Uniformly accurate numerical schemes for a class of dissipative problems

Philippe Chartier[1] Mohammed Lemou[2] Léopold Trémant[3]

[1] Univ Rennes, Inria - IRMAR - ENS Rennes

[2] Univ Rennes, CNRS - IRMAR - ENS Rennes

[3] Univ Rennes, Inria - IRMAR

CAN-J, 03/12/2020

Sommaire

Outline

Setting of the problem

We want to solve

$$\left\{\begin{array}{lll} \dot{x} = & a(x,z), & x(0) = x_0 \in \mathbb{R}^{d_x} \\ \dot{z} = & -\dfrac{\Lambda}{\varepsilon} z + b(x,z), & z(0) = z_0 \in \mathbb{R}^{d_z} \end{array}\right. \qquad (E_{xz})$$

on $[0, T]$ for $\varepsilon \in]0, \varepsilon^*]$ (not necessarily small), where a, b are analytic. The operator Λ is a relaxation term (from population dynamics, kinetic collisions, diffusion, etc).

Setting $u = \begin{pmatrix} x \\ z \end{pmatrix} \in \mathbb{R}^d$, this is written in the more compact form

$$\dot{u} = -\frac{A}{\varepsilon} u + f(u), \qquad u(0) = u_0. \qquad (E_u)$$

Assumptions

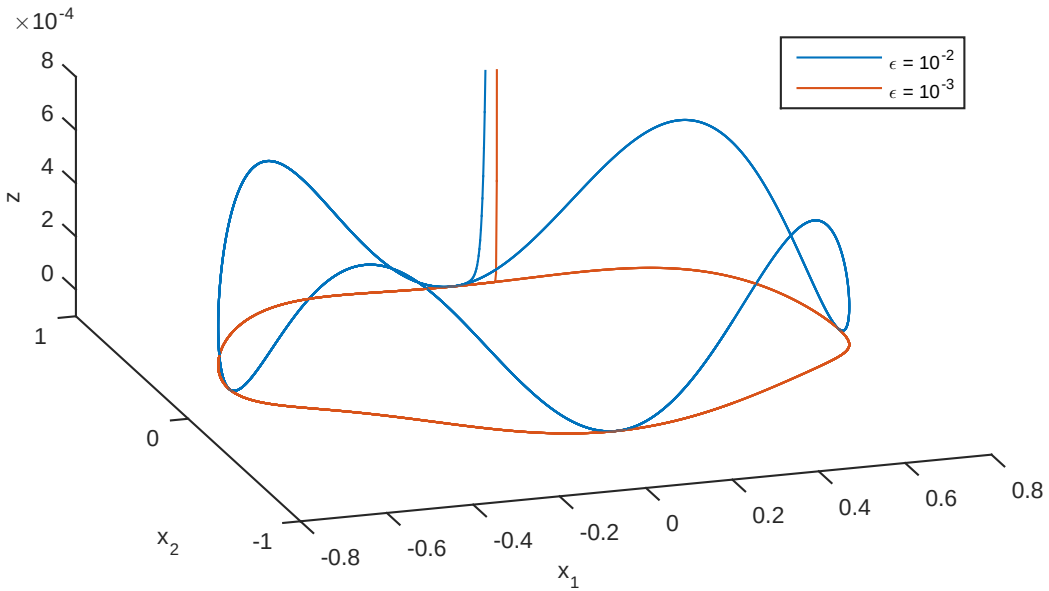
1. The vector field $u \mapsto f(u)$ is analytic;
2. The operator A is diagonal positive semi-definite;
3. The operator A has *integer* eigenvalues.

Behaviour

Center manifold theorem:
There exists a map $x \mapsto z = \varepsilon h^\varepsilon(x)$ such that the solution (x, z) of (E_{xz}) satisfies

$$\left|z(t) - \varepsilon h^\varepsilon\left(x(t)\right)\right| \lesssim e^{-t/\varepsilon}.$$

Therefore, for $t \gtrsim \varepsilon \log(1/\varepsilon)$, the dynamics are *non-stiff* and z is of size $\mathcal{O}(\varepsilon)$.



