Chapter 1

Introduction

For certain simulations it is of utmost importance to use methods that follow conservation laws, making sure that the model stays true to physical laws. While there are already established methods for this, none of them are as well known or as simple to set up as the Runge-Kutta method. Not only is the simple modification suggested in the article by Ranocha et al. [1] both simple and cheap, as we shall see it is also accurate.

Chapter 2

Preliminaries

2.1 ODE

We consider a time-dependent ODE of the form

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}u(t) &= f(t,u(t)), \quad t \in (0,T) \\ u(0) &= u^0\end{aligned}$$

for $u \in \mathcal{H}$ being a real Hilbert space with the inner product $\langle \cdot, \cdot \rangle$, inducing the norm $\|\cdot\|$.

2.2 Entropy

We denote by

$$\eta : \mathcal{H} \to \mathbb{R}$$

a smooth convex function in time. We call this entropy, while in other application it might instead represent some other form of non-increasing quantity, e.g. energy or momentum.

The change in entropy over time is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta(u(t)) = \langle \eta'(u(t)), f(t,u(t)) \rangle .$$

A entropy dissipative system will satisfy

$$\langle \eta'(u(t)), f(t,u(t)) \rangle \leq 0, \quad \forall u \in \mathcal{H}, t \in [0,T],$$

and a entropy conservative system will satisfy

$$\langle \eta'(u(t)), f(t,u(t)) \rangle = 0, \quad \forall u \in \mathcal{H}, t \in [0,T].$$

2.3 Classic Runge-Kutta

A general Runge-Kutta method with s stages is represented with the Butcher tableau

$$\begin{array}{c|c} c & A \\ \hline & b^{\mathrm{T}} \end{array},$$

with $A \in \mathbb{R}^{s \times s}$ and $b, c \in \mathbb{R}^s$.

The scheme of the method is

$$y_i = u^n + \Delta t \sum_{j=1}^{s} a_{i,j} f(t_n + c_j \Delta t, y_j), \quad i = 1, \ldots, s \tag{2.1}$$

$$u^{n+1} = u^n + \Delta t \sum_{i=1}^{s} b_i f(t_n + c_i \Delta t, y_i). \tag{2.2}$$

where y_i are the stage values.

For brevity we will use the notation

$$f_i := f(t_n + c_i \Delta t, y_i), \qquad f_0 := f(t_n, u^n).$$

Chapter 3

Relaxation Runge-Kutta Methods

3.1 Relaxation Runge-Kutta

The basic idea of the Relaxed Runge-Kutta method is to introduce a scaling factor to the weights b_i. We call this scaling factor $\gamma_n \in \mathbb{R}$ and construct a new scheme

$$u_\gamma^{n+1} = u^n + \gamma_n \Delta t \sum_{i=1}^{s} b_i f_i,$$

with $\gamma_n \in \mathbb{R}$.

Relaxed Runge-Kutta, RRK, interprets $u_\gamma^{n+1} \approx u(t_n + \gamma\Delta t)$. This means that using this method we will not be using a uniform step length and the number of steps might not be the same as with the regular RK method.

The incremental direction technique, or IDT-method, interprets $u_\gamma^{n+1} \approx u(t_n + \Delta t)$. This method retains the original step length, and takes as many steps to complete as the regular RK method it is based on.

In a earlier work one of the authors proposed choosing γ_n such that

$$\frac{\left\|u_\gamma^{n+1}\right\|^2 - \|u^n\|^2}{2} = \gamma_n \Delta t \sum_{i=1}^{s} b_i \langle y_i, f_i \rangle .$$

In this article however the suggestion is to instead use γ_n fulfilling the condition

$$\eta(u_\gamma^{n+1}) - \eta(u^n) = \gamma \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle .$$

This is done by finding a root of

$$r(\gamma) = \eta(u^n + \gamma_n \Delta t \sum_{i=1}^{s} b_i f_i) - \eta(u^n) - \gamma \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle . \tag{3.1}$$

The direction and entropy change

$$d^n := \sum_{i=1}^{s} b_i f_i$$

$$e := \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle$$

can both be computed on the fly during the RK method. This reduces finding the root of r to just a scalar root finding problem.

Note that since η is convex we have that r is convex in γ.

