

2011-2

## A Stability Study of a New Explicit Numerical Scheme for a System of Differential Equations with a Large Skew-Symmetric Component

Katharine Gurski  
*Howard University*

Stephen O'Sullivan  
*Technological University Dublin, [stephen.osullivan@tudublin.ie](mailto:stephen.osullivan@tudublin.ie)*

Follow this and additional works at: <https://arrow.tudublin.ie/scschmatart>



Part of the [Mathematics Commons](#)

---

### Recommended Citation

Gurski, K., O'Sullivan, S. (2011) A stability study of a new explicit numerical scheme for a system of differential equations with a large skew-symmetric component, *SIAM Journal on Numerical Analysis*, 49 (1), 368-386 doi:10.1137/090775804 doi/abs/10.1137/090775804

This Article is brought to you for free and open access by the School of Mathematics and Statistics at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [vera.kilshaw@tudublin.ie](mailto:vera.kilshaw@tudublin.ie).

# A STABILITY STUDY OF A NEW EXPLICIT NUMERICAL SCHEME FOR A SYSTEM OF DIFFERENTIAL EQUATIONS WITH A LARGE SKEW-SYMMETRIC COMPONENT.

KATHARINE GURSKI\* AND STEPHEN O’SULLIVAN†

**Abstract.** Explicit numerical methods for the solution of a system of stiff differential equations suffer from a time step size that approaches zero in order to satisfy stability conditions. Implicit schemes allow a larger time-step, but require more computations. When the differential equations are dominated by a skew-symmetric component, the problem is not stiffness in the sense that the size of the eigenvalues are unequal, rather the that the real eigenvalues are dominated by imaginary eigenvalues. We present and compare analytical results for stable time step limits for several explicit methods including the super-time-stepping method of Alexiades, Amiez, and Gremaud which is a explicit Runge-Kutta method for parabolic partial differential equations and a new method modeled on a predictor-corrector scheme with multiplicative operator splitting. This new explicit method, presented in regular and super-time-stepping form, increases stability without forcing the step size to zero.

**Key words.** explicit method, symmetric, skew-symmetric, multiplicative operator splitting, super-time-stepping, stability

**AMS subject classifications.** 65L06, 65L20

**1. Introduction.** In this paper we consider explicit time integration methods of parabolic equations. We assume space discretization has converted the partial differential equations into a system of ordinary differential equations. We therefore consider the initial value problem

$$\mathbf{B}'(t) = \mathbf{F}(\mathbf{B}(t)), \quad \mathbf{B}(t_0) = \mathbf{B}_0, \quad t \in [t_0, T], \quad \mathbf{B} \in \mathbb{R}^n. \quad (1.1)$$

We denote the current time as  $t^\ell$  and seek the solution at a later time  $t^{\ell+1} = t^\ell + \tau$ , where  $\tau$  is a positive and real time step. In particular we consider a system of ordinary differential equations that may be discretized as

$$\mathbf{B}^{\ell+1} = (\mathbf{I} - \tau\mathbf{M}) \mathbf{B}^\ell. \quad (1.2)$$

The general ordinary differential equation system (1.1) may be solved using an explicit or implicit scheme. An explicit scheme will calculate  $\mathbf{B}$  at the time  $t^{\ell+1}$  using  $\mathbf{B}(t^\ell)$ , while an implicit scheme will calculate  $\mathbf{B}$  by solving an equation involving both  $t^{\ell+1}$  and  $\mathbf{B}(t^\ell)$ . The solution to implicit schemes will require extra computations (thereby more computationally expensive) and are more difficult to implement. However, implicit schemes are often used for stiff problems when an explicit scheme requires the time step  $\tau$  to approach zero to keep the error bounded. For example, explicit schemes have to satisfy the Courant-Friedrichs-Lewy (CFL) stability condition [8] which requires a numerical scheme satisfy  $\rho(\mathbf{I} - \tau\mathbf{M}) < 1$  where  $\rho(\cdot)$  denotes the spectral radius. If  $\mathbf{M}$  is symmetric then this requires

$$\tau < \frac{2}{\lambda_M}. \quad (1.3)$$

We use the notation that the eigenvalues of an  $n \times n$  matrix  $\mathbf{X}$  are ordered  $\lambda_1(\mathbf{X}) \geq \lambda_2(\mathbf{X}) \geq \dots \geq \lambda_n(\mathbf{X})$  and  $\lambda_X$  refers to an eigenvalue in the interval  $[\lambda_n(\mathbf{X}), \lambda_1(\mathbf{X})]$ .

---

\*Department of Mathematics, Howard University, Washington, D.C.

†School of Physical Sciences, Dublin City University, Dublin, Ireland

As mentioned, the CFL stability restriction on the time step forces the stepsize  $\tau$  to be much smaller than the necessary size to satisfy the accuracy condition for the computation. There have been multiple approaches to increasing the time-step while continuing to maintain an explicit scheme. Our new method builds on two approaches, super-time-stepping (STS) [1] and multiplicative operator splitting [4].

The super-time-stepping (STS) method of Alexiades, Amiez, and Gremaud [1] for a symmetric  $\mathbf{M}$ , uses  $m$  intermediate steps

$$\mathbf{B}^{\ell+1} = \left( \prod_{j=1}^m (\mathbf{I} - \tau_j \mathbf{M}) \right) \mathbf{B}^{\ell}. \quad (1.4)$$

The full analysis shows that if  $\bar{\tau}_{\text{explicit}}$  is the time step for one time step of equation (1.4), then

$$\tau = \sum_{j=1}^m \tau_j = m^2 \bar{\tau}_{\text{explicit}}.$$

This actually gives an increased stepsize,

$$\tau < m \tau_{\text{explicit}},$$

over one step of the standard explicit scheme in equation (1.2). This larger stepsize from the STS method follows by enforcing the CFL condition on the exterior step rather than enforcing the CFL condition on each of the  $m$  interior steps. As a result larger time steps can be taken, which implies that the total number of steps is reduced significantly, compared with the standard explicit scheme.

One method that is considered semi-explicit is a multiplicative operator splitting approach called Lie splitting [13]. Using Lie splitting one could rephrase the problem as

$$\mathbf{B}^{\ell+1} = (\mathbf{I} - \tau \mathbf{M}_1)(\mathbf{I} - \tau \mathbf{M}_2) \mathbf{B}^{\ell}, \quad (1.5)$$

where  $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$ . This method can be used to split the time evolution into partial steps to separate the effects of convection,  $\mathbf{M}_1$ , and diffusion,  $\mathbf{M}_2$ . Typically the convection step is solved explicitly while the diffusion step is solved implicitly.

In this paper we consider a real, non-symmetric  $\mathbf{M}$ . Rather than seeking to separate the effects of convection and diffusion, we wish to split  $\mathbf{M}$  into symmetric and skew-symmetric components. This decomposition into symmetric and skew-symmetric components is  $\mathbf{P} = \frac{1}{2}(\mathbf{M} + \mathbf{M}^T)$  and  $\mathbf{S} = \frac{1}{2}(\mathbf{M} - \mathbf{M}^T)$  respectively. Thus, the system is

$$\mathbf{B}^{\ell+1} = \mathbf{G} \mathbf{B}^{\ell}, \quad (1.6)$$

where

$$\mathbf{G} = \mathbf{I} - \tau \mathbf{P} - \tau \mathbf{S}. \quad (1.7)$$

Systems of this type may be seen in models for weakly ionized plasmas with Hall and ambipolar diffusion [17] or other fluid dynamics applications when the advection term dominates diffusion due to small viscosity.