Consequence: Given a numerical error $e_i = u(t_i) - u_i$, the norm considered is

$$\boxed{|e_i|_\varepsilon := \left|\left(\mathrm{id} + \frac{A}{\varepsilon}\right) e_i\right|}$$

which rescales the z-component by $1/\varepsilon$ but makes no difference on the x-component.

Exponential RK schemes

Considering the original problem

$$\dot{u} = -\frac{1}{\varepsilon} A u + f(u), \qquad (E_u)$$

the stiff linear part is obviously a problem for explicit schemes.
However, IMEX schemes seem more appropriate. We focus on *exponential RK schemes*[1] which stem from the integral form

$$u(t_{i+1}) \approx e^{-\frac{\Delta t}{\varepsilon} A} u_i + \int_0^{\Delta t} e^{-\frac{1}{\varepsilon}(\Delta t - s)A} f\big(u(s)\big) \mathrm{d}s$$

For example the 1st-order scheme is

$$\boxed{u_{i+1} = e^{-\frac{\Delta t}{\varepsilon} A} u_i + \left(\int_0^{\Delta t} e^{-\frac{1}{\varepsilon}(\Delta t - s)A} \mathrm{d}s \right) f(u_i)}$$

This method, and higher order ones can be found in

- M. Hochbruck, A. Ostermann, *Explicit exponential Runge–Kutta methods for semilinear parabolic problems*. SIAM Journal on Numerical Analysis (2005)

[1] One could also consider IMEX-BDF schemes, for instance the first-order $\frac{1}{\Delta t}(u_{i+1} - u_i) = -\frac{1}{\varepsilon} A u_{i+1} + f(u_i)$. Lawson schemes, such as $u_{i+1} = \exp(-\frac{\Delta t}{\varepsilon} A) u_i + \Delta t \exp(-\frac{\Delta t}{\varepsilon} A) f(u_i)$ could also be studied. However for the former, the *uniform* convergence is of lower order, and the later has little to no literature on precise error bounds.

Order reduction – illustration

$$\begin{cases} \dot{x} = (1-z)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x, \\ \dot{z} = -\dfrac{1}{\varepsilon} z + x_1^2 x_2^2, \end{cases} \qquad \text{i.e.} \quad \partial_t \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = -\frac{1}{\varepsilon}\begin{pmatrix} 0 \\ 0 \\ u_3 \end{pmatrix} + \begin{pmatrix} -(1-u_3)u_2 \\ (1-u_3)u_1 \\ (u_1 u_2)^2 \end{pmatrix}.$$

