8-31-2024

Certifying stability in runge-kutta schemes: Algebraic conditions and semidefinite programming

Austin Juhl
New Jersey Institute of Technology, aj659@njit.edu

ABSTRACT

CERTIFYING STABILITY IN RUNGE-KUTTA SCHEMES: ALGEBRAIC CONDITIONS AND SEMIDEFINITE PROGRAMMING

by
Austin Juhl

Numerical stability is a critical property for a time-integration scheme. In the context of Runge-Kutta methods applied to stiff differential equations, A-stability is one of the most basic and practically important notions of stability. Dating back to the work of Dahlquist, it has been known that A-stability is equivalent to the Runge-Kutta stability function satisfying a particular convex feasibility problem. Specifically, up to a transformation, the stability function lies in the convex cone of positive functions. In recent years, sum-of-squares optimization and semidefinite programming have become valuable tools in developing rigorous certificates of stability in dynamical systems. Therefore, it is natural to employ these convex optimization tools for the purpose of rigorously certifying A- and $A(\alpha)$-stability in Runge-Kutta methods.

Two distinct convex feasibility problems defined by linear matrix inequalities are introduced. The first approach employs sum-of-squares programming applied to the Runge-Kutta E-polynomial, making it applicable to both A- and $A(\alpha)$-stability. The second approach refines the algebraic conditions for A-stability, as developed by Cooper, Scherer, Türke, and Wendler (CSTW), to incorporate the Runge-Kutta order conditions. The theoretical enhancement of the algebraic conditions facilitates the practical application of the refined conditions for certifying A-stability within a computational framework.

Additionally, a new theoretical perspective is provided, relating the algebraic conditions for A-stability to continued fraction approximations of the exponential. This perspective involves the introduction of a new transform defined in a recently established class of polynomials orthogonal with respect to a linear functional.

The E-polynomial and CSTW methodologies are utilized to obtain rigorous stability certificates for several implicit Runge-Kutta schemes proposed in the literature. Specific attention is given to certifying the implicit Runge-Kutta schemes utilized in the SUite of Nonlinear and DIfferential/ALgebraic equation Solvers (SUNDIALS).

CERTIFYING STABILITY IN RUNGE-KUTTA SCHEMES: ALGEBRAIC CONDITIONS AND SEMIDEFINITE PROGRAMMING

by
Austin Juhl

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey—Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT
Department of Mathematics and Computer Science, Rutgers-Newark

August 2024

APPROVAL PAGE

CERTIFYING STABILITY IN RUNGE-KUTTA SCHEMES: ALGEBRAIC CONDITIONS AND SEMIDEFINITE PROGRAMMING

Austin Juhl

Dr. David Shirokoff, Dissertation Advisor — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Christina Frederick, Committee Member — Date
Associate Professor of Mathematical Sciences, NJIT

Dr. Travis Askham, Committee Member — Date
Assistant Professor of Mathematical Sciences, NJIT

Dr. Simone Marras, Committee Member — Date
Assistant Professor of Mechanical and Industrial Engineering, NJIT

Dr. David Ketcheson, Committee Member — Date
Professor of Applied Mathematics and Computational Sciences,
King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

BIOGRAPHICAL SKETCH

Author: Austin Juhl

Degree: Doctor of Philosophy

Date: August 2024

Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences
  New Jersey Institute of Technology, Newark, NJ, 2024

- Bachelor of Science, Computer Science
  University of Kansas, Lawrence, KS, 2019

Major: Mathematical Sciences

Publications:

A. Juhl and D. Shirokoff. Algebraic conditions for stability in Runge-Kutta methods and their certification via semidefinite programming, 2024. arXiv: 2405.13921 [math.NA].

To my parents, Steve and Valerie, and my sister, Autumn, thank you for your unwavering support and for being pillars of hard work and dedication. Your example has laid the foundation for all of my accomplishments.

To Tom Klocek, the best teacher I have ever had, who focused my boundless third-grade energy into a lifelong love for mathematics.

To Anthony Ritz, who, in teaching me calculus, imparted far greater lessons in life. You have always had the timely gift of joining my journey at just the right moments.

To Lucas Schauer, who inspired my PhD pursuit and with whom I learned that 'getting low' can unlock the many mysteries of the universe.

To Jake Brusca, who picked me up in year one and carried me on his back across the finish line.

To Grace Scherschligt, who loved me when I was tired and hungry.

This accomplishment is a reflection of your collective impact on my life.

ACKNOWLEDGMENTS

I would like to begin by thanking my advisor, Professor David Shirokoff, for encouraging me to pursue all opportunities and for your endless patience. I'm grateful for your mentorship over the last five years and consider myself very lucky that you agreed to take me on as a student.

Thank you to my committee members: Professors David Ketcheson, Travis Askham, Christina Frederick, and Simone Marras, who generously gave their time and expertise. Special thanks to my external committee member, Professor David Ketcheson, for providing new and challenging schemes to test my certification methods.

I am grateful for the support from the National Science Foundation, under grants DMS-2012268 and DMS-2309727.

I extend my gratitude to the CoHRT-19 group: Atul Anurag, Prianka Bose, Jake Brusca, Nick Dubicki, Sam Evans, Fataou Maxwell, General Ozochiawaeze, José Pabon, Moshe Silverstein, and John Wu, for the long, sleepless nights spent working on homework and engaging in enriching conversations. I am thankful for the swift bond we formed, allowing us to support each other during the pandemic. I look forward to your many accomplishments.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

Numerical stability is a critical property for a time-integration scheme. In the context of Runge-Kutta (RK) methods applied to stiff differential equations, A-stability (or the related $A(\alpha)$-stability), introduced by Dahlquist [16], is one of the most basic and practically important notions of stability. As a result of the reduced degrees of freedom in the Butcher coefficients, Runge-Kutta methods that are A-stable with stability functions that are near-optimal order approximations have been well characterized [11, 26, 39, 48] (cf. [29] for algebraically stable methods). However, many existing and newly developed Runge-Kutta methods in the literature admit a diagonally implicit structure and lie outside the application of these previous results. While one strategy for verifying A-stability is via *Sturm sequences* [30], our approach here is rooted in convex optimization.

In recent years, sum-of-squares optimization and semidefinite programming have become valuable tools in developing rigorous certificates of stability in dynamical systems. Such certificates are useful in developing reliable algorithms and software. Examples range from stochastic linear-quadratic control [49], switched stability of nonlinear systems [1], reinforcement learning [31], stability in partial differential equations [23], and applications for robotic control [38]. Additionally, control systems have benefited from analogous certification techniques, including verifying the stability of reinforcement learning policies [19] or variable step-size convergence bounds for gradient descent [25]. In this work, we adopt a similar strategy to establish rigorous certificates for A- and $A(\alpha)$- stability in Runge-Kutta methods through the computational solution of linear matrix inequalities via semi-definite programming.

Chapter 1 will introduce the background of Runge-Kutta methods, covering the linear stability of RK schemes and the order conditions. These concepts are central to the main ideas explored in this dissertation. The chapter concludes with descriptions of linear matrix inequalities and semidefinite programming feasibility problems, the primary tools we use to obtain rigorous certificates of A-stability for RK schemes.

1.1 Runge-Kutta Background

Numerical time integration of an ordinary differential equation ($\mathbb{V} = \mathbb{R}^m$ or $\mathbb{C}^m$)

$$\boldsymbol{u}'(t) = f\big(t, \boldsymbol{u}(t)\big)\,, \quad \boldsymbol{u}(0) = \boldsymbol{u}_0\,; \quad \boldsymbol{u} \in \mathbb{V}\,, \quad f : \mathbb{R} \times \mathbb{V} \to \mathbb{V}\,, \tag{1.1}$$

results in a discrete-in-time dynamical system — one of the most common being a Runge-Kutta (RK) method. Runge-Kutta methods, discretize (1.1) with s stages as

$$\boldsymbol{h}_i = \boldsymbol{u}_n + \Delta t \sum_{j=1}^{s} a_{ij}\, f(t_n + c_j \Delta t, \boldsymbol{h}_j)\,, \quad i = 1, 2, \ldots, s \tag{1.2a}$$

$$\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \Delta t \sum_{j=1}^{s} b_j\, f(t_n + c_j \Delta t, \boldsymbol{h}_j)\,, \tag{1.2b}$$

where $\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$ and $t_n = n\Delta t$ denotes the nth time step. The RK scheme is defined by coefficients

$$\boldsymbol{A} = [a_{ij}]_{i,j=1}^{s}\,, \quad \boldsymbol{b} = [b_1, \ldots, b_s]^T, \quad \boldsymbol{c} = [c_1, \ldots, c_s]^T := \boldsymbol{A}\boldsymbol{e}\,; \quad \text{and} \quad \boldsymbol{e} = [1, \ldots, 1]^T.$$

1.1.1 Linear stability in RK schemes.

It is standard practice to assess the stability of a RK scheme by examining the scalar linear case where $f(\boldsymbol{u}) = \lambda \boldsymbol{u}$ with $\lambda \in \mathbb{C}$ and fixed Δt. The Runge-Kutta dynamics (1.2) applied to the scalar linear problem are expressed as:

$$\boldsymbol{u}_{n+1} = W(z)\boldsymbol{u}_n, \quad \text{where} \quad z := \lambda \Delta t\,.$$

Here $W(z)$ is the *stability function* of the scheme given by:

$$W(z) := 1 + z\boldsymbol{b}^T(\boldsymbol{I} - z\boldsymbol{A})^{-1}\boldsymbol{e} = \frac{\det(\boldsymbol{I} - z\boldsymbol{A} + z\boldsymbol{e}\boldsymbol{b}^T)}{\det(\boldsymbol{I} - z\boldsymbol{A})} = \frac{N(z)}{D(z)}. \tag{1.3}$$

In equation (1.3), $N(z)$ and $D(z)$ represent polynomials of degree at most s, sharing no common factors, and $\boldsymbol{I}$ is the identity matrix.

A method exhibits a *degenerate* stability function if $\deg N \leq s-1$ and $\deg D \leq s-1$; otherwise, the stability function is considered non-degenerate. Degenerate stability functions occur when $\det(\boldsymbol{I} - z\boldsymbol{A} + z\boldsymbol{e}\boldsymbol{b}^T)$ and $\det(\boldsymbol{I} - z\boldsymbol{A})$ share a common root.

The dynamics (1.2) are *stable* for a given z if $|W(z)| \leq 1$. We will primarily be concerned with numerical schemes $(\boldsymbol{A}, \boldsymbol{b})$ that are A-stable, i.e., those where

$$(A\text{-stability}) \qquad |W(z)| \leq 1 \qquad \text{for} \qquad z \in \mathbb{C}_- = \{z \in \mathbb{C} \ : \ \operatorname{Re} z \leq 0\}. \tag{1.4}$$

The A-stability criteria ensures that the discrete dynamics (1.2) are stable whenever the linear ODE (1.1) is stable. For schemes that are not A-stable, it is useful to characterize the largest $\alpha > 0$ for which the RK dynamics are stable for all z in the sector S_α opening into the left-half plane with angle α

$$(A(\alpha)\text{-stability}) \qquad |W(z)| \leq 1 \quad \text{for} \quad z \in S_\alpha = \{z \in \mathbb{C} \ : \ |\arg(-z)| \leq \alpha\}. \tag{1.5}$$

Note that A-stability is equivalent to $A(\alpha)$-stability with $\alpha = \frac{\pi}{2}$.

The primary goal of this dissertation is to refine and leverage existing theory and computational tools to obtain a rigorous certificate of A- or $A(\alpha)$-stability for RK schemes.

1.1.2 Accuracy and order conditions.

The notion of accuracy and order conditions will play a direct role in certifying the stability of RK methods within a computational framework. For an RK scheme to achieve *(classical) order* p on linear, autonomous problems, the stability function must approximate the exponential function to order p, such that

$$W(z) = e^z + \mathcal{O}(z^{p+1}) \quad \text{as} \quad z \to 0. \tag{1.6}$$

This approximation is achieved provided the RK coefficients $(\boldsymbol{A}, \boldsymbol{b})$ satisfy the *tall-tree* order conditions of order p:

$$\boldsymbol{b}^T \boldsymbol{A}^{j-1} \boldsymbol{e} = \frac{1}{j!} \quad \text{for} \quad 1 \leq j \leq p\,, \tag{1.7}$$

as outlined in [27]. Additional RK order conditions, i.e., the non-tall-tree conditions, are further required to achieve accuracy of order p on general (nonlinear) ODEs.

1.2 Linear Matrix Inequalities and Semidefinite Programming

The computational tools used to certify the stability of RK schemes involve using linear matrix inequalities to define a feasible set within a semidefinite program.

Given matrices $\boldsymbol{P}, \boldsymbol{N}_1, \ldots, \boldsymbol{N}_n$, in the set $\mathbb{S}^n$ of $n \times n$ real symmetric matrices, a *linear matrix inequality* (LMI) is defined as:

$$\boldsymbol{F}(\boldsymbol{\eta}) := \boldsymbol{P} + \sum_{j=1}^{d} \eta_j \boldsymbol{N}_j \succeq 0\,, \tag{1.8}$$

where $\boldsymbol{F} \succeq 0$ indicates that $\boldsymbol{F}$ is positive semi-definite ($\boldsymbol{F} \succ 0$ indicates that $\boldsymbol{F}$ is positive definite). The LMI (1.8) is *feasible* if there exists a vector $\boldsymbol{\eta}$ such that $\boldsymbol{F}(\boldsymbol{\eta}) \succeq 0$; otherwise the LMI is *infeasible*. The linearity of $\boldsymbol{F}(\boldsymbol{\eta})$ ensures that the set:

$$\mathcal{C} = \{\boldsymbol{\eta} \in \mathbb{R}^d \ : \ \boldsymbol{F}(\boldsymbol{\eta}) \succeq 0\}\,,$$

containing all $\boldsymbol{\eta}$ that satisfy the LMI (1.8), is convex. Thus, assessing the feasibility of $\boldsymbol{F}(\boldsymbol{\eta})$, in other words, determining whether $\mathcal{C}$ is non-empty, is a convex feasibility problem that can be solved via semidefinite programming.

Due to the matrix structure of (1.8), the feasible set $\mathcal{C}$ may lie in an affine plane with a dimension less than d, potentially resulting in $\mathcal{C}$ having an empty interior. The affine hull of a convex set is defined as the smallest affine space containing $\mathcal{C}$, meaning it is the intersection of all affine sets containing $\mathcal{C}$ [10]. The affine hull can also be defined as the set of all affine combinations of points in $\mathcal{C}$:

$$\operatorname{aff}(\mathcal{C}) := \left\{ \sum_{j=1}^{r} \mu_j x_j : r > 0,\ x_j \in \mathcal{C},\ \sum_{j=1}^{r} \mu_j = 1,\ \mu_j \in \mathbb{R} \right\}.$$

Note that in the definition of $\operatorname{aff}(\mathcal{C})$, the number of points r in $\mathcal{C}$ can be arbitrary. Furthermore, since $\operatorname{aff}(\mathcal{C})$ defines an affine space, by construction, it has the following form:

$$\operatorname{aff}(\mathcal{C}) = x_0 + \mathbb{V},$$

where x_0 is in $\mathcal{C}$ and $\mathbb{V}$ is a vector space. With this construction, the set $\mathcal{C}$ can be assigned a dimension as follows:

Definition 1.2.1 (Dimension of a convex set [10])**.** Let $\mathcal{C}$ be a convex set with affine hull $\operatorname{aff}(\mathcal{C})$ and associated vector space $\mathbb{V}$. Then the dimension of $\mathcal{C}$, written $\dim(\mathcal{C})$, is defined to be the dimension of $\mathbb{V}$.

A convex program characterized by a constant objective function and constrained by LMIs is called a *semidefinite programming* (*SDP*) *feasibility problem* and takes the form

$$\begin{aligned} &\text{Minimize:} && 1 \\ &\text{Subject to:} && \boldsymbol{F}(\boldsymbol{\eta}) := \boldsymbol{P} + \textstyle\sum_{j=1}^{d} \eta_j \boldsymbol{N}_j \succeq 0\,. \end{aligned} \tag{P1}$$

CHAPTER 2

ALGEBRAIC CONDITIONS FOR STABILITY

Dating back to the work of Dahlquist (e.g., [17]), it has been known that A-stability of (1.2) is equivalent to the RK stability function satisfying a particular convex feasibility problem. Specifically, up to a transformation, the stability function lies in the convex cone of positive functions [30, Chapter IV.5]. Subsequent convex feasibility conditions for A-stability include: (1) the RK E-polynomial lying in the convex cone of non-negative polynomials, and (2) the existence of a symmetric matrix that satisfies a set of algebraic conditions depending on the RK scheme coefficients. These algebraic conditions, referred to as the CSTW conditions, were developed by Cooper [13], Scherer and Türke [42], and Scherer and Wendler [43]. The CSTW conditions form a feasibility problem over the convex cone of semidefinite matrices. Semidefinite matrix conditions also exist for stronger notions of stability, such as algebraic stability (or the related concept of B-stability) for Runge-Kutta schemes [30, Chapter IV.12], and G-stability [18] for linear multistep methods.

This chapter focuses on algebraic conditions for stability in Runge-Kutta schemes. We begin by detailing the use of sum-of-squares optimization for certifying A- and $A(\alpha)$-stability via testing the non-negativity of the E-polynomial. We then explore the existing CSTW conditions. The main contributions of this chapter are twofold.

First, we provide a theoretical contribution that sharpens the CSTW conditions. The CSTW conditions do not account for the fact that the stability function is a pth order approximation to the exponential, which ultimately limits their practical use in providing rigorous certificates via computer-assisted means. Our theoretical result (Theorem 2.3.3 and Theorem 2.3.6) modifies the CSTW conditions to account for the

RK order conditions and singular coefficient matrix $\boldsymbol{A}$, thereby enabling the rigorous certification of stability via computational approaches.

Second, we introduce new algebraic conditions for A-stability derived from a new class of orthogonal polynomials. The chapter concludes with a brief review of Dahlquist's algebraic conditions for G-stability of Linear Multistep Methods.

2.1 E-polynomial Stability Conditions for $A-$ and $A(\alpha)$-stability

To certify the stability of Runge-Kutta schemes, we utilize linear matrix inequalities in two approaches. The first approach, suitable for both A- and $A(\alpha)$-stability, leverages the non-negativity of the (generalized) E-polynomial:

$$E(y;\alpha) = |D(ye^{-i\alpha})|^2 - |N(ye^{-i\alpha})|^2 \quad \text{where} \quad 0 < \alpha \leq \frac{\pi}{2}\,,\ y \in \mathbb{R}\,. \tag{2.1}$$

A scheme is then $A(\alpha)$-stable if (for $\alpha = \frac{\pi}{2}$ see [30, Chapter IV.3]) $W(z)$ is analytic in the interior of S_α and

$$E(y;\alpha) \geq 0 \qquad \text{for all} \qquad y \geq 0\,. \tag{2.2}$$

The condition (2.2) guarantees $A(\alpha)$-stability by ensuring $|W(z)| \leq 1$ for all $z \in \partial S_\alpha$. As $W(z)$ is analytic within the interior of S_α, the maximum modulus principle implies that $|W(z)|$ is maximized on ∂S_α, confirming $A(\alpha)$-stability.

2.1.1 Non-negative polynomials as SOS Linear Matrix Inequalities.

Condition (2.2) enables RK schemes to be certified A- and $A(\alpha)$-stable by determining whether a polynomial, the E-polynomial, is non-negative. This section uses the relationship between non-negative polynomials and sum-of-squares (SOS) to convert condition (2.2) into an equivalent LMI.

Let $\mathbb{R}[y]$ denote the set of single-variable polynomials with real coefficients. Two convex cones within $\mathbb{R}[y]$ include the set of non-negative polynomials, satisfying

$p(y) \geq 0$ for all $y \in \mathbb{R}$, and the set of SOS polynomials, where each polynomial can be expressed as $\sum_{j=1}^{\ell} q_j^2(y)$ for some $\{q_j\}_{j=1}^{\ell} \in \mathbb{R}[y]$. In one dimension, these two cones coincide, meaning that a polynomial $p(y) \in \mathbb{R}[y]$ is non-negative if and only if it can be decomposed into a sum-of-squares [37, Theorem 2.5]. For general multivariate polynomials, every SOS polynomial is non-negative; however, the converse is not true.

Determining whether a polynomial is SOS and non-negative can be formulated as a LMI. Consider the subspace $\mathcal{N}_m$ of $m \times m$ symmetric matrices $\mathbb{S}^m$ defined by:

$$\mathcal{N}_m := \left\{ \boldsymbol{N} \in \mathbb{S}^m \ : \ \boldsymbol{y}^T \boldsymbol{N} \boldsymbol{y} = 0 \right\}, \quad \text{where} \quad \boldsymbol{y} = [1, y, \ldots, y^{m-1}]^T . \tag{2.3}$$

The components $(n_{i,j})_{i,j=1}^{m}$ of a symmetric matrix $\boldsymbol{N} \in \mathcal{N}_m$ must satisfy:

$$\sum_{i+j=r}^{m} n_{i,j} = 0 \qquad \text{for} \quad r = 2,\ 3,\ \ldots,\ 2m\,.$$

Any polynomial $p \in \mathbb{R}[y]$ can then be expressed non-uniquely in factorized form as:

$$p(y) = p_0 + p_1 y + \ldots + p_{2m-2}\, y^{2m-2} = \boldsymbol{y}^T (\boldsymbol{P} + \boldsymbol{N}) \boldsymbol{y}\,, \tag{2.4}$$

where $\boldsymbol{N} \in \mathcal{N}_m$ and

$$\boldsymbol{P} = \begin{bmatrix} p_0 & \frac{1}{2}p_1 & & \\ \frac{1}{2}p_1 & p_2 & \ddots & \\ & \ddots & \ddots & \frac{1}{2}p_{2m-3} \\ & & \frac{1}{2}p_{2m-3} & p_{2m-2} \end{bmatrix} \in \mathbb{R}^{m\times m}\,. \tag{2.5}$$

The polynomial p, described in (2.4), is a sum-of-squares if and only if there exists an $\boldsymbol{N} \in \mathcal{N}_m$ such that $\boldsymbol{P} + \boldsymbol{N} \succeq 0$. If $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{P} + \boldsymbol{N}$ is a Cholesky factorization, then defining $q_j(y) = \boldsymbol{e}_j^T \boldsymbol{Q}\boldsymbol{y}$, where $\boldsymbol{e}_j$ is the jth unit vector, admits

$$p(y) = \|\boldsymbol{Q}\boldsymbol{y}\|^2 = \sum_{j=1}^{m} q_j^2(y)\,.$$

Conversely, if p is an SOS, the coefficients of q_j define a positive definite matrix $(\boldsymbol{P}+\boldsymbol{N})$.

To solve subsequent SDPs, it is useful to define a basis for $\mathcal{N}_m$, which has dimension

$$d := \frac{1}{2}(m-1)(m-2). \tag{2.6}$$

First, consider the index set

$$\mathcal{S}_m := \left\{(i,j) \in \mathbb{N}^2 \mid 1 \le i \le m-2\,, i+2 \le j \le m\right\} \qquad \text{for} \qquad m \ge 1\,,$$

which has d elements and is empty for $m = 1$ or 2. A basis for $\mathcal{N}_m$ is given by $\{\boldsymbol{N}_\ell\}_{\ell=1}^{d}$:

$$\boldsymbol{N}_\ell = \boldsymbol{e}_i\boldsymbol{e}_j^T + \boldsymbol{e}_j\boldsymbol{e}_i^T - \boldsymbol{e}_{\lfloor \frac{i+j}{2} \rfloor}\boldsymbol{e}_{\lceil \frac{i+j}{2} \rceil}^T - \boldsymbol{e}_{\lceil \frac{i+j}{2} \rceil}\boldsymbol{e}_{\lfloor \frac{i+j}{2} \rfloor}^T\,, \tag{2.7}$$

where

$$(i,j) \in \mathcal{S}_m \qquad \text{and} \qquad \ell = m(i-1) - \frac{1}{2}i(i+3) + j\,.$$

For example, $m = 3$ has the basis matrix:

$$\boldsymbol{N}_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

while $m = 4$ has three basis matrices:

$$\boldsymbol{N}_1 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{N}_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{N}_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

With these notations, a polynomial $p(y)$, given by (2.4), is non-negative if and only if there exists $\boldsymbol{\eta} \in \mathbb{R}^d$ such that the following LMI is satisfied:

$$\boldsymbol{F}(\boldsymbol{\eta}) := \boldsymbol{P} + \sum_{\ell=1}^{d} \eta_\ell \, \boldsymbol{N}_\ell \succeq 0 \,. \tag{2.8}$$

2.1.2 The E-polynomial LMI for $A(\alpha)$-stability.

An LMI for $A(\alpha)$-stability can be derived immediately by applying the LMI for general non-negative polynomials (2.8) to $E(y;\alpha)$ in (2.2).

In particular, if $\alpha = \frac{\pi}{2}$, the pth order conditions (1.7) imply that $E(y;\frac{\pi}{2})$ admits a factor of y^{2j} [30, Chapter IV.3]:

$$E\left(y;\frac{\pi}{2}\right) = \mathcal{O}(y^{2j}) \,, \qquad \text{as} \quad y \to 0 \quad \text{where,} \quad j \geq \kappa := \left\lfloor \frac{p}{2} \right\rfloor + 1 \,. \tag{2.9}$$

In general for $\alpha < \frac{\pi}{2}$, the polynomial (as implied by the calculations in [13])

$$E\,(y;\alpha) = \mathcal{O}(y) \,, \qquad \text{as} \quad y \to 0 \,. \tag{2.10}$$

Building on the asymptotics (2.9)–(2.10), let $F(y)$ be the polynomial from which the largest even monomial (ensured by the order conditions) has been factored out:

$$\begin{aligned} &\text{If } \alpha = \frac{\pi}{2} \ : \qquad F(y) := y^{-2\kappa}\, E(y;\frac{\pi}{2}) \,, \\ &\text{If } \alpha < \frac{\pi}{2} \ : \qquad F(y;\alpha) := y^{-2}\, E(y^2;\alpha) \,. \end{aligned}$$

In both cases, $F(y)$ is an even polynomial:

$$F(y) = p_0 + p_2\, y^2 + \ldots + p_{2m-2}\, y^{2m-2} \,, \tag{2.11}$$

where the coefficients of $F(y)$ are polynomial functions of the RK scheme coefficients and, in the case of $\alpha < \frac{\pi}{2}$, polynomial functions of $\beta := \cos(\alpha)$. $F(y)$ can then be expressed in a factorized form as in (2.4)

$$F(y) = \boldsymbol{y}^T(\boldsymbol{P} + \boldsymbol{N})\boldsymbol{y}$$

where $\boldsymbol{P}$ is a diagonal matrix whose diagonal components are the coefficients of $F(y)$ and $\boldsymbol{N} \in \mathcal{N}_m$ as defined in (2.3).

Combining the polynomial $F(y)$ and the LMI (2.8), it follows:

Lemma 2.1.1. *A RK scheme* $(\boldsymbol{A}, \boldsymbol{b})$ *is* $A(\alpha)$*-stable if:*

1. $\boldsymbol{A}$ *has no eigenvalues inside* S_α *(so that* $W(z)$ *is analytic in* S_α*); and*
2. *The LMI* (2.8) *is feasible for the polynomial* $F(y)$ *in* (2.11) *or equivalently the* E*-polynomial in* (2.1)*.*

2.1.3 Stability certification with the E-polynomial LMI.

The first example demonstrates the SOS LMI approach to verify A-stability for an idealized example that satisfies the tall tree order conditions and is known to be A-stable. The second example demonstrates the SOS LMI approach to verify $A(\alpha)$-stability for an idealized example that satisfies the tall tree order conditions.

Example 2.1.1 (A-stability certification for SDIRK(5,4))**.** The scheme is a Diagonally Implicit RK scheme, represented by the following Butcher tableau [30, Table 6.5, Chapter IV.6].

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{4} & & & & \\ \frac{1}{2} & \frac{1}{4} & & & \\ \frac{17}{50} & -\frac{1}{25} & \frac{1}{4} & & \\ \frac{371}{1360} & -\frac{137}{2720} & \frac{15}{544} & \frac{1}{4} & \\ \frac{25}{24} & -\frac{49}{48} & \frac{125}{16} & -\frac{85}{12} & \frac{1}{4} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \frac{25}{24} \\ -\frac{49}{48} \\ \frac{125}{16} \\ -\frac{85}{12} \\ \frac{1}{4} \end{bmatrix}. \tag{2.12}$$

The SDIRK(5,4) scheme in (2.12) has an E-polynomial

$$E(y) = y^6(9y^4 - 64y^2 + 512),$$

which after factoring out the largest monomial factor, yields

$$F(y) := 9y^4 - 64y^2 + 512 = \boldsymbol{y}^T\, \boldsymbol{F}(\eta)\, \boldsymbol{y}\,,$$

where

$$\boldsymbol{F}(\eta) = \begin{bmatrix} 512 & 0 & 0 \\ 0 & -64 & 0 \\ 0 & 0 & 9 \end{bmatrix} + \eta \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{2.13}$$

For this example, the space $\mathcal{N}_3$ (introduced in § 2.1.1) has dimension $d = 1$ and is spanned by the second matrix in (2.13). We then factorize $\boldsymbol{F}(\eta) = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$ with $\boldsymbol{L}, \boldsymbol{D} \in \mathbb{Q}^{3\times 3}$.

$$\boldsymbol{F}(-32) = \begin{bmatrix} 512 & 0 & -32 \\ 0 & 0 & 0 \\ -32 & 0 & 9 \end{bmatrix} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T\,,$$

where

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{16} & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{D} = \begin{bmatrix} 512 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 7 \end{bmatrix} \succeq 0\,.$$

This results in the following SOS representation of $E(y)$, certifying A-stability:

$$E(y) = y^6\, \left(7y^4 + 512(\tfrac{1}{16}y^2 - 1)^2\right)\,.$$

Example 2.1.2 ($A(\alpha)$-stability certification for ESDIRK(5,4))**.** The scheme is known to be $A(\alpha)$-stable for $\alpha \leq 89.5^o$ and is an Explicit first stage Singly Diagonally Implicit

Runge-Kutta scheme (ESDIRK), represented by the following Butcher tableau [45].

$$\boldsymbol{A} = \begin{bmatrix} 0 & & & & & \\ \frac{1}{6} & \frac{1}{6} & & & & \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & & & \\ \frac{11}{24} & -\frac{1}{4} & \frac{5}{8} & \frac{1}{6} & & \\ \frac{11}{36} & -\frac{1}{6} & \frac{11}{12} & -\frac{2}{9} & \frac{1}{6} & \\ \frac{1}{8} & \frac{3}{8} & \frac{3}{8} & -\frac{1}{12} & \frac{1}{24} & \frac{1}{6} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \frac{1}{8} \\ \frac{3}{8} \\ \frac{3}{8} \\ -\frac{1}{12} \\ \frac{1}{24} \\ \frac{1}{6} \end{bmatrix}. \tag{2.14}$$

The ESDIRK(5,4) scheme in (2.14) has an E-polynomial

$$E(y; \frac{\pi}{2}) = y^6(y^4 + 144y^2 - 2592),$$

which is negative near the origin and therefore not A-stable. The E-polynomial at $\beta = \frac{1}{8}$ ($\alpha \approx 82.8^o$) is

$$E(y^2, \tfrac{1}{8}) = y^2(y^{18} + \tfrac{15}{2}\, y^{16} + \tfrac{333}{2}\, y^{14} + \tfrac{5319}{4}\, y^{12} - \tfrac{4779}{16}\, y^{10} + \tfrac{2624643}{32}\, y^8 \\ + \tfrac{177147}{2}\, y^6 + 2125764\, y^4 + 1259712\, y^2 + 15116544)\,.$$

Removing the largest monomial factor then yields

$$F(y^2; \tfrac{1}{8}) := \boldsymbol{y}^T \boldsymbol{F}(\eta)\, \boldsymbol{y}\,, \quad \text{with} \quad \boldsymbol{F}(\boldsymbol{\eta}) := \boldsymbol{P} + \sum_{\ell=1}^{d} \eta_\ell\, \boldsymbol{N}_\ell\,,$$

where the diagonal matrix $\boldsymbol{P}$ is characterized by the coefficients of the polynomial $E(y^2, \frac{1}{8})$ and $\boldsymbol{N}_\ell$, defined by (2.7), are the basis matrices which span the space $\mathcal{N}_{10}$ with dimension $d = 36$. Even though the solution lives in a space with dimension 36, for this example, we can define a simple solution $\boldsymbol{F}(\eta)$ with a single basis matrix

$$\boldsymbol{F}(\tfrac{-4779}{32}) = \boldsymbol{P} + \tfrac{-4779}{32}\, \boldsymbol{N}_{15} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T,$$

where

$$\boldsymbol{L} = \begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \ddots & & \\ & & & & 1 & \\ -\frac{59}{497664} & & & & & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{D} = \operatorname{diag}\begin{bmatrix} 15116544 \\ 1259712 \\ 2125764 \\ \frac{177147}{2} \\ \frac{2624643}{32} \\ 0 \\ \frac{5319}{4} \\ \frac{333}{2} \\ \frac{15}{2} \\ \frac{193127}{196608} \end{bmatrix} \succeq 0.$$

This results in the following SOS representation of $E(y^2; \frac{1}{8})$, certifying $A(\alpha)$-stability for $\alpha \leq 82.8^o$:

$$E(y^2; \tfrac{1}{8}) = y^2\Big(\tfrac{193127}{196608}\, y^{18} + \tfrac{15}{2}\, y^{16} + \tfrac{333}{2}\, y^{14} + \tfrac{5319}{4}\, y^{12} + \tfrac{2624643}{32}\, y^8 + \tfrac{177147}{2}\, y^6 + 2125764\, y^4 + 1259712\, \big(y - \tfrac{59}{497664}\, y^9\big)^2 + 15116544\Big).$$

This example highlights the approach for bounding α. Subsequent values of β can be used to obtain bounds closer to the reported maximum α of 89.5^o.