We note that the  $\mathbf{G}$  scheme is not stable for a skew-dominated system. Our derivation in section 2 shows that the CFL condition requires

$$\tau < \frac{2 \operatorname{Re}(\lambda_G)}{|\lambda_G|^2}.$$

If the skew component of  $\mathbf{G}$  becomes dominant, then behavior of the system approaches

$$\mathbf{B}^{\ell+1} = (\mathbf{I} - \tau \mathbf{S}) \mathbf{B}^\ell, \quad (1.8)$$

as  $\ell$  becomes large. In such a system, the CFL condition requires the step size  $\tau$  to approach zero (as  $\operatorname{Re}(\lambda_G) \rightarrow 0$ ) to maintain stability. Hence we note that when the differential equations are dominated by a skew-symmetric component, the essence of the difficulty is not stiffness in the sense that the size of the eigenvalues are unequal, rather that the real eigenvalues are dominated by imaginary eigenvalues.

Anum-Addo, Darst and Gurski[2, 3] considered the effect of adding symmetric elements into the scheme with two methods,  $\mathbf{C}$  and  $\mathbf{D}$

$$\begin{aligned} \mathbf{C} &= \mathbf{I} - \tau \mathbf{P} - \tau \mathbf{S} + \tau^2 \mathbf{P}^2, \\ \mathbf{D} &= \mathbf{I} - \tau \mathbf{P} - \tau \mathbf{S} + \tau^2 \mathbf{S}^2. \end{aligned}$$

The analysis confirmed that adding symmetric elements did stabilize the system and adding a symmetric element composed of the skew-symmetric component was more stabilizing as the skew-symmetric component grew large in comparison to the symmetric component.

Hence we choose to add a symmetric element composed of the skew-symmetric component in a systematic manner. If we take one possible predictor-corrector approach [7, 21], we have

$$\mathbf{B}^{\ell+1} = \mathbf{B}^\ell - \tau \mathbf{S} \tilde{\mathbf{B}}^{\ell+1}, \quad (1.9)$$

where

$$\tilde{\mathbf{B}}^{\ell+1} = (\mathbf{I} - \tau \mathbf{S}) \mathbf{B}^\ell. \quad (1.10)$$

This yields the discretization scheme:

$$\mathbf{B}^{\ell+1} = (\mathbf{I} - \tau \mathbf{S} + \tau^2 \mathbf{S}^2) \mathbf{B}^\ell. \quad (1.11)$$

Since even powers of a skew operator are symmetric this correction adds a pure symmetric component back into the scheme and stabilizes it allowing a time step of

$$\tau < \frac{1}{|\lambda_S|}.$$

However, as there are several choices for using a predictor-corrector type scheme to add additional symmetric components for this problem, we have explored which scheme would be optimal. For example, one interpretation for combining multiplicative operator splitting with a predictor-corrector scheme is

$$\mathbf{F} = [\mathbf{I} - \tau \mathbf{P}] \times [\mathbf{I} - \tau \mathbf{S}] \times [\mathbf{I} - \tau \mathbf{S}].$$

While this scheme adds additional symmetric components, just as the  $\mathbf{C}$  scheme, it does not allow non-zero time-steps as  $\lambda_P$  approaches zero.

We introduce a new scheme which improves the stability of the scheme by allowing a larger time-step for the instances when skew-symmetric term dominates the symmetric term. This scheme incorporates multiplicative operator splitting [4] with a predictor-corrector scheme.

$$\mathbf{H} = \prod_{k=1}^m (\mathbf{I} - \tau_k \mathbf{P}) \times (\mathbf{I} - \tau_k \mathbf{S} + \tau_k^2 \mathbf{S}^2),$$

which is equivalent to a super-time-step version of the three step scheme:

$$\begin{aligned} \mathbf{B}^{\ell+1} &= (\mathbf{I} - \tau \mathbf{P}) \tilde{\mathbf{B}}^{\ell+1}, \\ \tilde{\mathbf{B}}^{\ell+1} &= \mathbf{B}^\ell - \tau \mathbf{S} \bar{\mathbf{B}}^{\ell+1}, \\ \bar{\mathbf{B}}^{\ell+1} &= (\mathbf{I} - \tau \mathbf{S}) \mathbf{B}^\ell. \end{aligned}$$

We note that the super-time-stepping algorithm stability conditions are shown analytically in [1] for symmetric  $\mathbf{M}$  with numerical results only for the non-symmetric  $\mathbf{M}$  case. Tilley and Balsara [18] note that the numerical results of [1] show a squeezing of the appropriate time-step for non-symmetric super-time-step method and that without an analytical proof, the super-time-step method may have unpredictable results. Hence we include an analytical proof for the stable time-step for the super-time-step version of the  $\mathbf{G}$  scheme,

$$\mathbf{T} = \prod_{k=1}^m [\mathbf{I} - \tau_k (\mathbf{P} + \mathbf{S})], \quad (1.12)$$

for comparison.

Since  $\mathbf{H}$  includes matrix products we cannot explicitly define  $\rho(\mathbf{H})$  in terms of  $\rho(\mathbf{P})$  and  $\rho(\mathbf{S})$  [11]. The eigenspaces of  $\mathbf{H}$  and  $\mathbf{T}$  are significantly more complicated than the eigenspace for  $\mathbf{G}$ . However, in this paper we find upper bounds on the CFL condition for  $\mathbf{H}$  and  $\mathbf{T}$  without the needing to define these in terms of  $\rho(\mathbf{P})$  and  $\rho(\mathbf{S})$ . We can summarize the relationship of the upper bounds on the stable time step in the following theorem, with the details of the proof given in later sections of this paper.

**THEOREM 1.1.** *The schemes  $\mathbf{H}$  and  $\mathbf{T}$  for the  $n \times n$  matrix  $\mathbf{G}$  with time-steps  $\tau_H$  and  $\tau_T$ , respectively, are equivalent if  $|\lambda_S| = 0$ . For  $|\lambda_S| = 0$  and  $\lambda_P > 0$ , we find that  $\tau_H = \tau_T = m\tau_G$ .*

*For  $|\lambda_S| \neq 0$ , and  $\lambda_P > 0$ , i.e. both symmetric and skew symmetric terms contribute and the skew component may dominate, the stable time steps are ordered as  $\tau_H > \tau_T > \tau_G$ .*

*For  $\lambda_P = 0$ , the  $\mathbf{G}$  and  $\mathbf{T}$  schemes are not stable for any stepsize. However, in this case  $\mathbf{H}$  is stable and allows a nonzero  $\tau_H$ .*

The proof of this theorem requires that we model the resolvents of  $\mathbf{H}$  and  $\mathbf{T}$ , as Chebyshev polynomials, taking advantage of optimal value results for Chebyshev polynomials from Markoff [15]. Our approach is similar to that for the symmetric super-time-stepping scheme of Alexiades, Amiez, and Gremaud[1].

Our paper is divided as follows. In section 2 we derive the stability conditions for **G**. In sections 3 and 4 we derive the stability conditions for the super-time-stepping scheme **T** and **H**, respectively. In section 5 we test the convergence of the stable scheme **H**. We include for comparison an alternate version of **H**, labeled **K**, which has an  $n + 1$  step predictor-corrector scheme combined with super-time-stepping. This scheme improves on **H** for matrices **M** of specific sizes. These results are included in section 6.

**2. Stability conditions for G.** We require the symmetric part of **G** to be positive definite. We will require this condition to be satisfied for all the schemes unless specifically mentioned.

To study the CFL condition for the **G** scheme, we need to find the eigenvalues and spectral radius of **G**. To this end we rely on the Bendixson-Hirsch-Toeplitz theorem [6, 10, 19] (an English version of the theorem may be found in [16]).