Solving this problem using exponential Runge-Kutta schemes, the convergence is

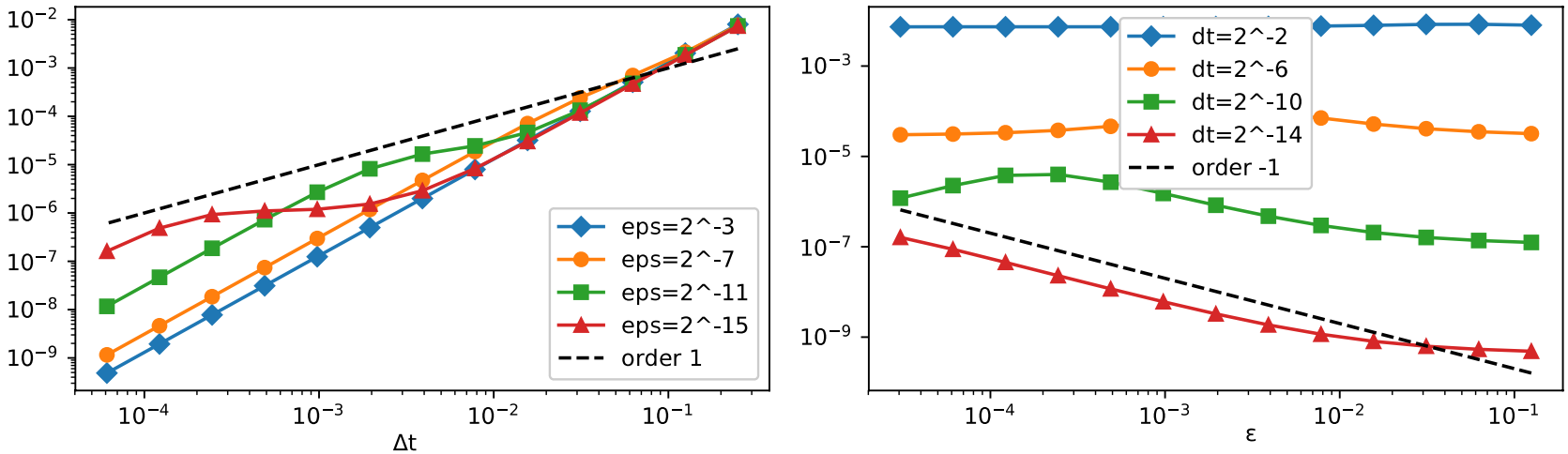


Figure: Errors on u^ε depending on Δt (left) and on ε (right) using an exponential RK2 scheme.

It appears that the scheme is asymptotic preserving[2] but not uniformly accurate.

[2] At fixed Δt, the scheme admits a limit $\varepsilon \searrow 0$ and this new scheme has the same order of convergence as before.

Order reduction – definition

Considering the original problem

$$\dot{u} = -\frac{1}{\varepsilon} A u + f(u), \qquad (E_u)$$

and a discretisation (t_i) from which the approximation $u_i \approx u(t_i)$ is computed, one observes the phenomenon of order reduction.

Definition: Order reduction

Solving this equation numerically with a scheme of order p at fixed ε, the error is bounded

$$\sup_i \left| u(t_i) - u_i \right|_\varepsilon \leq C_\varepsilon \Delta t^p$$

where $C_\varepsilon > 0$ is such that the *uniform error* is degraded, i.e.

$$\sup_\varepsilon \left\{ \sup_i |u(t_i) - u_i|_\varepsilon \right\} \leq C \Delta t^q \qquad \text{with } 0 \leq q \leq p.$$

For instance, exponential RK schemes reduce to order 1.

A method is said to be of *uniform accuracy* if one can take $q = p$.

Main result

We shall demonstrate how to find a micro-macro decomposition

$$u(t) = \Omega^{[n]}_{t/\varepsilon}(v(t)) + w(t)$$

where $\Omega^{[n]}_\tau(u) = e^{-\tau A}u + \mathcal{O}(\varepsilon)$, $\dot{v} = F^{[n]}(v)$ is non-stiff and $w = \mathcal{O}(\varepsilon^{n+1})$.

The (v, w) problem can be solved using an exponential RK scheme of order $n+1$, and the convergence is uniform of order $n+1$, i.e.

$$\left|u(t_i) - u_i\right|_\varepsilon = \left|\left(\mathrm{id} + \frac{1}{\varepsilon}A\right)\left(u(t_i) - u_i\right)\right| \leq C\Delta t^{n+1}$$

with C independent of ε and $u_i = \Omega^{[n]}_{t_i/\varepsilon}(v_i) + w_i$.

This result applies straightforwardly to ODEs, however with some care it is extended *partially* to some PDEs with ad-hoc adjustments.

Outline

The homological equation

Assume that one can write

$$u(t) = \Omega^{\varepsilon}_{t/\varepsilon}\left(v^{\varepsilon}(t)\right) + \mathcal{O}(\varepsilon^{\infty}) \quad \text{with} \quad \dot{v} = F^{\varepsilon}(v), \qquad (H_{\infty})$$

injecting this in (E_u), $\dot{u} = -\frac{1}{\varepsilon}Au + f(u)$ yields the homological equation

$$\partial_{\tau}\Omega^{\varepsilon}_{\tau}(u) + A\Omega^{\varepsilon}_{\tau}(u) = \varepsilon\Big(f \circ \Omega^{\varepsilon}_{\tau}(u) - \partial_u\Omega^{\varepsilon}_{\tau}(u)F^{\varepsilon}(u)\Big)$$

which can be solved recursively by setting

$$\big(\partial_{\tau} + A\big)\Omega^{[n+1]}_{\tau} = \varepsilon\Big(f \circ \Omega^{[n]}_{\tau} - \partial_u\Omega^{[n]}_{\tau} \cdot F^{[n]}\Big) \quad \text{and} \quad \Omega^{[0]}_{\tau} = e^{-\tau A}. \qquad (H_n)$$