We note also that since

$$r(0) = \eta(u^n + 0) - \eta(u^n) - 0 = 0 \tag{3.2}$$

r has a root at 0.

Using a non-positive γ would not be feasible. A zero-valued γ would halt the scheme, while a negative γ would amount to steps backwards in time.

3.2 Existence of a solution

In order to show that r has a positive root we need two things. Firstly that r' is negative at the root $\gamma = 0$ and that r' is positive at some point $\gamma > 0$.

Lemma 3.2.1. *Let a Runge-Kutta method be given such that $\sum_{i=1}^{s} b_i a_{i,j} > 0$. If $n''(u^n)(f_0, f_0) > 0$ then $r'(0) < 0$ for sufficiently small $\Delta t > 0$.*

Note that $\sum_{i=1}^{s} b_i a_{i,j} > 0$ is a reasonable assumption, since $\sum_{i=1}^{s} b_i a_{i,j} = 1/2$ is a condition for second-order accuracy.

Proof. By the definition of $r(\gamma)$ in Equation 3.1,

$$\frac{dr}{d\gamma} = \eta'(u^n + \gamma \Delta t \sum_{i=1}^{s} b_i f_i) \cdot \left(\Delta t \sum_{i=1}^{s} b_i f_i \right) - \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle .$$

Evaluating $r'(\gamma)$ at $\gamma = 0$ we get

$$\begin{aligned}
r'(0) &= \eta'(u^n) \cdot \left(\Delta t \sum_{i=1}^{s} b_i f_i \right) - \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle \\
&= \Delta t \sum_{i=1}^{s} b_i \left(\langle \eta'(u^n), f_i \rangle - \langle \eta'(y_i), f_i \rangle \right) \\
&= -\Delta t \sum_{i=1}^{s} b_i \left(\langle \eta'(y_i), f_i \rangle - \langle \eta'(u^n), f_i \rangle \right).
\end{aligned}$$

We expand $y_i = u^n + \Delta t \sum_{j=1}^{s} a_{i,j} f_j$ as defined in Equation 2.1 and get

$$r'(0) = -\Delta t \sum_{i=1}^{s} b_i \left(\left\langle \eta' \left(u^n + \Delta t \sum_{j=1}^{s} a_{i,j} f_j \right), f_i \right\rangle - \langle \eta'(u^n), f_i \rangle \right).$$

Then, by the fundamental theorem of calculus,

$$\begin{aligned}
r'(0) &= -\Delta t \sum_{i=1}^{s} b_i \int_0^1 \eta'' \left(u^n + v \Delta t \sum_{k=1}^{s} a_{i,k} f_k \right) \left(f_i, \Delta t \sum_{j=1}^{s} a_{i,j} f_j \right) \mathrm{d}v \\
&= -\Delta t^2 \sum_{i,j=1}^{s} b_i a_{i,j} \int_0^1 \eta'' \left(u^n + v \Delta t \sum_{k=1}^{s} a_{i,k} f_k \right) (f_i, f_j) \, \mathrm{d}v.
\end{aligned}$$

With Taylor expansions of $f_i, f_j = f_0 + \mathcal{O}(\Delta t)$,

$$\begin{aligned}
r'(0) &= -\Delta t^2 \sum_{i,j=1}^{s} b_i a_{i,j} \int_0^1 \eta'' \left(u^n + v \Delta t \sum_{k=1}^{s} a_{i,k} f_k \right) (f_0, f_0 + \mathcal{O}(\Delta t)) \, \mathrm{d}v \\
&= -\Delta t^2 \sum_{i,j=1}^{s} b_i a_{i,j} \int_0^1 \eta'' \left(u^n + v \Delta t \sum_{k=1}^{s} a_{i,k} f_k \right) (f_0, f_0) \, \mathrm{d}v + \mathcal{O}(\Delta t^3)
\end{aligned}$$

Thus, with the given assumptions, $r'(0) < 0$. □

Lemma 3.2.2. *Let a Runge-Kutta method be given such that* $\sum_{i,j=1}^{s} b_i(a_{i,j} - b_j) < 0$*. If* $\eta''(u^n)(f_0, f_0) > 0$ *then* $r'(1) > 0$ *for sufficiently small* $\Delta t > 0$*.*