**THEOREM 2.1.** (*Bendixson-Hirsch-Toeplitz*) *Let  $\mathbf{A} = \mathbf{B} + i\mathbf{C}$ , where  $\mathbf{B} = (A + A^H)/2$  and  $\mathbf{C} = (A - A^H)/2i$  are Hermitian. Let  $F(\mathbf{B}) = [\lambda_n(\mathbf{B}), \lambda_1(\mathbf{B})]$  and  $F(\mathbf{C}) = [\lambda_n(\mathbf{C}), \lambda_1(\mathbf{C})]$ . Then the eigenvalues of  $\mathbf{A}$  lie in the region  $F(\mathbf{B}) \times F(\mathbf{C})$  in the complex plane.*

The proof follows by letting  $x$  be an eigenvector of **A** of norm one and computing  $x^H \mathbf{A} x$ .

Let  $\mathbf{P}' = \mathbf{I} - \tau_G \mathbf{P}$ . If  $\mathbf{x}$  is an eigenvector of **P** then

$$(\mathbf{I} - \tau_G \mathbf{P})\mathbf{x} = (1 - \tau_G \lambda_P)\mathbf{x}.$$

Therefore by the Bendixson-Hirsch-Toeplitz theorem, the eigenvalues of  $\mathbf{P}'$  lie in the interval  $[1 - \tau_G \lambda_1(\mathbf{P}), 1 - \tau_G \lambda_n(\mathbf{P})]$ .

**LEMMA 2.2.** *Let  $\mathbf{G} = \mathbf{I} - \tau_G \mathbf{P} - \tau_G \mathbf{S}$ . Then the eigenvalues of **G** lie in the region  $[1 - \lambda_1(\mathbf{P}), 1 - \lambda_n(\mathbf{P})] \times [\tau_G |\lambda(\mathbf{S})|_n, \tau_G |\lambda(\mathbf{S})|_1]$ .*

*Proof.* We define  $\mathbf{A} = \mathbf{G}$ , then  $\mathbf{B} = (\mathbf{A} + \mathbf{A}^H)/2 = \mathbf{P}'$ ,  $\mathbf{C} = (\mathbf{A} - \mathbf{A}^H)/2i = -i\tau_G \mathbf{S}$ . Then  $F(\mathbf{B}) = [1 - \lambda_1(\mathbf{P}), 1 - \lambda_n(\mathbf{P})]$  and  $F(\mathbf{C}) = [\tau_G |\lambda(\mathbf{S})|_n, \tau_G |\lambda(\mathbf{S})|_1]$ . Then by the Bendixson-Hirsch-Toeplitz theorem the eigenvalues of **G** lie in the region  $F(\mathbf{B}) \times F(\mathbf{C})$  in the complex plane.  $\square$

Our stability results for the method **G** are summarized in the following theorem.

**THEOREM 2.3.** *The stepsize  $\tau_G$  for a stable scheme **G** is bounded above by*

$$\tau_G < \bar{\tau}_G = \frac{2\lambda_P}{\lambda_P^2 + |\lambda_S|^2},$$

where  $\lambda_P \in [\lambda_n(\mathbf{P}), \lambda_1(\mathbf{P})]$  and  $|\lambda_S| \in [|\lambda(\mathbf{S})|_n, |\lambda(\mathbf{S})|_1]$ .

*Proof.*

The  $L_2$  norm of any matrix **A** is  $\|\mathbf{A}\|_2 = \sqrt{\lambda_1(\mathbf{A}^H \mathbf{A})}$ . Therefore, by definition  $\rho(\mathbf{A}^H \mathbf{A}) = \|\mathbf{A}\|_2^2$  for all matrices **A**. We also have  $\rho(\mathbf{A}) \leq \|\mathbf{A}\|_2$  for any matrix **A**. Equality is attained when **A** is normal.

We use the notation used in the Bendixson-Hirsch-Toeplitz theorem and assume that  $\mathbf{x}$  is an eigenvector of **G**, and for  $\mathbf{B} = \frac{1}{2}(\mathbf{G} + (\mathbf{G})^H)$  and  $\mathbf{C} = \frac{1}{2i}(\mathbf{G} - (\mathbf{G})^H)$

$$\mathbf{x}^H \mathbf{B} \mathbf{x} = \frac{1}{2} \mathbf{x}^H (\mathbf{G} + (\mathbf{G})^H) \mathbf{x} = \frac{1}{2} (\mathbf{x}^H \mathbf{G} \mathbf{x} + (\mathbf{G} \mathbf{x})^H \mathbf{x}) = \lambda_B \mathbf{x}^H \mathbf{x},$$

where  $\lambda_B = \lambda_G + \lambda_{G^H}$ . Similarly,

$$\mathbf{x}^H i \mathbf{C} \mathbf{x} = \frac{1}{2i} \mathbf{x}^H i (\mathbf{G} - (\mathbf{G})^H) \mathbf{x} = \frac{1}{2} (\mathbf{x}^H \mathbf{G} \mathbf{x} - (\mathbf{G} \mathbf{x})^H \mathbf{x}) = i \lambda_C \mathbf{x}^H \mathbf{x},$$

where  $\lambda_C = -i(\lambda_G - \lambda_{G^H})$ . We can then define  $\lambda_G = \lambda_B + i\lambda_C$  and  $\lambda_{G^H} = \lambda_B - i\lambda_C$ . Then,

$$\begin{aligned}\mathbf{x}^H (\mathbf{G})^H \mathbf{G} \mathbf{x} &= \lambda_{G^H} \lambda_G \mathbf{x}^H \mathbf{x} = (\lambda_B + i\lambda_C) (\lambda_B - i\lambda_C) \mathbf{x}^H \mathbf{x} \\ &= (\lambda_B^2 + \lambda_C^2) \mathbf{x}^H \mathbf{x}.\end{aligned}$$

Thus the  $L_2$  norm of  $(\mathbf{G})^H \mathbf{G}$  is

$$\|(\mathbf{G})^H \mathbf{G}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|(\mathbf{G}\mathbf{x})^H (\mathbf{G}\mathbf{x})\|_2 = \max |\lambda_B^2 + \lambda_C^2|.$$

Since we have defined  $\mathbf{B} = \mathbf{I} - \tau\mathbf{P}$  and  $\mathbf{C} = -i\tau\mathbf{S}$ ,

$$\begin{aligned}\|(\mathbf{G})^H \mathbf{G}\|_2 &= \max |(1 - \tau\lambda_P)^2 - \tau^2\lambda_S^2| \\ &= \max(1 - \tau\lambda_P)^2 + \max(\tau^2|\lambda_S|^2)\end{aligned}$$

So we have

$$\rho(\mathbf{G}) \leq \sqrt{\max(1 - \tau_G\lambda_P)^2 + \max(\tau_G^2|\lambda_S|^2)}.$$

For a stable numerical scheme, we require  $\rho(\mathbf{G}) < 1$ . We enforce a stronger condition,

$$\begin{aligned}|1 - \tau_G\lambda_P|^2 + |\tau_G\lambda_S|^2 &< 1 \\ \text{or } -2\lambda_P\tau_G + \tau_G^2[\lambda_P^2 + |\lambda_S|^2] &< 0.\end{aligned}$$

Equation (2.1) simplifies to the CFL stability condition,

$$\tau_G < \bar{\tau}_G = \frac{2\lambda_P}{\lambda_P^2 + |\lambda_S|^2}. \quad (2.1)$$

□

**3. Super-Time-Stepping  $\mathbf{T}$ .** The cost savings in the super-time-stepping scheme comes from imposing the CFL condition on final step, not each intermediate step. The final step  $\tau_T$  equals  $m$  intermediary time steps. We seek to make  $\tau_T$  as large as possible. The stability proof for  $\mathbf{T}$  relies on Chebyshev polynomial relationships in combination with our results for  $\mathbf{G}$ .