Assuming that we have a projector $\langle \cdot \rangle$ parallel to the image of $\partial_{\tau} + A$, then

$$F^{[n]} = \big\langle \partial_u\Omega^{[n]} \big\rangle^{-1} \big\langle f \circ \Omega^{[n]} \big\rangle.$$

Is it therefore possible to compute $\Omega^{[n]}$ and $F^{[n]}$ up to any rank?

Solving iterations

We now want to solve

$$(\partial_\tau + A)\Omega_\tau^{[n+1]} = \varepsilon\Big(f \circ \Omega_\tau^{[n]} - \partial_u \Omega_\tau^{[n]} \cdot F^{[n]}\Big) \quad \text{and} \quad \Omega_\tau^{[0]} = e^{-\tau A}. \qquad (H_n)$$

For the first iteration, solving (H_1) yields

$$\Omega_\tau^{[1]} = e^{-\tau A}\Omega_0^{[1]} + \varepsilon \int_0^\tau e^{-(\tau-\sigma)A}\left(f \circ e^{-\sigma A} - e^{-\sigma A}F^{[0]}\right) \mathrm{d}\sigma.$$

$\Rightarrow$ In order to have only exponentials, $F^{[0]}$ is the $e^{-\sigma A}$-component of $f \circ e^{-\sigma A}$.

The projector $\langle \cdot \rangle$ basically extracts the $e^{-\tau A}$-component of any exponential map.

Reminder – Assumptions

1. The vector field $u \mapsto f(u)$ is analytic;
2. The operator A is diagonal positive semi-definite;
3. The operator A has *integer* eigenvalues.

The closure conditions – discussion

Because $\Omega^{[n]}$ is defined up to an $e^{-\tau A}$-component, at every iteration it can be written

$$\Omega_\tau^{[n]} = e^{-\tau A}\langle \Omega^{[n]} \rangle + \sum_{k \neq A} e^{-k\tau} \widehat{\Omega}_k^{[n]}$$

with $\widehat{\Omega}_k^{[n]} = \mathcal{O}(\varepsilon)$ given by (H_n) and $\langle \Omega^{[n]} \rangle$ a degree of freedom. A choice is needed, since the "slow" vector field $F^{[n]}$ is given by

$$F^{[n]} = \left\langle \partial_u \Omega^{[n]} \right\rangle^{-1} \langle f \circ \Omega^{[n]} \rangle.$$

Two main choices[3] arise:

- Standard: The simplest choice $\langle \Omega^{[n]} \rangle = \mathrm{id}$, which yields $F^{[n]} = \langle f \circ \Omega^{[n]} \rangle$, meaning there is *no costly inversion needed*. This is in the vein of the Chapman-Enskog expansion (for kinetic equations).
- Stroboscopic: A more involved choice $\Omega_0^{[n]} = \mathrm{id}$. This presents better *geometric properties* (preservation of invariants and Hamiltonian structure), but requires a costly inversion and some additional work besides solving (H_n).

[3] P. Chartier, M. Lemou, F. Méhats, G. Vilmart. *A new class of uniformly accurate numerical schemes for highly-oscillatory evolution equations*, FoCM (2019)

Properties of the asymptotic approximation

Set $\eta_\tau^{[n]}$ the defect, i.e. the error of approximation of $\Omega^{[n]}$, $F^{[n]}$ in the "exact" homological equation (H_∞),

$$\eta_\tau^{[n]} = \frac{1}{\varepsilon}(\partial_\tau + A)\Omega_\tau^{[n]} - \left(f \circ \Omega_\tau^{[n]} - \partial_u \Omega_\tau^{[n]} \cdot F^{[n]}\right).$$

Properties of the dissipative approximation[4]