2.2 The CSTW Algebraic Conditions for A-stability

Closely related to the E-polynomial approach is an LMI based on algebraic conditions for A-stability, laid out in a line of work by Cooper [13], Scherer and Türke [42] and Scherer and Wendler [43]. We contribute theory that sharpens these algebraic conditions, which enables their practical use within an SDP framework.

2.2.1 Cooper factorization of the E-polynomial.

Cooper [13] initially established sufficient algebraic conditions for A-stability by factorizing the E-polynomial into a quadratic form. For the purposes of factorizing the E-polynomial, let the vectors:

$$\boldsymbol{p} = [p_0, \ldots, p_{r-1}]^T \quad \text{and} \quad \boldsymbol{Z} = [1, z, \ldots, z^{r-1}]^T.$$

Additionally, define the matrices:

$$\boldsymbol{M} = [\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \ldots, \boldsymbol{A}^{r-1}\boldsymbol{e}] \quad \text{and} \quad \boldsymbol{P} = \begin{bmatrix} 1 & p_{r-1} & p_{r-2} & \cdots & p_2 & p_1 \\ 0 & 1 & p_{r-1} & \cdots & p_3 & p_2 \\ 0 & 0 & 1 & \cdots & p_4 & p_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & p_{r-1} \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

The integer r and the elements of the vector $\boldsymbol{p}$ come from expressing $\boldsymbol{A}^r\boldsymbol{e}$ as a linear combination of the preceding vectors in the sequence $\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \boldsymbol{A}^2\boldsymbol{e}, \ldots$. Specifically, r is the smallest positive integer such that:

$$\boldsymbol{A}^r\boldsymbol{e} + \boldsymbol{M}\boldsymbol{p} = 0.$$

The minimal polynomial p_e of $\boldsymbol{A}$ for $\boldsymbol{e}$, the smallest degree monic polynomial for which $p_e(\boldsymbol{A})\boldsymbol{e} = 0$, is then defined by:

$$p_e(z) = z^r + \boldsymbol{Z}^T\boldsymbol{p}$$

and the transformed polynomial $D(z)$, which is also the denominator of the rational form of the stability function $W(z)$, is given by:

$$D(z) = 1 + p_{r-1}z + \cdots + p_1 z^{r-1} + p_0 z^r.$$

Theorem 2.2.1 (Cooper [13])**.** *A method is A-stable if and only if $D(z)$ is not zero in $\overline{\mathbb{C}}_-$ and there exists a real symmetric matrix $\boldsymbol{R}$ such that $\boldsymbol{M}^T(\boldsymbol{R}\boldsymbol{e} - \boldsymbol{b}) = 0$ and*

$$E(y) := \boldsymbol{Z}^T(\bar{z})\boldsymbol{P}^T\boldsymbol{M}^T(\boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T)\boldsymbol{M}\boldsymbol{P}\boldsymbol{Z}(z) \geq 0 \quad \forall z = iy, \quad y \text{ real.}$$

This inequality is independent of $\boldsymbol{R}$.

2.2.2 Scherer–Türke, and the application of the KYP Lemma to RK stability functions.

Scherer and Türke [42] later re-derived almost identical conditions by applying the Kalman-Yakubovich-Popov (KYP) Lemma to the stability function $W(z)$. Note, a function f is *positive* if it is analytic in $\mathbb{C}_+$ and $\mathrm{Re}\,(f) > 0$ for all z satisfying $\mathrm{Re}\,(z) > 0$. The set of positive functions form a convex cone.

Lemma 2.2.2 (KYP)**.** *A nondegenerate rational function* $f(z) = \boldsymbol{u}^T(z\boldsymbol{I} + \boldsymbol{Y})^{-1}\boldsymbol{v}$ *(*$\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^s$*,* $\boldsymbol{Y} \in \mathbb{R}^{s\times s}$*) is positive if and only if there exists* $\boldsymbol{R}^T = \boldsymbol{R}$ $(\in \mathbb{R}^{s\times s})$ *such that*

$$\boldsymbol{R} \succeq 0 \qquad \textit{and} \qquad \boldsymbol{R}\boldsymbol{v} = \boldsymbol{u} \qquad \textit{and} \qquad \boldsymbol{R}\boldsymbol{Y} + \boldsymbol{Y}^T\boldsymbol{R} \succeq 0\,. \tag{2.15}$$

The application of the KYP lemma to the RK stability function provided necessary and sufficient algebraic conditions for RK schemes with nondegenerate stability functions.

Theorem 2.2.3. *(Scherer–Türke, Theorem 4.1 in [42]) Consider a RK scheme* $(\boldsymbol{A}, \boldsymbol{b})$ *having nondegenerate stability function* $W(z)$ *from* (1.3)*. Then the scheme is A-stable if and only if there exists a matrix* $\boldsymbol{R} \in \mathbb{S}^s$ *such that*

$$\begin{cases} \boldsymbol{R}\,\boldsymbol{e} = \boldsymbol{b}\,, \\ \boldsymbol{X} = \boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T\,, \\ \boldsymbol{R} \succ 0\,, \\ \boldsymbol{X} \succeq 0\,. \end{cases} \tag{2.16}$$

Corollary 4.2 in [42] states that the conditions (2.16) are sufficient for RK schemes with a degenerate stability function. Subsequent work by Scherer and Wendler [43] provided even more general algebraic conditions applicable to degenerate stability functions.

2.2.3 Scherer–Wendler and the Kalman decomposition for non-degenerate stability functions.

Degenerate stability functions are, in fact, used in practice. Recently, Runge-Kutta schemes with degenerate stability functions have been used to avoid order reduction via the weak stage order conditions (also known as parabolic order conditions) [7, 8, 9].

In this section, we utilize a concept from control theory called the Kalman decomposition, as in [43], to analyze the stability function. This decomposition provides a useful perspective in situations where the stability function is degenerate.

To begin, we introduce the concept of invariant subspaces. A subspace $V \subset \mathbb{R}^n$ is said to be A-*invariant* if $\boldsymbol{A}\boldsymbol{v} \in V$ for all $\boldsymbol{v} \in V$. Let the columns of a matrix $\boldsymbol{V}$ form a basis for an A-invariant space V, with $\boldsymbol{V}'$ being a complementary basis so that $\boldsymbol{W} := [\boldsymbol{V}\ \boldsymbol{V}']$ is a square invertible matrix. Then $\boldsymbol{A}$ has a block form:

$$\boldsymbol{W}^{-1}\boldsymbol{A}\boldsymbol{W} = \begin{bmatrix} \boldsymbol{A}_{vv} & \boldsymbol{A}_{v\bar{v}} \\ 0 & \boldsymbol{A}_{\bar{v}\bar{v}} \end{bmatrix}. \tag{2.17}$$

Next, we introduce two invariant subspaces related to the coefficient matrix $\boldsymbol{A}$ via the controllability matrix

$$\boldsymbol{Q}_{\text{con}} := \begin{bmatrix} \boldsymbol{e} & \boldsymbol{A}\boldsymbol{e} & \dots & \boldsymbol{A}^{s-1}\boldsymbol{e} \end{bmatrix} \tag{2.18}$$

and the observability matrix

$$\boldsymbol{Q}_{\text{obs}} := \begin{bmatrix} \boldsymbol{b}^T \\ \boldsymbol{b}^T\boldsymbol{A} \\ \vdots \\ \boldsymbol{b}^T\boldsymbol{A}^{s-1} \end{bmatrix}. \tag{2.19}$$

Let $\operatorname{col}\boldsymbol{Q}_{\text{con}}$ denote the column space of $\boldsymbol{Q}_{\text{con}}$ and $\ker\boldsymbol{Q}_{\text{obs}}$ denote the null space (or kernel) of $\boldsymbol{Q}_{\text{obs}}$. By construction,

$$V_1 := \operatorname{col}\boldsymbol{Q}_{\text{con}} \qquad \text{and} \qquad V_2 := \ker\boldsymbol{Q}_{\text{obs}} \tag{2.20}$$

are both A-invariant subspaces. It then follows that both $V_1 \cap V_2$ and $V_1 + V_2$ are A-invariant subspaces where

$$V_1 + V_2 := \{\boldsymbol{v}_1 + \boldsymbol{v}_2 \ : \ \boldsymbol{v}_1 \in V_1, \ \boldsymbol{v}_2 \in V_2\}.$$

This motivates the definition of four (not necessarily unique) vector spaces $X_1, X_2, X_3, X_4 \in \mathbb{R}^s$ defined as:

$$X_1 = \text{col}\, \boldsymbol{Q}_{\text{con}} \cap \ker \boldsymbol{Q}_{\text{obs}}, \tag{2.21}$$

$$X_1 \oplus X_2 = \text{col}\, \boldsymbol{Q}_{\text{con}}, \tag{2.22}$$

$$X_1 \oplus X_3 = \ker \boldsymbol{Q}_{\text{obs}}. \tag{2.23}$$

The space X_4 is chosen so that:

$$\mathbb{R}^s = X_1 \oplus X_2 \oplus X_3 \oplus X_4 \,.$$

The symbol $\oplus$ denotes an *algebraic direct sum* i.e., X_1, X_2, X_3, and X_4 are disjoint and their linear span is $\mathbb{R}^s$; every vector $\boldsymbol{x} \in \mathbb{R}^s$ has a unique decomposition

$$\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{x}_3 + \boldsymbol{x}_4$$

with $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4$ in X_1, X_2, X_3, X_4 respectively [4, Chapter I.3]. A basis for the subspaces in an algebraic direct sum must be linearly independent, but are not required to be orthogonal.

Consider a basis for X_1, X_2, X_3, X_4, denoted by matrices $\boldsymbol{T}_1, \boldsymbol{T}_2, \boldsymbol{T}_3, \boldsymbol{T}_4$ respectively, and set

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{T}_1 & \boldsymbol{T}_2 & \boldsymbol{T}_3 & \boldsymbol{T}_4 \end{bmatrix} \,. \tag{2.24}$$

With $\boldsymbol{T}$, we perform a change of basis for $\boldsymbol{A}$, resulting in

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} & \boldsymbol{A}_{14} \\ 0 & \boldsymbol{A}_{22} & 0 & \boldsymbol{A}_{24} \\ 0 & 0 & \boldsymbol{A}_{33} & \boldsymbol{A}_{34} \\ 0 & 0 & 0 & \boldsymbol{A}_{44} \end{bmatrix}. \tag{2.25}$$

The zero blocks in (2.25) follow from (2.17) combined with the fact that $X_1 \oplus X_2$ and $X_1 \oplus X_3$ are both A-invariant.

Furthermore, since $\boldsymbol{e} \in \operatorname{col} \boldsymbol{Q}_{\text{con}}$ and $\boldsymbol{b}^T \perp \ker \boldsymbol{Q}_{\text{obs}}$, we have:

$$\boldsymbol{T}^{-1}\boldsymbol{e} = \begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}^T\boldsymbol{T} = \begin{bmatrix} 0 & \boldsymbol{H}_2 & 0 & \boldsymbol{H}_4 \end{bmatrix} \quad \text{where} \quad \boldsymbol{H}_i = \boldsymbol{b}^T\boldsymbol{T}_i\,.$$

In situations where the stability function is degenerate, we can use the matrices produced by $\boldsymbol{T}$ to obtain a minimal/non-degenerate stability function. Starting from

$$\begin{aligned} W(z) &= 1 + z\boldsymbol{b}^T(I - z\boldsymbol{A})^{-1}\boldsymbol{e} \\ &= 1 + z\boldsymbol{b}^T\boldsymbol{T}\boldsymbol{T}^{-1}(I - z\boldsymbol{A})^{-1}\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{e} \\ &= 1 + z\boldsymbol{b}^T\boldsymbol{T}(I - z\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T})^{-1}\boldsymbol{T}^{-1}\boldsymbol{e}\,, \end{aligned}$$

we obtain

$$\begin{aligned} W(z) &= 1 + z\begin{bmatrix} 0 & \boldsymbol{H}_2 & 0 & \boldsymbol{H}_4 \end{bmatrix}\left(\boldsymbol{I} - z\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} & \boldsymbol{A}_{14} \\ 0 & \boldsymbol{A}_{22} & 0 & \boldsymbol{A}_{24} \\ 0 & 0 & \boldsymbol{A}_{33} & \boldsymbol{A}_{34} \\ 0 & 0 & 0 & \boldsymbol{A}_{44} \end{bmatrix}\right)^{-1}\begin{bmatrix} \boldsymbol{G}_1 \\ \boldsymbol{G}_2 \\ 0 \\ 0 \end{bmatrix} \\ &= 1 + z\boldsymbol{H}_2\left(\boldsymbol{I} - z\boldsymbol{A}_{22}\right)^{-1}\boldsymbol{G}_2. \end{aligned} \tag{2.26}$$

Thus, a scheme is A-stable if and only if the reduced (minimal) representation in (2.26) is A-stable. Applying the KYP Lemma 2.2.2 to the stability function (2.26), we get that a scheme is A-stable if and only if there exists a matrix $\boldsymbol{R} \in \mathbb{S}^s$ such that

$$\boldsymbol{R}\boldsymbol{G}_2 = \boldsymbol{H}_2^T \qquad \text{and} \qquad \boldsymbol{R}\boldsymbol{A}_{22} + \boldsymbol{A}_{22}^T\boldsymbol{R} - \boldsymbol{H}_2^T\boldsymbol{H}_2 \succeq 0\,.$$

A conceptually straightforward approach to testing degenerate schemes for A-stability is to first compute the Kalman decomposition to obtain the block matrices defining (2.26), and then test for A-stability on the minimal representation of $W(z)$. However, this approach may not always be computationally attractive.

The Scherer-Wendler result [43] requires only a partial Kalman decomposition. Instead of computing every block of $\boldsymbol{T}$, it suffices to compute a basis that includes the span of $\operatorname{col}[\boldsymbol{T}_1\,\boldsymbol{T}_2]$ and is contained in $\operatorname{col}[\boldsymbol{T}_1\,\boldsymbol{T}_2\,\boldsymbol{T}_3]$. Specifically, a basis for the space $\operatorname{col}\boldsymbol{Q}_{\text{obs}}$ can be computed to define a matrix $\boldsymbol{M}$, which can then be used to apply the KYP lemma.

Theorem 2.2.4. *(Scherer–Wendler, Theorem 6.1 in [43]) Let $\boldsymbol{M}$ be any matrix whose column space is equal to the span of $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \boldsymbol{A}^2\boldsymbol{e}, \ldots, \boldsymbol{A}^{s-1}\boldsymbol{e}]$. The RK scheme $(\boldsymbol{A}, \boldsymbol{b})$ is A-stable if and only if there exists a matrix $\boldsymbol{R} \in \mathbb{S}^s$ such that*

$$\begin{cases} \boldsymbol{R}\,\boldsymbol{e} = \boldsymbol{b}\,, \\ \boldsymbol{X} = \boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T\,, \\ \boldsymbol{M}^T\boldsymbol{R}\boldsymbol{M} \succeq 0\,, \\ \boldsymbol{M}^T\boldsymbol{X}\boldsymbol{M} \succeq 0\,. \end{cases} \tag{2.27}$$

Note that Theorem 2.2.4 does not require the stability function $W(z)$ to be nondegenerate. We refer to (2.27) as the CSTW conditions.

Remark. When $W(z)$ is nondegenerate, the matrix $\boldsymbol{M}$ can be the identity matrix $\boldsymbol{I}$. If $W(z)$ is degenerate, then (2.27) with $\boldsymbol{M} = \boldsymbol{I}$ provides a sufficient condition for A-stability, but it may not be necessary. In the degenerate case, $\boldsymbol{M}$ may be chosen as $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \ldots, \boldsymbol{A}^{r-1}\boldsymbol{e}]$, where r is the smallest number for which $\boldsymbol{A}^r\boldsymbol{e}$ can be expressed in terms of the vectors $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \ldots, \boldsymbol{A}^{r-1}\boldsymbol{e}]$.

2.2.4 Formulating the CSTW LMI.

The set of all matrices $\boldsymbol{R}$ that satisfy the CSTW conditions (2.27) is convex and the set is readily converted into an LMI by parameterizing the equality constraints. Let

$$\boldsymbol{B} := \operatorname{diag}\begin{bmatrix} b_1, & b_2, & \ldots, & b_s \end{bmatrix},$$

and

$$\boldsymbol{N}_{ij} := \boldsymbol{n}_{ij}\boldsymbol{n}_{ij}^T \qquad \text{where} \qquad \boldsymbol{n}_{ij} = \boldsymbol{e}_i - \boldsymbol{e}_j\,, \tag{2.28}$$

for $1 \le i < j \le s$ with $\boldsymbol{e}_i$ being the ith unit vector. Then by construction, $\boldsymbol{N}_{ij}$ is a basis for the vector space $\{\boldsymbol{N} \in \mathbb{S}^s : \boldsymbol{N}\boldsymbol{e} = 0\}$ (cf. [43]). For example, $s = 2$ has the basis matrix:

$$\boldsymbol{N}_{12} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

while $s = 3$ has three basis matrices:

$$\boldsymbol{N}_{12} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{N}_{13} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{N}_{23} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{bmatrix}.$$

Using the basis matrices $\boldsymbol{N}_{ij}$, we can parameterize $\boldsymbol{R}$ as:

$$\boldsymbol{R} = \boldsymbol{B} + \boldsymbol{N}(\boldsymbol{\eta}) \quad \text{where} \quad \boldsymbol{N}(\boldsymbol{\eta}) = \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \eta_{ij}\boldsymbol{N}_{ij}\,. \tag{2.29}$$

Combining the parameterization (2.29) with the inequalities from the CSTW conditions (2.27) yields the conditions in the form of an LMI:

$$\begin{bmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{B} - \boldsymbol{b}\boldsymbol{b}^T \end{bmatrix} + \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \eta_{ij} \begin{bmatrix} \boldsymbol{N}_{ij} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{N}_{ij}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{ij} \end{bmatrix} \succeq 0\,. \tag{2.30}$$

2.3 Sharpening the CSTW Conditions

In order to motivate the theoretical developments in this section, we must first introduce the A-stability certification algorithm at a high level. A detailed outline of the algorithm will be provided in Chapter 3.

Given a fixed pair $(\boldsymbol{A}, \boldsymbol{b})$, we define the following convex set:

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) := \left\{ \boldsymbol{\eta} \in \mathbb{R}^{s(s-1)/2} \ : \ \text{The LMI (2.30) holds} \right\}. \tag{2.31}$$

The CSTW conditions in Theorem 2.2.4 are equivalently stated in the following corollary.

Corollary 2.3.1. *A scheme* $(\boldsymbol{A}, \boldsymbol{b})$ *is A-stable if and only if* $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ *is non-empty.*

The A-stability certification algorithm uses computational optimization tools to find an element of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$. The algorithm involves six main steps:

1. Using the RK coefficients and CSTW conditions to create a symbolic version of the parameterized LMI (2.30).

2. Converting the symbolic LMI to double precision.

3. Employing the double precision LMI within the semidefinite feasibility problem (P1), implemented using a convex programming modeling system known as CVX [15].

4. Solving the feasibility problem using CVX, which outputs double precision parameters $\boldsymbol{\eta}^*$.

5. Passing the CVX output $\boldsymbol{\eta}^*$ back to the symbolic LMI matrix.

6. Certifying the symbolic LMI matrix as A-stable through a symbolic LDL factorization.

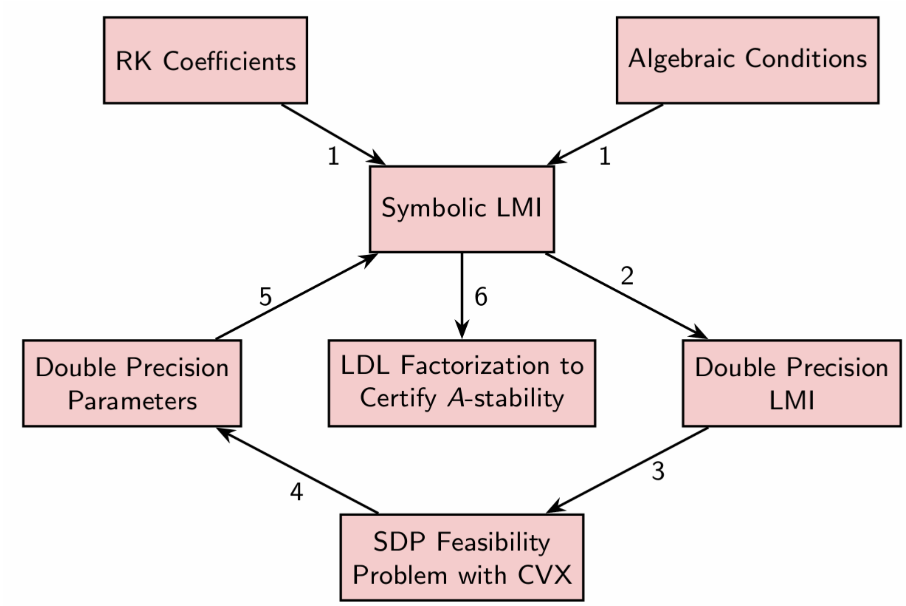

Figure 2.1 The diagram shows the general flow of the A-stability certification algorithm.

The convex set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ has a dimension of at most $s(s-1)/2$ (see Definition 1.2.1), since $\boldsymbol{R}$ lies in an affine space of symmetric matrices satisfying the linear constraint $\boldsymbol{Re} = \boldsymbol{b}$. Practically, we must parameterize the affine hull of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ in step 1 to ensure that steps 5 and 6 in the algorithm yield a rigorous certificate.

Unfortunately, the CSTW equality conditions (2.27) are insufficient to parameterize the affine hull of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ and therefore the upper bound on $\dim(\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}))$ of $s(s-1)/2$ is not sharp. Without characterizing the affine hull of the LMI (2.30) (or equivalently $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$), step 4 will yield an $\boldsymbol{\eta}^*$ in a larger dimensional space than the affine hull, and steps 5 and 6 are unlikely to generate a positive certificate. The phenomena described in this paragraph can be visualized in Figure 2.2.

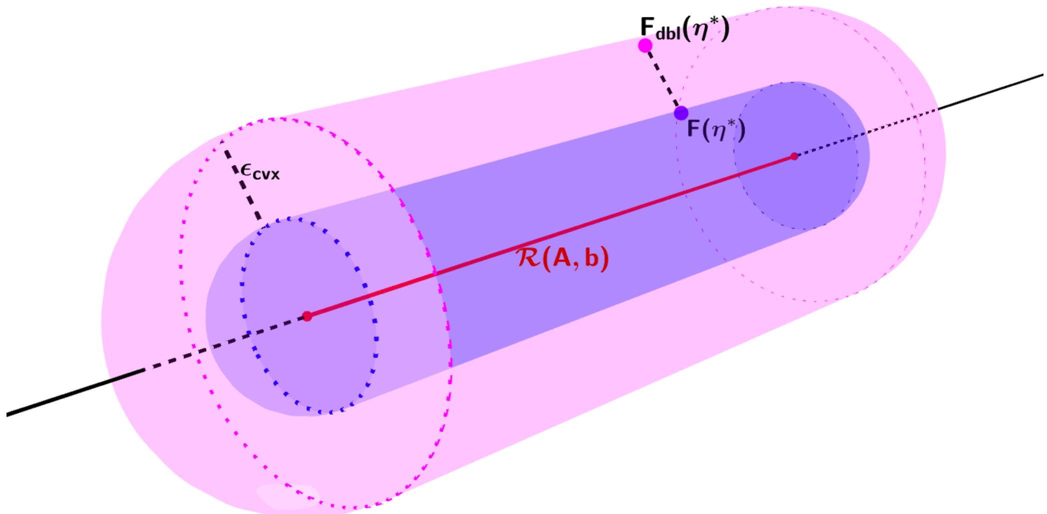

Figure 2.2 The blue capsule represents the affine hull parameterized by the equality constraints and bounded by the inequality constraints of the CSTW conditions (2.27). CVX produces a double precision $\boldsymbol{\eta}^*$ in an extension of the blue capsule (pink capsule). The CVX output $\boldsymbol{\eta}^*$ is passed to the symbolic LMI, F, projecting the solution onto the blue capsule. Since the set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is a lower dimensional set (red line) the algorithm did not successfully find an element of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ and has not produced a certificate of stability.

The following subsections describe additional structure in the CSTW conditions (2.27) that reduce the upper bound of the dimension of the feasible set. Identifying this structure allows us to sharpen the CSTW conditions (2.2.4) for use within the A-stability certification algorithm.

2.3.1 Null vectors associated with the Tall-Tree order conditions.

The authors in [43] observe, though do not resolve, that zero eigenvalues of $\boldsymbol{X}$ may limit the practical application of the algebraic conditions (2.27). The following lemma shows that the order conditions result in $\boldsymbol{X}$ always having a family of zero eigenvalues whenever the CSTW LMI is feasible.

Lemma 2.3.2. *Let the RK scheme* $(\boldsymbol{A}, \boldsymbol{b})$ *satisfy the tall-tree order conditions* (1.7) *with order* $p \geq 2$. *If* $\boldsymbol{R} \in \mathbb{S}^s$ *satisfies* $\boldsymbol{Re} = \boldsymbol{b}$ *and* $\boldsymbol{X} \succeq 0$ *where*

$$\boldsymbol{X} := \boldsymbol{RA} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{bb}^T\,,$$

then $\boldsymbol{X}$ *has the null vectors* $\boldsymbol{A}^{j-1}\boldsymbol{e}$*, that is:*

$$\boldsymbol{X}\boldsymbol{A}^{j-1}\boldsymbol{e} = 0 \qquad \textit{for} \qquad 1 \leq j \leq \left\lfloor \frac{p}{2} \right\rfloor .$$

The proof utilizes the fact that for any matrix $\boldsymbol{X} \succeq 0$,

$$\text{if} \qquad \boldsymbol{v}^T\boldsymbol{X}\boldsymbol{v} = 0\,, \qquad \text{then} \qquad \boldsymbol{X}\boldsymbol{v} = 0\,.$$

For instance, the Cholesky factorization $\boldsymbol{X} = \boldsymbol{Q}^T\boldsymbol{Q}$ shows that $\boldsymbol{v}^T\boldsymbol{X}\boldsymbol{v} = \|\boldsymbol{Q}\boldsymbol{v}\|^2 = 0$.

Proof. To simplify notation, let

$$\boldsymbol{v}_j = \boldsymbol{A}^j\boldsymbol{e} \qquad j \geq 0\,.$$

We show that for all

$$0 \leq n \leq \left\lfloor \frac{p}{2} \right\rfloor - 1\,,$$

the expression

$$\boldsymbol{v}_n^T\boldsymbol{X}\boldsymbol{v}_n = 0\,.$$

Since $\boldsymbol{X} \succeq 0$, it follows that $\boldsymbol{X}\boldsymbol{v}_n = 0$. For $n = 0$, the tall-tree conditions are $\boldsymbol{b}^T\boldsymbol{e} = 1$ and $\boldsymbol{b}^T\boldsymbol{v}_1 = \frac{1}{2}$, hence:

$$\begin{aligned}\boldsymbol{v}_0^T\boldsymbol{X}\boldsymbol{v}_0 &= \boldsymbol{b}^T\boldsymbol{v}_1 + \boldsymbol{v}_1^T\boldsymbol{b} - (\boldsymbol{e}^T\boldsymbol{b})^2 \\ &= 0\,.\end{aligned}$$

We now proceed by strong induction: Let $n \leq \lfloor \frac{p}{2} \rfloor - 1$ be any positive integer, and assume that $\boldsymbol{X}\boldsymbol{v}_k = 0$ for all $0 \leq k \leq n-1$. We show that $\boldsymbol{v}_n^T\boldsymbol{X}\boldsymbol{v}_n = 0$, which then completes the proof.

By hypothesis and the pth order tall-tree conditions, we have that

$$\boldsymbol{b}^T\boldsymbol{v}_j = \frac{1}{(j+1)!} \qquad \text{for} \qquad j = 0, \ldots, 2n+1\,, \tag{2.32}$$

as $2n + 1 \leq p - 1$ by the choice of n.

