tau\alpha,\tau\beta)] \tag{4.28}$$

Using Eqs. (4.27) and (4.28) we obtain the following result.

$$\int_0^\tau \mathrm{e}^{\xi\alpha}\xi^2\,\Phi(\xi\lambda_3-\xi\alpha,\xi\beta)\,\mathrm{d}\xi = \int_0^\tau \frac{\mathrm{e}^{\xi\lambda_3}-\mathrm{e}^{\xi\alpha}[\Psi(\xi\alpha,\xi\beta)+\xi\lambda_3\,\mathrm{sinc}(\xi\beta)]}{(\lambda_3-\alpha)^2+\beta^2}\,\mathrm{d}\xi \tag{4.29}$$

$$= \tau\left[\frac{\mathrm{q}(\tau\lambda_3)-\mathrm{e}^{\tau\alpha}[\mathrm{sinc}(\tau\beta)+\tau(\lambda_3-2\alpha)\,\Phi(-\tau\alpha,\tau\beta)]}{(\lambda_3-\alpha)^2+\beta^2}\right] \tag{4.30}$$

$$= \tau\mathrm{e}^{\tau\alpha}\left[\frac{[\mathrm{e}^{\tau\lambda_3}-1]\mathrm{e}^{-\tau\alpha}-\tau\lambda_3[\mathrm{sinc}(\tau\beta)+\tau(\lambda_3-2\alpha)\,\Phi(-\tau\alpha,\tau\beta)]}{\tau\lambda_3[(\lambda_3-\alpha)^2+\beta^2]}\right] \tag{4.31}$$

$$= \tau\mathrm{e}^{\tau\alpha}\left[\frac{\mathrm{e}^{\tau\lambda_3-\tau\alpha}-\Psi(\tau\alpha-\tau\lambda_3,\tau\beta)-\tau^2[(\lambda_3-\alpha)^2+\beta^2]\,\Phi(-\tau\alpha,\tau\beta)}{\tau\lambda_3[(\lambda_3-\alpha)^2+\beta^2]}\right] \tag{4.32}$$

$$= \tau^3\mathrm{e}^{\tau\alpha}\left[\frac{\Phi(\tau\lambda_3-\tau\alpha,\tau\beta)-\Phi(-\tau\alpha,\tau\beta)}{(\tau\lambda_3-\tau\alpha)-(-\tau\alpha)}\right] = \tau^3\mathrm{e}^{\tau\alpha}\,\Phi(\star,\tau\beta)[-\tau\alpha,\tau\lambda_3-\tau\alpha] \tag{4.33}$$

The results given in Eqs. (4.27), (4.28) and (4.33) are used to obtain $\mathbf{Q}(\tau,\mathbf{A})$ from $\mathbf{P}(\tau,\mathbf{A})$. Substituting $\lambda_3=0$ in Eq. (4.33) we get the following result.

$$\int_0^\tau \mathrm{e}^{\xi\alpha}\xi^2\,\Phi(-\xi\alpha,\xi\beta)\,\mathrm{d}\xi = \tau^3\mathrm{e}^{\tau\alpha}\,\Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha] \tag{4.34}$$

Using Eqs. (4.33) and (4.34) we arrive at the following result.

(4.35)
$$\int_0^\tau \mathrm{e}^{\xi\alpha}\xi^3\,\Phi(\star,\xi\beta)[-\xi\alpha;\xi\lambda_3-\xi\alpha]\,\mathrm{d}\xi = \int_0^\tau \mathrm{e}^{\xi\alpha}\xi^2\,\frac{\Phi(\xi\lambda_3-\xi\alpha,\xi\beta)-\Phi(-\xi\alpha,\xi\beta)}{\lambda_3}\,\mathrm{d}\xi$$

$$= \tau^3\mathrm{e}^{\tau\alpha}\left[\frac{\Phi(\star,\tau\beta)[-\tau\alpha;\tau\lambda_3-\tau\alpha]-\Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha]}{\lambda_3}\right] \tag{4.36}$$

$$= \tau^4\mathrm{e}^{\tau\alpha}\,\Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha;\tau\lambda_3-\tau\alpha] \tag{4.37}$$

The results given in Eqs. (4.27), (4.34) and (4.37) are used to obtain $\mathbf{R}(\tau,\mathbf{A})$ from $\mathbf{Q}(\tau,\mathbf{A})$.

REMARK: Note that the equations for $\mathbf{P}(\tau,\mathbf{A})$, $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ are expressed (3D problems where $n=3$) as the sum of three terms. Due to the properties of the polynomial in the Newton's form, the corresponding equations for $n=2$ (2D problems) can be obtained from the equations for $n=3$ by dropping out the third term.

4.3. Formulas for the eigenvalues of 3×3 matrices. Let $\det(\mathbf{A})$ and $\mathrm{tr}(\mathbf{A})$ denote the determinant and trace of the matrix $\mathbf{A}$, respectively. When $n=3$, the characteristic equation of the matrix $\mathbf{A}$ is given by the following.

$$\det(\mathbf{A}-\lambda\mathbf{I}) = 0 \tag{4.38}$$

$$\Rightarrow \lambda^3 - \mathrm{tr}(\mathbf{A})\lambda^2 + \frac{\mathrm{tr}(\mathbf{A})^2-\mathrm{tr}(\mathbf{A}^2)}{2}\lambda - \det(\mathbf{A}) = 0 \tag{4.39}$$

The solution of the above cubic equation can be found by Cardano's method (see [20]). The calculation steps of the same are summarized below.

$$\mathbf{B} := \mathbf{A} - \frac{\mathrm{tr}(\mathbf{A})}{3}\mathbf{I}, \quad Q := \frac{\mathrm{tr}(\mathbf{B}^2)}{6}, \quad R := \frac{\det(\mathbf{B})}{2} \tag{4.40}$$

$$\lambda_1 = \frac{\mathrm{tr}(\mathbf{A})}{3} + \sqrt[3]{R - \sqrt{R^2 - Q^3}}\, e^{-i(2\pi/3)} + \sqrt[3]{R + \sqrt{R^2 - Q^3}}\, e^{i(2\pi/3)} \tag{4.41}$$

$$\lambda_2 = \frac{\mathrm{tr}(\mathbf{A})}{3} + \sqrt[3]{R - \sqrt{R^2 - Q^3}}\, e^{-i(4\pi/3)} + \sqrt[3]{R + \sqrt{R^2 - Q^3}}\, e^{i(4\pi/3)} \tag{4.42}$$

$$\lambda_3 = \frac{\mathrm{tr}(\mathbf{A})}{3} + \sqrt[3]{R - \sqrt{R^2 - Q^3}} + \sqrt[3]{R + \sqrt{R^2 - Q^3}} \tag{4.43}$$

We follow the convention that the cube roots that appear in the above expressions are real and single valued. The three admissible solutions to the cube root function are already taken into consideration in the above formula.

Note that when the discriminant $(R^2 - Q^3) > 0$, we obtain complex eigenvalues. In this case, the formulas are already in a suitable format for implementation. When $(R^2 - Q^3) \leq 0$ we obtain real eigenvalues and the formulas for the same can be written in a form better suited for implementation as follows.

$$\theta := \arccos\left(\frac{R}{\sqrt{Q^3}}\right), \quad \lambda_n = \frac{\mathrm{tr}(\mathbf{A})}{3} + 2\sqrt{Q}\cos\left(\frac{2\pi n + \theta}{3}\right) \tag{4.44}$$

where arccos() denotes the inverse cosine function whose range is defined to be the closed interval $[0, \pi]$. The formula for the real eigenvalues given in Eq. (4.44) guarantees $\lambda_1 \leq \lambda_2 \leq \lambda_3$. This can be verified using the following results.

$$0 \leq \theta \leq \pi \Rightarrow \begin{array}{l} -1 \leq \cos\left(\frac{2\pi+\theta}{3}\right) \leq \frac{-1}{2} \\ \frac{-1}{2} \leq \cos\left(\frac{4\pi+\theta}{3}\right) \leq \frac{1}{2} \\ \frac{1}{2} \leq \cos\left(\frac{6\pi+\theta}{3}\right) \leq 1 \end{array} \Rightarrow \begin{array}{l} \frac{\mathrm{tr}(\mathbf{A})}{3} - 2\sqrt{Q} \leq \lambda_1 \leq \frac{\mathrm{tr}(\mathbf{A})}{3} - \sqrt{Q} \\ \frac{\mathrm{tr}(\mathbf{A})}{3} - \sqrt{Q} \leq \lambda_2 \leq \frac{\mathrm{tr}(\mathbf{A})}{3} + \sqrt{Q} \\ \frac{\mathrm{tr}(\mathbf{A})}{3} + \sqrt{Q} \leq \lambda_3 \leq \frac{\mathrm{tr}(\mathbf{A})}{3} + 2\sqrt{Q} \end{array} \tag{4.45}$$

Note that in the case of two equal eigenvalues, it will be either $\lambda_1 = \lambda_2$ or $\lambda_2 = \lambda_3$. In all the situations the eigenvalue λ_3 is always a real number.

5. Stable computation of formulas using finite precision.

5.1. Introduction. The issue with stable computation of formulas is best explained by an example. The example consists in the naïve computation of the second-order divided difference $\exp[1; 1+\varepsilon; 1+2\varepsilon]$. We denote by Formula1 the *as is* expression of the second-order divided difference.

$$x_1 = 1, \quad x_2 = 1 + \varepsilon, \quad x_3 = 1 + 2\varepsilon \tag{5.1}$$

$$\exp[x_1; x_2; x_3] = \frac{1}{x_3 - x_1}\left[\frac{e^{x_3} - e^{x_2}}{x_3 - x_2} - \frac{e^{x_2} - e^{x_1}}{x_2 - x_1}\right] \tag{5.2}$$

Using Eq. (4.5) the above equation can be rearranged in an algebraically equivalent form which we denote as Formula2.

$$\exp[x_1; x_2; x_3] = \frac{e^{x_1}}{(x_1 - x_2)(x_1 - x_3)} + \frac{e^{x_2}}{(x_2 - x_1)(x_2 - x_3)} + \frac{e^{x_3}}{(x_3 - x_1)(x_3 - x_2)} \tag{5.3}$$

TABLE 1
Loss of significant digits in the naïve computations of $\exp[1;1+\varepsilon;1+2\varepsilon]$.