When computing $\Omega^{[n]}$ and $F^{[n]}$ with the standard or stroboscopic choice,

1. The change of variable has derivatives $\partial_\tau^\alpha \left(\Omega_\tau^{[n]} - e^{-\tau A}\right)$ analytic and of size ε
2. The vector field $F^{[n]}$ is analytic and bounded independently of ε (i.e. it is non-stiff)
3. The defect $\partial_\theta^\alpha \eta^{[n]}$ is analytic of size ε^n with no component $e^{-\tau A}$

Furthermore, for the standard choice the following identities are met

1. $\Omega_\tau^{[n]}(u) = \Omega_0^{[n]}(e^{-\tau A}u)$;
2. $e^{-\tau A}F^{[n]}(u) = F^{[n]}(e^{-\tau A}u)$;
3. $\eta_\tau^{[n]}(u) = \eta_0^{[n]}(e^{-\tau A}u)$.

[4] P. Chartier, M. Lemou, L. T, *Uniformly accurate numerical schemes for a class of dissipative systems*, submitted (2020)

Micro-macro decomposition

After decomposing the solution u of the original problem (E_u) into

$$u(t) = \Omega^{[n]}_{t/\varepsilon}(v(t)) + w(t) \quad \text{where} \quad \dot{u} = -\frac{1}{\varepsilon}Au + f(u),$$

we now study the micro-macro problem in (v, w) which is

$$\begin{cases} \dot{v} = F^{[n]}(v) \\ \dot{w} = -\frac{1}{\varepsilon}A\left(\Omega^{[n]}_{t/\varepsilon}(v) + w\right) + f\left(\Omega^{[n]}_{t/\varepsilon}(v) + w\right) - \frac{\mathrm{d}}{\mathrm{d}t}\Omega^{[n]}_{t/\varepsilon}(v) \end{cases}$$

with initial conditions $v(0) = (\Omega^{[k]}_0)^{-1}(u_0)$, $w(0) = 0$. The red part can be written

$$\frac{1}{\varepsilon}A\Omega^{[n]}_{t/\varepsilon}(v) + \frac{\mathrm{d}}{\mathrm{d}t}\Omega^{[n]}_{t/\varepsilon}(v(t)) = \eta^{[n]}_{t/\varepsilon}(v) + f \circ \Omega^{[n]}_{t/\varepsilon}(v)$$

therefore the problem becomes

$$\boxed{\begin{cases} \dot{v} = F^{[n]}(v) \\ \dot{w} = -\frac{1}{\varepsilon}Aw + f\left(\Omega^{[n]}_{t/\varepsilon}(v) + w\right) - f(\Omega^{[n]}_{t/\varepsilon}(v)) - \eta^{[n]}_{t/\varepsilon}(v) \end{cases}}$$

Well-posedness of the micro-macro problem

The micro-macro problem can therefore be written

$$\begin{aligned}\dot{v} &= F^{[n]}(v)\\ \dot{w} &= -\frac{1}{\varepsilon}Aw + L\left(\Omega^{[n]}_{t/\varepsilon}(v), w\right)w - \eta^{[n]}_{t/\varepsilon}(v)\end{aligned}$$

with the essentially linear part $L^{[n]}$ given by

$$L(v,w)w = f(v+w) - f(v)$$

Properties of the new system

Up to a final time independent of ε,

(i) $v^{[n]}$ is regular and bounded

(ii) $\|w^{[n]}\|_\infty = \mathcal{O}(\varepsilon^{n+1})$

(iii) $\|\partial_t{}^\alpha E^{[n]}\|_\infty = \mathcal{O}(\varepsilon^{n-\alpha})$

(iv) $\|\partial_t^{\alpha+1} E^{[n]}\|_{L^1} = \mathcal{O}(\varepsilon^{n-\alpha})$

where $E^{[n]} = \partial_t w^{[n]} + \frac{1}{\varepsilon}Aw^{[n]}$.

