Hypocoercivity for Linear ODEs and Strong Stability for Runge–Kutta Methods

Franz Achleitner[1,b)], Anton Arnold[1,a),c)] and Ansgar Jüngel[1,d)]

[1] *Inst. f. Analysis u. Scientific Computing, Technische Universität Wien, Wiedner Hauptstr. 8, A-1040 Wien, Austria.*

[a)] Corresponding author: anton.arnold@tuwien.ac.at
[b)] franz.achleitner@tuwien.ac.at
[c)] URL: https://www.asc.tuwien.ac.at/arnold/
[d)] ansgar.juengel@tuwien.ac.at

Abstract. In this note, we connect two different topics from linear algebra and numerical analysis: hypocoercivity of semi-dissipative matrices and strong stability for explicit Runge–Kutta schemes. Linear autonomous ODE systems with a non-coercive matrix are called hypocoercive if they still exhibit uniform exponential decay towards the steady state. Strong stability is a property of time-integration schemes for ODEs that preserve the temporal monotonicity of the discrete solutions. It is proved that explicit Runge–Kutta schemes are strongly stable with respect to semi-dissipative, asymptotically stable matrices if the hypocoercivity index is sufficiently small compared to the order of the scheme. Otherwise, the Runge–Kutta schemes are in general not strongly stable. As a corollary, explicit Runge–Kutta schemes of order $p \in 4\mathbb{N}$ with $s = p$ stages turn out to be *not* strongly stable. This result was proved in [4], filling a gap left open in [8]. Here, we present an alternative, direct proof.

Hypocoercive ODEs

For linear autonomous ordinary differential equations (ODEs),

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \mathbf{L}u, \quad t > 0, \quad u(0) = u^0 \in \mathbb{C}^n, \tag{1}$$

with Lyapunov stable matrices $\mathbf{L} \in \mathbb{C}^{n\times n}$ (i.e., all eigenvalues of $\mathbf{L}$ have nonpositive real part and the purely imaginary eigenvalues are non-defective), we are concerned with characterizing their short-time decay behavior. To this end, we review first hypocoercivity properties of such systems [1, 10]:

Definition 1 *(a) The matrix* $\mathbf{L} \in \mathbb{C}^n$ *is called* dissipative *(resp.* semi-dissipative*) if its Hermitian part,* $\mathbf{L}_H := (\mathbf{L} + \mathbf{L}^*)/2$ *is negative definite (resp. negative semi-definite).*

(b) $-\mathbf{L}$ *is called* hypocoercive *(or* positive stable*) if there are constants* $\lambda > 0$ *and* $c \geq 1$ *such that the matrix exponential satisfies*

$$\|e^{\mathbf{L}t}\|_2 \leq ce^{-\lambda t}, \quad t \geq 0.$$

(c) Let $\mathbf{L} \in \mathbb{C}^n$ *be semi-dissipative. Its* hypocoercivity index (HC-index) m_{HC} *is defined as the smallest integer* $m \in \mathbb{N}_0$ *(if it exists) such that*

$$T_m := \sum_{j=0}^{m} \mathbf{L}_S^j \mathbf{L}_H (\mathbf{L}_S^*)^j < 0,$$

where $\mathbf{L}_S := (\mathbf{L} - \mathbf{L}^*)/2$ *denotes the skew-Hermitian part of* $\mathbf{L}$.

The HC-index of $\mathbf{L}$ characterizes the structural complexity of the interplay between $\mathbf{L}_H$ and $\mathbf{L}_S$, and it is bounded by

$$0 \leq \frac{n - \operatorname{rank} \mathbf{L}_\mathrm{H}}{\operatorname{rank} \mathbf{L}_\mathrm{H}} \leq m_{HC}(\mathbf{L}) \leq n - \operatorname{rank} \mathbf{L}_\mathrm{H} \leq n - 1;$$

see [1, 3, 6]. Hence the matrix $\mathbf{L}$ is dissipative if and only if $m_{HC}(\mathbf{L}) = 0$. For later use, we consider the example

$$\mathbf{L} = \begin{pmatrix} 0 & -1 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & -1 & \\ & & 1 & 0 & -1 \\ & & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{N\times N}. \tag{2}$$

Using the Kalman rank condition [1, Proposition 1], one easily verifies that $\mathbf{L}$ has (maximal) HC-index $N-1$. Moreover, the matrix $\mathbf{L}$ is asymptotically stable, i.e., it has only eigenvalues with negative real part.