ε	Formula1 computation	Exact 16 digits	Formula2 computation
10^{-01}	1.503 335 165 136 320	1.503 335 165 136 325	1.503 335 165 136 292
10^{-02}	1.372 811 947 550 877	1.372 811 947 550 820	1.372 811 947 550 871
10^{-03}	1.360 500 848 424 467	1.360 500 848 315 854	1.360 500 848 386 436
10^{-04}	1.359 276 824 430 971	1.359 276 836 249 607	1.359 276 831 150 054
10^{-05}	1.359 152 790 283 402	1.359 154 505 717 948	1.359 151 840 209 960
10^{-06}	1.359 135 026 857 928	1.359 142 273 371 229	1.359 375
10^{-07}	1.332 267 628 772 320	1.359 141 050 143 621	1.359 375
10^{-08}	2.220 446 084 949 470	1.359 140 927 820 931	4
10^{-09}	0	1.359 140 915 588 663	256
10^{-10}	0	1.359 140 914 365 436	0
10^{-11}	−2 220 445.681 810 107	1.359 140 914 243 114	−4 194 304
10^{-12}	−222 005 130.399 6447	1.359 140 914 230 881	−268 435 456
10^{-13}	0	1.359 140 914 229 658	17 179 869 184
10^{-14}	2 223 999 815 985.422	1.359 140 914 229 536	4 398 046 511 104
10^{-15}	0	1.359 140 914 229 523	0

All the expressions that appear in the formulas given by Nielson and Jung [15] are expressed in the above simplified form.

Table 1 illustrates the results of naïve computations of both formulas using double precision floating point arithmetic as $\varepsilon \to 0$. The exact values up to 16 digits of precision are given in the third column. The significant digits in both formula computations that coincide with the exact values are highlighted in green colour. We observe a gradual loss of significant digits in both formula computations which deteriorates as $\varepsilon \to 0$. For $\varepsilon \leq 10^{-8}$ we lose all the significant digits in both formula computations.

This example demonstrates two features: 1) algebraic rearrangements using Eq. (4.5) does not avoid loss of significance in Formula2 and 2) the bad fame of nearly confluent divided differeces is a blessing in disguise as it gives a priori warning about loss of significance in Formula1. In other words, due to the algebraic structure of Formula2 the loss of significance in the computations might go unnoticed or misdiagnosed.

In the analytical solution of the X-IVAS scheme, the following expressions might suffer from cancellation errors in a straight-forward (naïve) computation of the same using finite precision arithmetic.

$$\begin{aligned}(5.4)\quad & \exp[\tau\lambda_1;\tau\lambda_2],\quad \exp[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3],\quad \mathrm{q}(\tau\lambda_1),\quad \mathrm{q}[\tau\lambda_1;\tau\lambda_2],\quad \mathrm{q}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3],\\ & \mathrm{r}(\tau\lambda_1),\quad \mathrm{r}[\tau\lambda_1;\tau\lambda_2],\quad \mathrm{r}[\tau\lambda_1;\tau\lambda_2;\tau\lambda_3],\quad \Phi(-\tau\alpha,\tau\beta),\quad \Phi(\tau\lambda_3-\tau\alpha,\tau\beta),\\ & \Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha],\quad \Phi(\star,\tau\beta)[-\tau\alpha;\tau\lambda_3-\tau\alpha],\quad \Phi(\star,\tau\beta)[-\tau\alpha;-\tau\alpha;\tau\lambda_3-\tau\alpha]\end{aligned}$$

The above expressions can be identified as the elements of the following nested set of divided differences.

$$\begin{aligned}(5.5)\quad & \Big\{\big\{\exp[x_1;x_2],\mathrm{q}(x)\big\},\big\{\exp[x_1;x_2;x_3],\mathrm{q}[x_1;x_2],\mathrm{r}(x),\Phi(x,y)\big\},\\ & \big\{\mathrm{q}[x_1;x_2;x_3],\mathrm{r}[x_1;x_2],\Phi(\star,y)[x_1;x_2]\big\},\big\{\mathrm{r}[x_1;x_2;x_3],\Phi(\star,y)[x_1;x_1;x_2]\big\}\Big\}\end{aligned}$$

The order of the divided differences gradually increase from first-order in the first subset to fourth-order in the last subset. All elements of a subset are particular cases of the first element of that subset. For instance,

$$\mathrm{q}[x_1;x_2] = \exp[0;x_1;x_2], \quad \mathrm{r}(x) = \exp[0;0;x], \quad \Phi(x,y) = \exp[-\mathrm{i}y;\mathrm{i}y;x] \tag{5.6}$$

Following this line, it is possible to express all the divided differences in Eq. (5.5) as the divided differences of the exponential function; The details of the same are given in §5.5. As $\varepsilon \to 0$, the rate of loss of significant digits in a naïve computation of divided differences is generally equal to the order of the same. In the considered example, i.e. $\exp[x_1;x_2;x_3]$ we loose significant digits at a second order rate. Following this line, naïve computation of the third and the fourth subsets in Eq. (5.5) are meaningless for $\varepsilon \leq 10^{-5}$ and $\varepsilon \leq 10^{-4}$, respectively.

An algorithm for the accurate computation of divided differences of the exponential function was presented by McCurdy et.al [14]. Following this line, a similar algorithm for the accurate computation of divided differences of the auxiliary functions $\mathrm{q}()$ and $\mathrm{r}()$, cf. Eq. (4.10) and Eq. (4.11), was presented by Caliari [1]. These algorithms have a wider scope, i.e. they were designed to evaluate functions of $n \times n$ matrices appearing in exponential integrators for large systems of equations (ordinary or differential). A user who already has these algorithms implemented, might just invoke them to evaluate the divided differences listed in Eq. (5.4) and use them in the formulas for $\mathbf{P}(\tau,\mathbf{A})$, $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ given in §4.2. This would address the numerical stability issues in the formula computations.

In what follows, we present a simple yet stable piecewise definitions for divided differences. The methodology used to arrive at these piecewise definitions is of limited scope, i.e. this approach is not suitable for arbitrary kth order divided differences. Nevertheless, it is well suited for the at most fourth order divided differences found in the formulas for $\mathbf{P}(\tau,\mathbf{A})$, $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$.

5.2. Optimal series approximation of divided differences. In this section we establish optimal series approximation of divided differences of a given function $\mathrm{f}(x)$. Consider the sequence $\{x_1,x_2,\ldots,x_n\}$ and some definitions related to this sequence.

$$x_\mathrm{a} := \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \widetilde{x}_i := x_i - x_\mathrm{a}, \quad \mathscr{X} := \{\widetilde{x}_1,\widetilde{x}_2,\ldots,\widetilde{x}_n\} \tag{5.7}$$

$$\mathscr{X}_\mathrm{p} := \mathtt{choose}(\mathscr{X},2), \quad \binom{n}{k} := \frac{n!}{k!(n-k)!}, \quad x_\mathrm{p}^2 := \sum_{i=1}^{\binom{n}{2}} \prod_{j=1}^{2} \mathscr{X}_\mathrm{p}(i,j) \tag{5.8}$$

where x_a is the mean value of the sequence and $\widetilde{x}_i$ is the fluctuation of x_i about the mean. The function $\mathtt{choose}(\mathscr{X},2)$ returns a sequence $\mathscr{X}_\mathrm{p}$ consisting of pair-combinations ($2-$combinations) of elements from $\mathscr{X}$. In $\mathscr{X}_\mathrm{p}(i,j)$ the index i points to a combination and index j points to an element within this combination. The sum of the product of the pairs in $\mathscr{X}_\mathrm{p}$ is stored as the square of the auxiliary variable x_p. The result is stored as x_p^2 to highlight the fact that it is a second order term. Likewise, the triple, quadruple and quintuple combinations of $\mathscr{X}$ are denoted as $\mathscr{X}_\mathrm{t}$, $\mathscr{X}_\mathrm{q}$ and $\mathscr{X}_\mathrm{v}$, respectively. Further, the sum of the product of the triples, quadruples and quintuples are stored in x_t^3, x_q^4 and x_v^5, respectively. Following x_p^2, the superscripts (which are ordinary powers) in x_t^3, x_q^4 and x_v^5 highlight the fact that they are third,

fourth and fifth order terms, respectively. Thus,

$$\mathscr{X}_{\mathrm{t}} := \texttt{choose}(\mathscr{X},3), \quad \mathscr{X}_{\mathrm{q}} := \texttt{choose}(\mathscr{X},4), \quad \mathscr{X}_{\mathrm{v}} := \texttt{choose}(\mathscr{X},5) \tag{5.9}$$

$$x_{\mathrm{t}}^3 := \sum_{i=1}^{\binom{n}{3}} \prod_{j=1}^{3} \mathscr{X}_{\mathrm{t}}(i,j), \quad x_{\mathrm{q}}^4 := \sum_{i=1}^{\binom{n}{4}} \prod_{j=1}^{4} \mathscr{X}_{\mathrm{q}}(i,j), \quad x_{\mathrm{v}}^5 := \sum_{i=1}^{\binom{n}{5}} \prod_{j=1}^{5} \mathscr{X}_{\mathrm{v}}(i,j) \tag{5.10}$$

Using the above definitions, we can derive[1] the following identity for the divided differences of $\mathrm{f}(x)$. The mean value theorem guarantees the existence of a ξ in the smallest interval containing $\{x_1, x_2, \ldots, x_n\}$ such that,

$$\mathrm{f}^{(n)}(\xi) := \frac{\partial^n}{\partial \lambda^n} \mathrm{f}(\lambda)\bigg|_{\lambda=\xi}, \quad \mathrm{f}(x_n) = \mathrm{f}(\xi) + \sum_{n=1}^{\infty} (x_n - \xi)^n \frac{\mathrm{f}^{(n)}(\xi)}{n!} \tag{5.11}$$

$$\begin{aligned} \mathrm{f}[x_1; x_2; \ldots; x_n] = {} & \frac{\mathrm{f}^{(n-1)}(x_{\mathrm{a}})}{(n-1)!} - x_{\mathrm{p}}^2 \frac{\mathrm{f}^{(n+1)}(x_{\mathrm{a}})}{(n+1)!} + x_{\mathrm{t}}^3 \frac{\mathrm{f}^{(n+2)}(x_{\mathrm{a}})}{(n+2)!} + (x_{\mathrm{p}}^4 - x_{\mathrm{q}}^4) \frac{\mathrm{f}^{(n+3)}(x_{\mathrm{a}})}{(n+3)!} \\ & + (x_{\mathrm{v}}^5 - 2x_{\mathrm{p}}^2 x_{\mathrm{t}}^3) \frac{\mathrm{f}^{(n+4)}(\xi)}{(n+4)!} \end{aligned} \tag{5.12}$$