Well-posedness of the micro-macro problem

The micro-macro problem can therefore be written

$$\begin{aligned}\dot{v} &= F^{[n]}(v)\\ \dot{w} &= -\frac{1}{\varepsilon}Aw + L\left(\Omega^{[n]}_{t/\varepsilon}(v), w\right) w - \eta^{[n]}_{t/\varepsilon}(v)\end{aligned}$$

with the essentially linear part $L^{[n]}$ given by

$$L(v,w)w = \left(\int_0^1 \partial_u f\left(v + \mu w\right) \mathrm{d}\mu\right) w$$

Properties of the new system

Up to a final time independent of ε,

(i) $v^{[n]}$ is regular and bounded

(ii) $\|w^{[n]}\|_\infty = \mathcal{O}(\varepsilon^{n+1})$

(iii) $\|\partial_t{}^\alpha E^{[n]}\|_\infty = \mathcal{O}(\varepsilon^{n-\alpha})$

(iv) $\|\partial_t^{\alpha+1} E^{[n]}\|_{L^1} = \mathcal{O}(\varepsilon^{n-\alpha})$

where $E^{[n]} = \partial_t w^{[n]} + \frac{1}{\varepsilon} A w^{[n]}$.

Outline

Uniform accuracy of exponential schemes

Numerically, we solve the problem at times (t_i) with maximum time step Δt,

$$\dot{v} = F^{[n]}(v),$$
$$\dot{w} = -\frac{1}{\varepsilon} A w + \underbrace{L\left(\Omega^{[n]}_{t/\varepsilon}(v), w\right) w - \eta^{[n]}_{t/\varepsilon}(v)}_{E^{[n]}(t)}$$

Uniform accuracy of exponential RK schemes (P. Chartier, M. Lemou, L. T.)

Computing (v_i, w_i) with an exponential RK scheme of order $n+1$ and recovering an approximation $u_i = \Omega^{[n]}_{t_i/\varepsilon}(v_i) + w_i$ generates a uniform "modified" error on u of order $n+1$, i.e.

$$|u(t_i) - u_i|_\varepsilon \leq C \Delta t^{n+1}.$$

with C independent of ε and $0 \leq i \leq N$, and $|e|_\varepsilon = \left|\left(1 + \frac{1}{\varepsilon} A\right) e\right|$.

This stems from a result on exponential Runge-Kutta schemes in

- A. Hochbruck, M. Ostermann, *Exponential Runge–Kutta methods for parabolic problems*, App. Num. Math. (2004)

Oscillating example – expansion

We now look at the following "oscillating" system

$$\begin{cases} \dot{x} = (1-z)\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} x \\ \dot{z} = -\dfrac{1}{\varepsilon} z + x_1^2 x_2^2 \end{cases}$$

We compute the change of variable $\Omega^{[2]}$ and the vector field $F^{[2]}$ using symbolic computing, and obtain

$$\Omega_0^{[2]}(x,z) = \begin{pmatrix} x_1 - \varepsilon x_2 z - \frac{1}{2}\varepsilon^2 z^2 x_1 \\ x_2 + \varepsilon x_1 z - \frac{1}{2}\varepsilon^2 z^2 x_2 \\ z + \varepsilon (x_1 x_2)^2 - 2\varepsilon^2 x_1 x_2 (x_1^2 - x_2^2) \end{pmatrix},$$

$$F^{[2]}(x,z) = \begin{pmatrix} -x_2(1 - \varepsilon(x_1 x_2)^2 + 2\varepsilon^2 x_1 x_2 (x_1^2 - x_2^2)) \\ x_1(1 - \varepsilon(x_1 x_2)^2 + 2\varepsilon^2 x_1 x_2 (x_1^2 - x_2^2)) \\ 2\varepsilon z x_1 x_2 (x_1^2 - x_2^2) \end{pmatrix}.$$

and recover $\Omega_\tau^{[2]}(x,z) = \Omega_0^{[2]}(x, e^{-\tau} z)$, and $\eta_\tau^{[2]}$ satisfies

$$\eta_\tau^{[2]} = \frac{1}{\varepsilon}(\partial_\tau + A)\,\Omega_\tau^{[2]} + \partial_u \Omega_\tau^{[2]} \cdot F^{[2]} - f \circ \Omega_\tau^{[2]}.$$