The first step is to obtain an expression for $\boldsymbol{R}\boldsymbol{v}_n$. Substituting the definition of $\boldsymbol{X}$ in terms of $\boldsymbol{R}$ into the induction hypothesis $\boldsymbol{X}\boldsymbol{v}_k = 0$, yields the recursion relation

$$\boldsymbol{R}\boldsymbol{v}_{k+1} = \boldsymbol{b}\boldsymbol{b}^T\boldsymbol{v}_k - \boldsymbol{A}^T\boldsymbol{R}\boldsymbol{v}_k \qquad \text{for} \qquad 0 \le k \le n-1\,, \tag{2.33}$$

so that $\boldsymbol{R}\boldsymbol{v}_{k+1}$ is written in terms of $\boldsymbol{R}\boldsymbol{v}_k$. Setting $k = n-1$ in (2.33), we can use the conditions (2.32) on dot products $\boldsymbol{b}^T\boldsymbol{v}_k$ and iteratively eliminate $\boldsymbol{R}\boldsymbol{v}_j$ to express $\boldsymbol{R}\boldsymbol{v}_n$ in the basis $\{\boldsymbol{b}, \boldsymbol{A}^T\boldsymbol{b}, \ldots, (\boldsymbol{A}^T)^{n-1}\boldsymbol{b}\}$:

$$\boldsymbol{R}\boldsymbol{v}_n = \sum_{j=0}^{n} \frac{(-1)^j}{(n-j)!}(\boldsymbol{A}^T)^j\,\boldsymbol{b}\,. \tag{2.34}$$

We then have:

$$\begin{aligned}
\boldsymbol{v}_n^T\boldsymbol{X}\boldsymbol{v}_n &= \boldsymbol{v}_n^T\left(\boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T\right)\boldsymbol{v}_n\,,\\
&= 2\,\boldsymbol{v}_{n+1}^T\boldsymbol{R}\boldsymbol{v}_n - \frac{1}{(n+1)!^2} && (\text{since } \boldsymbol{A}\boldsymbol{v}_n = \boldsymbol{v}_{n+1})\,,\\
&= 2\sum_{j=0}^{n}\frac{(-1)^j}{(n-j)!(n+j+2)!} - \frac{1}{(n+1)!^2} && (\text{via (2.32) and (2.34)})\,,\\
&= \frac{(-1)^n}{(2n+2)!}\underbrace{\sum_{j=0}^{2n+2}\binom{2n+2}{j}1^{2n+2-j}(-1)^j}_{=(1-1)^{2n+2}}\,,\\
&= 0\,.
\end{aligned}$$

□

Remark. Lemma 2.3.2 can be viewed as a generalization of the fact that algebraically stable methods admit a set of null vectors in their algebraic stability matrix $\boldsymbol{B}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{B} - \boldsymbol{b}\boldsymbol{b}^T$. For further reference, see the proof of [30, Lemma 13.14].

The following theorem sharpens the original CSTW conditions to incorporate the null vectors of $\boldsymbol{X}$ from Lemma 2.3.2. Incorporating the null vectors from

Lemma 2.3.2 provides a tighter bound on the affine hull of the original CSTW conditions (2.27).

Theorem 2.3.3 (Modified CSTW Conditions)**.** *Let $\boldsymbol{M}$ be any matrix whose column space is equal to the span of $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \boldsymbol{A}^2\boldsymbol{e}, \ldots, \boldsymbol{A}^{s-1}\boldsymbol{e}]$. The pth order RK scheme $(\boldsymbol{A}, \boldsymbol{b})$ is A-stable if and only if there exists a matrix $\boldsymbol{R} \in \mathbb{S}^s$ that satisfies the CSTW conditions* (2.27) *and the condition matrix $\boldsymbol{X}$ satisfies*

$$\boldsymbol{X}\boldsymbol{A}^{j-1}\boldsymbol{e} = 0 \qquad \textit{for} \quad j = 1, \ldots, \left\lfloor \frac{p}{2} \right\rfloor . \tag{2.35}$$

Proof. Note that the proof of Lemma 2.3.2 holds if $\boldsymbol{X} \succeq 0$ is replaced with $\boldsymbol{M}^T\boldsymbol{X}\boldsymbol{M} \succeq 0$ for any matrix $\boldsymbol{M}$ whose columns span the null vectors $\boldsymbol{A}^{j-1}\boldsymbol{e}$ for $j \le \lfloor \frac{p}{2} \rfloor$. □

Remark. The benefit of Theorem 2.3.3 is that the additional equality constraints (2.35) sharpen the affine hull of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$.

An LMI for the modified CSTW conditions, defined by (2.27) and (2.35), can be obtained by parameterizing the η_{ij} variables in (2.29) to satisfy the additional affine constraints (2.35). The parameterization is then substituted back into (2.30).

2.3.2 Null vectors associated with singular coefficient matrix $\boldsymbol{A}$.

There is an additional null space that can arise for the matrix $\boldsymbol{X}$ in settings where $\boldsymbol{A}$ has a nontrivial null space. Let

$$J := \left\{ \boldsymbol{v} \in \mathbb{R}^s \; : \; \boldsymbol{b}^T\boldsymbol{v} = 0 \, , \boldsymbol{A}\boldsymbol{v} = 0 \right\} .$$

By the definition of J we have the following.

Lemma 2.3.4. *If $\boldsymbol{X} \succeq 0$, then $\boldsymbol{X}\boldsymbol{v} = 0$ for all $\boldsymbol{v} \in J$.*

Proof. Immediately one has

$$\boldsymbol{v}^T\boldsymbol{X}\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{R}\boldsymbol{A}\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{A}^T\boldsymbol{R}\boldsymbol{v} - (\boldsymbol{v}^T\boldsymbol{b})^2 = 0 \, ,$$

which, combined with $\boldsymbol{X} \succeq 0$, gives the desired result. □

Since DIRK coefficient matrices have non-zero entries along the diagonal, the corresponding subspace J is the trivial subspace, and Lemma 2.3.4 provides no new information.

The situation changes when an RK scheme has an explicit first stage; in other words, the first row of $\boldsymbol{A}$ is the zero vector. For example, an EDIRK has the first stage of the Butcher matrix explicit, such that $a_{11} = 0$ and $a_{jj} \neq 0$ for $2 \leq j \leq s$. In this instance, the coefficient matrix $\boldsymbol{A}$ has a nontrivial null vector $\boldsymbol{v} \in \ker(\boldsymbol{A})$. Moreover, the null vector $\boldsymbol{v}$ lies in the space spanned by $\{\boldsymbol{A}^j \boldsymbol{e}\}$ and may be constructed as follows.

First, let r be the smallest integer for which $\boldsymbol{A}^r \boldsymbol{e}$ can be written as a linear combination of smaller powers $\boldsymbol{A}^j \boldsymbol{e}$, i.e., for some set of coefficients p_j,

$$\left(\boldsymbol{A}^r + p_{r-1}\boldsymbol{A}^{r-1} + \cdots + p_1 \boldsymbol{A} + p_0 \boldsymbol{I}\right)\boldsymbol{e} = 0\,. \tag{2.36}$$

Taking the inner product of (2.36) with $\boldsymbol{e}_1$ from the left, and using the fact that, for RK schemes with an explicit first stage, $\boldsymbol{e}_1^T \boldsymbol{A} = 0$, yields $p_0 = 0$. Thus, one has

$$\boldsymbol{A}\underbrace{\left(\boldsymbol{A}^{r-1} + p_{r-1}\boldsymbol{A}^{r-2} + \cdots + p_1 \boldsymbol{I}\right)\boldsymbol{e}}_{=\boldsymbol{v}} = 0\,, \tag{2.37}$$

so that $\boldsymbol{A}\boldsymbol{v} = 0$. In (2.37), $\boldsymbol{v} \neq 0$, otherwise, this would violate the definition of r as being the smallest integer for which (2.36) holds.

These observations lead to the following lemma.

Lemma 2.3.5. *Given $(\boldsymbol{A}, \boldsymbol{b})$, let the first row of $\boldsymbol{A}$ be the zero vector, and let $\boldsymbol{b}$ be in the row space of $\boldsymbol{A}$. If $\boldsymbol{R} \in \mathbb{S}^s$ and $\boldsymbol{X} := \boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T \succeq 0$, then*

$$\boldsymbol{X}\boldsymbol{v} = 0 \quad \textit{with} \quad \boldsymbol{v} = \boldsymbol{M}\boldsymbol{p}_r\,,$$

where

$$\boldsymbol{M} = [\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \ldots, \boldsymbol{A}^{r-1}\boldsymbol{e}] \quad \textit{and} \quad \boldsymbol{p}_r = [1, p_{r-1}, p_{r-2}, \ldots, p_2, p_1]^T,$$

are the coefficients defining the (minimal) polynomial in (2.36)*.*

It is useful to look at the origin of the vector $\boldsymbol{v}$ in Lemma 2.3.4 in the context of the Kalman decomposition. In the Kalman decompositon, the space $\operatorname{sp}\{\boldsymbol{v}\}$ is a subspace of X_1, where

$$X_1 = \ker \boldsymbol{Q}_{\text{obs}} \cap \operatorname{sp} \boldsymbol{Q}_{\text{con}}.$$

The space X_1, and hence $\boldsymbol{v}$, can be projected out to obtain a minimal stability function involving only a subspace X_2. By computing the full Kalman decomposition first and then applying the CSTW conditions to a minimal stability function, introducing this zero eigenvalue could be avoided. However, as mentioned in Section 2.2.3, computing the full decomposition may not always be computationally attractive. Instead, we may add the null vectors from Lemma 2.3.5 to the modified CSTW conditions.

Theorem 2.3.6 (Modified CSTW Conditions for Explicit First Stage Schemes)**.** *Let* $(\boldsymbol{A}, \boldsymbol{b})$ *be a pth order RK scheme with the first row of* $\boldsymbol{A}$ *equal to the zero vector and* $\boldsymbol{b}$ *in the row space of* $\boldsymbol{A}$*. Let* $\boldsymbol{M}$ *be any matrix whose column space equals the span of* $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \boldsymbol{A}^2\boldsymbol{e}, \ldots, \boldsymbol{A}^{s-1}\boldsymbol{e}]$*. The scheme* $(\boldsymbol{A}, \boldsymbol{b})$ *is A-stable if and only if there exists a matrix* $\boldsymbol{R} \in \mathbb{S}^s$ *that satisfies the modified CSTW conditions* (2.27)*,* (2.35)*, and condition matrix* $\boldsymbol{X}$ *satisfies*

$$\boldsymbol{X}\boldsymbol{M}\boldsymbol{p}_r = 0 \quad \textit{where} \quad \boldsymbol{p}_r = [1, p_{r-1}, p_{r-2}, \ldots, p_2, p_1]^T \tag{2.38}$$

are the coefficients defining the (minimal) polynomial in (2.36)*.*

Proof. The proof follows directly from Lemma 2.3.5. □

2.3.3 Stability certification with the CSTW LMI.

In this section, we provide examples highlighting the significance of Lemmas 2.3.2 and 2.3.5. The examples demonstrate that without incorporating the affine constraints from Lemmas 2.3.2 and 2.3.2, computational approaches are unlikely to provide rigorous certifications for A-stability.

The convex set $\mathcal{R}$ has dimension at most $s(s-1)/2$ (i.e., the dimension of symmetric matrices minus the number of constraints imposed by $\boldsymbol{Re} = \boldsymbol{b}$). However, Theorem 2.3.3 indicates that combining the order conditions (1.7) with the inequality constraint $\boldsymbol{X} \succeq 0$ (or $\boldsymbol{M}^T\boldsymbol{X}\boldsymbol{M} \succeq 0$) further reduces the upper bound for the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$, and if $\boldsymbol{A}$ is singular then Theorem 2.3.6 indicates that the upper bound for the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is reduced further. Consequently, computational approaches that seek $\boldsymbol{\eta} \in \mathcal{R}$ will not be able to project the double precision solution back onto the exact feasible set without correctly characterizing the affine hull aff($\mathcal{R}$) — which is provided by Theorems 2.3.3 and 2.3.6.

The first example demonstrates how the zero eigenvalues of $\boldsymbol{X}$ reduce the upper bound for the dimension of the convex set $\mathcal{R}$. The second example demonstrates how a singular $\boldsymbol{A}$ reduces the upper bound for the dimension of $\mathcal{R}$ even further. The final example shows why $\boldsymbol{X} \succeq 0$ is a necessary hypothesis in both Lemmas.

Example 2.3.1 (SDIRK(3,2))**.** This example constructs the set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ defined by (2.31) for the following SDIRK 3-stage $p = 2$ method

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

The set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ consists of the set of parameters $\boldsymbol{\eta} = [\eta_{12}, \eta_{13}, \eta_{23}]^T$ which simultaneously satisfy the two LMIs $\boldsymbol{R}(\boldsymbol{\eta}) \succeq 0$ and $\boldsymbol{X}(\boldsymbol{\eta}) \succeq 0$ where

$$\begin{aligned}\boldsymbol{R}(\boldsymbol{\eta}) &= \boldsymbol{B} + \eta_{12}\,\boldsymbol{N}_{12} + \eta_{13}\,\boldsymbol{N}_{13} + \eta_{23}\,\boldsymbol{N}_{23}\,, \\ \boldsymbol{X}(\boldsymbol{\eta}) &= (\boldsymbol{B}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{B} - \boldsymbol{b}\boldsymbol{b}^T) + \eta_{12}\,(\boldsymbol{N}_{12}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{12}) \\ &\qquad + \eta_{13}\,(\boldsymbol{N}_{13}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{13}) + \eta_{23}\,(\boldsymbol{N}_{23}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{13})\,.\end{aligned}$$

Here $\boldsymbol{B} := \operatorname{diag} \boldsymbol{b} = \operatorname{diag}(1,\ -1,\ 1)$, while $\boldsymbol{N}_{ij}$ are defined in (2.28).

The CSTW conditions indicate that the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is at most 3. However, Theorem 2.3.3 further bounds the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ to be at most 1. Since $p = 2$, Theorem 2.3.3 implies that $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ may equivalently include the constraint $\boldsymbol{X}\boldsymbol{e} = 0$, i.e.,

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \left\{\boldsymbol{\eta} \in \mathbb{R}^3 \ :\ \boldsymbol{R}(\boldsymbol{\eta}) \succeq 0\,,\ \boldsymbol{X}(\boldsymbol{\eta}) \succeq 0\,, \boldsymbol{X}\boldsymbol{e} = 0\right\}.$$

The constraint $\boldsymbol{X}\boldsymbol{e} = 0$ imposes two independent linear equations on $\boldsymbol{\eta}$ whose solution forces $\eta_{12} = 3$ and $\eta_{23} = 2$. Thus,

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \left\{\boldsymbol{\eta} \in \mathbb{R}^3 \ :\ \eta_{12} = 3\,,\ \ \eta_{23} = 2\,,\ \ \boldsymbol{R}(\eta_{13}) \succeq 0\,,\ \boldsymbol{X}(\eta_{13}) \succeq 0\right\},$$

where

$$\begin{aligned}\boldsymbol{R}(\eta_{13}) &= \begin{bmatrix} 4 & -3 & 0 \\ -3 & 4 & -2 \\ 0 & -2 & 3 \end{bmatrix} \quad + \eta_{13} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \\ \boldsymbol{X}(\eta_{13}) &= \begin{bmatrix} 4 & -5 & 1 \\ -5 & 11 & -6 \\ 1 & -6 & 5 \end{bmatrix} + \eta_{13} \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & -1 \\ -1 & -1 & 2 \end{bmatrix}.\end{aligned}$$

Figure 2.3 visualizes the non-empty set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$. Note that $\dim(\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})) = 1$, and when parameterized in terms of $\eta_{12}, \eta_{13}, \eta_{23}$ it has an empty interior.

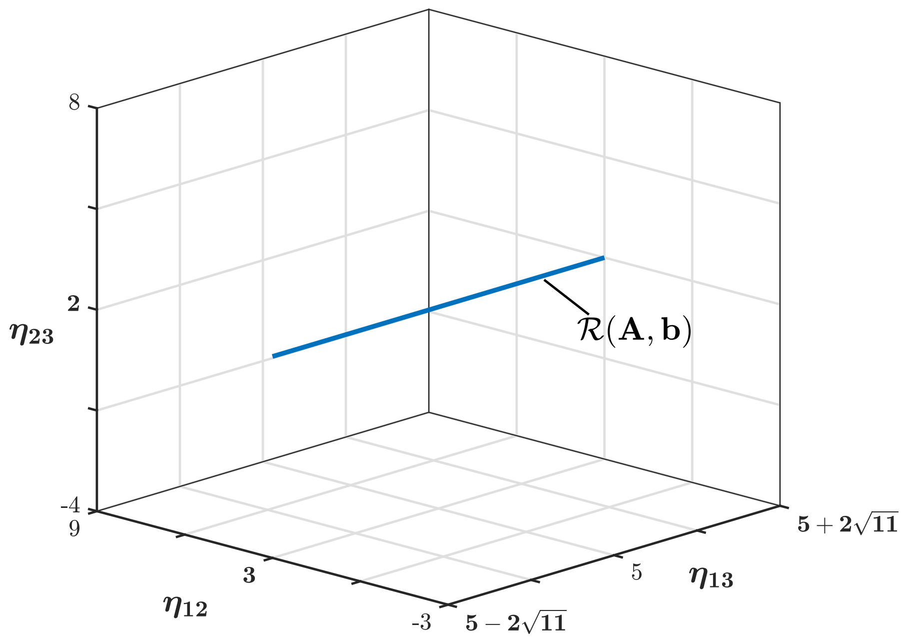

Figure 2.3 The blue line is a visualization of $\mathcal{R}(\boldsymbol{A},\boldsymbol{b})$ for SDIRK(3,2) in Example 2.3.1. Note the set $\mathcal{R}$ has dimension 1 (as defined in Definition 1.2.1) which is lower than the three-dimensional upper bound from Theorem 2.2.4. In this example, the null vectors for $\boldsymbol{X}$ in Theorem 2.3.3 provide a complete characterization of the affine hull of $\mathcal{R}$.

Example 2.3.2 (ARK2-DIRK-3-1-2 (M))**.** In this example, we analyze the set $\mathcal{R}(\boldsymbol{A},\boldsymbol{b})$ for an ESDIRK 3-stage, $p=2$ order scheme, with a singular matrix $\boldsymbol{A}$.

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{2-\sqrt{2}}{2} & \frac{2-\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{4} & \frac{\sqrt{2}}{4} & \frac{2-\sqrt{2}}{2} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \frac{\sqrt{2}}{4} \\ \frac{\sqrt{2}}{4} \\ \frac{2-\sqrt{2}}{2} \end{bmatrix}. \tag{2.39}$$

The set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is parameterized by $\boldsymbol{\eta} = [\eta_{12},\, \eta_{13},\, \eta_{23}]^T$ such that $\boldsymbol{R}(\boldsymbol{\eta}) \succeq 0$ and $\boldsymbol{X}(\boldsymbol{\eta}) \succeq 0$, where:

$$\begin{aligned} \boldsymbol{R}(\boldsymbol{\eta}) &= \boldsymbol{B} + \eta_{12}\, \boldsymbol{N}_{12} + \eta_{13}\, \boldsymbol{N}_{13} + \eta_{23}\, \boldsymbol{N}_{23}\,, \\ \boldsymbol{X}(\boldsymbol{\eta}) &= (\boldsymbol{B}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{B} - \boldsymbol{b}\boldsymbol{b}) + \eta_{12}\, (\boldsymbol{N}_{12}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{12}) \\ &\qquad + \eta_{13}\, (\boldsymbol{N}_{13}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{13}) + \eta_{23}\, (\boldsymbol{N}_{23}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{N}_{13})\,, \end{aligned}$$

with $\boldsymbol{B} = \mathrm{diag}(\boldsymbol{b})$. The vectors $\boldsymbol{e}$, $\boldsymbol{A}\boldsymbol{e}$, and $\boldsymbol{A}^2\boldsymbol{e}$ are linearly independent, thus $\boldsymbol{M}$ is the identity matrix.

According to the CSTW conditions, the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is at most 3. However, Theorem 2.3.3 includes the constraint $\boldsymbol{X}\boldsymbol{e} = 0$ for $p = 2$, therefore the dimension is at most 1:

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \left\{\boldsymbol{\eta} \in \mathbb{R}^3 \;:\; \boldsymbol{R}(\boldsymbol{\eta}) \succeq 0\,,\; \boldsymbol{X}(\boldsymbol{\eta}) \succeq 0\;, \boldsymbol{X}\boldsymbol{e} = 0\right\}.$$

The constraint $\boldsymbol{X}\boldsymbol{e} = 0$ results in the equations

$$\eta_{12} = \tfrac{\sqrt{2}}{2}\, \eta_{23} + \tfrac{1-\sqrt{2}}{4} \quad \text{and} \quad \eta_{13} = (1 - \sqrt{2})\, \eta_{23} + \tfrac{2\sqrt{2}-3}{2}.$$

Therefore, the set can be expressed as:

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \left\{ \boldsymbol{\eta} \in \mathbb{R}^3 \;:\; \begin{array}{c} \eta_{12} = \tfrac{\sqrt{2}}{2}\, \eta_{23} + \tfrac{1-\sqrt{2}}{4}, \\ \eta_{13} = (1 - \sqrt{2})\eta_{23} + \tfrac{2\sqrt{2}-3}{2}, \\ \boldsymbol{R}(\eta_{23}) \succeq 0,\; \boldsymbol{X}(\eta_{23}) \succeq 0 \end{array} \right\},$$

where

$$\boldsymbol{R}(\eta_{23}) = \begin{bmatrix} \frac{4\sqrt{2}-5}{4} & \frac{\sqrt{2}-1}{4} & \frac{3-2\sqrt{2}}{2} \\ \frac{\sqrt{2}-1}{4} & \frac{1}{4} & 0 \\ \frac{3-2\sqrt{2}}{2} & 0 & \frac{\sqrt{2}-1}{2} \end{bmatrix} + \eta_{23} \begin{bmatrix} \frac{2-\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & \sqrt{2}-1 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}+2}{2} & -1 \\ \sqrt{2}-1 & -1 & -\sqrt{2}-2 \end{bmatrix},$$

$$\boldsymbol{X}(\eta_{23}) = \begin{bmatrix} \frac{12\sqrt{2}-17}{8} & \frac{5\sqrt{2}-7}{8} & \frac{24-17\sqrt{2}}{8} \\ \frac{5\sqrt{2}-7}{8} & \frac{3-2\sqrt{2}}{4} & \frac{4-3\sqrt{2}}{8} \\ \frac{24-17\sqrt{2}}{8} & \frac{4-3\sqrt{2}}{8} & \frac{5\sqrt{2}-7}{2} \end{bmatrix} + \eta_{23} \begin{bmatrix} \frac{4-3\sqrt{2}}{2} & \frac{3-2\sqrt{2}}{2} & \frac{5\sqrt{2}-7}{2} \\ \frac{3-2\sqrt{2}}{2} & \frac{2-\sqrt{2}}{2} & \frac{3\sqrt{2}-5}{2} \\ \frac{5\sqrt{2}-7}{2} & \frac{3\sqrt{2}-5}{2} & 2\left(3-2\sqrt{2}\right) \end{bmatrix}.$$

Since $\boldsymbol{A}$ is singular, Theorem 2.3.6 includes the additional constraint $\boldsymbol{X M p}_r = 0$, reducing $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ to a single element:

$$\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \left\{ \boldsymbol{\eta} \in \mathbb{R}^3 \ : \ \eta_{12} = \tfrac{\sqrt{2}-2}{4}, \ \ \eta_{13} = \eta_{23} = \tfrac{4-3\sqrt{2}}{4}, \ \ \boldsymbol{R} \succeq 0 \, , \ \boldsymbol{X} \succeq 0 \right\},$$

where

$$\boldsymbol{R} = \begin{bmatrix} \frac{2-\sqrt{2}}{4} & \frac{2-\sqrt{2}}{4} & \frac{3\sqrt{2}-4}{4} \\ \frac{2-\sqrt{2}}{4} & \frac{2-\sqrt{2}}{4} & \frac{3\sqrt{2}-4}{4} \\ \frac{3\sqrt{2}-4}{4} & \frac{3\sqrt{2}-4}{4} & \frac{3\sqrt{2}-4}{\sqrt{2}} \end{bmatrix}, \qquad \boldsymbol{X} = \begin{bmatrix} \frac{17-12\sqrt{2}}{8} & \frac{17-12\sqrt{2}}{8} & \frac{12\sqrt{2}-17}{4} \\ \frac{17-12\sqrt{2}}{8} & \frac{17-12\sqrt{2}}{8} & \frac{12\sqrt{2}-17}{4} \\ \frac{12\sqrt{2}-17}{4} & \frac{12\sqrt{2}-17}{4} & \frac{17-12\sqrt{2}}{2} \end{bmatrix}$$

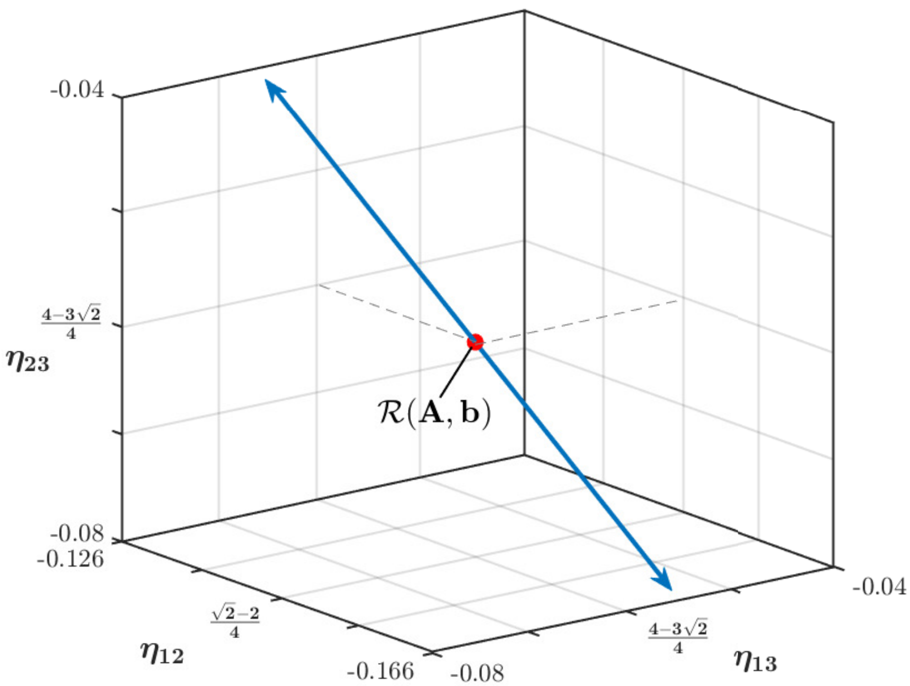

Figure 2.4 The set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ for ARK2-DIRK-3-1-2 (M) is visualized as a single point (red). Note that the original CSTW conditions bound the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ to be at most 3. The blue line visualizes the affine constraint on $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ characterized by Theorem 2.3.3, showing that $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ has at most dimension 1. Theorem 2.3.6 further restricts the dimension of $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ to be at most zero.

Example 2.3.3 (Hammer & Hollingsworth)**.** This example demonstrates why $\boldsymbol{X} \succeq 0$ is required as a hypothesis in Lemma 2.3.2 for $\boldsymbol{X}$ to have zero eigenvalues. The

Hammer & Hollingsworth two stage $p = 4$ method [27, Table II.7.3] is represented as

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} - \frac{\sqrt{3}}{6} \\ \frac{1}{4} + \frac{\sqrt{3}}{6} & \frac{1}{4} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}.$$

The scheme yields the following stability function and E-polynomial

$$W(z) = \frac{1 + z/2 + z^2/12}{1 - z/2 + z^2/12} \qquad \text{and} \qquad E(y) = 0\,.$$

Since $E(y) \geq 0$, the scheme is A-stable.

For this example, $s = 2$ so that the set $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b})$ is characterized by the one-dimensional set $\eta \in \mathbb{R}$ for which $\boldsymbol{X}(\eta) \succeq 0$ and $\boldsymbol{R}(\eta) \succeq 0$ where,

$$\boldsymbol{R}(\eta) = \frac{1}{2}\boldsymbol{I} + \eta \begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix}, \qquad \boldsymbol{X}(\eta) = -\frac{\sqrt{3}}{2}\eta \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

The only value of η for which $\boldsymbol{X}(\eta) \succeq 0, \boldsymbol{R}(\eta) \succeq 0$ is $\eta = 0$. Since $\boldsymbol{R}(0) = \frac{1}{2}\boldsymbol{I} \succeq 0$, we have that $\mathcal{R}(\boldsymbol{A}, \boldsymbol{b}) = \{0\}$. This shows that $\mathcal{R}$ is non-empty and provides a second proof that the scheme is A-stable.

This example highlights that the hypothesis of $\boldsymbol{X} \succeq 0$ in Lemma 2.3.2 is necessary. For any value $\eta \neq 0$, the matrix $\boldsymbol{X}$ is invertible and thus neither $\boldsymbol{e}$ or $\boldsymbol{Ae}$ are null vectors. When $\eta = 0$ and $\boldsymbol{X} \succeq 0$, Lemma 2.3.2 implies $\boldsymbol{X}$ has two null vectors $\{\boldsymbol{e}, \boldsymbol{Ae}\}$, and is consistent with the fact that $\boldsymbol{X}$ must be the zero matrix.

2.4 Algebraic Conditions for A-Stable Continued Fraction Approximations of the Exponential

Early work in the Runge-Kutta literature, as discussed in Hairer and Wanner's Chapter IV.3–5 [30], focused on characterizing which rational functions simultaneously approximate the exponential function and are A-stable. This includes the work of Birkhoff and Varga [6], Ehle [20], Ehle and Price [21], and Nørsett [39, 40].

In 1977 Butcher, parameterized all A-stable rational approximations of order $p \geq 2s-2$ [11]. Building on the work of Butcher, a general characterization of rational functions which are A-stable and pth order approximations of the exponential was provided by Hairer [26], and Hairer and Türke [28]. They showed that the stability function $W(z)$ is A-stable and a pth order approximation of the exponential if and only if it has a continued fraction form that mirrors the continued fraction form of the exponential up to a remainder function $\Psi_{\nu-1}(z)$.

We begin this section by briefly outlining the relevant analytic theory of continued fractions and its relation to positive functions and A-stability. We then combine several results from Hairer and Wanner [29, 30], introducing a general change of basis that allows us to express the remainder function $\Psi_{\nu-1}(z)$ as a matrix-vector expression. This framework enables us to assess the stability of RK schemes through another set of algebraic conditions.

2.4.1 Continued Fraction Approximations for the Exponential

In a slight abuse of notation, we use $W(z)$ in this subsection to denote an arbitrary rational function whose numerator and denominator have degrees $\leq s$, which is not necessarily defined in terms of a matrix-vector pair $(\boldsymbol{A}, \boldsymbol{b})$ as in (1.3).