We recall that the HC-index gives a precise characterization of the short-time behavior of the solutions to (1):

Proposition 1 ([2]) *Let the matrix $\mathbf{L} \in \mathbb{C}^{n\times n}$ be semi-dissipative. Then $\mathbf{L}$ is asymptotically stable (with HC-index $m_{HC} \in \mathbb{N}_0$) if and only if*

$$\|e^{t\mathbf{L}}\|_2 = 1 - ct^a + O(t^{a+1}) \quad \textit{for } t \in [0,\varepsilon), \tag{3}$$

for some $a, c, \varepsilon > 0$. In this case, necessarily $a = 2m_{HC} + 1$.

The sharp multiplicative factor c in (3) has been determined explicitly in [2, Theorem 2.7(b)].

Strong stability for Runge–Kutta methods

It is known that a matrix $\mathbf{L} \in \mathbb{C}^{n\times n}$ is *Lyapunov stable* if and only if there exists a positive definite Hermitian matrix $\mathbf{P}$ such that

$$\mathbf{L}^*\mathbf{P} + \mathbf{P}\mathbf{L} \le 0. \tag{4}$$

In this case, the solution $u(t)$ of (1) is nonincreasing in the norm $\|\cdot\|_{\mathbf{P}} := \sqrt{\langle\cdot,\mathbf{P}\cdot\rangle}$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u(t)\|_{\mathbf{P}}^2 = \langle u, (\mathbf{L}^*\mathbf{P} + \mathbf{P}\mathbf{L})u\rangle \le 0. \tag{5}$$

It is often desirable that a numerical scheme for (1) reproduces this decay behavior on the discrete level.

Following [4], we consider here explicit Runge–Kutta schemes, where u^k is an approximation for $u(k\tau)$, $k \in \mathbb{N}_0$, and τ is the uniform time step:

$$u^k = u^{k-1} + \tau\sum_{i=1}^{s} b_i K_i^k, \quad K_i^k = \mathbf{L}\Big(u^{k-1} + \tau\sum_{j=1}^{i-1} a_{ij}K_j^k\Big), \quad i = 1,\dots,s, \tag{6}$$

where $b_i \in \mathbb{C}$ are the weights, $a_{ij} \in \mathbb{C}$ are the coefficients of the Runge–Kutta matrix, and $s \in \mathbb{N}$ is the number of stages. This scheme can be rewritten in compact form as $u^k = R(\tau\mathbf{L})u^{k-1}$, using its *stability function* $R(z)$. For the scheme (6) to have order p, one needs at least $s \ge p$ stages. In this case, its stability function takes the form

$$R(z) = \sum_{j=0}^{p}\frac{z^j}{j!} + \sum_{j=p+1}^{s} c_j\frac{z^j}{j!}, \quad z \in \mathbb{C}, \quad c_{p+1} \ne 1, \tag{7}$$

with some constants $c_{p+1}, \dots, c_s \in \mathbb{C}$.

For the discrete analog of the monotonicity estimate (5), we shall use the following notions:

Definition 2 *(a) The Runge–Kutta scheme* (6) *is* strongly stable *if for all matrix dimensions $n \in \mathbb{N}$, for all Lyapunov stable matrices $\mathbf{L} \in \mathbb{C}^{n\times n}$, and for all Hermitian matrices $\mathbf{P} > 0$ such that* (4) *holds, the numerical solution to* (1) *satisfies $\|u^1\|_{\mathbf{P}} \le \|u^0\|_{\mathbf{P}}$ for all initial data $u^0 \in \mathbb{C}^n$ and sufficiently small time steps.*

(b) The Runge–Kutta scheme (6) *is* strongly stable w.r.t. a subset $\mathcal{L}_0$ *of Lyapunov stable matrices (of any dimension n), if the condition from (a) holds for all $\mathbf{L} \in \mathcal{L}_0$.*

Strong stability of explicit Runge–Kutta schemes was studied in, e.g., [9, §4] and [8]. We shall focus on explicit Runge–Kutta schemes with $s = p$ stages. Their stability function is given by the first sum in (7), and their stability behavior was analyzed in [8]:

$$\text{They are} \quad \begin{cases} \text{strongly stable} & \text{if } p \in 4\mathbb{N}_0 + 3\,, \\ \text{not strongly stable} & \text{if } p \in 4\mathbb{N}_0 + 1 \text{ or } p \in 4\mathbb{N}_0 + 2\,, \end{cases}$$

but the following instability result for the case $p \in 4\mathbb{N}$ was only found and proved recently in [4]:

Theorem 2 ([4]) *Explicit Runge–Kutta schemes of order $p \in 4\mathbb{N}$ with $s = p$ stages are* not *strongly stable.*

While the proof of this result in [4] was based on the quite technical result in Proposition 1, we shall give here an independent direct proof. In order to motivate our subsequent proof, we first cite another result from [4]. To this end, we define the following subset of asymptotically stable (and thus Lyapunov stable) matrices:

$$\mathcal{L}_{AS}^m := \{\mathbf{L} \text{ is semi-dissipative and asymptotically stable} : m_{HC}(\mathbf{L}) \le m\}, \quad m \in \mathbb{N}_0.$$