Note that the first term in the above series expansion provides a second-order approximation to $\mathrm{f}[x_1; x_2; \ldots; x_n]$. If the series is expanded with respect to any point other than x_{a}, the first-order terms are resurrected. Thus, the approximation is optimal for the choice x_{a}. For the first-order divided difference $\mathrm{f}[x_1; x_2]$, the above equation can be simplified and easily extended to any number of terms as shown below.

$$h := \frac{x_2 - x_1}{2}, \quad x_{\mathrm{p}}^2 = -h^2, \quad x_{\mathrm{t}}^2 = 0, \quad x_{\mathrm{q}}^2 = 0, \quad x_{\mathrm{v}}^2 = 0 \tag{5.13}$$

$$\mathrm{f}[x_1; x_2] = \mathrm{f}^{(1)}(x_{\mathrm{a}}) + h^2 \frac{\mathrm{f}^{(3)}(x_{\mathrm{a}})}{3!} + \cdots + h^{2n-2} \frac{\mathrm{f}^{(2n-1)}(x_{\mathrm{a}})}{(2n-1)!} + h^{2n} \frac{\mathrm{f}^{(2n+1)}(\xi)}{(2n+1)!} \tag{5.14}$$

Likewise, for the second-order divided difference $\mathrm{f}[x_1; x_2; x_3]$, Eq. (5.12) can be simplified to the following.

$$x_{\mathrm{p}}^2 = -\frac{3}{2} x_{\sigma}^2, \quad x_{\sigma}^2 := \frac{\widetilde{x}_1^2 + \widetilde{x}_2^2 + \widetilde{x}_3^2}{3}, \quad x_{\mathrm{t}}^3 = \widetilde{x}_1 \widetilde{x}_2 \widetilde{x}_3, \quad x_{\mathrm{q}}^2 = 0, \quad x_{\mathrm{v}}^2 = 0 \tag{5.15}$$

$$\mathrm{f}[x_1; x_2; x_3] = \frac{1}{2} \mathrm{f}^{(2)}(x_{\mathrm{a}}) + \frac{x_{\sigma}^2}{16} \mathrm{f}^{(4)}(x_{\mathrm{a}}) + \frac{x_{\mathrm{t}}^3}{120} \mathrm{f}^{(5)}(x_{\mathrm{a}}) + \frac{x_{\sigma}^4}{320} \mathrm{f}^{(6)}(x_{\mathrm{a}}) + 3 x_{\sigma}^2 x_{\mathrm{t}}^3 \frac{\mathrm{f}^{(7)}(\xi)}{7!} \tag{5.16}$$

where x_{σ} is the standard deviation of the considered sequence. It is possible to relate x_{p} and x_{σ} for all n and in this work we exploit this relationship as it reduces the number of arithmetic operations.

$$x_{\mathrm{p}}^2 = -\frac{n}{2} x_{\sigma}^2 \tag{5.17}$$

5.3. Double precision floating point numbers. We briefly describe how double precision floating point numbers are stored in a computer as per the IEEE 754 standard. Any

[1] As the algebra involved is overwhelming and error-prone, we have used the computer algebra system `Maple` to perform the simplifications and verifications. Thus, human intervention is dedicated to identify patterns and to discover abstract expressions such as x_{p}, x_{t}, x_{q}, etc.

decimal floating point number within the range of the double can be written in the normalized form as follows.

$$\text{Decimal form} \to (-1)^s\, 2^e\, 1.f \approx \begin{array}{|c|c|c|} \hline s & (e+1023)_{\mathrm{b}} & 0.f_{\mathrm{b}} \\ \hline \end{array} \leftarrow \text{Binary form} \tag{5.18}$$

$$\begin{array}{ccc} 1 & 11 & 52 \end{array} \leftarrow \text{No. of bits stored}$$

In the above equation the boolean $s \in \{0,1\}$ is called the *sign bit*, the integer e is called the *exponent*; $-1022 \le e \le 1023$ and the fraction f is called the *significand*. The numbers with a subscript b are expressed in the binary format. When the binary expression $0.f_{\mathrm{b}}$ is not exactly representable using 52 bits, it is rounded to the nearest representable number.

5.4. Piecewise definition of divided differences. To control (bound) the loss of significant digits in the computations of the divided differences in Eq. (5.5), we present piecewise definitions for the same. In this approach, we switch the computations to the corresponding series expansions of the same should the difference of the independent variables be less than some threshold. These threshold values are chosen such that we retain as many significant digits as possible. This methodology (technique) is simple and systematic. It is explained in full detail using an example by means of which we answer the three questions raised in this context by Kahan and Darcy [12]: 1) What value should be assigned to the threshold in this technique? 2) How many terms in the series approximation should this technique retain? and 3) How accurate is this technique?

The elements of the first subset of Eq. (5.5) can be computed [14] to machine precision by rearranging them to the following functional form[2].

$$\operatorname{sinhc}(x) := \begin{cases} \dfrac{\sinh(x)}{x} & \text{if } x \neq 0, \\ 1 & \text{if } x = 0. \end{cases} \tag{5.19}$$

$$\exp[x_1;x_2] = \mathrm{e}^{(x_1+x_2)/2} \operatorname{sinhc}\left(\frac{x_2-x_1}{2}\right), \quad \mathrm{q}(x) = \mathrm{e}^{x/2} \operatorname{sinhc}\left(\frac{x}{2}\right) \tag{5.20}$$

We now explain the methodology used to arrive at piecewise definitions of divided differences using $\exp[x_1;x_2;x_3]$ as an example. This term can be written as

$$\exp[x_1;x_2;x_3] = \mathrm{e}^{x_2} \exp[x_1-x_2;0;x_3-x_2] = \mathrm{e}^{x_2} \frac{\mathrm{q}(x_3-x_2)-\mathrm{q}(x_1-x_2)}{x_3-x_1} \tag{5.21}$$

where the function $\mathrm{q}(x)$ is evaluated as shown in Eq. (5.20). Without loss of generality, we assume $x_1 \le x_2 \le x_3$. Consequently we have,

$$\forall\, \xi \in [x_1,x_3], \quad |\xi - x_2| \le (x_3-x_1) \tag{5.22}$$

$$x_\sigma^2 \le (x_3-x_1)^2, \quad |x_{\mathrm{t}}^3| \le (x_3-x_1)^3 \tag{5.23}$$

In the computations of divided differences, the loss of significance is due to the subtractive cancellations that occur in the dependent variables which is brought to prominence after a division by the difference of the independent variables. Particularly, in Eq. (5.21) the loss of significant digits is due to the cancellations that occur in the term $\mathrm{q}(x_3-x_2)-\mathrm{q}(x_1-x_2)$. This term admits the following series expansion.

$$\mathrm{q}(x_3-x_2)-\mathrm{q}(x_1-x_2) = \frac{x_3-x_1}{2}[1+(x_{\mathrm{a}}-x_2)+\cdots] \tag{5.24}$$

[2]In this form the difference of the independent variables appear symbolically as input to a function that could be evaluated to machine precision

Let $x_3 - x_1 = 2^{-m}$ where $m \geq 1$ is an integer. Then, Eq. (5.22) implies that the higher order terms in Eq. (5.24) tend to zero as $m \to \infty$. Thus,

$$\left.\begin{array}{l} x_3 - x_1 = 2^{-m} \\ x_1 \leq x_2 \leq x_3 \end{array}\right\} \quad \Rightarrow \quad \mathrm{q}(x_3 - x_2) - \mathrm{q}(x_1 - x_2) = O(2^{-(m+1)}) \tag{5.25}$$

When written in the normalized decimal form (cf. Eq. (5.18)), the exponent of $\mathrm{q}(x_3 - x_2)$ and $\mathrm{q}(x_1 - x_2)$ will be 0 and -1, respectively. This can be inferred using Eq. (5.22) as shown below.

$$0 \leq x \leq 2^{-m} \Rightarrow 2^0 \leq \mathrm{q}(x) < 2^1, \quad -2^m \leq x \leq 0 \Rightarrow 2^{-1} \leq \mathrm{q}(x) < 2^0 \tag{5.26}$$

$$0 \leq (x_3 - x_2) \leq 2^{-m} \Rightarrow \mathrm{q}(x_3 - x_2) = (-1)^0 2^0 1.\widehat{f} \approx \boxed{0}\boxed{1023_\mathrm{b}}\boxed{0.\widehat{f}_\mathrm{b}} \tag{5.27}$$

$$-2^{-m} \leq (x_1 - x_2) \leq 0 \Rightarrow \mathrm{q}(x_1 - x_2) = (-1)^0 2^{-1} 1.\widetilde{f} \approx \boxed{0}\boxed{1022_\mathrm{b}}\boxed{0.\widetilde{f}_\mathrm{b}} \tag{5.28}$$

where $\widehat{f}$ and $\widetilde{f}$ denote the significands of $\mathrm{q}(x_3 - x_2)$ and $\mathrm{q}(x_1 - x_2)$, respectively. The subtraction $\mathrm{q}(x_3 - x_2) - \mathrm{q}(x_1 - x_2)$ can be described schematically as follows.

$$\begin{aligned} \mathrm{q}(x_3 - x_2) - \mathrm{q}(x_1 - x_2) &= (-1)^0 2^0 1.\widehat{f} - (-1)^0 2^{-1} 1.\widetilde{f} \\ &\approx \boxed{0}\boxed{1023_\mathrm{b}}\boxed{0.\widehat{f}_\mathrm{b}} - \boxed{0}\boxed{1022_\mathrm{b}}\boxed{0.\widetilde{f}_\mathrm{b}} && \text{normalized form} \\ &= \boxed{0}\boxed{1023_\mathrm{b}}\boxed{0.\widehat{f}_\mathrm{b}} - \boxed{0}\boxed{1023_\mathrm{b}}\boxed{0.1\widetilde{f}_\mathrm{b}} && \text{align radix points} \\ &= \boxed{0}\boxed{1023_\mathrm{b}}\boxed{0.\{0\}_m 1 f_\mathrm{b}} && O(2^{-(m+1)}) \\ &= \boxed{0}\boxed{(1022-m)_\mathrm{b}}\boxed{0.f_\mathrm{b}} && \text{normalized form} \end{aligned} \tag{5.29}$$

The notation $\{0\}_m$ means that the bits within the braces are repeated m times. We see that among the stored 52 bits of $\widehat{f}_\mathrm{b}$ and $\widetilde{f}_\mathrm{b}$, the first m bits are lost due to cancellation. After subtraction, the unit bit at the $m+1^{\text{th}}$ place will become the implicit bit of the result which is not stored, cf. Eq. (5.18). The exponent of the result will become $-(m+1)$. The significand of the result will become the remaining bits denoted in Eq. (5.29) as f_b of which only $51-m$ bits are significant.