The exponential function admits the following continued fraction expansion:

$$\mathrm{e}^z = 1 + \cfrac{z}{1 - \frac{1}{2}z + \cfrac{\xi_1^2 z^2}{1 + \cfrac{\xi_2^2 z^2}{1 + \ddots}}}\,, \quad \text{where} \quad \xi_n^2 := \frac{1}{4(4n^2-1)} \qquad (n \geq 1)\,. \tag{2.40}$$

For $W(z)$ to be a pth order approximation of the exponential, it must match the first convergents, or levels, of the continued fraction in (2.40). Specifically, as shown in [28, Lemma 3],

$$W(z) = \frac{1 + \frac{1}{2}\Psi_0(z)}{1 - \frac{1}{2}\Psi_0(z)}\,, \qquad \text{if } p = 1, 2\,, \tag{2.41}$$

and

$$W(z) = 1 + \cfrac{z}{1 - \frac{1}{2}z + \cfrac{\xi_1^2 z^2}{1 + \cfrac{\xi_2^2 z^2}{\ddots + \cfrac{\xi_{\nu-2}^2 z^2}{1 + \xi_{\nu-1}^2 z \Psi_{\nu-1}(z)}}}}\,, \qquad \text{for } p \geq 3\,, \tag{2.42}$$

where

$$\nu := \left\lfloor \frac{p+1}{2} \right\rfloor . \tag{2.43}$$

The functions $\Psi_{\nu-1}(z)$ can be obtained from $W(z)$ via repeated long division of polynomials. Since $W(z)$ is rational, $\Psi_{\nu-1}(z)$ is also rational with numerator and denominator polynomials of degree at most $s - \nu + 1$. Note that $\Psi_{\nu-1}(z)$ satisfies:

$$\Psi_{\nu-1}(z) = z + Cz^2 + \mathcal{O}(z^3) \qquad \text{as } z \to 0\,, \tag{2.44}$$

where $C \neq 0$ if p is odd and $C = 0$ if p is even.

The function $W(z)$ is a pth order approximation of the exponential if and only if it has the form (2.41)–(2.44). Therefore, whether $W(z)$ is A-stable and approximates the exponential to order p depends solely on the function $\Psi_{\nu-1}(z)$. Remarkably, characterizing A-stability is "compatible" with continued fractions of the form (2.42); historically, in the context of RK schemes, this characterization has been done in terms of a positive function.

Theorem 2.4.1 (Theorem 5.22 Chapter IV.5 [30])**.** *The function $W(z)$ given by* (2.41)–(2.42) *is A-stable if and only if $-\Psi_{\nu-1}(-1/z)$ is a positive function.*

Utilizing Theorem 2.4.1, known results on the theory of positive functions can then be used to prove A-stability.

We conclude with some basic intuition on Theorem 2.4.1. Starting with $W(z)$, it can always be written in the form (2.41) as the composition of a Möbius transformation, $z \to (1+\frac{1}{2}z)/(1-\frac{1}{2}z)$, which maps $\mathbb{C}^-$ into the open unit disk with $\Psi_0(z)$. Thus, $\Psi_0(z)$ is uniquely determined by $W(z)$, and a scheme is A-stable if and only if $-\Psi_0(-1/z)$ is positive. Equation (2.42) can be obtained by relating $\Psi_0(z)$ to $\Psi_{\nu-1}(z)$ through a sequence of functions:

$$\frac{1}{\Psi_k(z)} = 1 + \xi_{k+1}^2 z^2 \Psi_{k+1}(z)\,, \qquad k \geq 0\,. \tag{2.45}$$

Properties of positive functions can then be used to show that $-\Psi_k(-1/z)$ is positive if and only if $-\Psi_{k+1}(-1/z)$ is positive, from which Theorem 2.4.1 follows.

2.4.2 Block triangular matrices for rational functions that approximate the exponential.

We discuss here the known and partial results for which a matrix-vector pair $(\boldsymbol{A}, \boldsymbol{b})$ define an A-stable $W(z)$ that approximates e^z, i.e., satisfies (2.42) and Theorem 2.4.1.

For $p \geq 1$, let

$$\boldsymbol{Y} := \left[\begin{array}{cccccc} 0 & -\xi_1 & & & & \\ \xi_1 & 0 & -\xi_2 & & & \\ & \xi_2 & \ddots & \ddots & & \\ & & \ddots & 0 & -\xi_{\nu-1} & \\ \cline{5-6} & & & \xi_{\nu-1} & \multicolumn{2}{|c}{\boldsymbol{Y}_{\nu-1}} \end{array}\right] \in \mathbb{R}^{s\times s}\,, \tag{2.46}$$

where $\boldsymbol{Y}_{\nu-1}$ is a square matrix of size $s-\nu+1$ and $(\boldsymbol{Y}_{\nu-1})_{11} = 0$ if p is even. We collect several results from Hairer and Wanner and summarize them into one Lemma.

Lemma 2.4.2 (Hairer and Wanner [29]; cf. Chapter IV.5 [30])**.** *If there exists a matrix $\boldsymbol{U}$ that satisfies:*

$$\boldsymbol{U}^{-1}\boldsymbol{e} = \boldsymbol{e}_1\,, \qquad \boldsymbol{b}^T\boldsymbol{U} = \boldsymbol{e}_1^T\,, \qquad \boldsymbol{e}_1 = [1,\ 0,\ \cdots,0]^T\,, \tag{2.47}$$

and

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \boldsymbol{Y} + \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T\,, \tag{2.48}$$

where $\boldsymbol{Y}$ *has the form* (2.46)*, then the function* $W(z)$ *in* (1.3) *is a pth order approximation to* e^z *with the continued fraction form* (2.41)*–*(2.42) *where*

$$\Psi_0(z) = z\boldsymbol{e}_1^T\left(\boldsymbol{I} - z\boldsymbol{Y}\right)\boldsymbol{e}_1 \qquad \text{and} \qquad \Psi_{\nu-1}(z) = z\boldsymbol{e}_1^T(\boldsymbol{I} - z\boldsymbol{Y}_{\nu-1})\boldsymbol{e}_1\,. \tag{2.49}$$

Lemma 2.4.2 introduces a general change of basis that allows us to express the remainder function $\Psi_{\nu-1}(z)$ as a matrix-vector expression. Applying the generalized KYP Lemma to $\Psi_{\nu-1}(z)$ gives the following corollary.

Corollary 2.4.3. *Given the function* $\Psi_{\nu-1}(z)$ *defined in* (2.49)*,* $-\Psi_{\nu-1}(-1/z)$ *is a positive function if there exists an* $\boldsymbol{R}^T = \boldsymbol{R}$ *satisfying*

$$\boldsymbol{R}\boldsymbol{e}_1 = \boldsymbol{e}_1\,, \qquad \boldsymbol{M}^T\boldsymbol{R}\boldsymbol{M} \succeq 0\,, \qquad \boldsymbol{M}^T\left(\boldsymbol{R}\boldsymbol{Y}_{\nu-1} + \boldsymbol{Y}_{\nu-1}^T\boldsymbol{R} - \boldsymbol{e}_1\boldsymbol{e}_1^T\right)\boldsymbol{M} \succeq 0\,,$$

where $\boldsymbol{M}$ *is any matrix whose columns span* $[e_1, \boldsymbol{Y}_{\nu-1}\boldsymbol{e}_1, \ldots, \boldsymbol{Y}_{\nu-1}^{s-\nu+1}\boldsymbol{e}_1]$.

The framework presented in Corollary 2.4.3 enables us to assess the stability of RK schemes through another set of algebraic conditions.

2.5 Algebraic Conditions for A-Stability Derived from Orthogonal Polynomials

For RK schemes satisfying a set of *simplifying conditions* on $(\boldsymbol{A}, \boldsymbol{b})$, known in the literature as conditions $C(k)$, $D(l)$, Hairer and Wanner [29] constructed a change of basis matrix $\boldsymbol{W}$, referred to as the W-transform, that exactly satisfies the conditions in Lemma 2.4.2. The W-transform enabled a characterization of all B-stable (i.e., algebraically stable) RK schemes [30, Theorem 13.15], and a notion of equivalence for A- and B-stable schemes [28]. However, the condition $C(k)$, upon which the

W-transform relies, is somewhat restrictive for fully implicit (non-DIRK) RK schemes. DIRK schemes are limited to $C(1)$, with EDIRK schemes satisfying $C(2)$.

In this section, we introduce a change of basis on the matrix $\boldsymbol{A}$, defined in terms of a class of orthogonal polynomials with respect to a linear functional. The formal examination of these polynomials was first introduced in [8]. We show that the new basis, referred to as the HW-transform, converts the upper left block of $\boldsymbol{A}$ into a tridiagonal form. This transformation yields a representation of $W(z)$ as a continued fraction approximation to the exponential function and identifies the minimal variables in $(\boldsymbol{A}, \boldsymbol{b})$ responsible for A-stability.

Our approach parallels and produces a result analogous to the W-transform (defined in terms of the shifted Legendre polynomials) of Hairer and Wanner, but under a weaker set of assumptions. Here, we only assume $W(z)$ is a pth order approximation to the exponential, while the W-transform assumes a set of simplifying conditions (e.g., conditions $C(\eta)$, $D(\xi)$) which can only be satisfied for fully implicit RK schemes. Furthermore, our approach generalizes the characterization of A-stability by Butcher [11], which is valid under $2s - p \leq p \leq 2s$. In contrast, we make no assumption on p relative to s.

2.5.1 A class of polynomials orthogonal with respect to a linear functional.

Here, we introduce the polynomials $\{q_n\}_{n\geq 0}$ used to define the change of basis on $\boldsymbol{A}$. These polynomials are orthogonal with respect to a linear functional $\mathcal{L}(\cdot)$. Specifically, let $(\mu_n)_{n\geq 0}$ be a sequence of real numbers, and $\mathcal{L}(\cdot)$ be a linear functional on polynomials defined by $\mathcal{L}(x^n) = \mu_n$.

Definition 2.5.1. A family of polynomials $\{q_n\}_{n\geq 0}$ are *orthogonal with respect to a linear functional* $\mathcal{L}$ if:

1. $\deg(q_n) = n$ for $n \geq 0$

2. $\mathcal{L}(q_n q_m) = K_n \delta_{nm}$ for $n \neq m$, where $K_n \neq 0$.

A reference to polynomials orthogonal with respect to a linear functional can be found in [14, 47].

Let $\mathcal{L}(\cdot)$ be the linear functional on monomials x^n with *moments* $(\mu_n)_{n\geq 0}$ defined as

$$\mathcal{L}\left(x^n\right) := \mu_n, \text{ where } \mu_n := \frac{1}{(n+1)!}, \text{ for } n \geq 0 \,. \tag{2.50}$$

The choice of μ_n is motivated by the order conditions (1.7) and the generating function $(e^z - 1)/z$. For example,

$$\begin{aligned}\mathcal{L}\left(1+2x+3x^2\right) &= \mathcal{L}\left(1\right) + 2\mathcal{L}\left(x\right) + 3\mathcal{L}\left(x^2\right) \\ &= 1 + 2\frac{1}{2!} + 3\frac{1}{3!} = \frac{4}{5}\,.\end{aligned}$$

The functional $\mathcal{L}(\cdot)$ also naturally extends to products of polynomials. For instance,

$$\begin{aligned}\mathcal{L}\Big((1+x)(1-3x)\Big) &= \mathcal{L}\left(1\right) - 2\mathcal{L}\left(x\right) - 3\mathcal{L}\left(x^2\right) \\ &= -\frac{1}{2}\,.\end{aligned}$$

It is important to note that while $\mathcal{L}(p(x)q(x))$ does define a bilinear form on p and q, it is not positive definite and does not define an inner product. For example, $\mathcal{L}\left((1-2x)^2\right) = -\frac{1}{2}$.

Fixing $q_0(x) = 1$, we define $q_n(x)$ $(n \geq 0)$ to be the unique family of polynomials

$$q_n(x) = \alpha_{nn}x^n + \ldots + \alpha_{n0}\,, \qquad \text{with} \qquad \alpha_{nn} > 0\,. \tag{2.51}$$

The construction of $\{q_n\}_{n\geq 0}$ parallels the application of Gram-Schmidt on basis vectors $\{x^n\}_{n\geq 0}$ using $\mathcal{L}(q(x)p(x))$ in lieu of an inner product. Demanding that q_n be "orthogonal", with respect to $\mathcal{L}$, to every polynomial of degree less than n requires $\mathcal{L}(q_n x^j) = 0$ for $j \leq n-1$. This conditions yields a linear system on α_{nj} involving

Hankel matrices:

$$\boldsymbol{H}_n := (\mu_{i+j-2})_{i,j=1}^{n} = \begin{bmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \iddots & & \mu_{n+1} \\ \mu_2 & \ddots & \ddots & & \vdots \\ \vdots & & & \mu_{2n-2} & \mu_{2n-1} \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} & \mu_{2n} \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \qquad n \geq 0\,.$$

A general formula for monic polynomials $Q_n(x)$ that are orthogonal with respect to $\mathcal{L}$, can be obtained by setting $Q_0(x) = \mu_0$ and replacing the last column of $\boldsymbol{H}_n$ with monomials to be used in the following formula:

$$Q_n(x) := \frac{1}{\det \boldsymbol{H}_{n-1}} \det \begin{bmatrix} \mu_0 & \mu_1 & \mu_2 & \cdots & 1 \\ \mu_1 & \mu_2 & \iddots & & x \\ \mu_2 & \ddots & \ddots & & \vdots \\ \vdots & & & \mu_{2n-2} & x^{n-1} \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} & x^n \end{bmatrix} \qquad n \geq 1\,. \tag{2.52}$$

The $Q_n(x)$ are then orthogonal with respect to $\mathcal{L}$. The identity $\mathcal{L}(x^k Q_n) = 0$ $(0 \leq k \leq n-1)$ follows because $\mathcal{L}(x^k Q_n(x))$ is the ratio of two determinants as in (2.52), where the last column in the numerator determinant (i.e., the monomials) is replaced with the $(k+1)$th column. The determinant vanishes since it has two repeated columns, showing that $Q_n(x)$ is orthogonal to every polynomial of degree less than n, including $Q_j(x)$ for $j < n$.

Appendix B in [8] established several key quantities related to $Q_n(x)$. For $\mu_n = 1/(n+1)!$, the Hankel determinants are given by the explicit formula:

$$\det \boldsymbol{H}_n = \sigma(n)\frac{c(n)^2}{c(2n)}, \quad \text{where } c(n) := \Pi_{i=1}^{n-1} i!\,, \tag{2.53}$$

$$\text{and } \sigma(n) := \begin{cases} 1, & \text{if } n \equiv 0,1 \bmod 4, \\ -1, & \text{if } n \equiv 2,3 \bmod 4, \end{cases}.$$

A consequence of (2.53) is that $\det \boldsymbol{H}_n \neq 0$ and $Q_n(x)$ is well-defined.

In addition, the polynomials satisfy the normalization condition [8, Appendix B]

$$\mathcal{L}\big(Q_i Q_j\big) = \zeta_i \delta_{ij} \qquad \text{where} \qquad \zeta_n = \frac{\det \boldsymbol{H}_{n+1}}{\det \boldsymbol{H}_n}\,.$$

Using (2.53), the values of ζ_n can be computed as

$$|\zeta_n|^{1/2} = \frac{n!}{(2n)!}\frac{1}{\sqrt{2n+1}} \qquad \text{and} \qquad \operatorname{sign}(\zeta_n) = \frac{\sigma(n+1)}{\sigma(n)} = (-1)^n \qquad \text{for } n \geq 0\,. \tag{2.54}$$

The polynomials $q_n(x)$ are then the unique normalized polynomials with positive leading coefficients

$$q_n(x) := \frac{1}{|\zeta_n|^{1/2}} Q_n(x) = \frac{(2n)!}{n!}\sqrt{2n+1}Q_n(x)\,. \tag{2.55}$$

The three-term recurrence for $Q_n(x)$ was established in [8, Appendix B] as

$$Q_{n+1}(x) = x\,Q_n(x) + \xi_n^2 Q_{n-1}(x) \qquad n \geq 1\,. \tag{2.56}$$

Using the fact that

$$\frac{\zeta_n}{\zeta_{n-1}} = -\xi_n^2\,,$$

in (2.56), along with the definition of $q_n(x)$ in (2.55), yields

$$\xi_{n+1} q_{n+1}(x) = x q_n(x) + \xi_n q_{n-1}(x) \qquad \text{for } n \geq 1\,. \tag{2.57}$$

Lastly, the identity

$$x = \xi_1 q_1(x) + \frac{1}{2} q_0(x)\,, \tag{2.58}$$

follows from $x = Q_1(x) + \frac{1}{2}Q_0(x)$ combined with the normalization factor $\xi_1 = |\zeta_1|^{1/2}$ (since $\zeta_0 = \mathcal{L}(Q_0^2) = 1$).

In addition to $q_0(x) = 1$, the first few polynomials are as follows:

$$q_1(x) = 2\sqrt{3}\left(x - \tfrac{1}{2}\right), \qquad q_3(x) = 120\sqrt{7}\left(x^3 - \tfrac{1}{2}x^2 + \tfrac{1}{10}x - \tfrac{1}{120}\right),$$
$$q_2(x) = 12\sqrt{5}\left(x^2 - \tfrac{1}{2}x + \tfrac{1}{12}\right), \qquad q_4(x) = 5040\left(x^4 - \tfrac{1}{2}x^3 + \tfrac{3}{28}x^2 - \tfrac{1}{84}x + \tfrac{1}{1680}\right).$$

When $W(z)$ approximates e^z to order p, the order conditions (1.7) combined with (2.50) implies that

$$\boldsymbol{b}^T \boldsymbol{A}^j \boldsymbol{e} = \mathcal{L}(x^j) \qquad \text{for } 0 \le j \le p-1\,. \tag{2.59}$$

The identity (2.59) implies

$$\boldsymbol{b}^T q(\boldsymbol{A})\boldsymbol{e} = \mathcal{L}\big(q(x)\big) \qquad \text{for any } \deg q \le p-1\,,$$

and hence

$$\boldsymbol{b}^T q_i(\boldsymbol{A}) q_j(\boldsymbol{A})\boldsymbol{e} = \mathcal{L}(q_i q_j) = (-1)^i \delta_{ij} \qquad \text{if } i+j \le p-1\,. \tag{2.60}$$

A less obvious fact is that (2.60) fails for every pair $i+j=p$, i.e.,

$$\boldsymbol{b}^T q_i(\boldsymbol{A}) q_j(\boldsymbol{A})\boldsymbol{e} \ne (-1)^i \delta_{ij} \qquad \text{whenever } i+j = p\,. \tag{2.61}$$

In particular, the value of

$$\alpha := \boldsymbol{b}^T q_i(\boldsymbol{A}) q_{p-i}(\boldsymbol{A})\boldsymbol{e} \qquad \text{for any choice } 0 \le i \le p\,,$$

uniquely defines $\boldsymbol{b}^T \boldsymbol{A}^p \boldsymbol{e}$ since $\boldsymbol{b}^T \boldsymbol{A}^p \boldsymbol{e}$ can be written in terms of α, the coefficients of q_i, and values of $\boldsymbol{b}^T \boldsymbol{A}^j \boldsymbol{e}$ for $j \le p-1$, which are exactly the pth order conditions. The value α agrees with (2.60) if and only if the $(p+1)$th order condition holds $\boldsymbol{b}^T \boldsymbol{A}^p \boldsymbol{e} = 1/(p+1)!$. Thus, (2.61) follows from the definition of p as the largest integer in (1.7).

Remark. Several classical orthogonal polynomials have been used to study $W(z)$. These include Laguerre polynomials [39] and shifted Lengendre polynomials [30].

Unlike classical orthogonal polynomials, the q_n used here are not orthogonal with respect to an inner product. Instead, they are orthogonal with respect to a linear functional $\mathcal{L}(\cdot)$. As a result, the roots of q_n need not be real.

2.5.2 The HW-transform.

In this section, we use the polynomials $\{q_n\}_{n\geq 0}$ to define a transformation matrix $\boldsymbol{U}$ that satisfies (2.47)–(2.48). This will yield new formulas for the function $\Psi_{\nu-1}$ in terms of $\boldsymbol{A}$, which can then be used to test for A-stability.

The orthogonality relation (2.60) can first be re-cast into matrix form. For $j \geq 0$, let

$$\boldsymbol{h}_j := q_j(\boldsymbol{A})\, \boldsymbol{e}\,, \tag{2.62}$$

and

$$\boldsymbol{w}_j := (-1)^j\, q_j(\boldsymbol{A}^T)\, \boldsymbol{b}\,. \tag{2.63}$$

Define the matrices

$$\boldsymbol{H} := \begin{bmatrix} \boldsymbol{h}_0 & \boldsymbol{h}_1 & \dots & \boldsymbol{h}_{\nu-1} \end{bmatrix} \in \mathbb{R}^{s\times\nu}\,, \tag{2.64}$$

and

$$\boldsymbol{W} := \begin{bmatrix} \boldsymbol{w}_0 & \boldsymbol{w}_1 & \dots & \boldsymbol{w}_{\nu-1} \end{bmatrix} \in \mathbb{R}^{s\times\nu}\,. \tag{2.65}$$

Since the orthogonality relation (2.60) holds for $i, j \leq \nu - 1$, with ν defined in (2.43), we have

$$\boldsymbol{W}^T\boldsymbol{H} = \boldsymbol{I}\,. \tag{2.66}$$

Equation (2.66) also shows that the columns of both $\boldsymbol{H}$ and $\boldsymbol{W}$ are linearly independent. The linear independence follows because $\boldsymbol{H}\boldsymbol{x} = 0$, or analogously

$\boldsymbol{W}\boldsymbol{x} = 0$, admits only the trivial solution; for instance, $\boldsymbol{H}\boldsymbol{x} = 0$ implies $\boldsymbol{W}^T\boldsymbol{H}\boldsymbol{x} = \boldsymbol{x} = 0$.

Now introduce the $s \times s$ matrix

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{H} & \widetilde{\boldsymbol{H}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_0 & \boldsymbol{h}_1 & \dots & \boldsymbol{h}_{\nu-1} & \widetilde{\boldsymbol{H}} \end{bmatrix}, \tag{2.67}$$

which is obtained by augmenting $\boldsymbol{H}$ with any matrix $\widetilde{\boldsymbol{H}} \in \mathbb{R}^{s\times(s-\nu)}$ whose columns span the space orthogonal to $\mathrm{col}(\boldsymbol{W})$. The columns of $\widetilde{\boldsymbol{H}}$ are linearly independent, so that $\mathrm{rank}(\widetilde{\boldsymbol{H}}) = s-\nu$, and $\widetilde{\boldsymbol{H}}^T\boldsymbol{W} = 0$. We then have the following result for $\boldsymbol{U}^{-1}$.

Lemma 2.5.1 (Inverse of $\boldsymbol{U}$)**.** *Let $(\boldsymbol{A}, \boldsymbol{b})$ be an RK scheme with stability function $W(z)$ that is a pth ($p \geq 1$) order approximation to e^z. Let $\nu = \left\lfloor \frac{p+1}{2} \right\rfloor$ as defined in* (2.43)*. Then any matrix $\boldsymbol{U}$ of the form* (2.67) *is invertible and with inverse of the form*

$$\boldsymbol{U}^{-1} = \begin{bmatrix} \boldsymbol{W} & \widetilde{\boldsymbol{W}} \end{bmatrix}^T, \tag{2.68}$$

where $\widetilde{\boldsymbol{W}}$ satisfies $\widetilde{\boldsymbol{W}}^T\boldsymbol{H} = 0$ and $\widetilde{\boldsymbol{H}}^T\widetilde{\boldsymbol{W}} = \boldsymbol{I}$, with $\boldsymbol{H}$ and $\boldsymbol{W}$ defined as in (2.64)–(2.65).

Proof. We first show $\boldsymbol{U}\boldsymbol{x} = 0$ admits only the zero solution, thus ensuring that $\boldsymbol{U}^{-1}$ exists. In block form, $\boldsymbol{U}\boldsymbol{x} = 0$ is

$$\begin{bmatrix} \boldsymbol{H} & \widetilde{\boldsymbol{H}} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_u \\ \boldsymbol{x}_\ell \end{bmatrix} = 0 \qquad \text{where} \quad \boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_u \\ \boldsymbol{x}_\ell \end{bmatrix}. \tag{2.69}$$

Multiplying (2.69) through by $\boldsymbol{W}^T$, and using the fact that $\boldsymbol{W}^T\boldsymbol{H} = \boldsymbol{I}$ and $\boldsymbol{W}^T\widetilde{\boldsymbol{H}} = 0$, shows that $\boldsymbol{x}_u = 0$. Equation (2.69) then becomes $\widetilde{\boldsymbol{H}}\boldsymbol{x}_\ell = 0$. Since the columns of $\widetilde{\boldsymbol{H}}$ are linearly independent, we have $\boldsymbol{x}_\ell = 0$. Hence, $\boldsymbol{U}$ is invertible. The block form of $\boldsymbol{U}^{-1}$ in (2.69) then follows since $\boldsymbol{W}$ is the unique matrix which satisfies $\boldsymbol{W}^T\boldsymbol{H} = \boldsymbol{I}$ and $\boldsymbol{W}^T\widetilde{\boldsymbol{H}} = 0$. □

As outlined in the previous Lemma, the orthogonality of q_n demands that the top block of $\boldsymbol{U}^{-1}$ be prescribed in terms of q_j. The 3-term recurrence relation (2.57) provides another useful identity on $\boldsymbol{U}$ involving the columns of $\boldsymbol{H}$ and $\boldsymbol{W}$. Substituting $\boldsymbol{x} = \boldsymbol{A}$ into (2.57) – (2.58) and multiplying by $\boldsymbol{e}$ or $\boldsymbol{b}^T$ yields the identities:

$$\boldsymbol{A}\,\boldsymbol{h}_0 = \xi_1\,\boldsymbol{h}_1 + \frac{1}{2}\,\boldsymbol{h}_0 \qquad \text{and} \qquad \boldsymbol{A}^T\,\boldsymbol{w}_0 = -\xi_1\,\boldsymbol{w}_1 + \frac{1}{2}\,\boldsymbol{w}_0\,. \tag{2.70}$$

For all $j \geq 1$ (i.e., j is not limited to $\nu - 1$):

$$\boldsymbol{A}\,\boldsymbol{h}_j = \xi_{j+1}\,\boldsymbol{h}_{j+1} - \xi_j\,\boldsymbol{h}_{j-1}\,, \tag{2.71}$$

and

$$\boldsymbol{A}^T\,\boldsymbol{w}_j = \xi_j\,\boldsymbol{w}_{j-1} - \xi_{j+1}\,\boldsymbol{w}_{j+1}\,. \tag{2.72}$$

With these identities, $\boldsymbol{U}$ transforms the upper block of $\boldsymbol{A}$ into a tridiagonal matrix characterized as follows.

Lemma 2.5.2 (Formula for $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$)**.** *Let $W(z)$ be a pth order approximation to e^z defined by* (1.3) *with ν, $\boldsymbol{H}$, $\boldsymbol{W}$ and $\boldsymbol{U}$ defined in* (2.43) *and* (2.64)–(2.67)*. Then $\boldsymbol{U}$ satisfies the conditions of Lemma 2.4.2, that is*

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} = \left[\begin{array}{cccccc} \frac{1}{2} & -\xi_1 & & & & \\ \xi_1 & 0 & -\xi_2 & & & \\ & \xi_2 & \ddots & \ddots & & \\ & & \ddots & 0 & -\xi_{\nu-1} & \\ \hline & & & \xi_{\nu-1} & \multicolumn{2}{|c}{\boldsymbol{Y}_{\nu-1}} \end{array}\right], \qquad \text{if } p \geq 3\,(\nu \geq 2)\,, \tag{2.73}$$

and $\boldsymbol{U}^{-1}\boldsymbol{e} = \boldsymbol{e}_1$, *and* $\boldsymbol{b}^T\boldsymbol{U} = \boldsymbol{e}_1^T$. *Here,*

$$\boldsymbol{Y}_{\nu-1} = \begin{bmatrix} \boldsymbol{w}_{\nu-1}^T \\ \widetilde{\boldsymbol{W}}^T \end{bmatrix} \boldsymbol{A} \begin{bmatrix} \boldsymbol{h}_{\nu-1} & \widetilde{\boldsymbol{H}} \end{bmatrix} \tag{2.74}$$

is a square matrix of size $s - \nu + 1$.

Proof. Let $\boldsymbol{X} := \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$. Then the bottom $(s-\nu+1)\times(s-\nu+1)$ block of $\boldsymbol{X}$ is exactly $\boldsymbol{Y}_{\nu-1}$.

To show equation (2.73), expand $\boldsymbol{A}\boldsymbol{U} = \boldsymbol{U}\boldsymbol{X}$ in block form, e.g.,

$$\boldsymbol{A} \begin{bmatrix} \boldsymbol{h}_0 & \boldsymbol{h}_1 & \dots & \boldsymbol{h}_{\nu-1} & \widetilde{\boldsymbol{H}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}_0 & \boldsymbol{h}_1 & \dots & \boldsymbol{h}_{\nu-1} & \widetilde{\boldsymbol{H}} \end{bmatrix} \boldsymbol{X}\,.$$

The vector identities (2.70) and (2.71) define the first $(\nu - 1)$ columns of $\boldsymbol{X}$, which are exactly those given by the matrix on the right-hand side of (2.73).

Similarly, expanding out $\boldsymbol{U}^{-1}\boldsymbol{A} = \boldsymbol{X}\boldsymbol{U}^{-1}$ in block form yields:

$$\begin{bmatrix} \boldsymbol{w}_0 & \boldsymbol{w}_1 & \dots & \boldsymbol{w}_{\nu-1} & \widetilde{\boldsymbol{W}} \end{bmatrix}^T \boldsymbol{A} = \boldsymbol{X} \begin{bmatrix} \boldsymbol{w}_0 & \boldsymbol{w}_1 & \dots & \boldsymbol{w}_{\nu-1} & \widetilde{\boldsymbol{W}} \end{bmatrix}^T\,.$$

Again, the identities (2.70) and (2.72) prescribe the first $(\nu - 1)$ rows of $\boldsymbol{X}$. When written in matrix form, these rows are exactly the first $(\nu - 1)$ rows of the matrix on the right-hand side of (2.73). □

The top left value of $\boldsymbol{Y}_{\nu-1}$ will play a role in how $W(z)$ approximates e^z to order p. To unify the cases when $\nu = 1$ and $\nu > 1$, we introduce $\boldsymbol{Y}$ by subtracting the 1/2 appearing in the top left entry of $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$:

$$\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} - \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T\,. \tag{2.75}$$

Note that $\boldsymbol{Y}_j$ is the minor of $\boldsymbol{Y}$ with the first j $(0 \le j \le \nu - 1)$ columns and rows removed. With this notation, $\boldsymbol{Y}_0 = \boldsymbol{Y}$ and $\boldsymbol{Y}_{\nu-1}$ is defined in (2.74).