**THEOREM 3.1.** *The stepsize  $\tau_T$  for a stable scheme  $\mathbf{T}$  is bounded above by*

$$\tau_T \leq \frac{\bar{\tau}_G m}{2} \left( 1 + \sqrt{1 + \chi^2} \left[ \frac{-1}{\pi m(m-1)} + \frac{2}{\pi} \ln \left( \frac{(2m-1)^2}{m-1} \right) + \frac{2\gamma}{\pi} \right] + \frac{\pi}{8\sqrt{1 + \chi^2}} \right),$$

where  $\gamma$  is the Euler-Mascheroni constant and  $\chi = |\lambda_S|/\lambda_P$ . In the limit that  $\lambda_S$  is zero, the CFL condition for  $\mathbf{T}$  becomes

$$\tau_T < \frac{2m}{\lambda_P}.$$

In the limit that  $\lambda_P$  is zero, there is no real  $\tau_T$  that satisfies the CFL stability condition.

*Proof.* In the case that  $|\lambda_S| = 0$ , the scheme  $\mathbf{T}$  regresses to the super-time-stepping scheme with symmetric  $\mathbf{G}$  and the results are as in [1].

For the general  $\mathbf{T}$  scheme with  $\lambda_P > 0$  and  $|\lambda_S| \neq 0$ ,

$$\rho(\mathbf{T}) \leq \prod_{k=1}^m \sqrt{(1 - \tau_k \lambda_P)^2 + \tau_k^2 |\lambda_S|^2}. \quad (3.1)$$

We define  $\lambda = \lambda_P$ . Using  $\bar{\tau}_G$  from Theorem 2.3 we have  $|\lambda_S|_1^2 = 2\lambda_n(\mathbf{P})/\bar{\tau}_G - \lambda_1(\mathbf{P})^2$ . Therefore

$$|\lambda_S|^2 \leq |\lambda_S|_1^2 \leq (2\lambda_n(\mathbf{P})/\bar{\tau}_G - \lambda_1(\mathbf{P})^2) \leq \lambda(2/\bar{\tau}_G - \lambda).$$

Hence, our stability condition is satisfied if

$$\rho(\mathbf{T}) \leq \prod_{k=1}^m \sqrt{(1 - \tau_k \lambda)^2 + \tau_k^2 \lambda \left(\frac{2}{\bar{\tau}_G} - \lambda\right)}.$$

We may satisfy  $\rho(\mathbf{T}) < 1$  by requiring

$$\prod_{k=1}^m (1 - \tau_k \lambda)^2 + \tau_k^2 \lambda \left(\frac{2}{\bar{\tau}_G} - \lambda\right) < 1 \quad \forall \lambda \in [\lambda_n, \lambda_1]. \quad (3.2)$$

For strong stability we replace Eq. (3.2) by

$$\prod_{k=1}^m \left(1 + 2\tau_k \lambda \left(\frac{\tau_k}{\bar{\tau}_G} - 1\right)\right) \leq L_m \quad \forall \lambda \in [\mu, \lambda_1], \quad (3.3)$$

where  $\mu \in (0, \lambda_n]$  and  $0 < L_m < 1$ . We can define a polynomial

$$R_T(\lambda) = \sum_{k=0}^m \delta_k \lambda^k,$$

where  $\delta_0 = 1$  and  $\delta_k = 2\tau_k (\tau_k/\bar{\tau}_G - 1)$  for  $k > 0$ . We can phrase our problem as

$$\begin{aligned} |R_T(\lambda_P)| &\leq L_m & \forall \lambda \in [\mu, \lambda_1] & \text{(stability)} \\ |R_T'(0)| &= \left| \sum_{k=1}^m \delta_k \right| & \text{maximal} & \text{(optimality)} \end{aligned}$$

By the optimality properties of Chebyshev polynomials  $T_m(\cdot)$  of degree  $m$  we have if

$$L_m = \left[ T_m \left( \frac{\lambda_1 + \mu}{\lambda_1 - \mu} \right) \right]^{-1},$$

where  $L_m$  may be chosen arbitrary close to one by choosing  $\mu$  sufficiently small. The optimal value  $\tau_T$  can be derived from

$$R_T(\lambda) = \frac{T_m \left( \frac{\lambda_1 + \mu - 2\lambda}{\lambda_1 - \mu} \right)}{T_m \left( \frac{\lambda_1 + \mu}{\lambda_1 - \mu} \right)},$$

which gives

$$\delta_k = \frac{2}{\lambda_1} \left[ (\nu - 1) \cos \left( \frac{(2k-1)\pi}{2m} \right) + 1 + \nu \right]^{-1}. \quad (3.4)$$



In Eq. 3.4,  $\nu$  is a practical implementation of the damping factor  $L_m$  and can be manipulated directly to modify the time-step size. In the limit that  $\nu$  approaches 0, our definition of  $\delta_k$  gives

$$\frac{2}{\lambda_1 \phi_k} = \frac{2\tau_k^2}{\bar{\tau}_G} - 2\tau_k, \quad (3.5)$$

where

$$\phi_k = 1 - \cos\left(\frac{(2k-1)\pi}{2m}\right).$$

We may approximate  $\phi_k$  by

$$\phi_k \approx 1 - \left[1 - \frac{1}{2} \left(\frac{(2k-1)\pi}{2m}\right)^2\right] \approx \frac{(2k-1)^2 \pi^2}{8m^2}.$$

Then

$$\begin{aligned} \tau_T &\leq \sum_{k=1}^m \tau_k \leq \frac{\bar{\tau}_G m}{2} + \frac{1}{2} \sum_{k=1}^m \sqrt{\bar{\tau}_G^2 + \frac{32m^2 \bar{\tau}_G}{\lambda_1 (2k-1)^2 \pi^2}}, \\ &\leq \frac{\bar{\tau}_G m}{2} \left(1 + \frac{4}{\pi} \sum_{k=1}^m \sqrt{\frac{2}{\lambda_1 \bar{\tau}_G} \frac{1}{2k-1} \sqrt{1 + \frac{\lambda_1 (2k-1)^2 \pi^2 \bar{\tau}_G}{32m^2}}}\right). \end{aligned}$$

Note that

$$\lambda_1 \bar{\tau}_G = \frac{2\lambda_1^2}{\lambda_1^2 + |\lambda_S|^2} = \frac{2}{1 + \chi^2},$$

where  $\chi = |\lambda_S|/\lambda_1$ . Then

$$\tau_T \leq \frac{\bar{\tau}_G m}{2} \left(1 + \frac{4}{\pi} \sum_{k=1}^m \sqrt{1 + \chi^2} \left[\frac{1}{2k-1} + \frac{\pi^2 (2k-1)}{32m^2 (1 + \chi^2)}\right]\right).$$