Oscillating example – results

Finally, solving the micro-macro problem for $n = 1, 2$

$$\begin{cases} \dot{v} = F^{[n]}(v), \\ \dot{w} = -\dfrac{1}{\varepsilon} A w + L\left(\Omega^{[n]}_{t/\varepsilon}(v), w\right) w - \eta^{[n]}_{t/\varepsilon}(v) \end{cases}$$

with an ERK scheme and recovering $u_i = \Omega^{[n]}_{t_i/\varepsilon}(v_i) + w_i$, the error is summed up

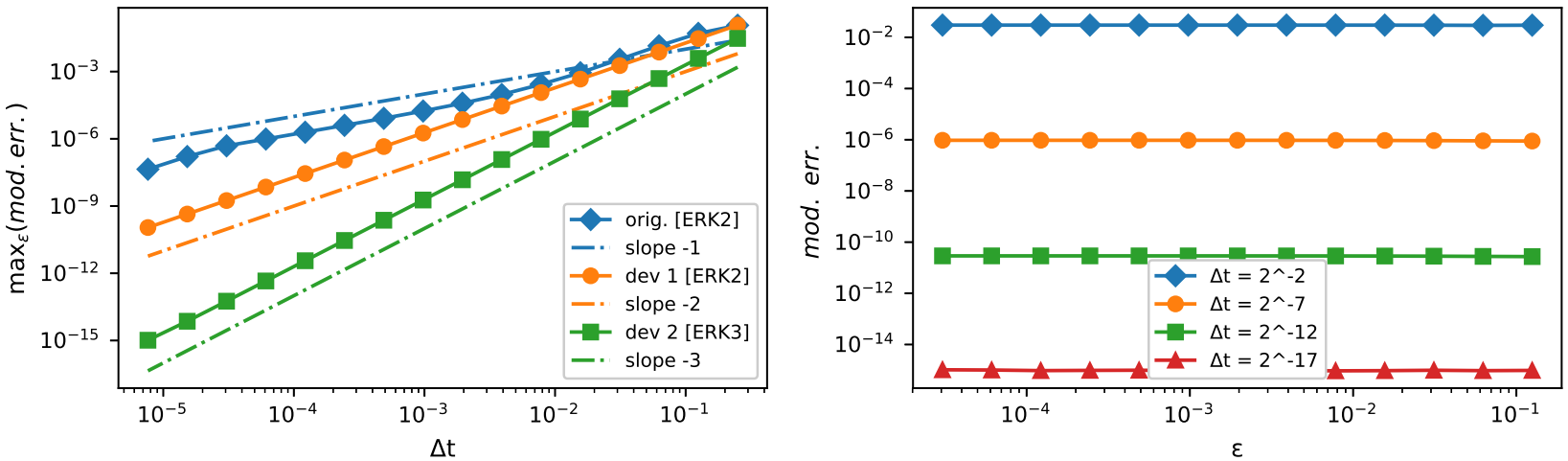


Figure: Maximum error for $\varepsilon = 2^{-k}$ ($k = 3, \ldots, 15$) as a function of Δt (left), and error as a function of ε for various values of Δt in the case $n = 2$ with an ERK3 scheme (right).

Outline

Prospects

- Extension to PDEs: Kinetic models with collisions (esp. BGK) can be written in this form, however the vector field then involves a differential operator, and the asymptotic developments lead to the apparition of a backwards heat equation. A workaround is possible using ad-hoc relaxations, but we hope to find a more robust approach.
- Near-equilibrium convergence: When choosing an initial condition $z_0 = 0$, the order of convergence improves. This is not surprising for the original problem, but this increase is also observed for the micro-macro model.
- Relax the assumption on A: Assuming that the eigenvalues of A are integers is a huge assumption, albeit a seemingly necessary one in the context of analytic function. If the vector field f were polynomial, this could probably be relaxed.
- Library: The micro-macro expansion can easily be computed up to any order using symbolic computing. A Python tool using SymPy would be an easy and useful addition to this result which would simplify implementations.

Merci pour votre attention !