Lemma 2.5.3. *Let $W(z)$ be a pth ($p \geq 1$) order approximation of e^z of the form* (1.3) *with $\boldsymbol{Y}$ defined by* (2.75)*. Then the top left entry of the minor $\boldsymbol{Y}_{\nu-1}$ of $\boldsymbol{Y}$ is*

$$(\boldsymbol{Y}_{\nu-1})_{11} \neq 0 \qquad \text{if } p \text{ is odd},$$

$$(\boldsymbol{Y}_{\nu-1})_{11} = 0 \qquad \text{if } p \text{ is even}.$$

Proof. The top left entry of $Y_{\nu-1}$ is:

$$(\boldsymbol{Y}_{\nu-1})_{11} = \boldsymbol{w}_{\nu-1}^T \boldsymbol{A} \boldsymbol{h}_{\nu-1} - \frac{1}{2}\delta_{\nu 1},$$

where $\delta_{\nu 1}$ is the Kronecker delta. Using the identity (2.70) when $\nu = 1$ and the 3-term recursion (2.71) when $\nu \geq 2$, we get:

$$(\boldsymbol{Y}_{\nu-1})_{11} = \begin{cases} \xi_1\big(\boldsymbol{w}_0^T\boldsymbol{h}_1\big) + \frac{1}{2}\big(\boldsymbol{w}_0^T\boldsymbol{h}_0\big) - \frac{1}{2} & \text{if } \nu = 1, \\ \xi_\nu\left(\boldsymbol{w}_{\nu-1}^T\boldsymbol{h}_\nu\right) - \xi_{\nu-1}\left(\boldsymbol{w}_{\nu-1}^T\boldsymbol{h}_{\nu-2}\right) & \text{if } \nu \geq 2, \end{cases} . \tag{2.76}$$

Since $2\nu - 3 \leq p - 1$, equation (2.60) implies the second large round brackets in (2.76) vanish. The first round brackets in (2.76) vanish when p is even, since $2\nu - 1 = p - 1$ and (2.60) applies, and does not vanish when p is odd, since $2\nu - 1 = p$ and (2.61) applies. □

We can summarize the results in the following theorem.

Theorem 2.5.4. *Let the RK scheme $(\boldsymbol{A}, \boldsymbol{b})$ be given with stability function $W(z)$ a pth order approximation of the exponential. Let $\boldsymbol{U}$ be the HW-transform. Then the RK scheme is A-stable if and only if there exists an $\boldsymbol{R} \succeq 0$ such that*

$$\boldsymbol{R}\boldsymbol{e}_1 = \boldsymbol{e}_1 \quad \textit{and} \quad \boldsymbol{R}\boldsymbol{Y}_{\nu-1} + \boldsymbol{Y}_{\nu-1}^T\boldsymbol{R} \succeq 0 \tag{2.77}$$

where $\boldsymbol{Y}_{\nu-1}$ is the bottom block of (2.73)*. Furthermore, when p is even, the condition* (2.77) *can be reduced to*

$$\boldsymbol{R}\boldsymbol{e}_1 = \boldsymbol{e}_1, \quad \left(\boldsymbol{R}\boldsymbol{Y}_{\nu-1} + \boldsymbol{Y}_{\nu-1}^T\boldsymbol{R}\right)\boldsymbol{e}_1 = 0, \quad \textit{and} \quad \boldsymbol{R}\boldsymbol{Y}_{\nu-1} + \boldsymbol{Y}_{\nu-1}^T\boldsymbol{R} \succeq 0.$$

Example 2.5.1 (SDIRK(3,2)). This example constructs an HW-transform matrix $\boldsymbol{U}$ and uses Theorem 2.5.4 to certify A-stability for the SDIRK 3-stage $p = 2$ scheme:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}.$$

First, we construct the $\boldsymbol{H}$ and $\boldsymbol{W}$ matrices as in (2.64) and (2.65), respectively, using the polynomial family $\{q_n\}_{n\geq 0}$ defined in Section 2.5.2:

$$\boldsymbol{H} = q_0(\boldsymbol{A})\boldsymbol{e} = \boldsymbol{e} \qquad \text{and} \qquad \boldsymbol{W} = q_0(\boldsymbol{A}^T)\boldsymbol{b} = \boldsymbol{b}.$$

Next, we construct $\widetilde{\boldsymbol{H}}$ such that $\boldsymbol{W}^T\widetilde{\boldsymbol{H}} = 0$:

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We then combine $\boldsymbol{H}$ and $\widetilde{\boldsymbol{H}}$ to get the HW-transform matrix

$$\boldsymbol{U} = [\boldsymbol{H} \;\; \widetilde{\boldsymbol{H}}] = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \qquad \text{with} \qquad \boldsymbol{U}^{-1} = [\boldsymbol{W} \;\; \widetilde{\boldsymbol{W}}]^T = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -1 \\ -1 & 1 & 0 \end{bmatrix}.$$

Applying the HW-transform yields:

$$\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} - \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T = \begin{bmatrix} 0 & -\frac{1}{2} & -\frac{1}{2} \\ 1 & 2 & 0 \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Since $p = 2$, we have that $\nu = 1$, and therefore $\boldsymbol{Y}_{\nu-1} = \boldsymbol{Y}_0 = \boldsymbol{Y}$. We then parameterize the equality constraints from Theorem 2.5.4 and get:

$$\boldsymbol{R}(\eta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{4}(\eta+1) & -\frac{1}{2}(\eta-1) \\ 0 & -\frac{1}{2}(\eta-1) & \eta \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{R}\boldsymbol{Y} + \boldsymbol{Y}^T\boldsymbol{R} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{1}{2}(\eta+3) & -\frac{1}{4}(3\eta-5) \\ 0 & -\frac{1}{4}(3\eta-5) & \eta \end{bmatrix}.$$

If we select $\eta = 1$, then Theorem 2.5.4 is satisfied, and the scheme is A-stable:

$$\boldsymbol{R}(1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \succ 0 \qquad \text{and}$$

$$\boldsymbol{R}(1)\boldsymbol{Y} + \boldsymbol{Y}^T\boldsymbol{R}(1) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{4} & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{7}{8} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{1}{4} \\ 0 & 0 & 1 \end{bmatrix} \succeq 0.$$

Example 2.5.2 (SDIRK(5,4))**.** This example constructs an HW-transform matrix $\boldsymbol{U}$ and then uses Theorem 2.5.4 to certify A-stability for the SDIRK 5-stage $p = 4$ scheme

$$\boldsymbol{A} = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & 0 & 0 \\ \frac{17}{50} & -\frac{1}{25} & \frac{1}{4} & 0 & 0 \\ \frac{371}{1360} & -\frac{137}{2720} & \frac{15}{544} & \frac{1}{4} & 0 \\ \frac{25}{24} & -\frac{49}{48} & \frac{125}{16} & -\frac{85}{12} & \frac{1}{4} \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} \frac{25}{24} \\ -\frac{49}{48} \\ \frac{125}{16} \\ -\frac{85}{12} \\ \frac{1}{4} \end{bmatrix}.$$

First, we construct matrices $\boldsymbol{H}$ and $\boldsymbol{W}$ using the polynomial family $\{q_n\}_{n\geq 0}$ defined in Section 2.5.2:

$$\boldsymbol{H} = [q_0(\boldsymbol{A})\boldsymbol{e} \;\; q_1(\boldsymbol{A})\boldsymbol{e}] = \begin{bmatrix} 1 & -\frac{\sqrt{3}}{2} \\ 1 & \frac{\sqrt{3}}{2} \\ 1 & \frac{\sqrt{3}}{10} \\ 1 & 0 \\ 1 & \sqrt{3} \end{bmatrix}, \quad \boldsymbol{W} = [q_0(\boldsymbol{A}^T)\boldsymbol{b} \;\; -q_1(\boldsymbol{A}^T)\boldsymbol{b}] = \begin{bmatrix} \frac{25}{24} & -\frac{41\sqrt{3}}{96} \\ -\frac{49}{48} & -\frac{17\sqrt{3}}{192} \\ \frac{125}{16} & \frac{25\sqrt{3}}{64} \\ -\frac{85}{12} & 0 \\ \frac{1}{4} & \frac{\sqrt{3}}{8} \end{bmatrix}.$$

We then construct $\widetilde{\boldsymbol{H}}$ such that $\boldsymbol{W}^T\widetilde{\boldsymbol{H}} = 0$:

$$\widetilde{\boldsymbol{H}} = \begin{bmatrix} -\frac{675}{1217} & \frac{1445}{1217} & \frac{243}{1217} \\ \frac{8625}{1217} & -\frac{6970}{1217} & \frac{546}{1217} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Combining $\boldsymbol{H}$ and $\widetilde{\boldsymbol{H}}$, we obtain the HW-transform matrix:

$$\boldsymbol{U} = [\boldsymbol{H} \ \ \widetilde{\boldsymbol{H}}] = \begin{bmatrix} 1 & -\frac{\sqrt{3}}{2} & -\frac{675}{1217} & \frac{1445}{1217} & \frac{243}{1217} \\ 1 & \frac{\sqrt{3}}{2} & \frac{8625}{1217} & -\frac{6970}{1217} & \frac{546}{1217} \\ 1 & \frac{\sqrt{3}}{10} & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & \sqrt{3} & 0 & 0 & 1 \end{bmatrix},$$

$$\boldsymbol{U}^{-1} = [\boldsymbol{W} \ \ \widetilde{\boldsymbol{W}}]^T = \begin{bmatrix} \frac{25}{24} & -\frac{49}{48} & \frac{125}{16} & -\frac{85}{12} & \frac{1}{4} \\ -\frac{41\sqrt{3}}{96} & -\frac{17\sqrt{3}}{192} & \frac{25\sqrt{3}}{64} & 0 & \frac{\sqrt{3}}{8} \\ -\frac{877}{960} & \frac{2011}{1920} & -\frac{887}{128} & \frac{85}{12} & -\frac{23}{80} \\ -\frac{25}{24} & \frac{49}{48} & -\frac{125}{16} & \frac{97}{12} & -\frac{1}{4} \\ \frac{23}{96} & \frac{247}{192} & -\frac{575}{64} & \frac{85}{12} & \frac{3}{8} \end{bmatrix}.$$

Next, we apply the HW-transform to get the matrix:

$$\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} - \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T = \begin{bmatrix} 0 & -\xi_1 & 0 & 0 & 0 \\ \xi_1 & 0 & -\frac{6225\sqrt{3}}{38944} & \frac{45475\sqrt{3}}{233664} & -\frac{1603\sqrt{3}}{77888} \\ 0 & \frac{\sqrt{3}}{600} & -\frac{13561}{77888} & \frac{447389}{778880} & \frac{218541}{3894400} \\ 0 & \frac{2\sqrt{3}}{255} & -\frac{585}{1217} & \frac{2099}{2434} & \frac{3297}{103445} \\ 0 & \frac{\sqrt{3}}{6} & \frac{18675}{38944} & -\frac{45475}{77888} & \frac{4809}{77888} \end{bmatrix}.$$

Since $p = 4$, we have $\nu = 2$, and therefore

$$\boldsymbol{Y}_{\nu-1} = \boldsymbol{Y}_1 = \begin{bmatrix} 0 & -\frac{6225\sqrt{3}}{38944} & \frac{45475\sqrt{3}}{233664} & -\frac{1603\sqrt{3}}{77888} \\ \frac{\sqrt{3}}{600} & -\frac{13561}{77888} & \frac{447389}{778880} & \frac{218541}{3894400} \\ \frac{2\sqrt{3}}{255} & -\frac{585}{1217} & \frac{2099}{2434} & \frac{3297}{103445} \\ \frac{\sqrt{3}}{6} & \frac{18675}{38944} & -\frac{45475}{77888} & \frac{4809}{77888} \end{bmatrix}.$$

We then parameterize the equality constraints from Theorem 2.5.4. By selecting $\boldsymbol{\eta} = \begin{bmatrix} \frac{46524}{125} & -\frac{2039}{131} & \frac{442}{609} \end{bmatrix}$, Theorem 2.5.4 is satisfied, indicating that the scheme is A-stable:

$$\boldsymbol{R}(\boldsymbol{\eta}) = \boldsymbol{L}_R \boldsymbol{D}_R \boldsymbol{L}_R^T \succ 0 \qquad \text{and} \qquad \boldsymbol{R}(\boldsymbol{\eta})\boldsymbol{Y}_1 + \boldsymbol{Y}_1^T \boldsymbol{R}(\boldsymbol{\eta}) = \boldsymbol{L}_Y \boldsymbol{D}_Y \boldsymbol{L}_Y^T \succeq 0.$$

See Appendix A.1.1 for the full LDL factorization.

Example 2.5.3 (ERK4, [27])**.** This example constructs and applies an HW-transform matrix $\boldsymbol{U}$ to the explicit RK 4-stage $p = 4$ scheme

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{6} \end{bmatrix}.$$

Following the same process as in previous examples, we have the HW-transform matrix

$$\boldsymbol{U} = \begin{bmatrix} 1 & -\sqrt{3} & 0 & 1 \\ 1 & 0 & -1 & -1 \\ 1 & 0 & 1 & 0 \\ 1 & \sqrt{3} & 0 & 1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{U}^{-1} = \begin{bmatrix} \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \\ -\frac{\sqrt{3}}{6} & 0 & 0 & \frac{\sqrt{3}}{6} \\ -\frac{1}{6} & -\frac{1}{3} & \frac{2}{3} & -\frac{1}{6} \\ \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Applying the transform yields

$$\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} - \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T = \begin{bmatrix} 0 & -\xi_1 & 0 & 0 \\ \xi_1 & 0 & \frac{\sqrt{3}}{6} & 0 \\ 0 & \frac{\sqrt{3}}{6} & -\frac{1}{2} & -\frac{1}{2} \\ 0 & \frac{\sqrt{3}}{6} & \frac{1}{2} & 0 \end{bmatrix}$$

$$\text{with} \qquad \boldsymbol{Y}_{\nu-1} = \boldsymbol{Y}_1 = \begin{bmatrix} 0 & \frac{\sqrt{3}}{6} & 0 \\ \frac{\sqrt{3}}{6} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{\sqrt{3}}{6} & \frac{1}{2} & 0 \end{bmatrix}.$$

Example 2.5.4 (Verner-8-5-6, [46])**.** This example constructs and applies an HW-transform matrix $\boldsymbol{U}$ to an explicit RK 8-stage $p = 6$ scheme

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{4}{75} & \frac{16}{75} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{5}{6} & -\frac{8}{3} & \frac{5}{2} & 0 & 0 & 0 & 0 & 0 \\ -\frac{165}{64} & \frac{55}{6} & -\frac{425}{64} & \frac{85}{96} & 0 & 0 & 0 & 0 \\ \frac{12}{5} & -8 & \frac{4015}{612} & -\frac{11}{36} & \frac{88}{255} & 0 & 0 & 0 \\ -\frac{8263}{15000} & \frac{124}{75} & -\frac{643}{680} & -\frac{81}{250} & \frac{2484}{10625} & 0 & 0 & 0 \\ \frac{3501}{1720} & -\frac{300}{43} & \frac{297275}{52632} & -\frac{319}{2322} & \frac{24068}{84065} & 0 & \frac{3850}{26703} & 0 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \frac{3}{40} \\ 0 \\ \frac{875}{2244} \\ \frac{23}{72} \\ \frac{264}{1955} \\ 0 \\ \frac{125}{11592} \\ \frac{43}{616} \end{bmatrix}.$$

Following the same process as in previous examples, we have the HW-transform matrix

$$
\boldsymbol{U} = \begin{bmatrix}
1 & -\sqrt{3} & \sqrt{5} & 0 & -\frac{3984}{4393} & 0 & \frac{29625}{61502} & -\frac{2881}{2674} \\
1 & -\frac{2\sqrt{3}}{3} & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & -\frac{7\sqrt{3}}{15} & -\frac{13\sqrt{5}}{75} & 0 & \frac{1672176}{3843875} & 0 & -\frac{84337}{430514} & \frac{1157173}{2339750} \\
1 & \frac{\sqrt{3}}{3} & -\frac{\sqrt{5}}{3} & 0 & -\frac{16272}{21965} & 0 & \frac{5675}{61502} & -\frac{7611}{13370} \\
1 & \frac{2\sqrt{3}}{3} & \frac{\sqrt{5}}{6} & 0 & 1 & 0 & 0 & 0 \\
1 & \sqrt{3} & \sqrt{5} & 0 & 0 & 1 & 0 & 0 \\
1 & -\frac{13\sqrt{3}}{15} & \frac{47\sqrt{5}}{75} & 0 & 0 & 0 & 1 & 0 \\
1 & \sqrt{3} & \sqrt{5} & 0 & 0 & 0 & 0 & 1
\end{bmatrix},
$$

$$
\boldsymbol{U}^{-1} = \begin{bmatrix}
\frac{3}{40} & 0 & \frac{875}{2244} & \frac{23}{72} & \frac{264}{1955} & 0 & \frac{125}{11592} & \frac{43}{616} \\
-\frac{3\sqrt{3}}{40} & 0 & -\frac{1225\sqrt{3}}{6732} & \frac{23\sqrt{3}}{216} & \frac{176\sqrt{3}}{1955} & 0 & -\frac{325\sqrt{3}}{34776} & \frac{43\sqrt{3}}{616} \\
\frac{11\sqrt{5}}{100} & 0 & -\frac{175\sqrt{5}}{4488} & -\frac{43\sqrt{5}}{360} & \frac{12\sqrt{5}}{425} & 0 & -\frac{25\sqrt{5}}{504} & \frac{43\sqrt{5}}{616} \\
-\frac{9}{40} & 1 & -\frac{5075}{6732} & -\frac{23}{216} & \frac{88}{1955} & 0 & -\frac{1025}{34776} & \frac{43}{616} \\
-\frac{1}{60} & 0 & \frac{175}{26928} & -\frac{187}{432} & \frac{1293}{1955} & 0 & \frac{3425}{69552} & -\frac{989}{3696} \\
-\frac{2}{5} & 0 & \frac{525}{1496} & -\frac{1}{24} & -\frac{1068}{1955} & 1 & \frac{1025}{3864} & -\frac{387}{616} \\
-\frac{461}{750} & 0 & -\frac{3325}{4488} & \frac{199}{600} & \frac{516}{48875} & 0 & \frac{1443}{1288} & -\frac{989}{9240} \\
-\frac{2}{5} & 0 & \frac{525}{1496} & -\frac{1}{24} & -\frac{1068}{1955} & 0 & \frac{1025}{3864} & \frac{229}{616}
\end{bmatrix}.
$$

Applying the transform yields

$$\boldsymbol{Y} = \boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U} - \frac{1}{2}\boldsymbol{e}_1\boldsymbol{e}_1^T = \begin{bmatrix} 0 & -\xi_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \xi_2 & 0 & -\xi_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \xi_2 & 0 & 0 & -\frac{23022\sqrt{5}}{768775} & 0 & \frac{33345\sqrt{5}}{1722056} & -\frac{65059\sqrt{5}}{1871800} \\ 0 & 0 & \frac{\sqrt{5}}{10} & 0 & -\frac{664}{4393} & 0 & \frac{9875}{123004} & -\frac{2881}{16044} \\ 0 & 0 & -\frac{149\sqrt{5}}{90} & \frac{55}{6} & -\frac{363387}{307510} & 0 & \frac{2574625}{20664672} & -\frac{4409779}{4492320} \\ 0 & 0 & \frac{137\sqrt{5}}{90} & -8 & \frac{3225602}{2306325} & 0 & -\frac{437635}{1722056} & \frac{16956577}{16846200} \\ 0 & 0 & -\frac{49\sqrt{5}}{150} & \frac{124}{75} & \frac{12605422}{19219375} & 0 & -\frac{1469581}{8610280} & \frac{19616299}{46795000} \\ 0 & 0 & \frac{1729\sqrt{5}}{1290} & -\frac{300}{43} & \frac{7596186}{6611465} & 0 & -\frac{6761975}{74048408} & \frac{319099}{374360} \end{bmatrix}$$

$$\text{with} \qquad \boldsymbol{Y}_{\nu-1} = \boldsymbol{Y}_2 = \begin{bmatrix} 0 & 0 & -\frac{23022\sqrt{5}}{768775} & 0 & \frac{33345\sqrt{5}}{1722056} & -\frac{65059\sqrt{5}}{1871800} \\ \frac{\sqrt{5}}{10} & 0 & -\frac{664}{4393} & 0 & \frac{9875}{123004} & -\frac{2881}{16044} \\ -\frac{149\sqrt{5}}{90} & \frac{55}{6} & -\frac{363387}{307510} & 0 & \frac{2574625}{20664672} & -\frac{4409779}{4492320} \\ \frac{137\sqrt{5}}{90} & -8 & \frac{3225602}{2306325} & 0 & -\frac{437635}{1722056} & \frac{16956577}{16846200} \\ -\frac{49\sqrt{5}}{150} & \frac{124}{75} & \frac{12605422}{19219375} & 0 & -\frac{1469581}{8610280} & \frac{19616299}{46795000} \\ \frac{1729\sqrt{5}}{1290} & -\frac{300}{43} & \frac{7596186}{6611465} & 0 & -\frac{6761975}{74048408} & \frac{319099}{374360} \end{bmatrix}.$$

2.6 Stability Conditions in Linear Multistep Methods

In this section, we revisit the problem of characterizing numerical stability for Linear Multistep Methods (LMMs). Similar to A- and $A(\alpha)$-stability for RK schemes, we can cast G-stability of LMMs as a Linear Matrix Inequality over the convex cone of semi-definite matrices. Dahlquist characterized how the order conditions reduce the dimension of the LMIs for stability. These observations allowed Dahlquist to characterize all two step G-stable methods.

We use the algebraic characterization of G-stable LMMs to parameterize the implicit part of a family of IMEX schemes introduced by Rosales, Seibold, Shirokoff and Zhou in [44].

2.6.1 Background on Linear Multistep Methods

Linear multistep methods (LMMs) are a common class of integrators which discretize (1.1) with k steps as:

$$\sum_{j=0}^{k} \alpha_j \, \boldsymbol{u}_{n+j} = \Delta t \sum_{j=0}^{k} \beta_j \, f(t_{n+j}, \boldsymbol{u}_{n+j}) \,. \tag{2.78}$$

and time stepping coefficients $\boldsymbol{\alpha} = [\alpha_k, \ldots, \alpha_0]^T$ and $\boldsymbol{\beta} = [\beta_k, \ldots, \beta_0]^T$. In the scheme (2.78), $\boldsymbol{u}_n \approx \boldsymbol{u}(t_n)$ represents the numerical approximation of the ODE solution at the nth timestep, where $t_n = n\Delta t$ denotes the time after n equally spaced intervals.

Linear Stability Given a set of fixed time-stepping coefficients, the stability of the dynamics defined by (2.78) is influenced by the function f and the time step size Δt, as detailed in [30].

For a linear function f, the multistep dynamics (2.78) admit linear mode solutions of the form $\boldsymbol{u}_n = \xi^n \boldsymbol{u}_0$, where ξ satisfies the *characteristic equation*:

$$\rho(\xi) - z\sigma(\xi) = 0, \quad \text{where} \quad z := \lambda \Delta t. \tag{2.79}$$

In the characteristic equation (2.79), ρ and σ are the *generating polynomials*, defined by the LMM coefficients as

$$\rho(\xi) = \sum_{j=0}^{k} \alpha_j \, \xi^j \,, \qquad \sigma(\xi) = \sum_{j=0}^{k} \beta_j \, \xi^j \,. \tag{2.80}$$

The linear dynamics (2.78) evolve with *growth factors* given by ξ for LMM's. The dynamics (2.78) are considered *stable* for a given z if all roots of the characteristic equation (2.79) meet the *root condition*, which requires that the roots lie within the unit disk with boundary roots being simple. The *region of absolute stability*, S, is the set of z values for which the linear dynamics (2.78) are stable:

$$S := \left\{ z \in \mathbb{C} \;:\; \rho(\xi) - z\,\sigma(\xi) \text{ satisfies the root condition} \right\}.$$

A numerical scheme, defined by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, is A-stable if S contains the left half of the complex plane, $\mathbb{C}_-$. The A-stability criteria ensures that the discrete dynamics, (2.78), remain stable whenever the linear ODE (1.1) is stable.

$$A\text{-stability:} \qquad S \supset \mathbb{C}_- = \{z \,:\, \operatorname{Re} z \leq 0\}\,, \tag{2.81}$$

For *irreducible* LMMs, i.e. ρ and σ share no common factors, a necessary and sufficient condition for A-stability is

$$\operatorname{Re}\left(\frac{\rho(\xi)}{\sigma(\xi)}\right) > 0 \quad \text{for} \quad |\xi| > 1\,.$$

Under condition (2.81), an LMM attains A-stability only if no solution ξ to the characteristic equation (2.79) satisfies $|\xi| > 1$ when z is within $\mathbb{C}_-$. Thus, $\rho(\xi)/\sigma(\xi)$ must lie outside $\overline{\mathbb{C}}_-$ whenever $|\xi| > 1$.

Accuracy and Order Conditions For an LMM characterized by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ to achieve order p, the local error must equal $\mathcal{O}(\Delta t^{p+1})$. The method is of order p, if and only if

$$\boldsymbol{e}^T\boldsymbol{\alpha} = 0 \quad \text{and} \quad \boldsymbol{\alpha}^T\boldsymbol{v}^q = q\,\boldsymbol{\beta}^T\boldsymbol{v}^{q-1} \quad \text{for} \quad q = 1,\,\ldots,\,p \tag{2.82}$$

are satisfied, where $\boldsymbol{v} = [k, k\text{-}1, \ldots, 0]^T$.

2.6.2 Nonlinear stability in LMMs

A stronger form of stability is achievable for a class of functions f, linear or nonlinear, that satisfy a one-sided Lipschitz condition:

$$\big\langle f(t, \boldsymbol{u}) - f(t, \hat{\boldsymbol{u}}), \boldsymbol{u} - \hat{\boldsymbol{u}} \big\rangle \leq 0 \quad \text{for all} \quad \boldsymbol{u}, \hat{\boldsymbol{u}} \in \mathbb{V}\,, t \geq 0\,. \tag{2.83}$$

Differential equations (1.1) with f satisfying condition (2.83) are *contractive*, meaning the distance between any two solutions does not increase over time. Numerical methods that preserve contractivity for associated discrete-in-time schemes are classified as $G-$ stable for LMMs.

G-stability for LMMs is defined in terms of the stability of the associated 1-leg method:

$$\sum_{j=0}^{k} \alpha_j \tilde{\boldsymbol{u}}_{n+j} = \Delta t\, f\left(\sum_{j=0}^{k} \beta_j \tilde{\boldsymbol{u}}_{n+j}\right), \quad \text{with} \quad \boldsymbol{U_n} := [\tilde{\boldsymbol{u}}_{n+k-1}^T, \ldots, \tilde{\boldsymbol{u}}_n^T]^T, \tag{2.84}$$

An LMM $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is *G-stable* if there exists a real, symmetric, and positive definite matrix $\boldsymbol{G}$ such that solution pairs $\boldsymbol{U}_n$ and $\hat{\boldsymbol{U}}_n$ (2.84) satisfy

$$||\boldsymbol{U}_{n+1} - \hat{\boldsymbol{U}}_{n+1}||_G \le ||\boldsymbol{U}_n - \hat{\boldsymbol{U}}_n||_G, \tag{2.85}$$

for all differential equations satisfying (2.83) and any step size $\Delta t > 0$. The G-norm is defined as

$$||\boldsymbol{U}_n||_G^2 := \sum_{i=1}^{k}\sum_{j=1}^{k} g_{ij} \langle \boldsymbol{u}_{n+i-1}, \boldsymbol{u}_{n+j-1} \rangle, \tag{2.86}$$

where $\langle \cdot, \cdot \rangle$ is the inner product on $\mathbb{V}$ used in (2.83). Dahlquist's work in [17] shows that while the original LMM dynamics $\{\boldsymbol{u}_n\}_{n\ge 0}$ may not be contractive, G-stability guarantees their stability.