We see that if $\exp[x_1;x_2;x_3]$ is evaluated using Eq. (5.21) we loose significant bits at a first order rate. We will call this form of computation as the *direct computation*. If $x_1 \leq x_2 \leq x_3$ and $x_3 - x_1 = 2^{-m}$ then in the direct computation of $\exp[x_1;x_2;x_3]$ we are left with $51-m$ significant bits in the significand.

Using Eq. (5.16) we can write the series expansion for $\exp[x_1;x_2;x_3]$ as follows.

$$\exp[x_1;x_2;x_3] = \frac{\mathrm{e}^{x_\mathrm{a}}}{2}\mathscr{S}, \quad \mathscr{S} := \left[1 + \frac{x_\sigma^2}{8} + \frac{x_\mathrm{t}^3}{60} + \frac{x_\sigma^4}{160} + \frac{x_\sigma^2 x_\mathrm{t}^3}{840} + \cdots\right] \tag{5.30}$$

The above form to evaluate $\exp[x_1;x_2;x_3]$ will be called as the *series computation*. Clearly, in the series computation we do not find removable singularities which imply that subtractive cancellations (if any) are not brought to prominence. However, the truncation of the series will introduce an error which will limit the number of significant digits in the series computation that match those in an exact computation. When the series $\mathscr{S}$ is truncated after the first n terms it will be denoted as $\mathscr{S}_n$. As $\exp(x_\mathrm{a})/2$ can be evaluated to machine precision, the number of significant digits in the series computation is essentially limited by the term

$\mathscr{S}_n$. The series $\mathscr{S}$ when written in the normalized decimal form has a zero exponent when $x_1 \le x_2 \le x_3$ and $x_3 - x_1 = 2^{-m}$. This can be inferred using Eqs. (5.16) and (5.23) as follows.

$$\exists\, \xi \in [x_1, x_3] \text{ such that, } \frac{e^{x_a}}{2}\left[1 + \frac{x_\sigma^2}{8} + \frac{x_t^3}{60} + \cdots\right] = \frac{e^{x_a}}{2} + \frac{x_\sigma^2}{16} e^{\xi} \tag{5.31}$$

$$\Rightarrow 1 + \frac{x_\sigma^2}{8} + \frac{x_t^3}{60} + \cdots = 1 + \frac{x_\sigma^2}{8} e^{\xi - x_a} \tag{5.32}$$

$$\Rightarrow 2^0 \le 1 + \frac{x_\sigma^2}{8} e^{\xi - x_a} \le 1 + 2^{-(2m+3)} e^{2^{-m}} < 2^1 \tag{5.33}$$

It follows that all terms except the first one contribute to the significand of $\mathscr{S}$. Thus,

$$\mathscr{S} = (-1)^0 2^0 1.f \approx \boxed{0}\boxed{1023_b}\boxed{0.f_b} \tag{5.34}$$

Hence, when $\mathscr{S}$ is replaced by $\mathscr{S}_n$, the associated truncation error can be understood as to limit the number of significant digits in the series computation. The truncation error associated to $\mathscr{S}_n$ is denoted as $\mathscr{E}_n$. From Eqs. (5.23) and (5.30) we infer,

$$\mathscr{E}_1 = O\left(\frac{x_\sigma^2}{8}\right) \le O(2^{-(2m+3)}), \quad \mathscr{E}_2 = O\left(\frac{x_t^3}{60}\right) \le O(2^{-(3m+6)}) \tag{5.35}$$

$$\mathscr{E}_3 = O\left(\frac{x_\sigma^4}{160}\right) \le O(2^{-(4m+8)}), \quad \mathscr{E}_4 = O\left(\frac{x_\sigma^2 x_t^3}{840}\right) \le O(2^{-(5m+10)}) \tag{5.36}$$

Expressing $\mathscr{S}_n = \mathscr{S} - \mathscr{E}_n$ in the double storage format we get,

$$\mathscr{S}_1 \approx \boxed{0}\boxed{1023_b}\boxed{0.f_b} - \boxed{0}\boxed{1023_b}\boxed{0.\{0\}_{2m+2}1\ldots} \tag{5.37}$$

$$\mathscr{S}_2 \approx \boxed{0}\boxed{1023_b}\boxed{0.f_b} - \boxed{0}\boxed{1023_b}\boxed{0.\{0\}_{3m+5}1\ldots} \tag{5.38}$$

$$\mathscr{S}_3 \approx \boxed{0}\boxed{1023_b}\boxed{0.f_b} - \boxed{0}\boxed{1023_b}\boxed{0.\{0\}_{4m+7}1\ldots} \tag{5.39}$$

$$\mathscr{S}_4 \approx \boxed{0}\boxed{1023_b}\boxed{0.f_b} - \boxed{0}\boxed{1023_b}\boxed{0.\{0\}_{5m+9}1\ldots} \tag{5.40}$$

where $\mathscr{E}_n$ is written after the alignment of radix points and the remaining digits in the significands are denoted by ellipsis. This implies that we have $(2m+2), (3m+5), (4m+7)$ and $(5m+9)$ significant digits in $\mathscr{S}_1, \mathscr{S}_2, \mathscr{S}_3$ and $\mathscr{S}_4$, respectively.

For each $\mathscr{S}_n$ we solve for m by matching the accuracy of the series computation with the one obtained in the direct computation. In this way, we obtain the threshold value of $(x_3 - x_1) = 2^{-m}$ and the lower bound for the number of significant digits `nsd` in a piecewise computation of $\exp[x_1; x_2; x_3]$. Thus,

$$\mathscr{S}_1: \quad 51 - m = 2m + 2 \quad \Rightarrow \quad m = 16, \quad \texttt{nsd} = 34 \text{ bits} \approx 11 \text{ decimal digits} \tag{5.41}$$

$$\mathscr{S}_2: \quad 51 - m = 3m + 5 \quad \Rightarrow \quad m = 12, \quad \texttt{nsd} = 39 \text{ bits} \approx 12 \text{ decimal digits} \tag{5.42}$$

$$\mathscr{S}_3: \quad 51 - m = 4m + 7 \quad \Rightarrow \quad m = 9, \quad \texttt{nsd} = 42 \text{ bits} \approx 13 \text{ decimal digits} \tag{5.43}$$

$$\mathscr{S}_4: \quad 51 - m = 5m + 9 \quad \Rightarrow \quad m = 7, \quad \texttt{nsd} = 44 \text{ bits} \approx 14 \text{ decimal digits} \tag{5.44}$$

In the above equations, the solution for m is rounded to the nearest integer. Using this rounded m we estimate `nsd` as the minimum of the number of significant digits found in the direct and the series computations. As the loss of significant digits is bounded, the piecewise computation of $\exp[x_1; x_2; x_3]$ is stable.

TABLE 2
Loss of significance controlled in the piecewise computations of $\exp[1;1+\varepsilon;1+2\varepsilon]$.

ε	Formula3 computation	Exact 16 digits	Formula4 computation
10^{-01}	1.503 335 165 136 323	1.503 335 165 136 325	1.503 335 165 136 323
10^{-02}	1.372 811 947 550 791	1.372 811 947 550 820	1.372 811 947 550 791
10^{-03}	1.360 500 848 316 010	1.360 500 848 315 854	1.360 500 848 315 854
10^{-04}	1.359 276 836 253 229	1.359 276 836 249 607	1.359 276 836 249 607
10^{-05}	1.359 154 505 691 532	1.359 154 505 717 948	1.359 154 505 717 948
10^{-06}	1.359 142 273 371 116	1.359 142 273 371 229	1.359 142 273 371 229
10^{-07}	1.359 141 050 143 620	1.359 141 050 143 621	1.359 141 050 143 622
10^{-08}	1.359 140 927 820 931	1.359 140 927 820 931	1.359 140 927 820 931
10^{-09}	1.359 140 915 588 663	1.359 140 915 588 663	1.359 140 915 588 663
10^{-10}	1.359 140 914 365 436	1.359 140 914 365 436	1.359 140 914 365 436
10^{-11}	1.359 140 914 243 114	1.359 140 914 243 114	1.359 140 914 243 114
10^{-12}	1.359 140 914 230 881	1.359 140 914 230 881	1.359 140 914 230 881
10^{-13}	1.359 140 914 229 658	1.359 140 914 229 658	1.359 140 914 229 658
10^{-14}	1.359 140 914 229 536	1.359 140 914 229 536	1.359 140 914 229 536
10^{-15}	1.359 140 914 229 524	1.359 140 914 229 523	1.359 140 914 229 524

The numerical test presented in §5.1 is repeated here and the computations of the piecewise definitions are shown in Table 2. The piecewise definitions considering $\mathscr{S}_1$ and $\mathscr{S}_4$ for the series computations are called Formula3 and Formula4, respectively. The significant digits in both formula computations that differ from the exact values are highlighted in green color. The lower bounds for the number of significant digits given in Eqs. (5.41) and (5.44) are reproduced in this test for Formula3 and Formula4, respectively.

5.5. Stable formulas for exponential divided differences. Following the methodology described in the previous section we present stable piecewise definitions of all the expressions that belong to the subsets in Eq. (5.5). In the series computation of each piecewise definition, we consider the first four terms in the corresponding series expansion. Recall that the exponential function is its own derivative. This result along with the abstraction (e.g. $x_{\mathrm{p}}, x_{\mathrm{t}}$ etc.) in the optimal series expansion permits us to use multiple terms in the series expansion without incurring substantial computational cost.

The elements of the first subset in Eq. (5.5) can be evaluated to machine precision without resorting to a series computation, cf. Eq. (5.20). The first element of the second subset, i.e. $\exp[x_1;x_2;x_3]$ was used as an example to describe the details of the piecewise computation technique in the previous section. The stable piecewise definition of the same when $x_1 \leq x_2 \leq x_3$ can be summarized as follows.