Proposition 3 ([4]) *All explicit Runge–Kutta schemes of order $p \in \mathbb{N}$ (with $s \ge p$ stages) are strongly stable w.r.t. $\mathcal{L}_{AS}^m$, if $m \in \mathbb{N}_0$ satisfies $2m + 1 \le p$.*

Proof (of Theorem 2). For each fixed $p \in 4\mathbb{N}$, we shall analyze an asymptotically stable matrix $\mathbf{L}$ as a counterexample. Due to Proposition 3, the HC-index of such $\mathbf{L}$ must be "large enough". More precisely, we choose $\mathbf{L}$ of the form (2) with $N := 1 + p/2$. It satisfies $\mathbf{L} \in \mathcal{L}_{AS}^m$ with $m = p/2$, which violates the index condition in Proposition 3, and may hence serve as a counterexample.

When choosing $\mathbf{P} = \mathbf{I}$, inequality (4) is satisfied, and it remains to show that

$$\|R(\tau\mathbf{L})\|_2 := \sup_{\|u_0\|=1} \|R(\tau\mathbf{L})u^0\| \le 1, \quad \text{where } R(z) = \sum_{j=0}^{p} \frac{z^j}{j!}, \tag{8}$$

does *not* hold on any interval $\tau \in [0, \varepsilon)$, with $\varepsilon > 0$ arbitrarily small, see Definition 2(a). Equivalently, we shall show that the matrix function $\mathbf{M}(\tau) := \mathbf{I} - R(\tau\mathbf{L})^* R(\tau\mathbf{L})$ has a negative determinant on $(0, \varepsilon)$ for sufficiently small $\varepsilon > 0$. Hence, $R(\tau\mathbf{L})^* R(\tau\mathbf{L})$ has at least one eigenvalue larger than one, and $\|R(\tau\mathbf{L})\|_2 > 1$ follows. We shall obtain the inequality $\det \mathbf{M}(\tau) < 0$, for τ small enough, from the subsequent lemma, thus closing this proof. □

Lemma 4 *Let $p \in 4\mathbb{N}$, and $\mathbf{L}$ of the form (2) with $N := 1 + p/2$. Then, $\mathbf{M}(\tau) := \mathbf{I} - R(\tau\mathbf{L})^* R(\tau\mathbf{L})$ satisfies*

$$\det \mathbf{M}(\tau) = c\tau^{N^2} + O(\tau^{N^2+1}) \quad \text{as } \tau \to 0\,, \tag{9}$$

with some $c < 0$ and $R(z)$ defined in (8).

Proof. We only give a sketch of the proof here, the full details are given in [5].

First, we consider the Runge–Kutta method with $p = s = 4$, and hence $N = 3$. In this case, we compute the matrix $\mathbf{M}(\tau)$ explicitly:

$$\mathbf{M}(\tau) = \begin{pmatrix} \tau^5/12 & \tau^4/4 & \tau^3/3 \\ \tau^4/4 & 2\tau^3/3 & \tau^2 \\ \tau^3/3 & \tau^2 & 2\tau \end{pmatrix} + \begin{pmatrix} O(\tau^6) & O(\tau^5) & O(\tau^4) \\ O(\tau^5) & O(\tau^4) & O(\tau^3) \\ O(\tau^4) & O(\tau^3) & O(\tau^2) \end{pmatrix} \quad \text{as } \tau \to 0\,.$$

The determinant of the first matrix yields the leading order coefficient:

$$\det \mathbf{M}(\tau) = \det \begin{pmatrix} \tau^5/12 & \tau^4/4 & \tau^3/3 \\ \tau^4/4 & 2\tau^3/3 & \tau^2 \\ \tau^3/3 & \tau^2 & 2\tau \end{pmatrix} + O(\tau^{10}) = -\frac{\tau^9}{216} + O(\tau^{10}) \quad \text{as } \tau \to 0.$$

Consequently, $\det \mathbf{M}(\tau) < 0$ for sufficiently small $\tau > 0$, which disproves (8) for $p = 4$.