We note that

$$\sum_{k=1}^m \frac{1}{2k-1} = \frac{1}{2} \gamma + \ln(2) + \psi^{(0)}(m + 1/2), \quad (3.6)$$

where  $\gamma$  represents the Euler-Mascheroni constant ( $\gamma \approx 0.5772$ ) and  $\psi^{(0)}(m + 1/2)$  is the digamma function

$$\psi^{(0)}(x) = \frac{d}{dx} \ln(\Gamma(x)).$$

Using the relationships for the digamma function

$$\begin{aligned} \psi^{(0)}(2m) &= \frac{1}{2} \psi^{(0)}(m + 1/2) + \frac{1}{2} \psi^{(0)}(m) + \ln 2, \\ \text{and } \psi^{(0)}(m) &= -\gamma + H_{m-1}, \end{aligned}$$

where  $H_m$  represents the  $m$ th partial sum of the harmonic series

$$H_m = \sum_{k=1}^m \frac{1}{k},$$

we may write eq. (3.6) as

$$\sum_{k=1}^m \frac{1}{2k-1} = H_{2m-1} - \frac{1}{2}H_{m-1}.$$

We have bounds on  $H_{2m-1}$  and  $H_{m-1}$  from [9, 22]:

$$\frac{1}{2(m+1)} < H_m - \ln m - \gamma < \frac{1}{2m}.$$

Hence we can bound our sum

$$\frac{-1}{4m(m-1)} + \ln\left(\frac{2m-1}{\sqrt{m-1}}\right) + \frac{1}{2}\gamma < \sum_{k=1}^m \frac{1}{2k-1} < \frac{1}{4m(2m-1)} + \ln\left(\frac{2m-1}{\sqrt{m-1}}\right) + \frac{1}{2}\gamma$$

and thus we can put the stronger condition on  $\tau_T$

$$\tau_T \leq \frac{\bar{\tau}_G m}{2} \left( 1 + \sqrt{1 + \chi^2} \left[ \frac{-1}{\pi m(m-1)} + \frac{2}{\pi} \ln\left(\frac{(2m-1)^2}{m-1}\right) + \frac{2\gamma}{\pi} \right] + \frac{\pi}{8\sqrt{1 + \chi^2}} \right). \quad (3.7)$$

In the case that  $\lambda_P = 0$ ,

$$\mathbf{T} = \prod_{k=1}^n (\mathbf{I} - \tau_k \mathbf{S}).$$

We note that

$$\|\mathbf{I} - \tau_k \mathbf{S}\|_2 = \max \sqrt{1 + \tau_k^2 |\lambda_S|^2}.$$

Thus, there is no real  $\tau_k$  that will satisfy the stability condition.  $\square$

**4. Stability conditions for  $\mathbf{H}$ .** In the previous section we noted that for systems that are not skew-symmetric dominated, increasing the number of intermediate time-steps will give improved convergence rates over the  $\mathbf{G}$  scheme. Hence, in this section, we incorporate the predictor-corrector skew-symmetric dominated schemes with a super-time-stepping scheme.

**THEOREM 4.1.** *The stepsize  $\tau_H$  for a stable scheme  $\mathbf{H}$  for a  $n \times n$  matrix  $\mathbf{G} = \mathbf{P} + \mathbf{S}$  and  $\lambda_1 = \lambda_1(\mathbf{P}) > 0$  is bounded above by*

$$\tau_H \leq \frac{\bar{\tau}_G}{2} \left( 1 + \sqrt{1 + \frac{4}{\bar{\tau}_G \lambda_1 (1 - \cos(\pi/12))}} \right),$$

when  $m = 1$ . For  $m > 1$ ,

$$\tau_H \leq \frac{\bar{\tau}_G m}{2} \left( 1 + \sqrt{1 + \chi^2} \left[ \frac{-6}{\pi m(m-1)} + \frac{12}{\pi} \ln\left(\frac{(2m-1)^2}{m-1}\right) + \frac{12\gamma}{\pi} \right] + \frac{\pi}{48\sqrt{1 + \chi^2}} \right), \quad (4.1)$$

where  $\chi = |\lambda_S|/\lambda_P$ .

In the limit that  $\lambda_S$  is zero,  $\mathbf{H}$  is identical to  $\mathbf{T}$ . In the limit that  $\lambda_P$  is zero, the CFL conditions become

$$\tau_H < \frac{1}{|\lambda_S|} \quad \text{for } m = 1,$$

and for  $m > 1$ ,

$$\tau_H < \frac{2m\sqrt{2}}{|\lambda_S|^2\pi} \left( \frac{-1}{4m(m-1)} + \ln \left( \frac{2m-1}{\sqrt{m-1}} \right) + \frac{\gamma}{2} \right). \quad (4.2)$$

*Proof.*

For  $\mathbf{H}$  and  $m = 1$ ,

$$\begin{aligned} \|\mathbf{I} - \tau\mathbf{S} + \tau^2\mathbf{S}^2\|_2 &= \|(\mathbf{I} + \tau^2\mathbf{S}^2) - \tau\mathbf{S}\|_2, \\ &= \max \sqrt{(1 - \tau^2|\lambda_S|^2)^2 + \tau^2|\lambda_S|^2}, \\ &= \sqrt{1 + \tau^4|\lambda_S|_n^4 + \tau^2(|\lambda_S|_1^2 - 2|\lambda_S|_n^2)}, \\ &\leq \sqrt{1 + \tau^4|\lambda_S|_n^4 + \tau^2|\lambda_S|_1^2}. \end{aligned}$$

Therefore

$$\rho(\mathbf{H}) \leq \left| (1 - \tau\lambda_P) \sqrt{1 + \tau^2|\lambda_S|_1^2 + \tau^4|\lambda_S|_1^4} \right|. \quad (4.3)$$

For strong stability we set

$$\left| (1 - \tau\lambda)^2 \left( 1 + \tau^2\lambda \left( \frac{2}{\bar{\tau}_G} - \lambda \right) + \tau^4\lambda^2 \left( \frac{2}{\bar{\tau}_G} - \lambda \right)^2 \right) \right| \leq L_6 \quad \forall \lambda \in [\mu, \lambda_1], \quad (4.4)$$

where  $\mu \in (0, \lambda_n]$  and  $0 < L_6 < 1$ . We can define a polynomial

$$R_H(\lambda) = \sum_{i=0}^6 \alpha_i \lambda^i,$$

where  $\alpha_0 = 1$  and  $\alpha_1 = -2\tau + 2\tau^2/\bar{\tau}_G$ . We can phrase our problem as thus

$$\begin{aligned} |R_H(\lambda_P)| &\leq L_6 \quad \forall \lambda \in [\mu, \lambda_1] \quad (\text{stability}) \\ |R'_H(0)| &= |\alpha_1| \quad \text{maximal} \quad (\text{optimality}) \end{aligned}$$

By the optimality properties of Chebyshev polynomials  $T_6(\cdot)$  of degree six we have if

$$L_6 = \left[ T_6 \left( \frac{\lambda_1 + \mu}{\lambda_1 - \mu} \right) \right]^{-1},$$

where  $L_6$  may be chosen arbitrary close to one by choosing  $\mu$  sufficiently small. The optimal value  $\tau_H$  can be derived from

$$\alpha_1 = \frac{2}{\lambda_1} \left[ (\nu - 1) \cos \left( \frac{\pi}{12} \right) + 1 + \nu \right]^{-1} = \frac{2\tau^2}{\bar{\tau}_G} - 2\tau. \quad (4.5)$$

Our optimal  $\tau_H$  is then, letting  $\nu \rightarrow 0$ ,

$$\tau_H = \frac{\bar{\tau}_G}{2} \left( 1 + \sqrt{1 + \frac{4}{\bar{\tau}_G \lambda_1 (1 - \cos(\pi/12))}} \right). \quad (4.6)$$

Following similar steps for  $\tau_J$  we expand for  $\chi = |\lambda_S|/\lambda_P > 1$  and find

$$\tau_H \approx \left( \frac{1}{2} + \frac{12|\lambda_S|}{\pi\lambda_P} \right) \bar{\tau}_G \quad \text{for } m = 1. \quad (4.7)$$

Building on our previous proof for  $m = 1$ , then for  $m > 1$  we have

$$\rho(\mathbf{H}) \leq \left| \prod_{k=1}^m (1 - \tau_k \lambda_P) \sqrt{1 + \tau_k^2 |\lambda_S|^2 + \tau_k^4 |\lambda_S|^4} \right|. \quad (4.8)$$