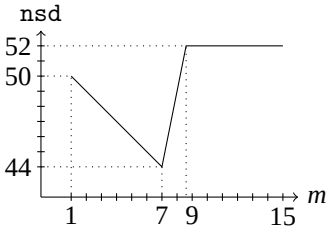
$$\exp[x_1;x_2;x_3] = \begin{cases} \mathrm{e}^{x_2}\dfrac{\mathrm{q}(x_3-x_2)-\mathrm{q}(x_1-x_2)}{x_3-x_1} & \text{if } (x_3-x_1) > 2^{-7} \\ \dfrac{\mathrm{e}^{x_{\mathrm{a}}}}{2}\left[1+\dfrac{x_\sigma^2}{8}+\dfrac{x_{\mathrm{t}}^3}{60}+\dfrac{x_\sigma^4}{160}\right] & \text{else} \end{cases} \tag{5.45}$$

It is essential to sort the arguments lest the series computation should incur a significant truncation error. The variation in the number of significant digits `nsd` in $\exp[x_1;x_2;x_3]$ with respect to m, where $(x_3-x_1) = 2^{-m}$, is denoted as $\mathtt{nsd}(\exp[x_1;x_2;x_3],m)$. Using the above

stable formula for $\exp[x_1;x_2;x_3]$ we obtain,

(5.46)

$$\texttt{nsd}(\exp[x_1;x_2;x_3],m) = \begin{cases} 51-m & \text{if } m<7 \\ 5m+9 & \text{if } 7\le m<8.6 \\ 52 & \text{if } m\ge 8.6 \end{cases},$$



Recall that $\mathrm{q}[x_1;x_2]=\exp[x_1;0;x_2]$, $\mathrm{r}(x)=\exp[0;0;x]$ and $\Phi(x,y)=\exp[-\mathrm{i}y;\mathrm{i}y;x]$. Let `sort` be a sorting function and $\widehat{x}_1\le\widehat{x}_2\le\widehat{x}_3$. Then, using Eq. (5.45) a stable formula for $\mathrm{q}[x_1;x_2]$ is

$$\{\widehat{x}_1,\widehat{x}_2,\widehat{x}_3\}=\texttt{sort}(\{x_1,0,x_2\}),\quad \mathrm{q}[x_1;x_2]=\exp[\widehat{x}_1;\widehat{x}_2;\widehat{x}_3]; \tag{5.47}$$

Likewise a stable formula for $\mathrm{r}(x)$ is

$$\{\widehat{x}_1,0,\widehat{x}_3\}=\texttt{sort}(\{0,0,x\}),\quad \mathrm{r}(x)=\exp[\widehat{x}_1;0;\widehat{x}_3]; \tag{5.48}$$

As $\exp[-\mathrm{i}y;\mathrm{i}y;x]$ involves complex numbers we give it special attention. Recall that the exponential function is holomorphic, i.e. it is complex differentiable in a neighbourhood of every point in its domain. This implies that it is infinitely differentiable and is equal to its own Taylor series. Thus, the optimal series approximation of divided differences presented in §5.2 naturally extends to $\exp[-\mathrm{i}y;\mathrm{i}y;x]$. Following this line, a stable formula for $\Phi(x,y)$ can be obtained as shown below.

$$z:=x+\mathrm{i}y,\quad z_{\mathrm{a}}:=\frac{x}{3},\quad z_\sigma^2:=2z_{\mathrm{a}}^2-\frac{2}{3}y^2,\quad z_{\mathrm{t}}^3:=2z_{\mathrm{a}}(z_{\mathrm{a}}^2+y^2) \tag{5.49}$$

$$\Phi(x,y)=\exp[-\mathrm{i}y;\mathrm{i}y;x]=\begin{cases}\dfrac{\mathrm{e}^{z/2}\operatorname{sinhc}(z^*/2)-\operatorname{sinc}(y)}{z} & \text{if } |z|>2^{-7} \\ \dfrac{\mathrm{e}^{z_{\mathrm{a}}}}{2}\left[1+\dfrac{z_\sigma^2}{8}+\dfrac{z_{\mathrm{t}}^3}{60}+\dfrac{z_\sigma^4}{160}\right] & \text{else}\end{cases} \tag{5.50}$$

The above definition assumes the availability of a complex math library which provides an interface for stable computation of common arithmetic operations, elementary and transcendental functions. This assumption holds for the `C++` programming language which is equipped with the standard math library `<complex>`.

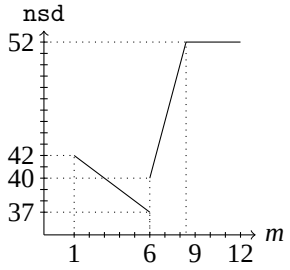
The divided difference $\exp[x_1;x_2;x_3;x_4]$ is a template for the elements of the third subset in Eq. (5.5). The proposed piecewise definition of $\exp[x_1;x_2;x_3;x_4]$ when $x_1\le x_2\le x_3\le x_4$ is

$$\exp[x_1;x_2;x_3;x_4]=\begin{cases}\dfrac{\exp[x_2;x_3;x_4]-\exp[x_1;x_2;x_3]}{x_4-x_1} & \text{if } (x_4-x_1)>2^{-6} \\ \dfrac{\mathrm{e}^{x_{\mathrm{a}}}}{3!}\left[1-\dfrac{x_{\mathrm{p}}^2}{20}+\dfrac{x_{\mathrm{t}}^3}{120}+\dfrac{x_{\mathrm{p}}^4-x_{\mathrm{q}}^4}{840}\right] & \text{else}\end{cases} \tag{5.51}$$

For the above piecewise definition of $\exp[x_1;x_2;x_3;x_4]$ we obtain

(5.52)

$$\texttt{nsd}(\exp[x_1;x_2;x_3;x_4],m)=\begin{cases}43-m & \text{if } m<6\\ 5m+10 & \text{if } 6\le m<8.4\\ 52 & \text{if } m\ge 8.4\end{cases},$$



It is straightforward to verify that $\mathrm{q}[x_1;x_2;x_3]=\exp[0;x_1;x_2;x_3]$, $\mathrm{r}[x_1;x_2]=\exp[0;0;x_1;x_2]$ and $\Phi(\star,y)[x_1;x_2]=\exp[-\mathrm{i}y;\mathrm{i}y;x_1;x_2]$. Using Eq. (5.51) a stable formula for $\mathrm{q}[x_1;x_2;x_3]$ is

(5.53) $$\{\widehat{x}_1,\widehat{x}_2,\widehat{x}_3,\widehat{x}_4\}=\texttt{sort}(\{0,x_1,x_2,x_3\}),\quad \mathrm{q}[x_1;x_2;x_3]=\exp[\widehat{x}_1;\widehat{x}_2;\widehat{x}_3;\widehat{x}_4];$$

Likewise a stable formula for $\mathrm{r}[x_1;x_2]$ is

(5.54) $$\{\widehat{x}_1,\widehat{x}_2,\widehat{x}_3,\widehat{x}_4\}=\texttt{sort}(\{0,0,x_1,x_2\}),\quad \mathrm{r}[x_1;x_2]=\exp[\widehat{x}_1;\widehat{x}_2;\widehat{x}_3;\widehat{x}_4];$$

The proposed piecewise definition of $\Phi(\star,y)[x_1;x_2]$ is

(5.55) $$\{\widehat{x}_1,\widehat{x}_2\}=\texttt{sortabs}(\{x_1,x_2\}),\quad z_1:=\widehat{x}_1+\mathrm{i}y,\quad z_2:=\widehat{x}_2+\mathrm{i}y,\quad z_{\mathrm{a}}:=\frac{\widehat{x}_1+\widehat{x}_2}{4}$$

(5.56) $$z_{\mathrm{p}}^2:=y^2+\widehat{x}_1\widehat{x}_2-6z_{\mathrm{a}}^2,\quad z_{\mathrm{t}}^3:=2z_{\mathrm{a}}(y^2-\widehat{x}_1\widehat{x}_2+4z_{\mathrm{a}}^2),\quad z_{\mathrm{q}}^4:=(y^2+z_{\mathrm{a}}^2)(\widehat{x}_1\widehat{x}_2-3z_{\mathrm{a}}^2)$$

(5.57) $$\Phi(\star,y)[x_1;x_2]=\begin{cases}\left[\dfrac{\exp[\widehat{x}_1;\widehat{x}_2]-\mathrm{e}^{z_1/2}\,\mathrm{sinhc}(z_1^*/2)}{z_2^*}-\Phi(\widehat{x}_1,y)\right]\dfrac{1}{z_2} & \text{if } |z_2|>2^{-6}\\ \dfrac{\mathrm{e}^{z_{\mathrm{a}}}}{3!}\left[1-\dfrac{z_{\mathrm{p}}^2}{20}+\dfrac{z_{\mathrm{t}}^3}{120}+\dfrac{z_{\mathrm{p}}^4-z_{\mathrm{q}}^4}{840}\right] & \text{else}\end{cases}$$

where `sortabs` is a function that sorts its arguments with respect to its absolute value, i.e. $|\widehat{x}_1|\le|\widehat{x}_2|$ in Eq. (5.55). The term $\Phi(\widehat{x}_1,y)$ is evaluated using the stable formula given in Eq. (5.50).