Note that the assumption $\sum_{i,j=1}^{s} b_i(a_{i,j} - b_j) < 0$ is reasonable. Since $\sum_{i,j=1}^{s} b_i b_j = 1$ and $\sum_{i=1}^{s} b_i a_{i,j} = 1/2$ is a condition for second-order accuracy we have that $\sum_{i,j=1}^{s} b_i(a_{i,j} - b_j) = -1/2$

Proof. By the definition of $r(\gamma)$ in Equation 3.1,

$$\frac{dr}{d\gamma} = \eta'(u^n + \gamma \Delta t \sum_{i=1}^{s} b_i f_i) \cdot \left(\Delta t \sum_{i=1}^{s} b_i f_i \right) - \Delta t \sum_{i=1}^{s} b_i \langle \eta'(y_i), f_i \rangle.$$

Evaluating $r'(\gamma)$ at $\gamma = 1$ we get

$$\begin{aligned} r'(1) &= \eta'(u^n + \Delta t \sum_{j=1}^{s} b_j f_j) \cdot \left(\Delta t \sum_{i=1}^{s} b_i f_i \right) - \Delta t \sum_{i=1}^{s} b_i \left\langle \eta'(y_i), f_i \right\rangle \\ &= \Delta t \sum_{i=1}^{s} b_i \left(\left\langle \eta'(u^n + \Delta t \sum_{j=1}^{s} b_j f_j), f_i \right\rangle - \left\langle \eta'(y_i), f_i \right\rangle \right) \\ &= -\Delta t \sum_{i=1}^{s} b_i \left(\left\langle \eta'(y_i), f_i \right\rangle - \left\langle \eta'(u^n + \Delta t \sum_{j=1}^{s} b_j f_j), f_i \right\rangle \right) \end{aligned}$$

Expanding $y_i = u^n + \Delta t \sum_{j=1}^{s} a_{i,j} f_j$ as defined in Equation 2.1 we get

$$r'(1) = -\Delta t \sum_{i=1}^{s} b_i \left(\left\langle \eta'(u^n + \Delta t \sum_{j=1}^{s} a_{i,j} f_j), f_i \right\rangle - \left\langle \eta'(u^n + \Delta t \sum_{j=1}^{s} b_j f_j), f_i \right\rangle \right).$$

We then make substitutions according to Equation 2.2, $u^{n+1} = u^n + \Delta t \sum_{j=1}^{s} b_j f_j$,

$$r'(1) = -\Delta t \sum_{i=1}^{s} b_i \left(\left\langle \eta' \left(u^{n+1} + \Delta t \sum_{j=1}^{s} (a_{i,j} - b_j) f_j \right), f_i \right\rangle - \left\langle \eta'(u^{n+1}), f_i \right\rangle \right).$$

Then, by the fundamental theorem of calculus,

$$r'(1) = -\Delta t \sum_{i=1}^{s} b_i \int_0^1 \eta'' \left(u^{n+1} + \Delta t \sum_{k=1}^{s} (a_{i,k} - b_k) f_k \right) \left(f_i, \Delta t \sum_{j=1}^{s} (a_{i,j} - b_j) f_j \right) \mathrm{d}v$$

$$r'(1) = -\Delta t^2 \sum_{i,j=1}^{s} b_i (a_{i,j} - b_j) \int_0^1 \eta'' \left(u^{n+1} + \Delta t \sum_{k=1}^{s} (a_{i,k} - b_k) f_k \right) (f_i, f_j) \, \mathrm{d}v$$

With Taylor expansions of $f_i, f_j = f_0 + \mathcal{O}(\Delta t)$,

$$r'(1) = -\Delta t^2 \sum_{i,j=1}^{s} b_i (a_{i,j} - b_j) \int_0^1 \eta'' \left(u^{n+1} + \Delta t \sum_{k=1}^{s} (a_{i,k} - b_k) f_k \right) (f_0, f_0 + \mathcal{O}(\Delta t)) \, \mathrm{d}v$$

$$r'(1) = -\Delta t^2 \sum_{i,j=1}^{s} b_i (a_{i,j} - b_j) \int_0^1 \eta'' \left(u^{n+1} + \Delta t \sum_{k=1}^{s} (a_{i,k} - b_k) f_k \right) (f_0, f_0) \, \mathrm{d}v + \mathcal{O}(\Delta t^3)$$

Thus, with the given assumptions, $r'(1) > 0$. □

Theorem 3.2.3. *Assume that the Runge-Kutta method satisfies* $\sum_{i,j=1}^{s} b_i a_{i,j} > 0$ *and* $\sum_{i,j=1}^{s} b_i (a_{i,j} - b_j) < 0$. *If* $\eta''(u^n)(f_0, f_0) > 0$ *then* r *has a positive root for sufficiently small* $\Delta t > 0$.