Note that $G-$stability implies $A-$stability. For instance, when $f(\boldsymbol{u}) = \lambda \boldsymbol{u}$ with $\operatorname{Re}\lambda \le 0$ then f satisfies condition (2.83), the 1-leg dynamics (2.84) are identical to the LMM dynamics and solutions $\{\boldsymbol{u}_n\}$ are bounded for all n. Futhermore,

Theorem 2.6.1. *(Dahlquist 1978) if a LMM is irreducible, then it is A-stable if and only if the corresponding one-leg method is G-stable.*

2.6.3 The G-Stability LMI

Theorem 2.6.2. *The method $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is G-stable, if and only if*

$$\boldsymbol{G} \succeq 0 \quad \textit{and} \quad (\boldsymbol{P} - \boldsymbol{Q}) \succeq 0, \tag{2.87}$$

where

$$\boldsymbol{P} = \boldsymbol{\alpha}^T\boldsymbol{\beta} + \boldsymbol{\beta}\boldsymbol{\alpha}^T, \quad \boldsymbol{Q} = \boldsymbol{R}_1\boldsymbol{G}\boldsymbol{R}_1^T - \boldsymbol{R}_0\boldsymbol{G}\boldsymbol{R}_0^T, \quad \boldsymbol{R}_1 = [\boldsymbol{I}_k, 0]^T, \quad \boldsymbol{R}_0 = [0, \boldsymbol{I}_k]^T.$$

Theorem 2.6.3. *Suppose that the pth order method* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *is G-stable with*

$$\boldsymbol{\beta}^T \boldsymbol{e} = \boldsymbol{\alpha}^T \boldsymbol{v} = 1. \tag{2.88}$$

Then

$$(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = 0 \quad \textit{for} \quad p \geq 1 \quad \textit{and} \quad (\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} = 0 \quad \textit{for} \quad p \geq 2. \tag{2.89}$$

Proof. In a slight abuse of notation, we will use $\boldsymbol{e}$ to represent the vector of ones in all dimensions. Applying $\boldsymbol{e}$ to both sides of $\boldsymbol{P} - \boldsymbol{Q}$, we have

$$\boldsymbol{e}^T(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = \underbrace{\boldsymbol{e}^T\boldsymbol{\alpha}}_{0}\boldsymbol{\beta}^T\boldsymbol{e}^T + \boldsymbol{e}\boldsymbol{\beta}\underbrace{\boldsymbol{\alpha}^T\boldsymbol{e}}_{0} - \boldsymbol{e}^T\boldsymbol{G}\boldsymbol{e} + \boldsymbol{e}^T\boldsymbol{G}\boldsymbol{e} = 0. \quad \text{(via (2.82))}$$

Since the scheme is G-stable $(\boldsymbol{P} - \boldsymbol{Q}) \succeq 0$, therefore $(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = 0$ and we get the following condition on $\boldsymbol{G}$.

$$\begin{aligned} &(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = \boldsymbol{\alpha}\underbrace{\boldsymbol{\beta}^T\boldsymbol{e}}_{1} + \boldsymbol{\beta}\underbrace{\boldsymbol{\alpha}^T\boldsymbol{e}}_{0} - \boldsymbol{R}_1\boldsymbol{G}\boldsymbol{e} + \boldsymbol{R}_2\boldsymbol{G}\boldsymbol{e} = 0 \quad \text{(via (2.82) and (2.88))} \\ &\iff (\boldsymbol{R}_1 - \boldsymbol{R}_0)\boldsymbol{G}\boldsymbol{e} = \boldsymbol{\alpha}. \end{aligned} \tag{2.90}$$

For the case when $p \geq 2$, we will use the following notation.

$$\boldsymbol{v}_0 = \boldsymbol{R}_0\boldsymbol{v} = [k-1, k-2, \ldots, 0]^T \quad \text{and} \quad \boldsymbol{v}_1 = \boldsymbol{v}_0 + \boldsymbol{e} = \boldsymbol{R}_1\boldsymbol{v} = [k, k-1, \ldots, 1]^T.$$

Applying $\boldsymbol{v}$ to both sides of $\boldsymbol{P} - \boldsymbol{Q}$, we have

$$\begin{aligned} \boldsymbol{v}^T(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} &= \underbrace{\boldsymbol{v}^T\boldsymbol{\alpha}}_{1}\boldsymbol{\beta}^T\boldsymbol{v} + \boldsymbol{v}\boldsymbol{\beta}\underbrace{\boldsymbol{\alpha}^T\boldsymbol{v}}_{1} - \boldsymbol{v}^T\boldsymbol{R}_1\boldsymbol{G}\boldsymbol{R}_1^T\boldsymbol{v}^T + \boldsymbol{v}^T\boldsymbol{R}_0\boldsymbol{G}\boldsymbol{R}_0^T\boldsymbol{v} \\ &= 2\boldsymbol{\beta}^T\boldsymbol{v} - (\underbrace{2\boldsymbol{v}_0 + \boldsymbol{e}}_{\boldsymbol{v}_1^2 - \boldsymbol{v}_0^2})^T\boldsymbol{G}\boldsymbol{e} \\ &= \boldsymbol{\alpha}^T\boldsymbol{v}^2 - \boldsymbol{v}^{2^T}\underbrace{(\boldsymbol{R}_1 - \boldsymbol{R}_0)\boldsymbol{G}\boldsymbol{e}}_{\boldsymbol{\alpha}} \quad \text{(via (2.82) and (2.90))} \\ &= 0. \end{aligned}$$

Therefore, $(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} = 0$. ◻

The complete LMI for G-stability is then

Theorem 2.6.4 (Dahlquist Conditions with order conditions)**.** *Let* $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ *be a pth order LMM with* $\boldsymbol{\beta}^T \boldsymbol{e} = 1$*. Then the method is G-stable if and only if there exists a matrix* $\boldsymbol{G} \in \mathbb{S}^s$ *such that*

$$\begin{cases} \boldsymbol{G} \succeq 0\,, \\ (\boldsymbol{P} - \boldsymbol{Q}) \succeq 0\,, \\ (\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = 0 \quad \textit{for} \quad p \geq 1\,, \\ (\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} = 0 \quad \textit{for} \quad p \geq 2\,. \end{cases} \tag{2.91}$$

2.6.4 Examples

Example 2.6.1 (BDF2)**.** The first example demonstrates how to determine if the second order Backward Differentiation Formula (BDF2) is G-stable by finding a positive semidefinite matrix $\boldsymbol{G}$ that satisfies Theorem 2.6.4. The scheme coefficients are as follows.

$$\boldsymbol{\alpha} = \begin{bmatrix} \frac{3}{2} & -2 & \frac{1}{2} \end{bmatrix}^T \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$$

First, we will satisfy the condition

$$(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = \left(\begin{bmatrix} 3 & -2 & \frac{1}{2} \\ -2 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} - \begin{bmatrix} g_{1,1} & g_{1,2} & 0 \\ g_{1,2} & g_{2,2} - g_{1,1} & -g_{1,2} \\ 0 & -g_{1,2} & -g_{2,2} \end{bmatrix} \right) \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0$$

resulting in

$$g_{1,1} = \frac{3}{2} - g_{1,2} \qquad \text{and} \qquad g_{2,2} = -g_{1,2} - \frac{1}{2}.$$

Next we will satisfy

$$(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} = \left(\begin{bmatrix} 3 & -2 & \frac{1}{2} \\ -2 & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} - \begin{bmatrix} \frac{3}{2} - g_{1,2} & g_{1,2} & 0 \\ g_{1,2} & -2 & -g_{1,2} \\ 0 & -g_{1,2} & g_{1,2} + \frac{1}{2} \end{bmatrix} \right) \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = 0$$

resulting in

$$\boldsymbol{G} = \begin{bmatrix} \frac{5}{2} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P} - \boldsymbol{Q} = \boldsymbol{w}\boldsymbol{w}^T \quad \text{where} \quad \boldsymbol{w} = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\sqrt{2} & \frac{1}{\sqrt{2}} \end{bmatrix}^T.$$

Since $g_{1,1} > 0$ and $\det(\boldsymbol{G}) = \frac{1}{4} > 0$, by Sylvester's criterion $\boldsymbol{G} \succ 0$. The only nonzero eigenvalue of $(\boldsymbol{P} - \boldsymbol{Q})$ is $\boldsymbol{w}^T\boldsymbol{w} = 3$, therefore $(\boldsymbol{P} - \boldsymbol{Q}) \succeq 0$ and BDF2 is G-stable.

Example 2.6.2 (Rosales, Seibold, Shirokoff, Zhou 2nd Order IMEX [44])**.** The second example demonstrates how to determine if the implicit part of a family of second-order IMEX schemes is G-stable by finding a positive semidefinite matrix $\boldsymbol{G}$ that satisfies Theorem 2.6.4. The normalized scheme coefficients are

$$\boldsymbol{\alpha} = \begin{bmatrix} -\frac{\delta - 4}{2\,\delta} & \frac{2\,\delta - 4}{\delta} & -\frac{3\,\delta - 4}{2\,\delta} \end{bmatrix}^T \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \frac{1}{\delta^2} & \frac{2\,\delta - 2}{\delta^2} & \frac{(\delta - 1)^2}{\delta^2} \end{bmatrix}^T$$

for $\delta \in (0, 1]$. First, we will satisfy the condition $(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{e} = 0$ with

$$\boldsymbol{P} = \begin{bmatrix} \frac{-\delta+4}{\delta^3} & \frac{-\delta^2+7\,\delta-8}{\delta^3} & \frac{-(\delta-2)^3}{2\,\delta^3} \\ \frac{-\delta^2+7\,\delta-8}{\delta^3} & \frac{2\,(2\,\delta-2)\,(2\,\delta-4)}{\delta^3} & \frac{2\,\delta^3-11\,\delta^2+17\,\delta-8}{\delta^3} \\ \frac{-(\delta-2)^3}{2\,\delta^3} & \frac{2\,\delta^3-11\,\delta^2+17\,\delta-8}{\delta^3} & \frac{-(3\,\delta-4)\,(\delta-1)^2}{\delta^3} \end{bmatrix}$$

resulting in

$$g_{1,1} = -\frac{\delta + 2\,\delta\,g_{1,2} - 4}{2\,\delta} \qquad \text{and} \qquad g_{2,2} = -\frac{2\,\delta\,g_{1,2} - 3\,\delta + 4}{2\,\delta}.$$

Next we will satisfy $(\boldsymbol{P} - \boldsymbol{Q})\boldsymbol{v} = 0$ resulting in

$$\boldsymbol{G}(\delta) = \begin{bmatrix} \frac{(10\,\delta^2-3)^2+1}{20\,\delta^2} & -\frac{\delta^2-5\,\delta+5}{\delta^2} \\ -\frac{\delta^2-5\,\delta+5}{\delta^2} & \frac{5\,\delta^2-14\,\delta+10}{2\,\delta^2} \end{bmatrix} \quad \text{and} \quad \boldsymbol{P} - \boldsymbol{Q} = \left(\frac{2 - \delta}{\delta}\right)^3 \boldsymbol{w}\boldsymbol{w}^T$$

Since $g_{1,1} > 0$ and $\det(\boldsymbol{G}(\delta)) = \frac{(\delta-2)^2}{4\,\delta^2} > 0$, by Sylvester's criterion $\boldsymbol{G} \succ 0$. The only nonzero eigenvalue of $(\boldsymbol{P} - \boldsymbol{Q})$ is $\left(\frac{2-\delta}{\delta}\right)^3 \boldsymbol{w}^T\boldsymbol{w} = 3\left(\frac{2-\delta}{\delta}\right)^3$ which is positive for $\delta \in (0, 2)$, therefore $(\boldsymbol{P} - \boldsymbol{Q}) \succeq 0$ and IMEX is G-stable.

CHAPTER 3

CERTIFYING STABILITY VIA SEMIDEFINITE PROGRAMMING

In this chapter, the algebraic conditions represented as LMIs in Chapter 2 are used within a convex feasibility problem to rigorously certify stability for several recently devised schemes in the literature. We begin with an overview in Section 3.1, detailing the A-stability certification algorithm. Next, we apply the algorithm to idealized examples that satisfy the tall-tree order conditions exactly in Section 3.2.

Additionally, we examine A-stability for several recently devised schemes developed via numerical software [2, 9]. Although these schemes do not satisfy the tall-tree order conditions exactly, they exhibit a residual of typical size $\mathcal{O}(10^{-15})$. The chapter concludes with examples that establish rigorous bounds on α for $A(\alpha)$-stability.

3.1 Computational Details for Rigorous Certification.

This section describes the computational details for rigorously verifying stability via the feasibility of an LMI, proceeding through three main steps.

Building the symbolic LMI functions

First, we initialize as symbolic variables the scheme coefficients $\boldsymbol{A}, \boldsymbol{b}$ and other necessary variables such as $\boldsymbol{B}, \boldsymbol{I}, \boldsymbol{M}, \boldsymbol{e}$, etc. Next, the LMI's are parameterized. The process for building the symbolic LMI functions differs slightly between the E-polynomial and CSTW SDP approaches.

For the E-polynomial SDP, the symbolic stability function $W(z)$ is created with the given symbolic coefficient matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$. The stability function is converted to a rational form that produces the symbolic numerator and denominator functions, $N(z)$ and $D(z)$, as in (1.3). The symbolic E-polynomial is then constructed with $N(z)$, $D(z)$, and its coefficients are used to define the symbolic diagonal matrix $\boldsymbol{P}$ from (2.5). The free variables $\boldsymbol{\eta}$ for the E-polynomial LMI are introduced via the general symbolic symmetric matrix $\boldsymbol{N}_E$ which is constructed in the manner described in (2.7), satisfying $\boldsymbol{y}^T \boldsymbol{N}_E \boldsymbol{y} = 0$. The matrix $\boldsymbol{N}_E$ is then converted into a symbolic function $\boldsymbol{N}_E(\boldsymbol{\eta})$. Combining $\boldsymbol{P} + \boldsymbol{N}_E(\boldsymbol{\eta})$ gives the E-polynomial symbolic LMI function.

In the CSTW approach, the free variables $\boldsymbol{\eta}$ are introduced via the general symbolic symmetric matrix $\boldsymbol{N}$ which is constructed as in (2.29) so that $\boldsymbol{N}\boldsymbol{e} = 0$. The matrix $\boldsymbol{N}$ is then used to define the symbolic matrices $\boldsymbol{R} = \boldsymbol{B} + \boldsymbol{N}$ and $\boldsymbol{X} = \boldsymbol{R}\boldsymbol{A}^T + \boldsymbol{A}\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T$. The equality conditions (i.e. the null space of $\boldsymbol{X}$) are parameterized by solving $\boldsymbol{X}\boldsymbol{v} = 0$ for the free variables introduced by $\boldsymbol{N}$, where $\boldsymbol{v}$ is defined by the null vectors in Theorems 2.3.3 and 2.3.6. The matrices $\boldsymbol{R}$ and $\boldsymbol{X}$ are then converted into symbolic functions $\boldsymbol{R}(\boldsymbol{\eta})$ and $\boldsymbol{X}(\boldsymbol{\eta})$, which are the CSTW symbolic LMI functions.

Solving the SDP

We then use the symbolic LMI functions to create double-precision LMI functions that are used as constraints within optimization software. Specifically, CVX, a package for specifying and solving convex programs [15, 24], is used to numerically solve the relevant semi-definite programming problems:

To demonstrate the non-negativity of the E-polynomial, i.e., assess the feasibility of (2.8), we solve:

$$\begin{array}{ll} \text{Minimize:} & 1 \\ \text{Subject to:} & \boldsymbol{F}(\boldsymbol{\eta}) = \boldsymbol{P} + \sum_{j=1}^{d} \eta_j \, \boldsymbol{N}_j \succeq 0 \, , \end{array} \tag{P2}$$

where $\boldsymbol{P} + \sum_{j=1}^{d} \eta_j \, \boldsymbol{N}_j$ is represented in Matlab as a double-precision function.

For assessing the feasibility of the modified CSTW approach (2.35), we solve:

$$\begin{array}{lll}
\text{Minimize:} & 1 & \\
\text{Subject to:} & \boldsymbol{R}\boldsymbol{e} = \boldsymbol{b}\,, & \\
& \boldsymbol{X} = \boldsymbol{R}\boldsymbol{A} + \boldsymbol{A}^T\boldsymbol{R} - \boldsymbol{b}\boldsymbol{b}^T\,, & \\
& \boldsymbol{X}\boldsymbol{A}^{j-1}\boldsymbol{e} = 0\,, & \text{for} \quad j = 1,\, \ldots,\, \left\lfloor \frac{p}{2} \right\rfloor\,, \\
& \boldsymbol{M}^T\boldsymbol{R}\boldsymbol{M} \succeq 0\,, & \\
& \boldsymbol{M}^T\boldsymbol{X}\boldsymbol{M} \succeq 0\,, &
\end{array} \tag{P3}$$

where the constraints of (P3) are entirely represented in Matlab by double precision functions. Additionally, $\boldsymbol{M}$ is any matrix containing the columns where r is the largest integer for which $\boldsymbol{A}^r\boldsymbol{e}$ can be written as a linear combination of $[\boldsymbol{e}, \boldsymbol{A}\boldsymbol{e}, \ldots, \boldsymbol{A}^{r-1}\boldsymbol{e}]$.

While not a proof, a positive output from CVX provides numerical evidence that the convex set in question is feasible.

LDL Factorization

Finally, the output of CVX, $\boldsymbol{\eta}^*$, is used to construct a rigorous certificate of stability in settings where $\boldsymbol{P} \in \mathbb{Q}^{m\times m}$, or $\boldsymbol{A} \in \mathbb{Q}^{s\times s}, \boldsymbol{b} \in \mathbb{Q}^s$. The double-precision output $\boldsymbol{\eta}^*$ is passed to the symbolic LMI functions, producing symbolic rational matrics. Exact symbolic LDL factorizations are then performed to yield matrices $\boldsymbol{L}, \boldsymbol{D}$ with rational entries. A rigorous certificate is ensured provided $\boldsymbol{D} \succeq 0$.

For example, CVX yields $\boldsymbol{\eta}^* \in \mathbb{Q}^d$ in the case of the E-polynomial; substituting $\boldsymbol{\eta}^*$ back into the LMI and symbolically computing an LDL factorization yields:

$$\boldsymbol{F}(\boldsymbol{\eta}^*) = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T \quad \text{with} \quad \boldsymbol{L}, \boldsymbol{D} \in \mathbb{Q}^{m\times m}\,.$$

For presentation and simplicity, the output values of CVX η_j^* are often rounded to nearby $\bar{\eta}_j \in \mathbb{Q}$ with integer numerators and denominators having fewer digits, yet still yielding positive certificates of feasibility.

Remark. Throughout this section, coefficient matrices are presented wherever possible. However, in several practical examples, the coefficients of various matrices, e.g., $\boldsymbol{P}$, $\boldsymbol{X}$, etc., as well as the LDL factorizations of $\boldsymbol{F}$ or $\boldsymbol{X}, \boldsymbol{R}$, admit rational values exceeding 100 digits.

Remark. If the coefficients of $\boldsymbol{P}$ or $\boldsymbol{A}, \boldsymbol{b}$ lie in a field extension $\mathbb{F}$ of $\mathbb{Q}$, the LDL factorization also admits matrices $\boldsymbol{L}, \boldsymbol{D}$ in $\mathbb{F}$. Therefore, this approach generalizes as long as one can determine the sign of any element $x \in \mathbb{F}$, such as with interval arithmetic. Additionally, schemes $\boldsymbol{A}, \boldsymbol{b}$ with irrational entries may have E-polynomials with rational entries and coefficient matrices $\boldsymbol{P}$, allowing this approach to apply without further modification (e.g., the scheme (3.5) has coefficients in $\mathbb{Q}[\sqrt{2}]$ but an E-polynomial with coefficients in $\mathbb{Q}$).

The following code are general Matlab implementations of the two SDPs. Note, the function `ldls()` performs a symbolic LDL factorization, but it is not a native Matlab function. For details on the `ldls()` function see Appendix A.3.

E-polynomial SDP Code:

```
% Initialize Variables
A = sym(...); b = sym(...);            % Symbolic coefficients
s = length(b);  p =... ;               % Stage number and order
e = sym(ones(s,1));  I = sym(eye(s));

% E-polynomial Conditions
syms z; syms y real;

% Determine N and D functions
[N(z),D(z)] = numden(1 + z*b'*(I - z*A)^(-1)*e);

% Create E-polynomial
E(y) = collect(expand(D(1i*y)*D(-1i*y) - N(1i*y)*N(-1i*y)));

% Create symbolic P matrix
k = coeffs(E(y)); P = diag(k); m = length(k);

% Create monomial vector
ys = sym(ones(m,1));
for i = 1:r-1
    ys(i+1) = y^i;
end

% Define N such that y'Ny==0
n = triu(sym('n',[m m],'real'),2);
N = sym(zeros(m,m));
```

```
Ir = sym(eye(m));
for i = 1:m-2
    for j = i+2:m
        c = ceil((i+j)/2);
        f = floor((i+j)/2);
        N = N + n(i,j)*(Im(:,i)*Im(:,j)' + Im(:,j)*Im(:,i)' -
           Im(:,f)*Im(:,c)' - Im(:,c)*Im(:,f)');
    end
end

% Create symbolic function for N
inputs = sort(n(n ~= 0));
Ns = symfun(N,inputs);

% Create double precision matrix P and function N
Pd = double(P); Nd = matlabFunction(N);

% E-poly SDP
cvx_precision high
cvx_begin sdp quiet
    variable eta
    minimize 1
    subject to
        in = num2cell(eta);
        Pd+Nd(in{:}) >= 0;
cvx_end

% Symbolic LDL factorization of F
F = P + Ns(in{:});
[LF,DF] = ldls(F);
```

CSTW SDP Code:

```
% Initialize Variables
A = sym(...); b = sym(...);           % Symbolic coefficients
s = length(b);  p =... ;              % Stage number and order
e = sym(ones(s,1));  I = sym(eye(s));
B = diag(b);
M = sym(ones(s));
for i = 1:s-1
    M(:,i+1) = A^i*e;
end
r = rank(M);

% Parameterize Ne=0
n = triu(sym('n',[s s],'real'),1);
N = sym(zeros(s,s));
for i = 1:s-1
    for j = i+1:s
        v = I(:,i) - I(:,j);
```

```
        N = N + n(i,j).*(v*v');
    end
end

% Define R and X wrt N
R = B + N;
X = R*A + A'*R - b*b';

% Parameterize null vectors Xv=0 and update N,R,X
% Null vectors from order conditions
X_nullity = floor(p/2);
for i = 1:X_nullity
    in = struct2cell(solve(X*M(:,i), n(i,i+1:s)));
    Ns = symfun(N, n(i,i+1:s));
    N = Ns(in{:});
    R = B + N;
    X = R*A + A'*R - b*b';
    n(i,i+1:s) = 0 ;
end

% Null vectors from EDIRK
if all(A(1,:)==0)
    X_nullity = X_nullity+1;
    k = sym('k',[r 1],'real');
    k = cell2sym(struct2cell(solve(A^r*e+M(:,1:r)*k)));
    p_r = flip(k(2:r));
    in = struct2cell(solve(X*M*p_r, n(X_nullity,X_nullity+1:s)
       ));
    Ns = symfun(N, n(X_nullity,X_nullity+1:s));
    N = Ns(in{:});
    R = B + N;
    X = R*A + A'*R - b*b';
    n(X_nullity,X_nullity+1:s) = 0 ;
end

% Degrees of freedom
dof = nchoosek(s-X_nullity,2);

% Create symbolic functions for R,X
inputs = sort(n(n ~= 0));
Rs = symfun(R,inputs);
Xs = symfun(X,inputs);

% Create double precision MATLAB functions for R,X
Rd = matlabFunction(R);
Xd = matlabFunction(X);

% SDP feasibility problem
cvx_begin sdp quiet
    variable eta(dof,1)
    minimize 1
```

```
    subject to
        in = num2cell(eta);
        Rd(in{:}) >= 0;
        Xd(in{:}) >= 0;
cvx_end

% Symbolic LDL factorization of X,R
X = Xs(in{:});
[LX,DX] = ldls(X);
R = Rs(in{:});
[LR,DR] = ldls(R);
```

3.2 Certifying Idealized Schemes via SDP.

To demonstrate the hybrid computational-analytic LMI solution approach to verify A-stability, we begin with idealized examples that satisfy the tall-tree order conditions exactly and are known to be A-stable.

Example 3.2.1 (SDIRK(5,4))**.** First, we revisit the SDIRK(5,4) scheme represented by the Butcher tableau (2.12) [30, Table 6.5, Chapter IV.6]. We provide two proofs: one using the E-polynomial LMI and a second based on modified CSTW LMI conditions.

E-ploynomial SDP

The SDIRK(5,4) scheme in (2.12) has an E-polynomial

$$E(y, \frac{\pi}{2}) = y^6(9y^4 - 64y^2 + 512),$$

which after factoring out the largest monomial factor, yields

$$F(y) := 9y^4 - 64y^2 + 512 = \boldsymbol{y}^T\, \boldsymbol{F}(\eta)\, \boldsymbol{y}\,,$$

where

$$\boldsymbol{F}(\eta) = \begin{bmatrix} 512 & 0 & 0 \\ 0 & -64 & 0 \\ 0 & 0 & 9 \end{bmatrix} + \eta \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3.1}$$

For this example, the space $\mathcal{N}_3$ (introduced in § 2.1.1) has dimension $d = 1$ and is spanned by the second matrix in (3.1). The numerical solution (P2) for $\boldsymbol{F}(\eta)$ yields the following output:

$$\eta^* = -61.786375823904734\,.$$

The fact that CVX obtains a solution is numerical evidence suggesting $E(y) \geq 0$, indicating that the scheme is A-stable. This numerical evidence can be converted into a rigorous proof by substituting η^* into the symbolic LMI $\boldsymbol{F}(\eta)$ and factorizing exactly $\boldsymbol{F}(\eta^*) = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$ with $\boldsymbol{L}, \boldsymbol{D} \in \mathbb{Q}^{3\times 3}$ to yield $\boldsymbol{D} \succeq 0$.

We would like to remark that the value of η^* lies in the interior of the convex set defined by the LMI (P2), and is not the only value that yields a positive certificate of stability. For example, values of η close to η^* also yield positive certificates as demonstrated by the following example where $\eta = -60$ is used:

$$\boldsymbol{F}(-60) = \begin{bmatrix} 512 & 0 & -60 \\ 0 & 56 & 0 \\ -60 & 0 & 9 \end{bmatrix} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T\,,$$

where

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{15}{128} & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{D} = \begin{bmatrix} 512 & 0 & 0 \\ 0 & 56 & 0 \\ 0 & 0 & \frac{63}{32} \end{bmatrix} \succeq 0\,.$$

This results in the following SOS representation of $E(y)$, certifying A-stability:

$$E(y) = y^6 \left(512 \left(\frac{15}{128} y^2 - 1 \right)^2 + 56y^2 + \frac{63}{32} y^4 \right).$$

Modified CSTW SDP

Using the SDIRK(5,4) scheme in (2.12), this example highlights several key differences in the modified CSTW approach relative to the E-polynomial.

Since $p = 4$, Theorem 2.3.3 adds two linear constraints on $\boldsymbol{X}$ (corresponding to null vectors $\boldsymbol{e}$ and $\boldsymbol{A}\boldsymbol{e}$) into the associated LMI. Consequently, problem (P3) has three degrees of freedom, denoted by the vector $\boldsymbol{\eta}$. The rounded numerical solution of (P3) yields matrices

$$\boldsymbol{X}^* = \begin{bmatrix} \frac{729823}{97920} & -\frac{348733}{195840} & \frac{875727}{21760} & -\frac{334871}{7200} & \frac{237}{400} \\ -\frac{348733}{195840} & \frac{170083}{391680} & -\frac{1259867}{130560} & \frac{160397}{14400} & -\frac{57}{400} \\ \frac{875727}{21760} & -\frac{1259867}{130560} & \frac{5678645}{26112} & -\frac{241217}{960} & \frac{16}{5} \\ -\frac{334871}{7200} & \frac{160397}{14400} & -\frac{241217}{960} & \frac{1045211}{3600} & -\frac{1479}{400} \\ \frac{237}{400} & -\frac{57}{400} & \frac{16}{5} & -\frac{1479}{400} & \frac{19}{400} \end{bmatrix}, \quad \boldsymbol{R}^* = \begin{bmatrix} \frac{195061}{16320} & -\frac{42157}{10880} & \frac{416905}{6528} & -72 & \frac{11}{10} \\ -\frac{42157}{10880} & \frac{131641}{65280} & -\frac{324335}{13056} & \frac{1259}{48} & -\frac{11}{20} \\ \frac{416905}{6528} & -\frac{324335}{13056} & \frac{4888637}{13056} & -\frac{99107}{240} & \frac{73}{10} \\ -72 & \frac{1259}{48} & -\frac{99107}{240} & \frac{34459}{75} & -\frac{391}{50} \\ \frac{11}{10} & -\frac{11}{20} & \frac{73}{10} & -\frac{391}{50} & \frac{11}{50} \end{bmatrix},$$

which then admits LDL factorizations of the form

$$\boldsymbol{X}^* = \boldsymbol{L}_X \boldsymbol{D}_X \boldsymbol{L}_X^T\,, \qquad \boldsymbol{R}^* = \boldsymbol{L}_R \boldsymbol{D}_R \boldsymbol{L}_R^T\,,$$

where the matrices have coefficients in $\mathbb{Q}$ with

$$\boldsymbol{D}_X = \operatorname{diag}\left[\frac{729823}{97920} \;\; \frac{2466451}{280252032} \;\; \frac{7352143}{246645100} \;\; 0 \;\; 0 \right],$$

$$\boldsymbol{D}_R = \operatorname{diag}\left[\frac{195061}{16320} \;\; \frac{28479739}{37451712} \;\; \frac{3647946461}{341756868} \;\; \frac{3800443925}{43775357532} \;\; \frac{103805}{2104052} \right].$$

The fact that $\boldsymbol{D}_X$ admits two zero eigenvalues is by construction and follows from Theorem 2.3.3. The certification of A-stability is thus confirmed as the pair $\boldsymbol{X}^*, \boldsymbol{R}^*$ satisfy the CSTW condition in exact arithmetic.

Example 3.2.2 (ESDIRK32I5L2SA-5-2-3)**.** Next, we examine the ESDIRK32I5L2SA-5-2-3 scheme represented by the following coefficients:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \frac{9}{40} & \frac{9}{40} & 0 & 0 & 0 \\ \frac{19}{72} & \frac{14}{45} & \frac{9}{40} & 0 & 0 \\ \frac{3337}{11520} & \frac{233}{720} & \frac{207}{1280} & \frac{9}{40} & 0 \\ \frac{7415}{34776} & \frac{9920}{30429} & \frac{4845}{9016} & -\frac{5827}{19320} & \frac{9}{40} \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} \frac{7415}{34776} \\ \frac{9920}{30429} \\ \frac{4845}{9016} \\ -\frac{5827}{19320} \\ \frac{9}{40} \end{bmatrix} \tag{3.2}$$

E-ploynomial SDP

The scheme (3.2) has an E-polynomial

$$E(y; \tfrac{\pi}{2}) = \gamma y^4 \left(y^4 - \tfrac{6684332800}{387420489} y^2 + \tfrac{21314560000}{129140163}\right),$$

with $\gamma = 387420489$. After factoring, we obtain

$$F(y) := y^4 - \tfrac{6684332800}{387420489} y^2 + \tfrac{21314560000}{129140163} = \boldsymbol{y}^T \boldsymbol{F}(\eta)\, \boldsymbol{y},$$

where

$$\boldsymbol{F}(\eta) = \begin{bmatrix} \frac{21314560000}{129140163} & 0 & 0 \\ 0 & -\frac{6684332800}{387420489} & 0 \\ 0 & 0 & 1 \end{bmatrix} + \eta \begin{bmatrix} 0 & 0 & 1 \\ 0 & -2 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3.3}$$

The numerical solution (P2) for $\boldsymbol{F}(\eta)$ yields:

$$\eta^* = -8.626715661390847\,.$$

CVX's solution suggests $E(y) \geq 0$, indicating A-stability.

We round η^* to $\overline{\eta} = -\frac{3342166400}{387420489}$ and obtain:

$$\boldsymbol{F}(-\tfrac{3342166400}{387420489}) = \begin{bmatrix} \frac{21314560000}{129140163} & 0 & -\frac{3342166400}{387420489} \\ 0 & 0 & 0 \\ -\frac{3342166400}{387420489} & 0 & 1 \end{bmatrix} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T\,,$$

where

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1044427}{19982400} & 0 & 1 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{D} = \begin{bmatrix} \frac{21314560000}{129140163} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{2656838970463}{4838494487121} \end{bmatrix} \succeq 0\,.$$

This results in the following SOS representation of $E(y)$, certifying A-stability:

$$E(y) = \gamma y^4 \left(\tfrac{2656838970463\,y^4}{4838494487121} + \tfrac{21314560000}{129140163} \left(\tfrac{1044427}{19982400}\,y^2 - 1\right)^2 \right)\,.$$

Modified CSTW SDP

Since $p = 3$, Theorem 2.3.3 adds the linear constraint $\boldsymbol{X}\boldsymbol{e} = 0$ to the LMI. The scheme is stiffly accurate (i.e., $\boldsymbol{b}^T$ is equal to the last row of $\boldsymbol{A}$) and $\boldsymbol{A}$ is singular; therefore, Theorem 2.3.6 adds a second constraint $\boldsymbol{X}\boldsymbol{M}\boldsymbol{p}_r = 0$ to the LMI. Consequently, problem (P3) has three degrees of freedom, denoted by the vector $\boldsymbol{\eta}$. The rounded numerical solution of (P3) yields matrices

$$\boldsymbol{X}^* = \begin{bmatrix} \frac{11146204166203}{40789637296128000} & \frac{-8407940321693}{35690932634112000} & \frac{-1709258255729}{705006076723200} & \frac{6382644745231}{1616256052992000} & \frac{-65346943}{41828572800} \\ \frac{-8407940321693}{35690932634112000} & \frac{3860202003949}{7807391513712000} & \frac{10728355669307}{3084401585664000} & \frac{-2435250853211}{353556011592000} & \frac{2306359393}{732000024000} \\ \frac{-1709258255729}{705006076723200} & \frac{10728355669307}{3084401585664000} & \frac{2487396506157}{88441622528000} & \frac{-122606540519369}{2374499633408000} & \frac{1217622669}{54222224000} \\ \frac{6382644745231}{1616256052992000} & \frac{-2435250853211}{353556011592000} & \frac{-122606540519369}{2374499633408000} & \frac{1903581383743}{19080800625600} & \frac{-525072881}{11619048000} \\ \frac{-65346943}{41828572800} & \frac{2306359393}{732000024000} & \frac{1217622669}{54222224000} & \frac{-525072881}{11619048000} & \frac{508689}{24056000} \end{bmatrix}$$

$$\boldsymbol{R}^* = \begin{bmatrix} \frac{21573785741}{201072844800} & \frac{34973549813}{307892793600} & \frac{117719341}{3923763200} & -\frac{1029173}{30544920} & -\frac{443}{111600} \\ \frac{34973549813}{307892793600} & \frac{37811272673}{269406194400} & \frac{109440859}{858323200} & -\frac{2726113}{30544920} & \frac{3773}{111600} \\ \frac{117719341}{3923763200} & \frac{109440859}{858323200} & \frac{12760225893}{27466342400} & -\frac{9}{34} & \frac{9}{50} \\ -\frac{1029173}{30544920} & -\frac{2726113}{30544920} & -\frac{9}{34} & \frac{227522041}{987619080} & -\frac{14}{97} \\ -\frac{443}{111600} & \frac{3773}{111600} & \frac{9}{50} & -\frac{14}{97} & \frac{47959}{300700} \end{bmatrix},$$

which then admits LDL factorizations of the form

$$\boldsymbol{X}^* = \boldsymbol{L}_X \boldsymbol{D}_X \boldsymbol{L}_X^T, \qquad \boldsymbol{R}^* = \boldsymbol{L}_R \boldsymbol{D}_R \boldsymbol{L}_R^T,$$

where the matrices have coefficients in $\mathbb{Q}$ with

$$\boldsymbol{D}_X = \operatorname{diag}\begin{bmatrix} \frac{11146204166203}{40789637296128000} \\ \frac{947079461269957489}{3250768152664772544000} \\ 0 \\ \frac{37459168241287633057}{35861184932577354052288} \\ 0 \end{bmatrix}, \quad \boldsymbol{D}_R = \operatorname{diag}\begin{bmatrix} \frac{21573785741}{201072844800} \\ \frac{6894617926234393}{343114320952869984} \\ 0 \\ \frac{3311480780425235177}{43042455579626615400} \\ \frac{1642980924432777139}{26491846243401881416} \end{bmatrix}.$$

The fact that $\boldsymbol{D}_X$ admits two zero eigenvalues is by construction and follows from Theorems 2.3.3 and 2.3.6. The certification of A-stability is thus confirmed as the pair $\boldsymbol{X}^*, \boldsymbol{R}^*$ satisfy the CSTW condition in exact arithmetic.