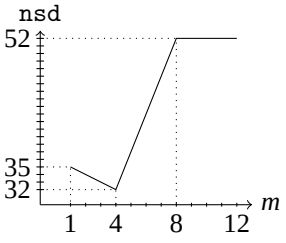
The divided difference $\exp[x_1;x_2;x_3;x_4;x_5]$ is a template for the elements of the third subset in Eq. (5.5). The proposed piecewise definition of $\exp[x_1;x_2;x_3;x_4;x_5]$ when $x_1\le x_2\le x_3\le x_4\le x_5$ is

(5.58) $$\exp[x_1;x_2;x_3;x_4;x_5]=\begin{cases}\dfrac{\exp[x_2;x_3;x_4;x_5]-\exp[x_1;x_2;x_3;x_4]}{x_5-x_1} & \text{if } (x_5-x_1)>2^{-4}\\ \dfrac{\mathrm{e}^{x_{\mathrm{a}}}}{4!}\left[1-\dfrac{x_{\mathrm{p}}^2}{30}+\dfrac{x_{\mathrm{t}}^3}{210}+\dfrac{x_{\mathrm{p}}^4-x_{\mathrm{q}}^4}{1680}\right] & \text{else}\end{cases}$$

For the above piecewise definition of $\exp[x_1;x_2;x_3;x_4;x_5]$ we obtain

(5.59)

$$\texttt{nsd}(\exp[x_1;x_2;x_3;x_4;x_5],m)=\begin{cases}36-m & \text{if } m<4\\ 5m+12 & \text{if } 4\le m<8\\ 52 & \text{if } m\ge 8\end{cases},$$



It is straightforward to verify that $\mathrm{r}[x_1;x_2;x_3] = \exp[0;0;x_1;x_2;x_3]$ and $\Phi(\star,y)[x_1;x_2;x_3] = \exp[-\mathrm{i}y;\mathrm{i}y;x_1;x_2;x_3]$. Using Eq. (5.58) a stable formula for $\mathrm{r}[x_1;x_2;x_3]$ can be written as,

$$\{\widehat{x}_1,\widehat{x}_2,\widehat{x}_3,\widehat{x}_4,\widehat{x}_5\} = \mathtt{sort}(\{0,0,x_1,x_2,x_3\}) \tag{5.60}$$

$$\mathrm{r}[x_1;x_2;x_3] = \exp[\widehat{x}_1;\widehat{x}_2;\widehat{x}_3;\widehat{x}_4;\widehat{x}_5]; \tag{5.61}$$

The proposed piecewise definition of $\Phi(\star,y)[x_1;x_2;x_3]$ is

$$\{\widehat{x}_1,\widehat{x}_2,\widehat{x}_3\} = \mathtt{sortabs}(\{x_1,x_2,x_3\}), \quad z_3 := \widehat{x}_3 + \mathrm{i}y \tag{5.62}$$

$$z_\mathrm{a} := \frac{\widehat{x}_1+\widehat{x}_2+\widehat{x}_3}{5}, \quad z_\mathrm{q}^4 := 6z_\mathrm{a}^4 + \frac{1}{2}[11z_\mathrm{a}^2 - (\widehat{x}_1^2+\widehat{x}_2^2+\widehat{x}_3^2)](y^2+3z_\mathrm{a}^2) - 2z_\mathrm{a}\widehat{x}_1\widehat{x}_2\widehat{x}_3 \tag{5.63}$$

(5.64)

$$z_\mathrm{p}^2 := \frac{1}{2}[2y^2+5z_\mathrm{a}^2-(\widehat{x}_1^2+\widehat{x}_2^2+\widehat{x}_3^2)], \quad z_\mathrm{t}^3 := \frac{z_\mathrm{a}}{2}[4y^2-35z_\mathrm{a}^2+3(\widehat{x}_1^2+\widehat{x}_2^2+\widehat{x}_3^2)] + \widehat{x}_1\widehat{x}_2\widehat{x}_3$$

(5.65)

$$\Phi(\star,y)[x_1;x_2;x_3] = \begin{cases} \left[\dfrac{\exp[\widehat{x}_1;\widehat{x}_2;\widehat{x}_3]-\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2]}{z_3^*} - \Phi(\star,y)[\widehat{x}_1;\widehat{x}_2]\right]\dfrac{1}{z_3} & \text{if } |z_3| > 2^{-4} \\ \dfrac{\mathrm{e}^{z_\mathrm{a}}}{4!}\left[1-\dfrac{z_\mathrm{p}^2}{30}+\dfrac{z_\mathrm{t}^3}{210}+\dfrac{z_\mathrm{p}^4-z_\mathrm{q}^4}{1680}\right] & \text{else} \end{cases}$$

where the term $\Phi(\star,y)[\widehat{x}_1;\widehat{x}_2]$ is evaluated using the stable formula given in Eq. (5.57). Note that in the stable formula for $\Phi(\star,y)[\widehat{x}_1;\widehat{x}_2]$, just the direct computation of $\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2]$ is sufficient as the threshold value of $|z_2|$ to switch to a series computation is larger for the former than the latter. This means that the series computation of $\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2]$ will never be used in the stable computation of $\Phi(\star,y)[\widehat{x}_1;\widehat{x}_2]$. On the contrary, in the stable computation of $\Phi(\star,y)[x_1;x_2;x_3]$ the switch to the series computation is governed by some threshold value of $|z_3|$ which includes the possibility $|z_2| \to 0$. Therefore, in Eq. (5.65) it is necessary to evaluate the term $\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2]$ in a piecewise manner.

Following Eqs. (5.45) and (5.50), a stable formula for $\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2]$ can be obtained as follows.

$$z_1 := \widehat{x}_1 + \mathrm{i}y, \quad z_2 := \widehat{x}_2 + \mathrm{i}y, \quad z_\mathrm{a} := \frac{\widehat{x}_1+\widehat{x}_2+\mathrm{i}y}{3}, \quad \widetilde{z}_1 := \mathrm{i}y - z_\mathrm{a} \tag{5.66}$$

$$\widetilde{z}_2 := \widehat{x}_1 - z_\mathrm{a}, \quad \widetilde{z}_3 := \widehat{x}_2 - z_\mathrm{a}, \quad z_\sigma^2 := \frac{\widetilde{z}_1^2+\widetilde{z}_2^2+\widetilde{z}_3^2}{3}, \quad z_\mathrm{t}^3 := \widetilde{z}_1\widetilde{z}_2\widetilde{z}_3 \tag{5.67}$$

$$\exp[\mathrm{i}y;\widehat{x}_1;\widehat{x}_2] = \begin{cases} \dfrac{\exp[\widehat{x}_1;\widehat{x}_2]-\mathrm{e}^{z_1/2}\,\mathrm{sinhc}(z_1^*/2)}{z_2^*} & \text{if } |z_2| > 2^{-7} \\ \dfrac{\mathrm{e}^{z_\mathrm{a}}}{2}\left[1+\dfrac{z_\sigma^2}{8}+\dfrac{z_\mathrm{t}^3}{60}+\dfrac{z_\sigma^4}{160}\right] & \text{else} \end{cases} \tag{5.68}$$

6. Examples. We present two examples to validate the numerical stability in the computation of the proposed formulas for the X-IVAS scheme. In these examples the eigenvalues of the matrix $\mathbf{A}$ gradually tends to zero. The symbolic computation of the formulas for the chosen eigenvalues are done using Maple and the first 16 significant decimal digits are stored as reference solutions. These reference solutions are used to measure the relative error in the formula computations using double precision floating point arithmetic.

6.1. Example 1. Consider the case when two of the eigenvalues of the matrix $\mathbf{A}$ are complex numbers. Let $\varepsilon := 10^{-n}$ and choose $n \in \{1,2,3,\ldots,15\}$. For each ε define the

TABLE 3
Relative errors in the usual and stable computation when $\lambda = \{\varepsilon \pm \mathrm{i}2\varepsilon, 4\varepsilon\}$.

ε	$\frac{\|\mathbf{R}^{\mathrm{u}}-\mathbf{R}\|}{\|\mathbf{R}\|}$	$\frac{\|\mathbf{R}^{\mathrm{s}}-\mathbf{R}\|}{\|\mathbf{R}\|}$	$\frac{\|\mathbf{Q}^{\mathrm{u}}-\mathbf{Q}\|}{\|\mathbf{Q}\|}$	$\frac{\|\mathbf{Q}^{\mathrm{s}}-\mathbf{Q}\|}{\|\mathbf{Q}\|}$	$\frac{\|\mathbf{P}^{\mathrm{u}}-\mathbf{P}\|}{\|\mathbf{P}\|}$	$\frac{\|\mathbf{P}^{\mathrm{s}}-\mathbf{P}\|}{\|\mathbf{P}\|}$
10^{-01}	$6.1\ 10^{-15}$	$7.4\ 10^{-15}$	$5.3\ 10^{-16}$	$1.1\ 10^{-15}$	$1.1\ 10^{-16}$	$2.6\ 10^{-16}$
10^{-02}	$6.7\ 10^{-12}$	$4.8\ 10^{-13}$	$6.9\ 10^{-14}$	$7.5\ 10^{-15}$	$1.1\ 10^{-15}$	$9.7\ 10^{-17}$
10^{-03}	$5.7\ 10^{-09}$	$4.3\ 10^{-17}$	$6.1\ 10^{-12}$	$0.0\ 10^{+00}$	$1.1\ 10^{-14}$	$9.9\ 10^{-17}$
10^{-04}	$4.9\ 10^{-06}$	$6.9\ 10^{-17}$	$5.2\ 10^{-10}$	$0.0\ 10^{+00}$	$9.8\ 10^{-14}$	$0.0\ 10^{+00}$
10^{-05}	$9.4\ 10^{-04}$	$1.2\ 10^{-16}$	$1.0\ 10^{-08}$	$0.0\ 10^{+00}$	$1.8\ 10^{-13}$	$2.1\ 10^{-21}$
10^{-06}	$2.4\ 10^{+00}$	$2.1\ 10^{-16}$	$2.6\ 10^{-06}$	$5.9\ 10^{-17}$	$4.9\ 10^{-12}$	$9.9\ 10^{-17}$
10^{-07}	$2.8\ 10^{+03}$	$3.0\ 10^{-17}$	$3.0\ 10^{-04}$	$8.3\ 10^{-17}$	$5.7\ 10^{-11}$	$9.9\ 10^{-17}$
10^{-08}	$2.8\ 10^{+06}$	$0.0\ 10^{+00}$	$3.0\ 10^{-02}$	$8.3\ 10^{-17}$	$5.7\ 10^{-10}$	$9.9\ 10^{-17}$
10^{-09}	$2.2\ 10^{+09}$	$3.0\ 10^{-17}$	$2.4\ 10^{+00}$	$5.9\ 10^{-17}$	$4.5\ 10^{-09}$	$9.9\ 10^{-17}$
10^{-10}	$7.2\ 10^{+11}$	$0.0\ 10^{+00}$	$7.7\ 10^{+01}$	$0.0\ 10^{+00}$	$1.4\ 10^{-08}$	$0.0\ 10^{+00}$
10^{-11}	$7.2\ 10^{+13}$	$0.0\ 10^{+00}$	$7.7\ 10^{+02}$	$1.2\ 10^{-27}$	$1.4\ 10^{-08}$	$2.0\ 10^{-27}$
10^{-12}	$1.9\ 10^{+18}$	$2.1\ 10^{-16}$	$2.0\ 10^{+06}$	$5.9\ 10^{-17}$	$3.8\ 10^{-06}$	$9.9\ 10^{-17}$
10^{-13}	$2.7\ 10^{+21}$	$2.1\ 10^{-16}$	$2.9\ 10^{+08}$	$5.9\ 10^{-17}$	$5.4\ 10^{-05}$	$9.9\ 10^{-17}$
10^{-14}	$6.9\ 10^{+23}$	$6.2\ 10^{-31}$	$7.4\ 10^{+09}$	$0.0\ 10^{+00}$	$1.4\ 10^{-04}$	$1.9\ 10^{-30}$
10^{-15}	$6.9\ 10^{+25}$	$3.0\ 10^{-17}$	$7.4\ 10^{+10}$	$1.4\ 10^{-31}$	$1.4\ 10^{-04}$	$0.0\ 10^{+00}$

matrix $\mathbf{A}$ and a corresponding auxiliary matrix $\mathbf{Z}$ as follows.