For the general case $p \geq 8$, we insert the stability matrix (7),

$$\mathbf{M}(\tau) = \mathbf{I} - \Big(\sum_{j=0}^{p} \frac{\tau^j}{j!}(\mathbf{L}^*)^j\Big)\Big(\sum_{\ell=0}^{p} \frac{\tau^\ell}{\ell!}\mathbf{L}^\ell\Big),$$

expand the matrix coefficients M_{ij} of $\mathbf{M}(\tau)$ in powers of τ and use properties of the matrices $\mathbf{L}$ and $\mathbf{L}_H$, leading to

$$M_{ij} = \bar{m}_{ij}\tau^{p+3-(i+j)} - \bar{n}_{ij}\tau^{p+1} + O(\tau^{p+4-(i+j)}),$$

where $\bar{m}_{ij}$ and $\bar{n}_{ij}$ are numbers depending on p and $\mathbf{L}_H$. The determinant of $\mathbf{M}(\tau)$ can be computed by using the Leibniz formula for determinants and definition $N = p/2 + 1$:

$$\det \mathbf{M}(\tau) = \det(\bar{m}_{ij} - \bar{n}_{ij})_{1\leq i,j\leq N}\tau^{N^2} + O(\tau^{N^2+1}).$$

The remaining determinant can be calculated by taking into account the explicit formula of the Hankel determinant:

$$c := \det(\bar{m}_{ij} - \bar{n}_{ij})_{1\leq i,j\leq N} = 2^N\Big(\prod_{i=1}^{N} \frac{1}{(p/2-i+1)!}\Big)^2 \frac{\prod_{1\leq i<j\leq N}(i-j)^2}{\prod_{i,j=1}^{N}(i+(j-1))}\Big(1-\binom{p}{p/2}\Big).$$

We deduce from $\binom{p}{p/2} > 1$ that $c < 0$ as claimed in Lemma 4. In particular, for sufficiently small $\tau > 0$, $\det \mathbf{M}(\tau)$ is negative, which also finishes the proof of Theorem 2. □

Finally, we note that $\|R(\tau\mathbf{L})\|_2 > 1$, the condition to violate strong stability, can also be tested numerically. But only for $p = 4$ the computation can be carried out in the standard double precision of Matlab: It shows that $\varepsilon = 0.304$ can be used, yielding $\max_{[0,\varepsilon]} \|R(\tau\mathbf{L})\|_2 - 1 \approx 1.3\text{E-}6$. For larger values of p, this computation is numerically so sensitive that we used octuple precision: For $p = 8$ one can use $\varepsilon = 0.027$, yielding $\max_{[0,\varepsilon]} \|R(\tau\mathbf{L})\|_2 - 1 \approx 9.0\text{E-}22$, and for $p = 12$, $\varepsilon = 0.027$ with $\max_{[0,\varepsilon]} \|R(\tau\mathbf{L})\|_2 - 1 \approx 7.2\text{E-}46$.

ACKNOWLEDGMENTS

The authors were supported by the Austrian Science Fund (FWF) via the FWF-funded SFB # F65. The third author acknowledges support from the FWF, grant P33010. This work received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme, ERC Advanced Grant NEUROMORPH, no. 101018153.

REFERENCES

[1] F. Achleitner, A. Arnold, and E. Carlen. On multi-dimensional hypocoercive BGK models. *Kinet. Relat. Models* 11 (2018), 953–1009.

[2] F. Achleitner, A. Arnold, and E. A. Carlen. The hypocoercivity index for the short time behavior of linear time-invariant ODE systems. *J. Differ. Equ.* 371 (2023), 83–115.

[3] F. Achleitner, A. Arnold, and E. Carlen. The hypocoercivity index for the intermediate and large time behavior of ODEs. *In preparation* (2023).

[4] F. Achleitner, A. Arnold, A. Jüngel. Necessary and sufficient conditions for strong stability of explicit Runge–Kutta methods. Submitted 2023. https://arxiv.org/abs/2308.05689

[5] F. Achleitner, A. Arnold, A. Jüngel. Hypocoercivity for Linear ODEs and Strong Stability for Runge-Kutta Methods. arxiv.org/abs/2310.19758

[6] F. Achleitner, A. Arnold, and V. Mehrmann. Hypocoercivity and hypocontractivity concepts for linear dynamical systems. *Electron. J. Linear Algebra* 39 (2023) 33–61.

[7] C. Krattenthaler. Advanced determinant calculus. A complement. *Lin. Alg. Appl.* 41 (2005), 68–166.

[8] Z. Sun and C.-W. Shu. Strong stability of explicit Runge–Kutta time discretizations. *SIAM J. Numer. Anal.* 57 (2019), 1158–1182.

[9] E. Tadmor. From semidiscrete to fully discrete: Stability of Runge–Kutta schemes by the energy method. II. In: D. Estep and S. Tavener (eds.), *Collected Lectures on the Preservation of Stability under Discretization*. Proc. Appl. Math. 109, pp. 25–49. SIAM, Philadelphia, 2002.

[10] C. Villani. Hypocoercivity. *Memoirs Amer. Math. Soc.*, vol. 202. Amer. Math. Soc., Providence, 2009.