Proof. Since $r(0) = 0$ and $r'(0) < 0$ we have that $r(\gamma) < 0$ for small $\gamma > 0$. Because $r'(1) > 0$ and r is convex we have that r' is monotone. Thus, there must be a positive root of r. □

3.3 Accuracy

Theorem 3.3.1. *Let a given RK-method be of order p. Consider the IDT and RRK methods based on them and suppose that $\gamma_n = 1 + \mathcal{O}(\Delta t^{p-1})$, then*

1. *The IDT method interpreting $u_\gamma^{n+1} \approx u(t_n + \Delta t)$ has order $p - 1$.*

2. *The RRK method interpreting $u_\gamma^{n+1} \approx u(t_n + \gamma\Delta t)$ has order p.*

Theorem 3.3.2. *Let $\mathcal{W}$ be a Banach space, $\varPhi : [0, T] \times \mathcal{H} \to \mathcal{W}$ a smooth function and b_i, c_i coefficients of a Runge-Kutta method of order p. Then*

$$\sum_{i=1}^{s} b_i \varPhi(t_n + c_i\Delta t, y_i) = \sum_{i=1}^{s} b_i \varPhi(t_n + c_i\Delta t, u(t_n + c_i\Delta t)) + \mathcal{O}(\Delta t^p).$$

Corollary 3.3.3. *If η is smooth and the given Runge-Kutta method is p-order accurate, $r(\gamma = 1) = \mathcal{O}(\Delta t^{p+1})$.*

Theorem 3.3.4. *Assume there exists a positive root γ_n of r. Consider the IDT/RRK methods based on a given Runge-Kutta method that is p-order accurate. Then*

1. *The IDT method interpreting $u_\gamma^{n+1} \approx u(t_n + \Delta t)$ has order $p - 1$.*

2. *The RRK method interpreting $u_\gamma^{n+1} \approx u(t_n + \gamma\Delta t)$ has order p.*

Proof. We have that $r(1) = \mathcal{O}(\Delta t^{p+1})$ and $r'(1) = c\Delta t^2 + \mathcal{O}(\Delta t^3)$ for some $c > 0$. Thus there is a root $\gamma_n = 1 + \mathcal{O}(\Delta t^{p-1})$ of r. Applying to earlier theorem yield the desired result. □

Chapter 4

Numerical Examples

Since the stepsize of RRK is unknown apriori, the last step of the method is always made using the ordinary RK method. This was applied to the IDT methods as well, although not necessary. Because of this we are in our convergence studies regarding the second to last value instead of the last.

We are studying the effects of values of Δt close to 1 in the examples. These are not reasonable choices for Δt, as the roots of r become less predictable. For the scalar root finding the method `scipy.optimize.root_scalar` was used within the hardcoded range $[0.5, 2]$ was used.

Note that fewer RK-methods were used for Problem 1 than in Problem 2. This is because the higher order Runge-Kutta methods resulted in an error from the method `isClose` used somewhere in `Assimulo`, a library of solvers that ours were written to inherit from.