Table 3.1 consists of a set of idealized RK schemes from the SUNDIALS library of implicit RK methods. They have been certified A-stable via observation of positive E-polynomial coefficients, or both SDPs (P2) and (P3).

Table 3.1 Idealized A-stable Schemes and the Method of Certification

Scheme	**Certification Method**
ARK2-DIRK-3-1-2 (M)[22]	+ coefficients
ARK2-DIRK-3-1-2 (E)[22]	+ coefficients
ARK436L2SA-DIRK-6-3-4 (M)[32]	E-poly & CSTW SDPs
ARK436L2SA-DIRK-6-3-4 (E)[32]	+ coefficients
ESDIRK325L2SA-5-2-3 (M)[33]	E-poly & CSTW SDPs
ESDIRK325L2SA-5-2-3 (E)[33]	E-poly & CSTW SDPs
ESDIRK32I5L2SA-5-2-3 (M)[33]	E-poly & CSTW SDPs
ESDIRK32I5L2SA-5-2-3 (E)[33]	+ coefficients

Scheme	Certification Method
ESDIRK436L2SA-6-3-4 (M)[33]	E-poly & CSTW SDPs
SDIRK-2-1-2 (M)	+ coefficients
SDIRK-2-1-2 (E)	+ coefficients
SDIRK-5-3-4 (M)[30]	E-poly & CSTW SDPs
TRBDF2-3-3-2 (M)[3]	+ coefficients

We formalize the results in the following Lemma.

Lemma 3.2.1. *The RK schemes listed in Table 3.1 satisfy the tall-tree order conditions, have nonnegative E-polynomials, and satisfy Theorem 2.3.3 or Theorem 2.3.6. Therefore, the schemes are A-stable.*

3.3 Schemes Failing to Satisfy the Tall-Tree Order Conditions.

Building on the idealized example from the previous subsection, we now focus on certifying stability for RK schemes developed from numerical solutions of the order conditions (i.e., using numerical optimization software). The schemes considered here have coefficients that do not exactly satisfy the tall-tree order conditions (1.7). Instead, the coefficients produce the right-hand side of (1.7) with a small residual. The tall-tree conditions play a special role when certifying A-stability (which is a property of the stability function $W(z)$) as they are exactly the conditions that ensure $W(z)$ is a pth order approximation to e^z. Failing to satisfy the tall-tree order conditions leads to two complications: the E-polynomial no longer satisfies (2.9) but instead includes low-degree monomials with small coefficients; additionally, the matrix $\boldsymbol{X}$, used in the CSTW conditions, no longer admits exact zero eigenvalues, but rather small nonzero ones.

We test schemes from two sources:

- Diagonally implicit Runge-Kutta schemes with weak stage order [9] (cf. [7, 8]). These schemes were developed to alleviate the effects of *order reduction* on stiff problems, primarily arising from spatial discretizations of linear partial differential equations. The schemes are denoted as WSO DIRK(s,p,q), where s, p, and q denote the number of stages, classical order, and weak stage order, respectively.

- (Very high order) Diagonally implicit Runge-Kutta schemes with additional practical properties, developed in [2].

It is worth noting that while both [2, 9] provide strong numerical evidence for A-stability (e.g., a solution with small residual was provided for the CSTW LMI in [2]) neither work provides a rigorous certificate in the form of an exact solution to one of the LMI's.

We adopt two strategies for testing A-stability, outlined as follows:

Strategy 1

The Butcher coefficients $(\boldsymbol{A}, \boldsymbol{b})$ are reported as decimal expansions, typically to 16 digits of accuracy (e.g., the accuracy of double-precision floating-point arithmetic from which they were obtained). Treating the coefficients as rational values, we test for the non-negativity of the associated E-polynomial, which also has rational coefficients. This strategy determines whether the dynamics (1.2) are A-stable despite the coefficients not exactly satisfying the tall-tree order conditions.

In contrast to the E-polynomial approach, challenges arise when using the CSTW approach to obtain a rigorous certificate of A-stability. The modified CSTW conditions depend on the exact satisfaction of the tall-tree order conditions. However, in cases where the Butcher coefficients approximate these conditions with a small residual, the null vectors of $\boldsymbol{X}$ are no longer applicable constraints in the SDP, and the original CSTW conditions must be used. The feasible set of the CSTW conditions may then take the form of a tubular domain of a low-dimensional set, thereby introducing computational challenges to the numerical solution.

Strategy 2

To overcome the challenges in Strategy 1 due to approximations in the tall-tree order conditions, we assess the A-stability of perturbed schemes $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ that simultaneously:

- Have coefficients in $\mathbb{Q}$, satisfy the tall-tree order conditions (1.7) in exact arithmetic, and in the case where $\boldsymbol{A}$ is singular satisfy $\boldsymbol{b}^T \boldsymbol{M} \boldsymbol{p}_r = 0$;
- Are perturbations of the reported scheme $(\boldsymbol{A}, \boldsymbol{b})$ in the literature, satisfying an error bound

$$|b_i - \tilde{b}_i| < \epsilon_b \quad \text{and} \quad |a_{ij} - \tilde{a}_{ij}| < \epsilon_A \quad \text{for} \quad i, j = 1, \ldots, s\,. \tag{3.4}$$

- Satisfy the pth non-tall tree order with residuals of size $\mathcal{O}(\epsilon_b) + \mathcal{O}(\epsilon_A) + \mathcal{O}(\epsilon)$ where ϵ is the residuals of the unperturbed scheme $(\boldsymbol{A}, \boldsymbol{b})$.

The general approach to creating the scheme $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ involves initializing the vector $\tilde{\boldsymbol{b}}$ with symbolic free variables and then solving the conditions you would like to enforce exactly, such as the tall-tree order conditions. Structural components must also be considered when perturbing the scheme $(\boldsymbol{A}, \boldsymbol{b})$. For example, SDIRKs must preserve the same value along the diagonal of $\tilde{\boldsymbol{A}}$, stiffly accurate schemes must ensure that $\tilde{\boldsymbol{b}}^T$ is the last row of $\tilde{\boldsymbol{A}}$, and zero entries must remain zero.

Since the perturbed schemes, $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$, satisfy the tall-tree order conditions, their E-polynomials satisfy (2.9) and the associated $\boldsymbol{X}$ matrix in the CSTW approach admits null vectors (in exact arithmetic) characterized by Theorem 2.3.3. Thus, both the E-polynomial and CSTW approaches provide pathways for certifying A-stability for $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$.

Several schemes that fail to achieve rigorous certificates of A-stability after Strategy 1 are shown to be (ϵ_A, ϵ_b)-close to schemes that attain rigorous A-stability certificates by Strategy 2.

3.3.1 Certification of A-stability via Strategy 1

Here, we verify A-stability of three schemes using Strategy 1. Since the schemes do not satisfy the order conditions exactly, the E-polynomial ($\alpha = \frac{\pi}{2}$) has the form:

$$E(y) = y^2 \boldsymbol{y}^T \boldsymbol{F}(\eta) \boldsymbol{y}, \qquad \text{with} \qquad \boldsymbol{F}(\eta) = \boldsymbol{P} + \sum_{j=1}^{d} \eta_j \boldsymbol{N}_j \,,$$

where $\boldsymbol{P}$ and $\boldsymbol{N}_j$ are as in § 2.1.1. The details for each scheme are as follows:

1. WSO DIRK(12,5,4) developed in [9]: A symbolic computation of $E(y)$ yields a diagonal matrix $\boldsymbol{P}$, containing the coefficients of $E(y)$. Since $\boldsymbol{P} \succeq 0$, it follows that $\boldsymbol{F}(\eta) \succeq 0$ trivially when $\boldsymbol{\eta} = 0$. Consequently, no SDP is required, as the LMI is immediately satisfied.

2. WSO DIRK(7,4,4), also developed in [9]: A symbolic computation of $E(y) = y^2 \boldsymbol{y}^T \boldsymbol{P} \boldsymbol{y}$, yields a diagonal matrix $\boldsymbol{P} \in \mathbb{Q}^{7\times 7}$ with two negative coefficients

$$p_6 < 0 \,, \qquad p_8 < 0 \,.$$

   Since $\boldsymbol{F}(0)$ does not satisfy the LMI, we seek solutions to the LMI (which has dimension $d = 15$) via an SDP. The E-polynomial SDP identifies a candidate feasible solution $\boldsymbol{\eta}^*$, which for computational simplicity, we round to $\bar{\boldsymbol{\eta}}$, resulting in the matrix:

$$\boldsymbol{F}(\bar{\boldsymbol{\eta}}) = \boldsymbol{P} - 108420\boldsymbol{N}_{10} + 20\boldsymbol{N}_{12} - 3420\boldsymbol{N}_{13} - 30\boldsymbol{N}_{15} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T,$$

   with $\boldsymbol{L}, \boldsymbol{D} \in \mathbb{Q}^{7\times 7}$ and $\boldsymbol{D} \succ 0$.

3. DIRK(13,8)[1]A[(14,6)A] developed in [2]: Similar to the previous case, symbolic computation of the E-polynomial yields $\boldsymbol{P} \in \mathbb{Q}^{13\times 13}$ with two negative coefficients

$$p_{14} < 0\,, \qquad p_{22} < 0\,,$$

requiring an SDP to find a potential sum of squares representation for $E(y)$. The E-polynomial LMI has dimension $d = 66$, and the corresponding SDP identifies a feasible solution $\boldsymbol{\eta}^*$, which upon rounding yields $\bar{\boldsymbol{\eta}}$ (reported in Appendix A.1.2), resulting in the matrix $\boldsymbol{F}(\bar{\boldsymbol{\eta}}) = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$, where the diagonal matrix $\boldsymbol{D} \succ 0$.

Table 3.2 contains RK schemes from [9] and [2], as well as schemes from the SUNDIALS library that fail to satisfy the tall-tree conditions. They were certified A-stable with Strategy 1.

Table 3.2 A-stable Schemes that Approximate the Tall-Tree Conditions and the Method of Certification

Scheme	**Certification Method**
ARK437L2SA-DIRK-7-3-4 (M)[35]	E-poly SDP
ARK548L2SA-ESDIRK-8-4-5 (M)[33]	+ coefficients
ARK548L2SAb-DIRK-8-4-5 (M)[35]	+ coefficients
ARK548L2SAb-DIRK-8-4-5 (E)[35]	+ coefficients
Billington-3-3-2 (M)[5]	+ coefficients
Cash-5-2-4 (M)[12]	+ coefficients
Cash-5-3-4 (M)[12]	+ coefficients
Cash-5-3-4 (E)[12]	+ coefficients
ESDIRK324L2SA-4-2-3 (M)[34]	+ coefficients
ESDIRK324L2SA-4-2-3 (E)[34]	+ coefficients

Scheme	Certification Method
ESDIRK437L2SA-7-3-4 (M)[34]	E-poly SDP
Kvaerno-4-2-3 (E)[36]	+ coefficients
Kvaerno-5-3-4 (M)[36]	+ coefficients
Kvaerno-5-3-4 (E)[36]	+ coefficients
Kvaerno-7-4-5 (M)[36]	+ coefficients
Kvaerno-7-4-5 (E)[36]	+ coefficients
WSO DIRK (7,4,4)[9]	E-poly SDP
WSO DIRK(12,5,4)[9]	+ coefficients
DIRK(13,8)[1]A[(14,6)A][2]	E-poly SDP

The complete set of coefficients for $\bar{\boldsymbol{\eta}}, \boldsymbol{L}, \boldsymbol{D}$, as well as generating Matlab code, can be found in the supplemental material.

The results are formalized with the following Lemma.

Lemma 3.3.1. *The E-polynomials for the schemes listed in Table 3.2 are non-negative, and therefore, the schemes are A-stable.*

3.3.2 Certification of A-stability via Strategy 2

We now apply Strategy 2 to certify A-stability for schemes where Strategy 1 fails due to the E-polynomial for $(\boldsymbol{A}, \boldsymbol{b})$ being negative near the origin or as $|y| \to \infty$.

Throughout, tildes denote quantities for the perturbed scheme $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$, e.g., $\tilde{\boldsymbol{P}}$ is the matrix containing the coefficients of the E-polynomial $\tilde{E}(y)$.

1. DIRK(6,6)[1]A[(7,5)A] developed in [2]: The original scheme's E-polynomial, $E(y) = y^2 \boldsymbol{y}^T \boldsymbol{P} \boldsymbol{y}$, possesses three negative coefficients in $\boldsymbol{P}$

$$p_0 < 0, \qquad p_2 < 0, \qquad p_4 < 0\,.$$

Given that $p_0 < 0$, $E(y)$ is negative for values of y near the origin, indicating that the scheme is not A-stable. This inability to achieve A-stability directly follows from the failure to satisfy the tall-tree order conditions — which, if satisfied, would ensure $p_0 = p_2 = p_4 = 0$.

We introduce the perturbed scheme

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \frac{33128226}{109158329} & & & & & \\ -\frac{254432096}{909477001} & \frac{51289103}{102571593} & & & & \\ \frac{33289838}{118645151} & -\frac{825218320}{1881654059} & \frac{130993323}{602959172} & & & \\ -\frac{13583292}{200438515} & \frac{156154430}{158643099} & -\frac{65409371}{245235917} & \frac{81765600}{330141853} & & \\ \frac{16354062}{130133299} & -\frac{247816720}{248961507} & \frac{169383005}{222482121} & -\frac{49241043}{234166886} & \frac{79900588}{92184791} & \\ -\frac{34719176}{94331171} & -\frac{155737141}{155748342} & \frac{42945649}{80312134} & -\frac{50573402}{289227347} & \frac{678237381}{1102812170} & \frac{31879369}{45767530} \end{bmatrix},$$

where the coefficient vector $\tilde{\boldsymbol{b}} \in \mathbb{Q}^6$ are chosen to exactly satisfy the tall-tree conditions (1.7) with $p = 6$. The pair $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ satisfy the error bound (3.4) with $\epsilon_A = 5 \cdot 10^{-17}$, and $\epsilon_b = 6 \cdot 10^{-15}$.

A-stability for the perturbed scheme is certified via two approaches:

- The E-polynomial $\tilde{E}(y) = y^8 \boldsymbol{y}^T \tilde{\boldsymbol{P}} \boldsymbol{y} \geq 0$ follows immediately since $\tilde{\boldsymbol{P}} \in \mathbb{Q}^{3\times 3}$ satisfies $\tilde{\boldsymbol{P}} \succeq 0$.
- The modified CSTW approach (P3) yields an LMI in 3 variables. The numerical solution produces a solution pair:

$$\tilde{\boldsymbol{X}}^* = \tilde{\boldsymbol{L}}_X \tilde{\boldsymbol{D}}_X \tilde{\boldsymbol{L}}_X^T \quad \text{and} \quad \tilde{\boldsymbol{R}}^* = \tilde{\boldsymbol{L}}_R \tilde{\boldsymbol{D}}_R \tilde{\boldsymbol{L}}_R^T,$$

  with $\tilde{\boldsymbol{D}}_X \succeq 0$ ($\tilde{\boldsymbol{D}}_X$ has 4 zero eigenvalues by Theorem 2.3.3), and $\tilde{\boldsymbol{D}}_R \succeq 0$.

2. SDIRK(9,6)[1]SAL[(9,5)A] developed in [2]: Similar to the previous example, this scheme is not A-stable as the E-polynomial is negative near the origin.

This is reflected in the matrix $\boldsymbol{P} \in \mathbb{Q}^{9\times 9}$ having negative coefficients

$$p_0 < 0, \qquad p_2 < 0, \qquad p_4 < 0\,,$$

a consequence of failing to satisfy the tall-tree order conditions. A perturbed scheme $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$, defined in Appendix A.1.3, satisfies the error bound (3.4) with $\epsilon_A = \epsilon_b = 8 \cdot 10^{-15}$.

Similar to DIRK(6,6), A-stability for the perturbed scheme $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ is certified through two approaches:

- The E-polynomial $\tilde{E}(y) = y^8 \boldsymbol{y}^T \tilde{\boldsymbol{P}} \boldsymbol{y} \geq 0$, with $\tilde{\boldsymbol{P}} \in \mathbb{Q}^{6\times 6}$, is immediately non-negative since $\tilde{\boldsymbol{P}} \succeq 0$, avoiding the need for solving an SDP.
- The modified CSTW approach, which utilizes an LMI with 15-degree-of-freedom, is solved via SDP. The numerical solver identifies optimal pairs $(\tilde{\boldsymbol{R}}^*, \tilde{\boldsymbol{X}}^*)$, yielding exact LDL factorizations over $\mathbb{Q}$, with all matrices $\boldsymbol{D}$ being positive semidefinite.

3. WSO DIRK(12,5,5) developed in [9]: Similar to the previous two examples, this scheme also fails to be A-stable due to the lowest order terms in the E-polynomial having negative coefficients

$$p_0 < 0, \qquad p_2 < 0,$$

implying that $E(y) < 0$ for small y. A perturbed scheme $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ is defined in Appendix A.1.3 and satisfies (3.4) with $\epsilon_A = \epsilon_b = 9 \cdot 10^{-15}$.

For $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$, the E-polynomial has the form $\tilde{E}(y) = y^6 \boldsymbol{y}^T \tilde{\boldsymbol{F}}(\boldsymbol{\eta}) \boldsymbol{y}$, where $\tilde{\boldsymbol{P}} \in \mathbb{Q}^{10\times 10}$ is not positive definite. The associated LMI for the non-negativity of $\tilde{\boldsymbol{F}}$ contains $\dim \mathcal{N}_{10} = 36$ degrees of freedom. The Appendix A.1.3 presents a solution $\bar{\boldsymbol{\eta}}$ that yields $\tilde{\boldsymbol{F}}(\bar{\boldsymbol{\eta}}) = \tilde{\boldsymbol{L}} \tilde{\boldsymbol{D}} \tilde{\boldsymbol{L}}^T$, with $\tilde{\boldsymbol{D}} \succeq 0$, thereby certifying A-stability.

Table 3.3 contains schemes that satisfy the tall-tree conditions and are ϵ-close (as described in (3.4)) to RK schemes from [9] and [2], as well as schemes from the SUNDIALS library. They were certified A-stable with Strategy 2.

Table 3.3 A-stable ϵ-Schemes, the Order of Perturbation, and the Method of Certification

ϵ-Scheme	$\boldsymbol{\epsilon}$	**Certification Method**
ARK324L2SA-DIRK-4-2-3 (M)[32]	$9 \cdot 10^{-27}$	+ coefficients
ARK324L2SA-DIRK-4-2-3 (E)[32]	$2 \cdot 10^{-26}$	+ coefficients
ARK437L2SA-DIRK-7-3-4 (M)[35]	$6 \cdot 10^{-25}$	E-poly & CSTW SDPs
ARK548L2SA-ESDIRK-8-4-5 (E)[33]	$5 \cdot 10^{-25}$	+ coefficients
Cash-5-2-4 (E)[12]	$4 \cdot 10^{-13}$	+ coefficients
ESDIRK436L2SA-6-3-4 (E)[33]	$5 \cdot 10^{-26}$	E-poly & CSTW SDPs
ESDIRK437L2SA-7-3-4 (E)[34]	$3 \cdot 10^{-24}$	E-poly & CSTW SDPs
ESDIRK547L2SA-7-4-5 (M)[33]	$3 \cdot 10^{-25}$	E-poly & CSTW SDPs
ESDIRK547L2SA-7-4-5 (E)[33]	$2 \cdot 10^{-25}$	+ coefficients
ESDIRK547L2SA2-7-4-5 (M)[34]	$1 \cdot 10^{-24}$	E-poly & CSTW SDPs
ESDIRK547L2SA2-7-4-5 (E)[34]	$7 \cdot 10^{-25}$	E-poly & CSTW SDPs
Kvaerno-4-2-3 (M)[36]	$3 \cdot 10^{-15}$	+ coefficients
QESDIRK436L2SA-6-3-4 (M)[33]	$2 \cdot 10^{-25}$	E-poly & CSTW SDPs
QESDIRK436L2SA-6-3-4 (M)[33]	$9 \cdot 10^{-24}$	E-poly & CSTW SDPs
ESDIRK43I6L2SA-6-3-4 (M)[33]	$2 \cdot 10^{-25}$	E-poly SDP
DIRK(6,6)[1]A[(7,5)A][2]	$6 \cdot 10^{-15}$	+ coefficients & CSTW SDP
SDIRK(9,6)[1]SAL[(9,5)A][2]	$8 \cdot 10^{-15}$	+ coefficients & CSTW SDP
WSO DIRK(12,5,5)[9]	$9 \cdot 10^{-15}$	E-poly SDP

We formalize the result in the following Lemma.

Lemma 3.3.2. *There exist perturbed schemes* $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ *for all RK schemes listed in Table 3.3 which satisfy the error bound* (3.4) *with* $\epsilon_A, \epsilon_b = \mathcal{O}(10^{-15})$*. These schemes simultaneously satisfy the tall-tree conditions and are A-stable.*

3.4 Certifying $A(\alpha)$-Stability.

We now shift attention and establish rigorous bounds on α, for $A(\alpha)$-stability of two schemes that are not A-stable.

First note that $E(y^2;\beta)$, where $\beta := \cos(\alpha)$, is a bivariate polynomial in y, β with rational coefficients. An initial rational value of β is fixed, and the associated E-polynomial SDP (P2) is solved to determine whether the associated feasible set is nonempty. The value of β is incrementally decreased until (P2) indicates that the feasible set is empty. The last β value before the feasible set is reported non-empty is β^*. The corresponding $\alpha^* = \cos^{-1}(\beta^*)$ is the lower bound for the maximal angle α at which the scheme is confirmed to be $A(\alpha)$-stable. An exact rational certificate in the form of an LDL factorization is then reported.

The result is formalized as follows:

Lemma 3.4.1. *The schemes* (3.5) *and* (3.8) *are* $A(\alpha)$*-stable for some maximum angle* α *that is bounded below by the angle* α^**, where* $\alpha^* = \cos^{-1}\beta^*$ *and* β^* *is defined in* (3.7) *and* (3.9) *respectively.*

Example 3.4.1 (The IRK(4,4) Scheme of Ramos and Vigo)**.** The following scheme, developed by Ramos and Vigo, is a 4-stage, 4th order fully implicit method based on a BDF-type Chebyshev approximation [41]:

$$\mathbf{A} = \begin{bmatrix} \frac{22-\sqrt{2}}{96} & \frac{5-8\sqrt{2}}{48} & \frac{22-7\sqrt{2}}{96} & \frac{-1}{16} \\ \frac{4+3\sqrt{2}}{24} & \frac{1}{6} & \frac{4-3\sqrt{2}}{24} & 0 \\ \frac{22+7\sqrt{2}}{96} & \frac{5+8\sqrt{2}}{48} & \frac{22+\sqrt{2}}{96} & \frac{-1}{16} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \\ 0 \end{bmatrix}. \tag{3.5}$$

By defining $\beta := \cos(\alpha)$, the generalized E-polynomial (2.1) becomes the following bivariate polynomial in y and β:

$$\begin{aligned} E(y^2;\beta) =& y^2\Big(y^{14} + 22\beta y^{12} + (272\beta^2 - 16)y^{10} + (1920\beta^3 + 96\beta)y^8 \\ &+ (6144\beta^4 + 3840\beta^2)y^6 + (36864\beta^3 + 10752\beta)y^4 + 73728\beta y^2 + 294912\beta\Big). \end{aligned} \tag{3.6}$$

When $\beta = 0$ (corresponding to $\alpha = \frac{\pi}{2}$), the E-polynomial

$$E\left(y;\frac{\pi}{2}\right) = y^6\left(y^2 - 16\right),$$

is negative for $|y| < 4$, and the scheme is not A-stable.

Instead, we determine a lower bound for the maximal value of α for which $E(y^2;\beta) \geq 0$. The last β value before reaching an empty feasible set is the bound

$$\beta^* = \frac{19699132}{4466212691} \qquad (\alpha^* \approx 89.74728^o). \tag{3.7}$$

Solving Problem (P2) with the value β^* (details provided in Appendix A.2.1) produces the positive certificate

$$E(y^2;\beta^*) = y^2\boldsymbol{y}^T\boldsymbol{F}(\bar{\boldsymbol{\eta}})\boldsymbol{y} \geq 0$$

where $\boldsymbol{F}(\bar{\boldsymbol{\eta}}) = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T$, and the diagonal matrix $\boldsymbol{D} \succ 0$, certifying the scheme is $A(\alpha)$-stable for maximal angle $\alpha \geq \alpha^*$.

Example 3.4.2 (The ESDIRK(8,6) Scheme of Skvortsov)**.** We examine the 8-stage, 6th order ESDIRK scheme of Skvortsov [45] characterized by its Butcher matrix:

$$
\boldsymbol{A} = \begin{bmatrix}
0 & & & & & & & \\
\frac{1}{6} & \frac{1}{6} & & & & & & \\
\frac{11}{96} & -\frac{1}{32} & \frac{1}{6} & & & & & \\
\frac{1}{12} & -\frac{1}{4} & \frac{1}{2} & \frac{1}{6} & & & & \\
-\frac{2015}{15072} & -\frac{6987}{5024} & \frac{3271}{1884} & \frac{175}{471} & \frac{1}{6} & & & \\
-\frac{326531}{573678} & -\frac{114988}{31871} & \frac{1208156}{286839} & \frac{132950}{286839} & \frac{68}{203} & \frac{1}{6} & & \\
-\frac{331717945}{2106545616} & -\frac{480525599}{416107776} & \frac{2240951089}{1404363744} & \frac{394951619}{2808727488} & -\frac{5160553}{26834976} & \frac{35815}{352512} & \frac{1}{6} & \\
\frac{16264655341}{73026914688} & \frac{9786099235}{14425069568} & -\frac{34306812733}{48684609792} & -\frac{15985588007}{97369219584} & \frac{37652437}{930279168} & -\frac{340747}{12220416} & \frac{1}{26} & \frac{1}{6} \\
\frac{7}{90} & 0 & 0 & 0 & \frac{16}{45} & -\frac{4}{45} & \frac{2}{15} \; \frac{16}{45} & \frac{1}{6}
\end{bmatrix} \tag{3.8}
$$

The scheme is stiffly accurate, with $\boldsymbol{b}^T$ being the last row of $\boldsymbol{A}$.

The generalized E-polynomial is

$$E(y, \beta) = y^2 \sum_{j=0}^{15} q_{2j}(\beta)\, y^{2j} \qquad \text{where} \qquad \beta = \cos(\alpha)\,,$$

with polynomial coefficients of β:

$$
\begin{aligned}
q_{30}(\beta) &= 1 \\
q_{28}(\beta) &= 96\beta \\
q_{26}(\beta) &= 4032\beta^2 + \tfrac{61415271}{616225} \\
q_{24}(\beta) &= 96768\beta^3 + \tfrac{73235232}{3925}\beta \\
q_{22}(\beta) &= 1451520\beta^4 + \tfrac{3567255552}{3925}\beta^2 - \tfrac{91554624}{3925} \\
q_{20}(\beta) &= 13934592\beta^5 + \tfrac{14120096256}{785}\beta^3 + \tfrac{9673437312}{3925}\beta \\
q_{18}(\beta) &= 83607552\beta^6 + \tfrac{158718486528}{785}\beta^4 + \tfrac{70172863488}{785}\beta^2 - \tfrac{3175034112}{3925} \\
q_{16}(\beta) &= 286654464\beta^7 + \tfrac{1149206704128}{785}\beta^5 + \tfrac{985309774848}{785}\beta^3 + \tfrac{136916137728}{785}\beta \\
q_{14}(\beta) &= 429981696\beta^8 + \tfrac{5111514906624}{785}\beta^6 + \tfrac{8045011279872}{785}\beta^4 + \tfrac{694870576128}{157}\beta^2 - \tfrac{4152010752}{785} \\
q_{12}(\beta) &= \tfrac{10416951558144}{785}\beta^7 + \tfrac{41311620145152}{785}\beta^5 + \tfrac{34328744275968}{785}\beta^3 + \tfrac{5012232804864}{785}\beta \\
q_{10}(\beta) &= \tfrac{650132324352}{5}\beta^6 + 232190115840\beta^4 + 121899810816\beta^2 \\
q_8(\beta) &= \tfrac{3064909529088}{5}\beta^5 + 835884417024\beta^3 + 121899810816\beta \\
q_6(\beta) &= 2298682146816\beta^4 + 1671768834048\beta^2 \\
q_4(\beta) &= 6269133127680\beta^3 + 1253826625536\beta \\
q_2(\beta) &= 9403699691520\beta^2 \\
q_0(\beta) &= 5642219814912\beta\,.
\end{aligned}
$$

Similar to the scheme of Ramos and Vigo, this scheme fails to be A-stable. When $\beta = 0$ ($\alpha = \frac{\pi}{2}$), the E-polynomial simplifies to:

$$
E\left(y; \frac{\pi}{2}\right) = y^8 \left(y^8 + \tfrac{61415271}{616225}y^6 - \tfrac{91554624}{3925}y^4 - \tfrac{3175034112}{3925}y^2 - \tfrac{4152010752}{785}\right)\,.
$$

Since the lowest order monomial term is negative, e.g., $-4152010752/785 < 0$, the polynomial $E(y;\pi/2)$ is negative for values of y near the origin.

After iterating through values of β, we establish the bound:

$$
\beta^* = \frac{2218472195}{100000000000} \qquad (\alpha^* \approx 88.7288^o)\,. \tag{3.9}
$$

This value is consistent with the α bound reported in [45] and is now accompanied by a rigorous certificate.