For strong stability we set

$$\left| \prod_{k=1}^m (1 - \tau_k \lambda)^2 \left( 1 + \tau_k^2 \lambda \left( \frac{2}{\bar{\tau}_G} - \lambda \right) + \tau_k^4 \lambda^2 \left( \frac{2}{\bar{\tau}_G} - \lambda \right)^2 \right) \right| \leq L_{6m} \quad \forall \lambda \in [\mu, \lambda_1], \quad (4.9)$$

where  $\mu \in (0, \lambda_n]$  and  $0 < L_{6m} < 1$ . We can define the polynomial

$$R_H(\lambda) = \sum_{i=0}^6 \sum_{k=1}^m \alpha_{i,k} \lambda^{i \cdot k},$$

where  $\sum_{k=1}^m \alpha_{0,k} = 1$  and  $\alpha_{1,k} = 2\tau_k(\tau_k/\bar{\tau}_G - 1)$ . We can phrase our problems as

$$\begin{aligned} |R_H(\lambda_P)| &\leq L_{6m} & \forall \lambda \in [\mu, \lambda_1] & \text{ (stability)} \\ |R'_H(0)| &= \left| \sum_{k=1}^m \alpha_{1,k} \right| & \text{maximal} & \text{ (optimality)} \end{aligned}$$

The optimal value  $\tau_H$  can be derived from

$$\frac{2}{\lambda_1 u_k} = \frac{2\tau_k^2}{\bar{\tau}_G} - 2\tau_k, \quad (4.10)$$

where

$$u_k = 1 - \cos\left(\frac{(2k-1)\pi}{12m}\right) \approx \frac{(2k-1)^2 \pi^2}{288m^2}.$$

The rest of the proof for the  $m > 1$  case follows by the same logic for  $\mathbf{T}$ , we find  $\mathbf{H}$  satisfies Eqs (4.1).

Next we tackle the case when  $\lambda_P = 0$ .

If  $\lambda_P = 0$  and  $m = 1$ ,  $\lambda(\mathbf{H}_{\text{sym}}) = \lambda(\mathbf{I} + \tau_H^2 \mathbf{S}^2)$  and  $\lambda(\mathbf{H}_{\text{skew}}) = \lambda(-\tau_H \mathbf{S})$ . So  $\lambda(\mathbf{H}_{\text{sym}}) = 1 - \tau_H^2 |\lambda_S|^2$  and  $\lambda(\mathbf{H}_{\text{skew}}) = -\tau_H \lambda_S$ , where  $|\lambda_S| \in [|\lambda(\mathbf{S})|_n, |\lambda(\mathbf{S})|_1]$ . For numerical stability,  $\rho(\mathbf{H}) < 1$ , which may be enforced by requiring  $(1 - \tau_H^2 |\lambda_S|^2)^2 + \tau_H^2 |\lambda_S|^2 < 1$ . This inequality is satisfied when  $\tau_H < 1/|\lambda_S|$ .

If  $\lambda_P = 0$  and  $m > 1$ , we have

$$\rho(\mathbf{H}) \leq \prod_{k=1}^m \left| \sqrt{1 - \tau_k^2 |\lambda_S|^2 + \tau_k^4 |\lambda_S|^4} \right|.$$

Thus by inspection we can see that for  $m = 1$ ,  $\tau_H < \frac{1}{|\lambda_S|}$ .

For  $m > 1$ , we define a polynomial

$$R_H = \sum_{i=0}^2 \sum_{k=1}^m (-1)^{i+1} \alpha_{i,k} \lambda^{i \cdot k},$$

where  $\lambda = |\lambda_S|^2$ ,  $\alpha_{0,k} = 1$ , and  $\alpha_{1,k} = \tau_k^2$ .

As before we find that we will optimize our problem when  $2/(\lambda\phi_k) = \alpha_{1,k}$ , where  $\phi_k$  is defined as in section 3. This results in the inequality,

$$\sum_{k=1}^m \tau_k < \frac{2m\sqrt{2}}{|\lambda_S|\pi} \sum_{k=1}^m \frac{1}{2k-1}.$$

Substituting in for  $\sum_{k=1}^m \frac{1}{2k-1}$  from section 3, we have

$$\tau_H < \frac{2m\sqrt{2}}{\pi|\lambda_S|} \left[ \frac{-1}{4m(m-1)} + \ln \left( \frac{2m-1}{\sqrt{m-1}} \right) + \frac{\gamma}{2} \right].$$

□

This concludes our proof Theorem 1.1, which summarizes the relationship of the sizes needed for the **G** and **H** schemes.

**5. Convergence.** Let us assume that the **G** scheme converges. We know that if  $\rho(\mathbf{G}) < 1$ , then  $\rho(\mathbf{H}) < 1$  and  $\rho(\mathbf{T}) < 1$  and then all of these schemes converge. In this section we show that these schemes converge to the **G** scheme.

We begin with convergence of the scheme **H** to **G**. We define  $\mathbf{x}^{\ell+1}$  as

$$\mathbf{x}^{\ell+1} = \mathbf{G}\mathbf{x}^\ell = [\mathbf{G}]^{\ell+1}\mathbf{x}^0.$$

Let us define  $\mathbf{y}^{\ell+1} = \mathbf{H}\mathbf{y}^\ell$  with and  $\mathbf{y}^0 = \mathbf{x}^0$ .

LEMMA 5.1. *The **H** scheme converges to the **G** scheme. That is for  $N$  large,*

$$\frac{\|\mathbf{y}^{\ell+1} - \mathbf{x}^{\ell+1}\|_2}{\|\mathbf{x}^0\|_2} \rightarrow 0,$$

as  $\ell \rightarrow N$ .

*Proof.* Let us define  $\mathbf{z}^{\ell+1} = \mathbf{T}\mathbf{z}^\ell$  with and  $\mathbf{z}^0 = \mathbf{x}^0$ . Also let

$$\mathbf{H}_{1,k} = (\mathbf{I} - \tau_k \mathbf{P}) (\mathbf{I} - \tau_k \mathbf{S} + \tau_k^2 \mathbf{S}^2) \quad \text{and} \quad \mathbf{G}_k = \mathbf{I} - \tau_k \mathbf{P} - \tau_k \mathbf{S}.$$

We note that

$$\mathbf{y}^1 = \mathbf{H}\mathbf{x}^0 = (\mathbf{H} - \mathbf{T})\mathbf{x}^0 + \mathbf{T}\mathbf{x}^0 = (\mathbf{H} - \mathbf{T})\mathbf{x}^0 + \mathbf{z}^1.$$

We can prove by induction that

$$\begin{aligned} \mathbf{y}^{\ell+1} &= \mathbf{z}^{\ell+1} + \sum_{j=0}^{\ell} (\mathbf{H})^{\ell-j} (\mathbf{H} - \mathbf{T}) \mathbf{z}^j, \\ &= \mathbf{z}^{\ell+1} + \sum_{j=0}^{\ell} (\mathbf{H})^{\ell-j} (\mathbf{H} - \mathbf{T}) (\mathbf{T})^j \mathbf{x}^0. \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\|\mathbf{y}^{\ell+1} - \mathbf{z}^{\ell+1}\|}{\|\mathbf{x}^0\|} &= \left\| \sum_{j=0}^{\ell} (\mathbf{H})^{\ell-j} (\mathbf{H} - \mathbf{T}) (\mathbf{T})^j \right\|, \\ &\leq \sum_{i=0}^{\ell} |\lambda_H|_1^{\ell-j} \|\mathbf{H} - \mathbf{T}\| |\lambda_T|_1^j. \end{aligned}$$