4.1 Problem 1 - Conserved exponential entropy

First we consider the system

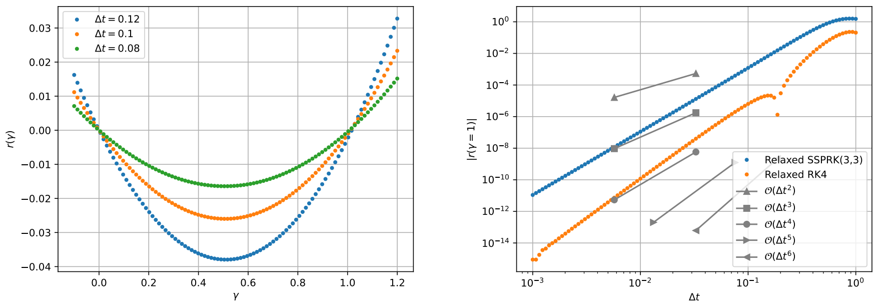
$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}u_1(t)\\u_2(t)\end{bmatrix} = \begin{bmatrix}-\exp(u_2(t))\\ \exp(u_1(t))\end{bmatrix}, \qquad u^0 = \begin{bmatrix}1\\0.5\end{bmatrix},$$

with exponential entropy

$$\eta(u) = \exp(u_1) + \exp(u_2), \qquad \eta'(u) = \begin{bmatrix}\exp(u_1)\\ \exp(u_2)\end{bmatrix},$$

and analytic solution

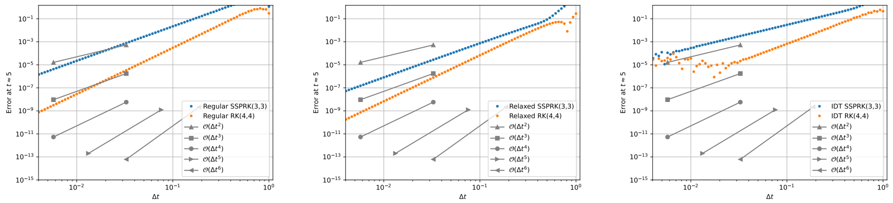
$$u(t) = \left(\log\left(\frac{\mathrm{e}+\mathrm{e}^{3/2}}{\sqrt{\mathrm{e}}+\mathrm{e}^{(\sqrt{\mathrm{e}}+\mathrm{e})t}}\right), \log\left(\frac{\mathrm{e}^{(\sqrt{\mathrm{e}}+\mathrm{e})t}(\sqrt{\mathrm{e}}+\mathrm{e})}{\sqrt{\mathrm{e}}+\mathrm{e}^{(\sqrt{\mathrm{e}}+\mathrm{e})t}}\right)\right)^{\mathrm{T}}.$$



(a) $r(\gamma)$ for SSPRK(3,3)

(b) $|r(\gamma = 1)|$ for some RK methods

Figure 4.1: Numerical results for r at the first time step of problem 1.

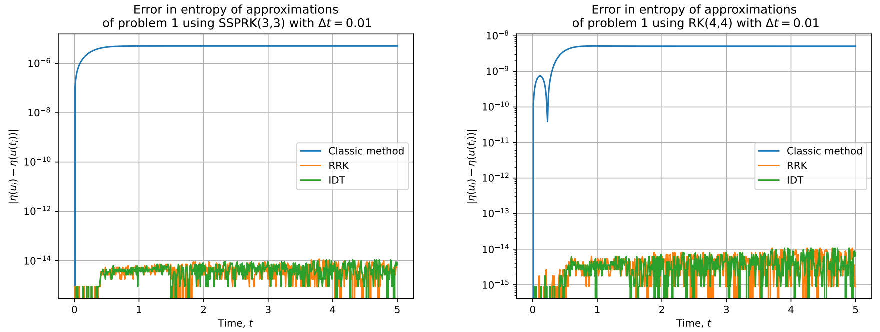


(a) Classic methods

(b) RRK Methods

(c) IDT methods

Figure 4.2: Convergence study for problem 1.



(a) SSPK(3,3)

(b) RK(4,4)

Figure 4.3: Error of entropy for problem 1.

4.2 Problem 2 - Dissipated exponential entropy

Consider the ODE

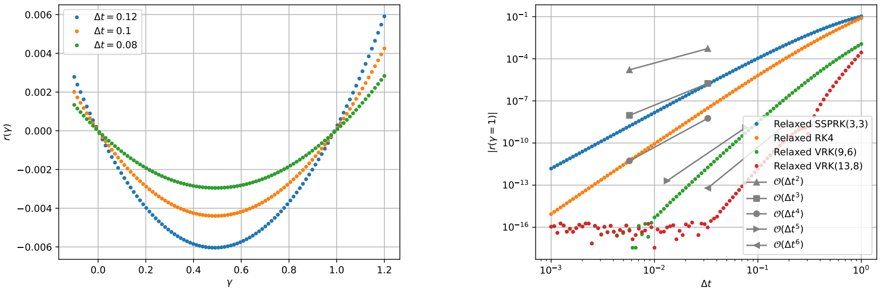
$$\frac{\mathrm{d}}{\mathrm{d}t}u(t) = -\exp(u(t)), \qquad u^0 = 0.5,$$