$$\mathbf{A} := \begin{bmatrix} a+\varepsilon & -2\varepsilon & b \\ 2\varepsilon & a+\varepsilon & c \\ 0 & 0 & a+d\varepsilon \end{bmatrix} \Rightarrow \mathtt{eigs}(\mathbf{A}) := \begin{bmatrix} \alpha \pm \mathrm{i}\beta \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} a+\varepsilon \pm \mathrm{i}2\varepsilon \\ a+d\varepsilon \end{bmatrix} \tag{6.1}$$

$$\mathbf{Z} := (\mathbf{A}-\alpha\mathbf{I})^2+\beta^2\mathbf{I} = \varepsilon \begin{bmatrix} 0 & 0 & b(d-1)+c\varepsilon \\ 0 & 0 & c(d-1)+b\varepsilon \\ 0 & 0 & d(d-2)\varepsilon+5\varepsilon \end{bmatrix} \Rightarrow \begin{matrix} \mathbf{Z}\mathbf{A} = (a+d\varepsilon)\varepsilon\mathbf{Z} \\ \mathbf{Z}\mathbf{A}^2 = (a+d\varepsilon)^2\varepsilon\mathbf{Z} \end{matrix} \tag{6.2}$$

where $\mathtt{eigs}(\mathbf{A})$ represents the eigenvalues of $\mathbf{A}$. We can drive all the eigenvalues and/or the gap betwen them to zero by appropriately choosing the parameters a and d. For each $\mathbf{A}$ we compute $\mathbf{R}(\tau,\mathbf{A})$ using the stable formulas summarized in the previous section. The matrices $\mathbf{P}(\tau,\mathbf{A})$ and $\mathbf{Q}(\tau,\mathbf{A})$ are computed from $\mathbf{R}(\tau,\mathbf{A})$ using the relationships given in Eqs. (3.16) and (3.18).

Table 3 illustrates the relative errors in the computations of the matrices $\mathbf{P}(\tau,\mathbf{A})$, $\mathbf{Q}(\tau,\mathbf{A})$ and $\mathbf{R}(\tau,\mathbf{A})$ choosing $\tau = 1$, $a = 0$, $b = c = 1$ and $d = 4$. Therein $\mathbf{R}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{s}}$ denotes the usual (naïve) and the stable (piecewise) computations of the matrix $\mathbf{R}$, respectively. The Frobenius norm is used in $\|\mathbf{R}\|$. The relative errors $\|\mathbf{P}^{\mathrm{s}}-\mathbf{P}\|/\|\mathbf{P}\|$, $\|\mathbf{Q}^{\mathrm{s}}-\mathbf{Q}\|/\|\mathbf{Q}\|$ and $\|\mathbf{R}^{\mathrm{s}}-\mathbf{R}\|/\|\mathbf{R}\|$ are found to be within the guaranteed computation accuracies established for the same and reflect the robustness of the stable formulas. The gradual loss of significance as $\varepsilon \to 0$ is reflected as a gradual increase in the relative error (from machine epsilon to values intolerably high) in the usual computations of the considered matrices. The maximum relative error in the computations of $\mathbf{P}^{\mathrm{u}}$, $\mathbf{Q}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{u}}$ are of the order of 10^{-4}, 10^{10} and 10^{25}, respectively. As $\varepsilon \to 0$ we observe $(\|\mathbf{P}^{\mathrm{u}}-\mathbf{P}\|/\|\mathbf{P}\|)$ is $O(\varepsilon)$ times smaller than $(\|\mathbf{Q}^{\mathrm{u}}-\mathbf{Q}\|/\|\mathbf{Q}\|)$ which in turn

TABLE 4
Relative errors in the usual and stable computation when $\lambda = \{\varepsilon, 2\varepsilon, 3\varepsilon\}$.

ε	$\frac{\|\mathbf{R}^{\mathrm{u}}-\mathbf{R}\|}{\|\mathbf{R}\|}$	$\frac{\|\mathbf{R}^{\mathrm{s}}-\mathbf{R}\|}{\|\mathbf{R}\|}$	$\frac{\|\mathbf{Q}^{\mathrm{u}}-\mathbf{Q}\|}{\|\mathbf{Q}\|}$	$\frac{\|\mathbf{Q}^{\mathrm{s}}-\mathbf{Q}\|}{\|\mathbf{Q}\|}$	$\frac{\|\mathbf{P}^{\mathrm{u}}-\mathbf{P}\|}{\|\mathbf{P}\|}$	$\frac{\|\mathbf{P}^{\mathrm{s}}-\mathbf{P}\|}{\|\mathbf{P}\|}$
10^{-01}	$3.9\ 10^{-14}$	$7.1\ 10^{-14}$	$1.7\ 10^{-15}$	$3.2\ 10^{-15}$	$5.4\ 10^{-16}$	$4.0\ 10^{-16}$
10^{-02}	$1.0\ 10^{-11}$	$7.1\ 10^{-14}$	$1.2\ 10^{-13}$	$2.1\ 10^{-14}$	$1.0\ 10^{-15}$	$2.1\ 10^{-15}$
10^{-03}	$1.1\ 10^{-07}$	$1.3\ 10^{-16}$	$6.8\ 10^{-11}$	$7.8\ 10^{-17}$	$6.6\ 10^{-14}$	$1.1\ 10^{-16}$
10^{-04}	$3.8\ 10^{-05}$	$1.3\ 10^{-16}$	$1.4\ 10^{-09}$	$7.9\ 10^{-17}$	$9.3\ 10^{-14}$	$8.2\ 10^{-17}$
10^{-05}	$3.5\ 10^{-02}$	$5.2\ 10^{-17}$	$4.2\ 10^{-07}$	$0.0\ 10^{+00}$	$8.1\ 10^{-12}$	$0.0\ 10^{+00}$
10^{-06}	$8.0\ 10^{+01}$	$0.0\ 10^{+00}$	$4.4\ 10^{-05}$	$5.5\ 10^{-17}$	$5.8\ 10^{-11}$	$0.0\ 10^{+00}$
10^{-07}	$4.5\ 10^{-02}$	$1.7\ 10^{-16}$	$6.7\ 10^{-05}$	$7.9\ 10^{-17}$	$7.8\ 10^{-11}$	$8.2\ 10^{-17}$
10^{-08}	$4.0\ 10^{+07}$	$5.2\ 10^{-17}$	$4.7\ 10^{-01}$	$0.0\ 10^{+00}$	$7.5\ 10^{-09}$	$8.2\ 10^{-17}$
10^{-09}	$4.0\ 10^{+10}$	$0.0\ 10^{+00}$	$5.5\ 10^{+01}$	$5.5\ 10^{-17}$	$1.0\ 10^{-07}$	$8.2\ 10^{-17}$
10^{-10}	$3.4\ 10^{-01}$	$3.0\ 10^{-17}$	$8.3\ 10^{-02}$	$0.0\ 10^{+00}$	$1.5\ 10^{-08}$	$0.0\ 10^{+00}$
10^{-11}	$3.4\ 10^{-01}$	$3.0\ 10^{-17}$	$8.3\ 10^{-02}$	$0.0\ 10^{+00}$	$1.5\ 10^{-08}$	$0.0\ 10^{+00}$
10^{-12}	$4.0\ 10^{+19}$	$3.0\ 10^{-17}$	$5.5\ 10^{+07}$	$7.9\ 10^{-17}$	$1.0\ 10^{-04}$	$8.2\ 10^{-17}$
10^{-13}	$4.0\ 10^{+22}$	$0.0\ 10^{+00}$	$5.5\ 10^{+09}$	$5.5\ 10^{-17}$	$1.0\ 10^{-03}$	$8.2\ 10^{-17}$
10^{-14}	$8.0\ 10^{+25}$	$3.0\ 10^{-17}$	$5.5\ 10^{+11}$	$0.0\ 10^{+00}$	$8.0\ 10^{-03}$	$8.2\ 10^{-17}$
10^{-15}	$8.0\ 10^{+28}$	$4.2\ 10^{-17}$	$5.5\ 10^{+13}$	$0.0\ 10^{+00}$	$6.1\ 10^{-02}$	$1.1\ 10^{-16}$

is $O(\varepsilon)$ times smaller than $(\|\mathbf{R}^{\mathrm{u}} - \mathbf{R}\|/\|\mathbf{R}\|)$. The following results explain this behaviour.

$$\mathbf{Z}\mathbf{A} = 4\varepsilon^2\mathbf{Z}, \quad \mathbf{Z}\mathbf{A}^2 = 16\varepsilon^3\mathbf{Z} \tag{6.3}$$

$$\mathbf{R}^{\mathrm{u}} - \mathbf{R} \approx \big(\Phi^{\mathrm{u}}(\star, 2\varepsilon)[-\varepsilon; -\varepsilon; 3\varepsilon] - \Phi(\star, 2\varepsilon)[-\varepsilon; -\varepsilon; 3\varepsilon]\big)\mathbf{Z} \tag{6.4}$$

$$\mathbf{Q}^{\mathrm{u}} - \mathbf{Q} = (\mathbf{R}^{\mathrm{u}} - \mathbf{R})\mathbf{A}, \quad \mathbf{P}^{\mathrm{u}} - \mathbf{P} = (\mathbf{R}^{\mathrm{u}} - \mathbf{R})\mathbf{A}^2 \tag{6.5}$$

Equation (6.4) holds because $\Phi(\star, 2\varepsilon)[-\varepsilon; -\varepsilon; 3\varepsilon]$ is the highest-order divided difference term in Eq. 4.25 and its computation error dominantes over the rest. The relative error in the usual (naïve) computation of $\Phi(\star, 2\varepsilon)[-\varepsilon; -\varepsilon; 3\varepsilon]$ is approximately 10^{40} when $\varepsilon = 10^{-15}$. Recall that the matrices $\mathbf{P}(\tau, \mathbf{A})$ and $\mathbf{Q}(\tau, \mathbf{A})$ govern the evolution of the particle positions. Likewise, the matrices $\mathbf{Q}(\tau, \mathbf{A})$ and $\mathbf{R}(\tau, \mathbf{A})$ govern the evolution of the particle velocities.