Since

$$\begin{aligned} \mathbf{H} - \mathbf{T} &= \prod_{k=1}^m \mathbf{H}_{1,k} - \prod_{k=1}^m \mathbf{G}_k, \\ &= - \prod_{k=1}^m \mathbf{G}_k \sum_{n \neq k, n=1}^m \tau_n^2 (\mathbf{P}\mathbf{S} - \mathbf{S}^2) + O(\tau_n^3), \end{aligned}$$

then

$$|\lambda(\mathbf{H} - \mathbf{T})|_1 \leq \tau_G^2 (m-1) (\lambda_1(\mathbf{P})|\lambda(\mathbf{S})|_1 + |\lambda(\mathbf{S})|_1^2) \prod_{k=1}^m |\lambda(\mathbf{G}_k)|_1,$$

where  $\phi = \lambda_1(\mathbf{P})/|\lambda(\mathbf{S})|_1$ . Then

$$\frac{\|\mathbf{y}^{\ell+1} - \mathbf{z}^{\ell+1}\|}{\|\mathbf{x}^0\|} \leq \frac{4(m-1)(\phi+1)}{(\phi^2+1)^2} \prod_{k=1}^m |\lambda(\mathbf{G}_k)|_1 \sum_{i=0}^{\ell} |\lambda_H|_1^{\ell-j} |\lambda_T|_1^j.$$

For a skew dominated system,  $\phi_1 < 1$ . Note that  $|\lambda_H|_1 < 1$  and  $|\lambda_T| < 1$  if the  $\mathbf{T}$  and  $\mathbf{H}$  schemes converge. For  $N$  large,

$$\lim_{\ell \rightarrow N} \frac{\|\mathbf{y}^{\ell+1} - \mathbf{z}^{\ell+1}\|}{\|\mathbf{x}^0\|} \rightarrow 0,$$

thereby proving that the  $\mathbf{H}$  converges to the  $\mathbf{T}$  scheme.

We also have

$$\frac{\|\mathbf{y}^{\ell+1} - \mathbf{x}^{\ell+1}\|}{\|\mathbf{x}^0\|} \leq \frac{\|\mathbf{y}^{\ell+1} - \mathbf{z}^{\ell+1}\|}{\|\mathbf{x}^0\|} + \frac{\|\mathbf{z}^{\ell+1} - \mathbf{x}^{\ell+1}\|}{\|\mathbf{x}^0\|}.$$

Now

$$\begin{aligned} \frac{\|\mathbf{z}^{\ell+1} - \mathbf{x}^{\ell+1}\|}{\|\mathbf{x}^0\|} &= \|(\mathbf{T})^{\ell} - (\mathbf{G})^{\ell}\|, \\ &\leq |\lambda_T|_1^{\ell} + |\lambda_G|_1^{\ell}. \end{aligned}$$

As  $\ell \rightarrow \infty$ , since  $|\lambda_T|_1 < 1$  and  $|\lambda_G|_1 < 1$ , then

$$\frac{\|\mathbf{z}^{\ell+1} - \mathbf{x}^{\ell+1}\|}{\|\mathbf{x}^0\|} \rightarrow 0.$$

□

LEMMA 5.2. *The  $\mathbf{H}$  scheme has a second order convergence.*

*Proof.* We know that  $\exp(-\mathbf{M}t)$  is the true solution to

$$\frac{d\mathbf{y}}{dt} = -\mathbf{M}\mathbf{y}.$$

We say that  $\mathbf{y}(\ell\tau)$  is the exact solution at the  $\ell$ th step. First we look at the error between the  $\mathbf{H}$  scheme solution and the exact solution at the  $(\ell + 1)$ th step.

$$\begin{aligned} \|E_H(\ell + 1)\| &= \|\mathbf{y}^{\ell+1} - \mathbf{y}((\ell + 1)\tau)\|, \\ &= \|\mathbf{H}\mathbf{y}^\ell - e^{-\mathbf{M}\tau}\mathbf{y}(\ell\tau)\|, \\ &\leq \|\mathbf{H}(\mathbf{y}^\ell - \mathbf{y}(\ell\tau)) + (\mathbf{H} - e^{-\mathbf{M}\tau})\mathbf{y}(\ell\tau)\|, \\ &\leq \|\mathbf{H}\| \|E_H(\ell)\| + \|\mathbf{H} - e^{-\mathbf{M}\tau}\| \|\mathbf{y}(\ell\tau)\|. \end{aligned}$$

We have

$$\|\mathbf{y}(\ell\tau)\| = \|e^{-\mathbf{M}\ell\tau}\mathbf{y}(0)\| = \|e^{-\mathbf{P}\ell\tau}\| \|e^{-\mathbf{S}\ell\tau}\| \|\mathbf{y}(0)\|.$$

Since  $\mathbf{P}$  is positive definite and  $\mathbf{S}$  is skew,  $\|e^{-\mathbf{P}\ell\tau}\| < 1$  and  $\|e^{-\mathbf{S}\ell\tau}\| \leq 1$ , therefore  $\|\mathbf{y}(\ell\tau)\| < \|\mathbf{y}(0)\|$ . Then we have

$$\begin{aligned} \|E_H(\ell + 1)\| &\leq |\lambda_H| \|E_H(\ell)\| + \|\mathbf{H} - e^{-\mathbf{M}\tau}\| \|\mathbf{y}(0)\|. \\ &\leq (\ell + 1) \|\mathbf{H} - e^{-\mathbf{M}\tau}\| \|\mathbf{y}(0)\|. \end{aligned} \quad (5.1)$$

Since  $\mathbf{P}$  and  $\mathbf{S}$  do not commute,

$$e^{-\mathbf{M}\tau} = \mathbf{I} - (\mathbf{P} + \mathbf{S})\tau + (\mathbf{P} + \mathbf{S})(\mathbf{P} + \mathbf{S})\tau^2 + O(\tau^3).$$

Expanding and only keeping the lowest order terms,

$$\begin{aligned} \|\mathbf{H} - e^{-\mathbf{M}\tau}\| &= \left\| \prod_{k=1}^m (\mathbf{I} - \tau_k \mathbf{P}) \times (\mathbf{I} - \tau_k \mathbf{S} + \tau_k^2 \mathbf{S}^2) \right. \\ &\quad \left. - \mathbf{I} - (\mathbf{P} + \mathbf{S})\tau + (\mathbf{P} + \mathbf{S})(\mathbf{P} + \mathbf{S})\tau^2 + O(\tau^3) \right\|, \\ &= \left\| \left( \mathbf{I} - (\mathbf{P} + \mathbf{S}) \sum_{k=1}^m \tau_k + (\mathbf{P}^2 + \mathbf{P}\mathbf{S} + \mathbf{S}\mathbf{P} + \mathbf{S}^2) \sum_{k=1}^m \tau_k^2 + O(\tau_k^3) \right) \right. \\ &\quad \left. - (\mathbf{I} - (\mathbf{P} + \mathbf{S})\tau + (\mathbf{P} + \mathbf{S})(\mathbf{P} + \mathbf{S})\tau^2) \right\|, \\ &= \|(\mathbf{P}^2 + \mathbf{P}\mathbf{S} + \mathbf{S}\mathbf{P} + \mathbf{S}^2) \left[ \sum_{k=1}^m \tau_k^2 - \tau^2 \right] + O(\tau_k^3)\|, \\ &\leq (\lambda_P + |\lambda_S|)^2 \left( \sum_{k=1}^m \tau_k^2 - \tau^2 \right) + O(\tau_k^3), \\ &\leq 2(\lambda_P + |\lambda_S|)^2 \tau^2 + O(\tau_k^3). \end{aligned}$$

So we have

$$\|E_H(\ell + 1)\| \leq (\ell + 1) (\lambda_P + |\lambda_S|)^2 \tau^2 \|\mathbf{y}(0)\| + O(\tau^3). \quad (5.2)$$

□

**6. An  $n+1$  step predictor-corrector scheme.** We introduce another variant of the predictor-corrector, multiplicative split operator, super-time stepping scheme, **K**. This is a  $n+1$ -step scheme with intermediate time steps of size  $\tau_i$ .

$$\mathbf{K} = \prod_{i=1}^m (\mathbf{I} - \tau_i \mathbf{P}) \times \prod_{j=1}^n (\mathbf{I} - \tau_i \mathbf{S}_j), \quad (6.1)$$

where  $\mathbf{S}_j$  is zero in all rows except the  $j$ th row and that is the  $j$ th row of  $\mathbf{S}$ . From the following theorem we find that if  $n \leq 4$  then  $\tau_H > \tau_K$ ; if  $n = 5$  then  $\tau_H = \tau_K$ ; if  $n \geq 5$ , then  $\tau_H < \tau_K$ .