with the exponential entropy

$$\eta(u) = \exp(u), \qquad \eta'(u) = \exp(u),$$

and analytical solution

$$u(t) = -\log(\mathrm{e}^{-1/2} + t).$$



(a) $r(\gamma)$ for SSPRK(3,3)

(b) $|r(\gamma = 1)|$ for some RK methods

Figure 4.4: Numerical results for r at the first time step of problem 2.

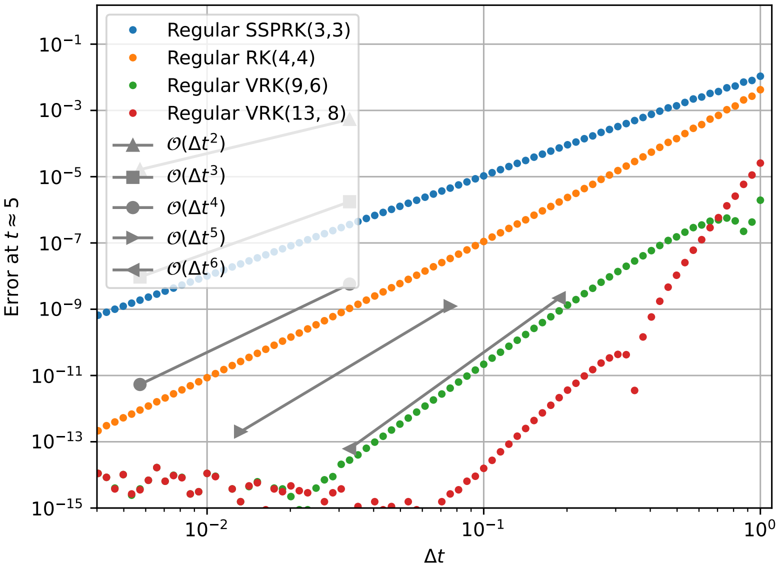


Figure 4.5: Convergence study for problem 2, classic methods.

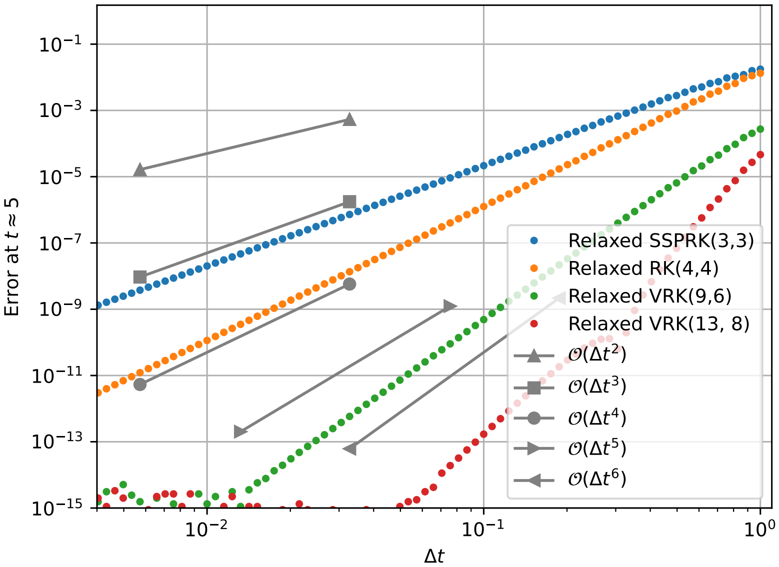


Figure 4.6: Convergence study for problem 2, relaxed methods.