6.2. Example 2. Here the details differ from the previous example only in the definitions of the matrices $\mathbf{A}$ and $\mathbf{Z}$.

$$\mathbf{A} := \begin{bmatrix} a+\varepsilon & b & c \\ 0 & a+2\varepsilon & d \\ 0 & 0 & a+3\varepsilon \end{bmatrix} \Rightarrow \mathtt{eigs}(\mathbf{A}) := \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} a+\varepsilon \\ a+2\varepsilon \\ a+3\varepsilon \end{bmatrix} \tag{6.6}$$

$$\mathbf{Z} := (\mathbf{A} - \lambda_1\mathbf{I})(\mathbf{A} - \lambda_2\mathbf{I}) = \begin{bmatrix} 0 & 0 & bd+ch \\ 0 & 0 & 2dh \\ 0 & 0 & 2h^2 \end{bmatrix} \Rightarrow \begin{matrix} \mathbf{Z}\mathbf{A} = (a+3\varepsilon)\mathbf{Z} \\ \mathbf{Z}\mathbf{A}^2 = (a+3\varepsilon)^2\mathbf{Z} \end{matrix} \tag{6.7}$$

By construction, all the eigenvalues and the gap between them can be driven to zero with decreasing values of ε for appropriate choice of the parameter a.

Table 4 illustrates the relative errors in the computations of the matrices $\mathbf{P}(\tau, \mathbf{A})$, $\mathbf{Q}(\tau, \mathbf{A})$ and $\mathbf{R}(\tau, \mathbf{A})$ choosing $\tau = 1$, $a = 0$ and $b = c = d = 1$. The behaviour of the usual and stable computations are similar to what is observed in the previous example. The maximum relative error in the usual computations of $\mathbf{P}^{\mathrm{u}}$, $\mathbf{Q}^{\mathrm{u}}$ and $\mathbf{R}^{\mathrm{u}}$ are of the order of 10^{-2}, 10^{13} and 10^{28}, respectively. As before $(\|\mathbf{P}^{\mathrm{u}} - \mathbf{P}\|/\|\mathbf{P}\|)$ is $O(\varepsilon)$ times smaller than $(\|\mathbf{Q}^{\mathrm{u}} - \mathbf{Q}\|/\|\mathbf{Q}\|)$

which in turn is $O(\varepsilon)$ times smaller than $(\|\mathbf{R}^{\mathrm{u}}-\mathbf{R}\|/\|\mathbf{R}\|)$. The following results explain this behaviour.

$$\mathbf{ZA}=3\varepsilon\mathbf{Z},\quad \mathbf{ZA}^2=9\varepsilon^2\mathbf{Z} \tag{6.8}$$

$$\mathbf{R}^{\mathrm{u}}-\mathbf{R}\approx\big(\mathrm{r}^{\mathrm{u}}[\varepsilon;2\varepsilon;3\varepsilon]-\mathrm{r}[\varepsilon;2\varepsilon;3\varepsilon]\big)\mathbf{Z} \tag{6.9}$$

$$\mathbf{Q}^{\mathrm{u}}-\mathbf{Q}=(\mathbf{R}^{\mathrm{u}}-\mathbf{R})\mathbf{A},\quad \mathbf{P}^{\mathrm{u}}-\mathbf{P}=(\mathbf{R}^{\mathrm{u}}-\mathbf{R})\mathbf{A}^2 \tag{6.10}$$

Equation (6.9) holds because $\mathrm{r}[\varepsilon;2\varepsilon;3\varepsilon]$ is the highest-order divided difference term in Eq. (4.14) and its computation error dominates over the rest. The relative error in the usual (naïve) computation of $\mathrm{r}[\varepsilon;2\varepsilon;3\varepsilon]$ is approximately 10^{28} when $\varepsilon=10^{-15}$.

7. Conclusions. Formula computations in the neighbourhood of removable singularities suffer loss of significance when they are done using finite precision arithmetic. Formulas for the solution of the X-IVAS scheme involve many removable singularities. Hence, the use of numerically stable formulas for the same is a criteria for robustness.

We have proposed numerically stable formulas for the closed-form analytical solution of the X-IVAS scheme. Therein, the Newton form of the polynomial interpolation definition is used for the functions of matrices which appear in the formulas. In this definition, removable singularities and the terms/expressions that participate to yield a finite limit at these points are grouped together as divided differences. In other algebraically equivalent forms, these terms/expressions get dispersed. The poor reputation of nearly confluent divided differences with respect to the loss of significance in floating point computations is a blessing in disguise. We get *a priori* warning about possible numerical instabilities in formula computations. To control the loss of significance, we have presented piecewise definitions for these divided differences. The piecewise definitions switch the computations to the respective series approximations of the divided differences should the gap between the independent variables be less than a specified threshold. These divided differences are expressible as the divided difference of the exponential function of an appropriate order less than or equal to four. For the terms involving the second, third and fourth order divided differences, the double precision floating-point computation of their piecewise definitions guarantee at least 14, 12 and 10 significant decimal digits to be exact, respectively. The implementation of these piecewise definitions is simple and the computations are stable.

8. Acknowledgement. I thank Mr. Guillermo Casas-González for reading the manuscript in draft form and suggesting improvements.

REFERENCES

[1] MARCO CALIARI, *Accurate evaluation of divided differences for polynomial interpolation of exponential propagators*, Computing, 80 (2007), pp. 189–201.

[2] MARCO CALIARI, ALEXANDER OSTERMANN, AND STEFAN RAINER, *Meshfree Exponential Integrators*, SIAM Journal on Scientific Computing, 35 (2013), pp. A431–A452.

[3] DARIN P. DIACHIN AND JAMES A. HERZOG, *Analytic streamline calculations on linear tetrahedra*, in 13th Computational Fluid Dynamics Conference, Reston, Virigina, June 1997, American Institute of Aeronautics and Astronautics, pp. 733–742.

[4] FREEMAN GILBERT AND GEORGE E. BACKUS, *Propagator matrices in elastic wave and vibration problems*, Geophysics, 31 (1966), pp. 326–332.

[5] OPENCFD LTD (ESI GROUP), *OpenFOAM—The open source CFD toolbox*. `http://www.openfoam.org/`.

[6] NICHOLAS J. HIGHAM, *Accuracy and Stability of Numerical Algorithms*, Society for Industrial and Applied Mathematics, Jan. 2002.

[7] ———, *Functions of Matrices: Theory and Computation*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2008.

[8] MARLIS HOCHBRUCK, CHRISTIAN LUBICH, AND HUBERT SELHOFER, *Exponential Integrators for Large Systems of Differential Equations*, SIAM Journal on Scientific Computing, 19 (1998), pp. 1552–1574.

[9] SERGIO R. IDELSOHN, JULIO MARTI, PABLO BECKER, AND EUGENIO OÑATE, *Analysis of multifluid flows with large time steps using the particle finite element method*, International Journal for Numerical Methods in Fluids, 75 (2014), pp. 621–644.

[10] SERGIO R. IDELSOHN, NORBERTO NIGRO, ALEJANDRO LIMACHE, AND EUGENIO OÑATE, *Large time-step explicit integration method for solving problems with dominant convection*, Computer Methods in Applied Mechanics and Engineering, 217-220 (2012), pp. 168–185.

[11] SERGIO R. IDELSOHN, EUGENIO OÑATE, AND FACUNDO DEL PIN, *The particle finite element method: a powerful tool to solve incompressible flows with free-surfaces and breaking waves*, International Journal for Numerical Methods in Engineering, 61 (2004), pp. 964–989.

[12] WILLAM KAHAN AND JOSEPH D. DARCY, *How Java's Floating-Point Hurts Everyone Everywhere*. ACM 1998 Workshop on Java for High-Performance Network Computing. `http://www.cs.berkeley.edu/~wkahan/JAVAhurt.pdf`, 1998. Online; accessed: 05-02-2014.

[13] PETER KIPFER, FRANK RECK, AND GUNTHER GREINER, *Local Exact Particle Tracing on Unstructured Grids*, Computer Graphics Forum, 22 (2003), pp. 133–142.

[14] ALLAN CHARLES MCCURDY, KWOK CHOI NG, AND BERESFORD NEILL PARLETT, *Accurate computation of divided differences of the exponential function*, Mathematics of Computation, 43 (1984), pp. 501–501.

[15] GREGORY M. NIELSON AND IL-HONG JUNG, *Tools for computing tangent curves for linearly varying vector fields over tetrahedral domains*, IEEE Transactions on Visualization and Computer Graphics, 5 (1999), pp. 360–372.

[16] EUGENIO OÑATE, SERGIO R. IDELSOHN, FACUNDO DEL PIN, AND ROMAIN AUBRY, *The particle finite element method. An overview.*, International Journal of Computational Methods, 1 (2004), pp. 267–307.

[17] ALEXANDER OSTERMANN, MECHTHILD THALHAMMER, AND WILLIAM M. WRIGHT, *A Class of Explicit Exponential General Linear Methods*, BIT Numerical Mathematics, 46 (2006), pp. 409–431.

[18] BERESFORD NEILL PARLETT, *A recurrence among the elements of functions of triangular matrices*, Linear Algebra and its Applications, 14 (1976), pp. 117–121.

[19] JAMES F. PRICE, *Lagrangian and Eulerian Representations of Fluid Flow: Kinematics and the Equations of Motion*. Woods Hole Oceanographic Institution. `http://www.whoi.edu/science/PO/people/jprice/class/ELreps.pdf`, 2006. Online; accessed: 10-12-2012.

[20] ERIC W. WEISSTEIN, *Cubic Formula*. From MathWorld—A Wolfram Web Resource. `http://mathworld.wolfram.com/CubicFormula.html`. Online; accessed: 23-10-2013.