The E-polynomial produced by β^* can be written as

$$E(y^2;\beta^*) = y^2 \boldsymbol{y}^T \boldsymbol{F}(\boldsymbol{\eta}) \boldsymbol{y}\,, \quad \text{with} \quad \boldsymbol{F}(\boldsymbol{\eta}) = \boldsymbol{P} + \sum_{j=1}^{105} \eta_j \boldsymbol{N}_j\,,$$

where $\boldsymbol{P} \in \mathbb{Q}^{16\times16}$ is a diagonal matrix containing three negative coefficients and $d = 105$ is the dimension (2.6). Numerically solving Problem (P2) yields a solution $\boldsymbol{\eta}^*$, which again we round for simplicity to a nearby rational value $\bar{\boldsymbol{\eta}}$ presented in Appendix A.2.2. Substituting $\bar{\boldsymbol{\eta}}$ into the LMI and factorizing yields

$$E(y^2;\beta^*) = y^2 \boldsymbol{y}^T \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T \boldsymbol{y} \geq 0\,, \tag{3.10}$$

where $\boldsymbol{D}, \boldsymbol{L} \in \mathbb{Q}^{16\times16}$ with $\boldsymbol{D}$ being diagonal with non-negative entries.

Remark. Even though $\boldsymbol{\eta}^*$ has been rounded to a nearby rational value $\bar{\boldsymbol{\eta}}$ with fewer digits, which we report in the Appendix, the matrix and $\boldsymbol{D}$ in (3.10) contains integer denominators with up to 548 digits.

Data Availability

Certificates of stability and the supporting numerical code are available online from https://github.com/ajuhl/Algebraic-Conditions-for-Stability-Certified-via-SDP.

CHAPTER 4

CONCLUSION

Contributions to the stability theory and practical application in Runge-Kutta methods are presented in this dissertation. Two key theoretical enhancements are provided. First, the CSTW conditions were modified to account for the Runge-Kutta order conditions in Theorem 2.3.3 and the singular coefficient matrix $\boldsymbol{A}$ in Theorem 2.3.6. These modifications address practical limitations introduced by zero eigenvalues of CSTW conditions matrix $\boldsymbol{X}$, enabling rigorous certification of stability through computational approaches. Second, new algebraic conditions for A-stability were introduced, derived from a novel class polynomials orthogonal with respect to a linear functional. This development leads to a representation of $W(z)$ as a continued fraction approximation to the exponential function and identifies the minimal variables in $(\boldsymbol{A}, \boldsymbol{b})$ responsible for A-stability.

Incorporating the sum-of-squares representation of the E-polynomial and the modified CSTW conditions into linear matrix inequalities enabled rigorous certification of stability via semidefinite programming. This approach has practical applications, particularly in validating and constructing stable time-stepping schemes for potential implementation in industrial software.

Looking ahead, the approaches and perspectives developed in this dissertation could be useful in certifying stability in other time-integration schemes. For example, Dahlquist's algebraic conditions for G-stability of Linear Multistep Methods were explored in Section 2.6. Semidefinite programming can be applied in other settings where algebraic conditions for stability are known, such as the algebraic stability of general linear methods. Alternatively, there are settings, such as A-stability in general linear methods, where, to the best of our knowledge, algebraic conditions for stability have yet to be formulated, providing opportunities for future work.

APPENDIX A

ADDITIONAL MATERIALS FOR RK METHODS

A.1 Supplemental Details for A-stable Schemes

Here, we present coefficients verifying A-stability.

A.1.1 SDIRK(5,4)

The HW-transform was used to verify A-stability of the SDIRK(5,4) scheme in Example 2.5.2. The following are the full LDL factorizations:

$$\boldsymbol{L}_R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -\frac{1749629661547257}{1466793482401430} & 1 & 0 \\ 0 & \frac{14609948848175}{293358696480286} & -\frac{166559936112545213}{1322607221174555535} & 1 \end{bmatrix},$$

$$\boldsymbol{D}_R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{146679348240143}{561186228540} & 0 & 0 \\ 0 & 0 & \frac{112421613799837220475}{374154375230100448976} & 0 \\ 0 & 0 & 0 & \frac{1131787522590904}{15560084954994771} \end{bmatrix},$$

$$\boldsymbol{L}_Y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -\frac{112954617902918094189}{96686712121760382200} & 1 & 0 \\ 0 & \frac{4850830510515448919}{96686712121760382200} & \frac{829170978394227987318402689}{22937559852461082330738635621} & 1 \end{bmatrix},$$

$$\boldsymbol{D}_Y = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{16670122779613859}{75361505118144} & 0 & 0 \\ 0 & 0 & \frac{1349268226615357784161096213}{32911228613810865201210880000} & 0 \\ 0 & 0 & 0 & \frac{827873853161678860082674738}{530262413059835609175310811709} \end{bmatrix}.$$

A.1.2 DIRK (13,8)[1]A[(14,6)]A

The scheme has E-polynomial $E(y) = y^2 \boldsymbol{y}^T \boldsymbol{F}(\bar{\boldsymbol{\eta}}) \boldsymbol{y}$ where

$$\boldsymbol{F}(\bar{\boldsymbol{\eta}}) = \boldsymbol{P} + \sum_{j=1}^{66} \bar{\eta}_j \boldsymbol{N}_j \succeq 0\,, \tag{A.1}$$

with coefficients of $\bar{\boldsymbol{\eta}}$ given by:

$\bar{\eta}_1 = -8470700$

$\bar{\eta}_2 = 0$

$\bar{\eta}_3 = -3700$

$\bar{\eta}_4 = -17219137300$

$\bar{\eta}_5 = -2500$

$\bar{\eta}_6 = 1552662793500$

$\bar{\eta}_7 = 1100$

$\bar{\eta}_8 = -291771102600$

$\bar{\eta}_9 = 3000$

$\bar{\eta}_{10} = 5801136200$

$\bar{\eta}_{11} = -10101540029400$

$\bar{\eta}_{12} = -949900$

$\bar{\eta}_{13} = 45726880755500$

$\bar{\eta}_{14} = 244100$

$\bar{\eta}_{15} = -2222445564500$

$\bar{\eta}_{16} = -3000$

$\bar{\eta}_{17} = -1556643190544700$

$\bar{\eta}_{18} = 1744900$

$\bar{\eta}_{19} = 292506767260200$

$\bar{\eta}_{20} = -1976300$

$\bar{\eta}_{21} = -5815692149200$

$\bar{\eta}_{22} = -13689040894555700$

$\bar{\eta}_{23} = -2465700$

$\bar{\eta}_{24} = 665333207306300$

$\bar{\eta}_{25} = -639200$

$\bar{\eta}_{26} = -27185350128356800$

$\bar{\eta}_{27} = -58914800$

$\bar{\eta}_{28} = 540480365078400$

$\bar{\eta}_{29} = -3060809665186300$

$\bar{\eta}_{30} = 933800$

$\bar{\eta}_{31} = -101765843690800$

$\bar{\eta}_{32} = 0$

$\bar{\eta}_{33} = -16794400$

$\bar{\eta}_{34} = 0$

$\bar{\eta}_{35} = 3600$

$\bar{\eta}_{36} = 19168235300$

$\bar{\eta}_{37} = 0$

$\bar{\eta}_{38} = -14572100$

$\bar{\eta}_{39} = 0$

$\bar{\eta}_{40} = 18000$

$\bar{\eta}_{41} = 16836266900$

$\bar{\eta}_{42} = 0$

$\bar{\eta}_{43} = -3637800$

$\bar{\eta}_{44} = -5738476103900$

$\bar{\eta}_{45} = 16400$

$\bar{\eta}_{46} = 4362777200$

$\bar{\eta}_{47} = 0$

$\bar{\eta}_{48} = 2843900$

$\bar{\eta}_{49} = -1564541946300$

$\bar{\eta}_{50} = -4500$

$\bar{\eta}_{51} = 338036200$

$\bar{\eta}_{52} = 26403742783600$

$\bar{\eta}_{53} = -61100$

$\bar{\eta}_{54} = -20078540400$

$\bar{\eta}_{55} = 0$

$\bar{\eta}_{56} = -256500$

$\bar{\eta}_{57} = 294637386500$

$\bar{\eta}_{58} = 600$

$\bar{\eta}_{59} = -63663500$

$\bar{\eta}_{60} = -1286411886200$

$\bar{\eta}_{61} = 200$

$\bar{\eta}_{62} = 978576800$

$\bar{\eta}_{63} = 0$

$\bar{\eta}_{64} = -5871439100$

$\bar{\eta}_{65} = 0$

$\bar{\eta}_{66} = 1269200$

A.1.3 The perturbed schemes from § 3.3.2

In this appendix, we report the coefficients of the perturbed schemes $(\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{b}})$ for which the lemmas in § 3.3.2 were established.

SDIRK(9,6)[1]SAL[(9,5)A]: The perturbed scheme coefficients are

$$\tilde{\boldsymbol{A}} = \begin{bmatrix} \frac{87518253}{401224696} \\ -\frac{109147862}{1208036163} & \frac{87518253}{401224696} \\ \frac{70447391}{407323275} & -\frac{60128027}{170018875} & \frac{87518253}{401224696} \\ \frac{258011928}{503929669} & \frac{32776647}{1131632696} & -\frac{13636148}{946751265} & \frac{87518253}{401224696} \\ \frac{2139251}{459753907} & -\frac{18361775}{242765601} & \frac{13889605}{63926963} & -\frac{17789601}{861400843} & \frac{87518253}{401224696} \\ \frac{48479320}{54097599} & \frac{222681723}{1598951647} & -\frac{6683180}{35754039} & \frac{14834219}{220428796} & -\frac{67840169}{193336343} & \frac{87518253}{401224696} \\ \frac{67080581}{121311880} & -\frac{401007739}{912707597} & \frac{169517869}{507988720} & -\frac{10552416}{310889555} & -\frac{54396621}{357996284} & \frac{12212839}{571158716} & \frac{87518253}{401224696} \\ \frac{176730716}{279920507} & \frac{185311713}{255696311} & -\frac{40314204}{93283073} & \frac{92464054}{154464243} & -\frac{281536253}{397040384} & -\frac{15560941}{32151589} & \frac{170501635}{450595763} & \frac{87518253}{401224696} \\ 0 & \tilde{b}_2 & \tilde{b}_3 & \tilde{b}_4 & \tilde{b}_5 & \tilde{b}_6 & \tilde{b}_7 & \tilde{b}_8 & \frac{87518253}{401224696} \end{bmatrix}$$

where $\tilde{b}_j$ $(3 \leq j \leq 8)$ is provided by solving for the tall tree conditions (1.7) (for $p = 6$), leaving a free variable $\tilde{b}_2$ which is selected to minimize the ℓ_2 difference $\boldsymbol{b} - \tilde{\boldsymbol{b}}$.

WSO DIRK(12,5,5) The perturbed scheme coefficients are

$$
\tilde{\boldsymbol{A}} = \begin{bmatrix}
\frac{10747729}{261281103} \\
\frac{39255733}{244819013} & \frac{19264472}{289086473} \\
-\frac{63980223}{186836899} & \frac{70601555}{81544818} & \frac{48699329}{492234648} \\
\frac{757656751}{80284215} & -\frac{5452955845}{500830197} & \frac{261637874}{98954371} & \frac{86994158}{471217857} \\
-\frac{290059410}{846787661} & \frac{142027951}{274596637} & \frac{63720572}{69536691} & \frac{8719292}{166871841} & \frac{230505997}{1977768146} \\
-\frac{96530823}{46089059} & \frac{278501442}{108044467} & \frac{186959114}{327745145} & \frac{31757051}{261668409} & -\frac{65608216}{138056009} & \frac{50822223}{96152122} \\
\frac{110645970}{326232259} & -\frac{52198210}{186593643} & \frac{48069176}{46243347} & \frac{36028733}{602611029} & -\frac{42230947}{197997752} & \frac{11234921}{134641567} & \frac{339062341}{1406835502} \\
\frac{432515173}{73254485} & \frac{1037694284}{327224921} & -\frac{1241126819}{100347987} & -\frac{76295713}{152911958} & \frac{155195477}{71832145} & \frac{268359096}{140054533} & \frac{309243168}{155550259} & \frac{266682747}{1194765721} \\
\frac{131193951}{284188360} & -\frac{74904386}{387416395} & -\frac{136922649}{1129220324} & \frac{46345993}{695639065} & \frac{61016892}{143403385} & \frac{74333617}{94618599} & \frac{157610940}{188314681} & \frac{31846751}{198449271} & \frac{305893355}{845914548} \\
-\frac{82024283}{115728139} & \frac{91013047}{140744863} & \frac{63214225}{132835881} & -\frac{178521351}{694495505} & \frac{96297873}{85736426} & \frac{97069417}{174996915} & \frac{25465272}{79767817} & \frac{11384038}{31516593} & \frac{232759857}{396742103} & \frac{83426711}{354434193} \\
\frac{100055236}{234642175} & \frac{213090564}{161088509} & \frac{96986289}{228435568} & -\frac{174464145}{68947184} & -\frac{67238506}{859605737} & \frac{2068579853}{1961737581} & \frac{170367931}{366730407} & \frac{198092237}{172991608} & \frac{312091183}{725567705} & \frac{159786147}{106558690} & \frac{9565123}{660600961} \\
\frac{5486027}{454369097} & \frac{9757227}{18810635} & \frac{62278071}{555407431} & -\frac{1694527}{341651848} & -\frac{133046372}{98916929} & \frac{323864293}{952870301} & \frac{261379716}{320347663} & \tilde{b}_8 & \tilde{b}_9 & \tilde{b}_{10} & \tilde{b}_{11} & \tilde{b}_{12}
\end{bmatrix}
$$

The coefficients $\tilde{b}_j \in \mathbb{Q}$ $(8 \leq j \leq 12)$ are chosen to satisfy the order conditions exactly (1.7) for $p = 5$.

The E-polynomial $E(y) = y^2 \boldsymbol{y}^T \boldsymbol{F}(\overline{\boldsymbol{\eta}})\boldsymbol{y} = \boldsymbol{LDL}^T \geq 0$ where $\overline{\boldsymbol{\eta}} \cdot 10^{-3}$ has coefficients:

$\overline{\eta}_1 = -6922561820555$	$\overline{\eta}_{13} = -550$	$\overline{\eta}_{25} = 692417$
$\overline{\eta}_2 = -6159041$	$\overline{\eta}_{14} = -1017731680875$	$\overline{\eta}_{26} = -2928211$
$\overline{\eta}_3 = 401988060958$	$\overline{\eta}_{15} = -1687736$	$\overline{\eta}_{27} = -2$
$\overline{\eta}_4 = 209390$	$\overline{\eta}_{16} = 8353405937$	$\overline{\eta}_{28} = -5711$
$\overline{\eta}_5 = -3199255341$	$\overline{\eta}_{17} = 14765$	$\overline{\eta}_{29} = 3$
$\overline{\eta}_6 = -4095$	$\overline{\eta}_{18} = -10433650$	$\overline{\eta}_{30} = 143788$
$\overline{\eta}_7 = 3934894$	$\overline{\eta}_{19} = -40414145977$	$\overline{\eta}_{31} = -10$
$\overline{\eta}_8 = -6775128853059$	$\overline{\eta}_{20} = -57578$	$\overline{\eta}_{32} = -40848$
$\overline{\eta}_9 = -4571676$	$\overline{\eta}_{21} = 139245638$	$\overline{\eta}_{33} = 0$
$\overline{\eta}_{10} = 142233029108$	$\overline{\eta}_{22} = 147$	$\overline{\eta}_{34} = 859$
$\overline{\eta}_{11} = 198557$	$\overline{\eta}_{23} = -553788073$	$\overline{\eta}_{35} = 0$
$\overline{\eta}_{12} = -490076976$	$\overline{\eta}_{24} = -636$	$\overline{\eta}_{36} = -3$

A.2 Supplemental Details for $A(\alpha)$-stable Schemes

Here, we present numerical coefficients verifying $A(\alpha)$-stability.

A.2.1 The IRK(4,4) scheme of Ramos and Vigo

The scheme has E-polynomial $E(y^2, \beta^*) = y^2 \boldsymbol{y}^T \boldsymbol{F}(\overline{\boldsymbol{\eta}})\boldsymbol{y}$ where

$$\boldsymbol{F}(\overline{\boldsymbol{\eta}}) = \boldsymbol{P} + \sum_{j=1}^{21} \overline{\eta}_j \boldsymbol{N}_j \succeq 0\,, \tag{A.2}$$

with coefficients of $\overline{\boldsymbol{\eta}}$ given by: $\overline{\eta}_2 = \overline{\eta}_4 = \overline{\eta}_6 = \overline{\eta}_8 = \overline{\eta}_{10} = \overline{\eta}_{13} = \overline{\eta}_{15} = \overline{\eta}_{17} = \overline{\eta}_{20} = 0$

$\overline{\eta}_1 = -\frac{343818785}{387257}$	$\overline{\eta}_7 = -\frac{1002782638}{963823}$	$\overline{\eta}_{12} = -\frac{140623753}{190944}$	$\overline{\eta}_{18} = \frac{1407711}{121108}$
$\overline{\eta}_3 = \frac{44352332}{270307}$	$\overline{\eta}_9 = \frac{26195675}{165379}$	$\overline{\eta}_{14} = \frac{19205029}{233487}$	$\overline{\eta}_{19} = -\frac{23169437}{338293}$
$\overline{\eta}_5 = -\frac{7044484}{1620291}$	$\overline{\eta}_{11} = -\frac{526928}{268115}$	$\overline{\eta}_{16} = -\frac{55546025}{187031}$	$\overline{\eta}_{21} = -\frac{2727674}{410745}$

A.2.2 ESDIRK(8,6) Skvortsov scheme in §3.4.2

We present coefficients for the scheme in subsection 3.4.2.

Solution coefficients $\bar{\boldsymbol{\eta}}$ for non-negativity of $E(y, \beta^*)$.

$$\overline{\eta}_1 = -\frac{544417542815}{1496}$$
$$\overline{\eta}_3 = \frac{29842486335}{2687}$$
$$\overline{\eta}_5 = -\frac{49740795}{931}$$
$$\overline{\eta}_7 = -\frac{28918182864}{1151}$$
$$\overline{\eta}_9 = \frac{699763151}{1893}$$
$$\overline{\eta}_{11} = -\frac{611431235}{606}$$
$$\overline{\eta}_{13} = \frac{4436247}{898}$$
$$\overline{\eta}_{15} = -\frac{20553023}{989}$$
$$\overline{\eta}_{17} = -\frac{184449}{1142}$$
$$\overline{\eta}_{19} = -\frac{5040480296101}{64}$$
$$\overline{\eta}_{21} = -\frac{20253725215145}{857}$$
$$\overline{\eta}_{23} = \frac{240883590679}{566}$$
$$\overline{\eta}_{25} = \frac{4696001199805}{754}$$
$$\overline{\eta}_{27} = -\frac{26257051251763}{262}$$
$$\overline{\eta}_{28} = -\frac{12225979229715}{323}$$
$$\overline{\eta}_{30} = \frac{433008073700}{13}$$
$$\overline{\eta}_{32} = -\frac{8424611000191}{98}$$
$$\overline{\eta}_{34} = -\frac{25273367886229}{395}$$
$$\overline{\eta}_{36} = \frac{9262966388511}{292}$$
$$\overline{\eta}_{38} = -\frac{11180085898009}{131}$$
$$\overline{\eta}_{40} = -\frac{10519983461815}{799}$$
$$\overline{\eta}_{42} = -\frac{13505692581791}{106}$$
$$\overline{\eta}_{44} = \frac{13635317782915}{474}$$
$$\overline{\eta}_{46} = -\frac{22015444860842}{179}$$
$$\overline{\eta}_{48} = -\frac{23644956443647}{379}$$
$$\overline{\eta}_{50} = -\frac{2159073125275}{1597}$$
$$\overline{\eta}_{51} = \frac{58606387358}{645}$$
$$\overline{\eta}_{53} = -\frac{2209952042}{1439}$$
$$\overline{\eta}_{55} = -\frac{3023415560208}{431}$$
$$\overline{\eta}_{57} = \frac{667361119609}{1153}$$
$$\overline{\eta}_{59} = -\frac{15606394145}{932}$$
$$\overline{\eta}_{61} = \frac{70764854}{907}$$
$$\overline{\eta}_{63} = -\frac{6003091166555}{1293}$$
$$\overline{\eta}_{65} = \frac{145331883768}{577}$$
$$\overline{\eta}_{67} = -\frac{2769061905}{757}$$
$$\overline{\eta}_{69} = \frac{2674611184492}{349}$$
$$\overline{\eta}_{70} = -\frac{435375081998}{497}$$
$$\overline{\eta}_{72} = \frac{28638357917}{941}$$
$$\overline{\eta}_{74} = -\frac{18964563}{119}$$
$$\overline{\eta}_{76} = -\frac{883878192398}{307}$$
$$\overline{\eta}_{78} = \frac{38657713081}{326}$$
$$\overline{\eta}_{80} = -\frac{1721374819}{1269}$$
$$\overline{\eta}_{82} = \frac{2813400132854}{117}$$
$$\overline{\eta}_{84} = -\frac{958434399311}{491}$$
$$\overline{\eta}_{85} = \frac{11434535023}{204}$$
$$\overline{\eta}_{87} = -\frac{220302170}{863}$$
$$\overline{\eta}_{89} = \frac{9715693027109}{1173}$$
$$\overline{\eta}_{91} = -\frac{261194717165}{607}$$
$$\overline{\eta}_{93} = \frac{5649757805}{927}$$
$$\overline{\eta}_{95} = -\frac{5175464369937}{284}$$
$$\overline{\eta}_{96} = \frac{1361933288513}{864}$$
$$\overline{\eta}_{98} = -\frac{161915093932}{3421}$$
$$\overline{\eta}_{100} = \frac{405541643}{1810}$$
$$\overline{\eta}_{102} = -\frac{4892528411838}{1501}$$
$$\overline{\eta}_{103} = \frac{127826367589}{733}$$
$$\overline{\eta}_{105} = -\frac{1622486899}{642}$$

All $\overline{\eta}_j = 0$ for any $j = 1, \ldots, 105$ not defined above.

A.3 `ldls()` Definition

```
function [L,D] = ldls(M)
%Symbolic LDL factorization

    % Initialize L and D
    l = length(M);
    L = sym(eye(l));
    D = sym(zeros(l));

    % LDL^T decomposition
    for j = 1:l
        % Diagonal D entries
        sumD = 0;
        for k = 1:j-1
            sumD = sumD + L(j,k)^2*D(k,k);
        end
        D(j,j) = M(j,j) - sumD;

        % Off-diagonal L entries
        for i = j+1:l
            sumL = 0;
            for k = 1:j-1
                sumL = sumL + L(i,k)*L(j,k)*D(k,k);
            end
            if(D(j,j)~=0)
                L(i,j) = (M(i,j) - sumL)/D(j,j);
            end
        end
    end
end
```

REFERENCES

[1] A. A. Ahmadi and R. M. Jungers. Switched stability of nonlinear systems via SOS-convex Lyapunov functions and semidefinite programming. In *52nd IEEE Conference on Decision and Control*, pages 727–732, 2013.

[2] Y. Alamri and D. Ketcheson. Very high-order A-stable stiffly accurate diagonally implicit Runge-Kutta methods with error estimators, 2024. arXiv: 2211.14574.

[3] R. Bank, W. Coughran, W. Fichtner, E. Grosse, D. Rose, and R. Smith. Transient simulation of silicon devices and circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 4(4):436–451, 1985.

[4] R. Bhatia. *Matrix analysis*, volume 169. Springer Science & Business Media, 2013.

[5] S. Billington. *Type Insensitive Codes for the Solution of Stiff and Non-stiff Systems of Ordinary Differential Equations*. Master's thesis, University of Manchester, 1983.

[6] G. Birkhoff and R. S. Varga. Discretization errors for well-set Cauchy problems. *Journal of Mathematics and Physics*, 44:1–23, 1965.

[7] A. Biswas, D. Ketcheson, S. Roberts, B. Seibold, and D. Shirokoff. Explicit Runge-Kutta methods that alleviate order reduction, 2023. arXiv:2310.02817.

[8] A. Biswas, D. Ketcheson, B. Seibold, and D. Shirokoff. Algebraic structure of the weak stage order conditions for Runge–Kutta methods. *SIAM Journal on Numerical Analysis*, 62(1):48–72, 2024.

[9] A. Biswas, D. Ketcheson, B. Seibold, and D. Shirokoff. Design of DIRK schemes with high weak stage order. *Communications in Applied Mathematics and Computer Science*, 18:1–28, 2023.

[10] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004, pages 127–187.

[11] J. Butcher. On A-stable implicit Runge-Kutta methods. *BIT Numerical Mathematics*, 17:375–378, 4, 1977.

[12] J. R. CASH. Diagonally Implicit Runge-Kutta Formulae with Error Estimates. *IMA Journal of Applied Mathematics*, 24(3):293–301, Nov. 1979. ISSN: 0272-4960. eprint: https://academic.oup.com/imamat/article-pdf/24/3/293/1930344/24-3-293.pdf.

[13] G. Cooper. An algebraic condition for A-stable Runge-Kutta methods. *Numerical Analysis*:32–46, 1986.

[14] S. Corteel, J. S. Kim, and D. Stanton. Moments of orthogonal polynomials and combinatorics. *Recent trends in combinatorics*:545–578, 2016.

[15] I. CVX Research. CVX: matlab software for disciplined convex programming, version 2.0. http://cvxr.com/cvx, Aug. 2012.

[16] G. Dahlquist. A special stability problem for linear multistep methods. *BIT Numerical Mathematics*, 3:27–43, 1963.

[17] G. Dahlquist. Positive functions and some applications to stability questions for numerical methods. In C. DE BOOR and G. H. GOLUB, editors, *Recent Advances in Numerical Analysis*, pages 1–29. Academic Press, 1978.

[18] G. Dahlquist. On stability and error analysis for stiff non-linear problems part I. Technical report, CM-P00069396, 1975.

[19] Y. Drori and M. Teboulle. Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145:1436–4646, 2014.

[20] B. Ehle. A-stable methods and Padé approximations to the exponential. *SIAM Journal on Mathematical Analysis*, 4:671–680, 1973.

[21] B. Ehle and Z. Picel. Two-parameter, artibtrary order, exponential approximations for stiff equations. *Mathematics of Computation*, 29(130):501–511, 1975.

[22] R. W. Freund. A transpose-free quasi-minimal residual algorithm for non-hermitian linear systems. *SIAM Journal on Scientific Computing*, 14(2):470–482, 1993. eprint: https://doi.org/10.1137/0914029.

[23] A. Gahlawat and G. Valmorbida. A semi-definite programming approach to stability analysis of linear partial differential equations. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pages 1882–1887, 2017.

[24] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008.

[25] B. Grimmer. Provably faster gradient descent via long steps, 2024. arxiv.org/abs/2307.06324.

[26] E. Hairer. Constructive characterization of A-stable approximations to exp z and its connection with algebraically stable Runge-Kutta methods. *Numerische Mathematik*, 39:247–258, 1982.

[27] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I (2nd Revised. Ed.): Nonstiff Problems*. Springer-Verlag, New York, 1993. ISBN: 0-387-56670-8.

[28] E. Hairer and H. Türke. The equivalence of B-stability and A-stability. *BIT Numerical Mathematics*, 24:520–528, 1984.

[29] E. Hairer and G. Wanner. Algebraically stable and implementable Runge-Kutta methods of high order. *SIAM Journal on Numerical Analysis*, 18:1098–1108, 1981.

[30] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II(2nd Revised. Ed.): Stiff and Differential-Algebraic Problems*. Springer Berlin Heidelberg, 1996.

[31] M. Jin and J. Lavaei. Stability-certified reinforcement learning: a control-theoretic perspective. *IEEE Access*, 8:229086, 2020.

[32] C. T. Kelley. *Iterative Methods for Linear and Nonlinear Equations*. Society for Industrial and Applied Mathematics, 1995. eprint: https://epubs.siam.org/doi/pdf/10.1137/1.9781611970944.

[33] C. A. Kennedy and M. H. Carpenter. Additive Runge-Kutta schemes for convection-diffusion–reaction equations. *Applied Numerical Mathematics*, 44(1):139–181, 2003. ISSN: 0168-9274.

[34] C. A. Kennedy and M. H. Carpenter. Diagonally implicit Runge-Kutta methods for stiff ODEs. *Applied Numerical Mathematics*, 146:221–244, 2019. ISSN: 0168-9274.

[35] C. A. Kennedy and M. H. Carpenter. Higher-order additive Runge-Kutta schemes for ordinary differential equations. *Applied Numerical Mathematics*, 136:183–205, 2019. ISSN: 0168-9274.

[36] A. Kværnø. Singly diagonally implicit Runge-Kutta methods with an explicit first stage. *BIT Numerical Mathematics*, 44:489–502, 2004.

[37] J. B. Lasserre. *Moments, Positiv Polynomials and Their Applications*. Imperial College Press, 2010.

[38] A. Majumdar, A. A. Ahmadi, and R. Tedrake. Control design along trajectories with sums of squares programming. In *2013 IEEE International Conference on Robotics and Automation*, pages 4054–4061, 2013.

[39] S. Nørsett. C-polynomials for rational approximations to the exponential function. *Numerische Mathematik*, 25:39–56, 1975.

[40] S. Nørsett. Restricted Padé approximations to the exponential function. *SIAM Journal on Numerical Analysis*, 15(5):1008–1029, 1978.

[41] H. Ramos and J. Vigo-Aguiar. A fourth-order Runge-Kutta method based on BDF-type chebyshev approximations. *Journal of Computational and Applied Mathematics*, 204:124–136, 2007.

[42] R. Scherer and H. Turke. Algebraic characterization of A-stable Runge-Kutta methods. *Applied Numerical Mathematics*, 5:133–144, 1989.

[43] R. Scherer and W. Wendler. Complete algebraic characterization of A-stable Runge-Kutta methods. *SIAM Journal on Numerical Analysis*, 31(2):540–551, 1994.

[44] B. Seibold, D. Shirokoff, and D. Zhou. Unconditional stability for multistep ImEx schemes: practice. *Journal of Computational Physics*, 376:295–321, 2019.

[45] L. Skortsov. Diagonally implicit Runge-Kutta methods for stiff problems. *Computational Mathematics and Mathematical Physics*, 46:2110–2123, 2006.

[46] J. H. Verner. Explicit Runge–Kutta methods with estimates of the local truncation error. *SIAM Journal on Numerical Analysis*, 15(4):772–790, 1978. eprint: https://doi.org/10.1137/0715051.

[47] G. Viennot. *Une théorie combinatoire des polynômes orthogonaux généraux*. Departement de Mathématiques et d'Informatique, Université du Québec à Montréal, 1983.

[48] G. Wanner. Characterization of all A-stable methods of order 2m-4. *BIT Numerical Mathematics*, 20:367–374, 1980.

[49] D. D. Yao, S. Zhang, and X. Y. Zhou. Stochastic linear-quadratic control via semidefinite programming. *SIAM Journal on Control and Optimization*, 40(3):801–823, 2001.