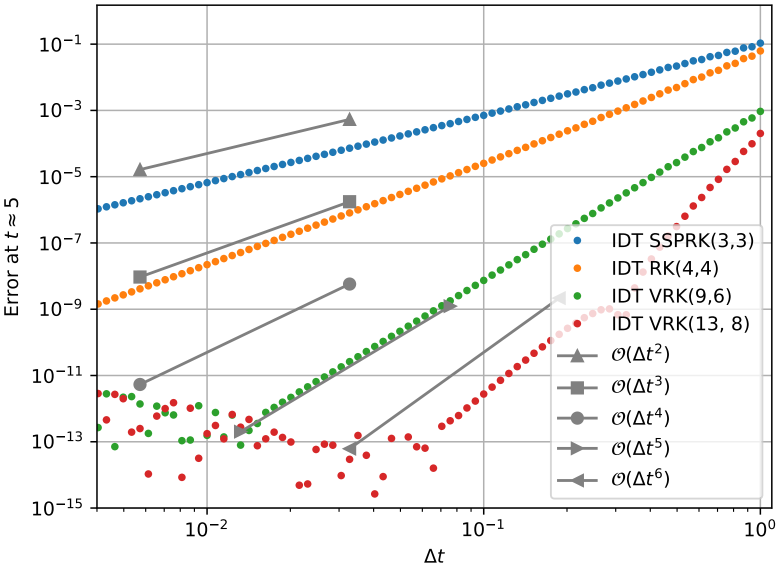
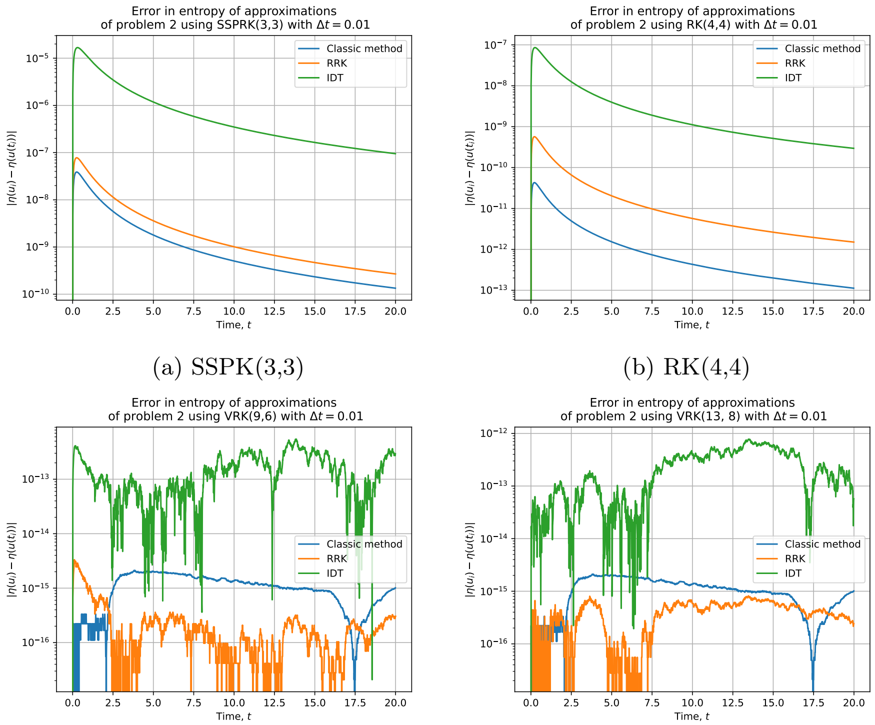


Figure 4.7: Convergence study for problem 2, IDT methods.



(a) SSPK(3,3)

(b) RK(4,4)

(c) VRK(9,6)

(d) VRK(13,8)

Figure 4.8: Error of entropy for problem 2.

Chapter 5

Conclusion

- Simple
- Cheap
- Physically correct
- Requires η

One drawback is that the method requires prior knowledge of η. This is however a reasonable assumption as someone interested in conserving the entropy of their simulation should have some insight into the entropy of their problem.

Bibliography

[1] Hendrik Ranocha, Mohammed Sayyari, Lisandro Dalcin, Matteo Parsani, and David I. Ketcheson. Relaxation runge–kutta methods: Fully discrete explicit entropy-stable schemes for the compressible euler and navier–stokes equations. *SIAM Journal on Scientific Computing*, 42(2):A612–A638, jan 2020.