**THEOREM 6.1.** *The stepsize  $\tau_K$  for a stable scheme  $\mathbf{K}$  for a  $n \times n$  matrix  $\mathbf{G} = \mathbf{P} + \mathbf{S}$  and  $\lambda_1 = \lambda_1(\mathbf{P}) > 0$  is bounded above by*

$$\tau_K \leq \frac{\bar{\tau}_G}{2} \left( 1 + \sqrt{1 + \frac{4}{\bar{\tau}_G \lambda_1 (1 - \cos(\pi/(2(n+2))))}} \right),$$

when  $m = 1$ . For  $m > 1$ ,

$$\begin{aligned} \tau_K \leq \frac{\bar{\tau}_G m}{2} \left( 1 + \sqrt{1 + \chi^2} \left[ \frac{-(n+2)}{\pi(m-1)} + \frac{2m(n+2)}{\pi} \ln \left( \frac{(2m-1)^2}{m-1} \right) \right. \right. \\ \left. \left. + \frac{2m(n+2)\gamma}{\pi} \right] + \frac{\pi}{8m(n+2)\sqrt{1+\chi^2}} \right), \end{aligned} \quad (6.2)$$

where  $\chi = |\lambda_S|/\lambda_P$ .

In the limit that  $\lambda_S$  is zero,  $\mathbf{K}$ , is identical to  $\mathbf{H}$  and  $\mathbf{T}$ .

In the limit that  $\lambda_P$  is zero, the CFL condition becomes

$$\tau_K < \frac{2n\sqrt{2}}{|\lambda_S|\pi}$$

for  $m = 1$  and for  $m > 1$

$$\tau_K < \frac{2mn\sqrt{2}}{\pi|\lambda_S|} \left[ \frac{-1}{4m(m-1)} + \ln \left( \frac{2m-1}{\sqrt{m-1}} \right) + \frac{\gamma}{2} \right].$$

*Proof.* For  $m = 1$ ,

$$\begin{aligned} \prod_{j=1}^n (\mathbf{I} - \tau \mathbf{S}_j) &= \mathbf{I} - \tau \mathbf{S} + \tau^2 \sum_{i,j=1}^n \mathbf{S}_i \mathbf{S}_j + \mathbf{Q}, \\ &= \mathbf{I} - \tau \mathbf{S} - \tau^2 \sum_{i,j=1}^n s_{ij}^2 + \mathbf{Q}, \end{aligned}$$

where  $s_{ij}$  is the element of  $\mathbf{S}$  in the  $i$ th row and  $j$ th column and

$$\mathbf{Q} = O(\tau^2 \sum_{i,j,k=1, i < j < k}^n S_i S_j S_k).$$



Since  $s_{ii} = 0$ , for all  $1 \leq i \leq n$ , we can replace the product by

$$\begin{aligned} \prod_{j=1}^n (\mathbf{I} - \tau \mathbf{S}_j) &= \mathbf{I} - \tau^2 \sum_{i,j=1}^n s_{ij}^2 + \mathbf{Q}, \\ &= (\mathbf{I} - \tau^2 \sum_{i,j=1}^n s_{ij}^2 + \mathbf{Q}_{\text{sym}}) + (-\tau \mathbf{S} + \mathbf{Q}_{\text{skew}}) \end{aligned}$$

and

$$\begin{aligned} \left\| \prod_{j=1}^n (\mathbf{I} - \tau \mathbf{S}_j) \right\|_2 &\leq \left[ (1 - \tau^2 \sum_{i,j=1}^n s_{ij}^2 + \lambda_{\text{sym}}(\mathbf{Q}))^2 + \tau^2 |\lambda_S|^2 + \right. \\ &\quad \left. \tau \lambda (\mathbf{S} \mathbf{Q}_{\text{skew}} + \mathbf{Q}_{\text{skew}} \mathbf{S}) + \lambda (\mathbf{Q}_{\text{skew}})^2 \right]^{1/2}. \end{aligned}$$

We note that

$$\sum_{i,j=1}^n s_{ij}^2 = \|\mathbf{S}\|_F$$

where  $\|\mathbf{S}\|_F$  represents the Frobenius norm of  $\mathbf{S}$ . We have that

$$\|\mathbf{A}\|_2 \leq \|\mathbf{A}\|_F \leq \sqrt{n} \|\mathbf{A}\|_2,$$

for any  $n \times n$  matrix  $\mathbf{A}$ . From our earlier work with the  $\mathbf{H}$  scheme, we know that both the skew and symmetric pieces of  $\lambda(\mathbf{Q})$  and eigenvalues of matrix products including  $\mathbf{Q}$  will be of order  $O(\lambda^3)$  where  $\lambda = \lambda_P$  as before. Therefore we will lump these terms together and

$$\left\| \prod_{j=1}^n (\mathbf{I} - \tau \mathbf{S}_j) \right\|_2 \leq \left[ (1 - \tau^2 |\lambda_S|_n^2)^2 + \tau^2 |\lambda_S|_1^2 + O(\lambda^3) \right]^{1/2}.$$

Hence,

$$\rho(\mathbf{K}) \leq \left| (1 - \tau \lambda) (1 + \tau^2 |\lambda_S|_1^2 + O(\lambda^3))^{1/2} \right|. \quad (6.3)$$

For strong stability we set

$$\left| (1 - \tau \lambda)^2 (1 + \tau^2 \lambda (2/\bar{\tau}_G - \lambda) + O(\lambda^3)) \right| \leq L_{n+2} \quad \forall \lambda \in [\mu, \lambda_1], \quad (6.4)$$

where  $\mu \in (0, \lambda_n]$  and  $0 < L_{n+2} < 1$ . We can define a polynomial

$$R_K(\lambda) = \sum_{j=0}^{n+2} \gamma_j \lambda^j,$$

where  $\gamma_0 = 1$  and  $\gamma_1 = -2\tau + 2\tau^2/\bar{\tau}_G$ . We can phrase our problem as thus

$$\begin{array}{ll} |R_K(\lambda)| \leq L_{n+2} & \forall \lambda \in [\mu, \lambda_1] \quad (\text{stability}) \\ |R'_K(0)| = |\gamma_1| & \text{maximal} \quad (\text{optimality}) \end{array}$$

While this problem looks nearly identical to the strong stability version of the **H** scheme for  $m = 1$  we realize that our equation for  $\tau$  will be different as  $\gamma_1$  has a different definition.

$$\gamma_1 = \frac{2}{\lambda_1} \left[ (\nu - 1) \cos\left(\frac{\pi}{2(n+2)}\right) + 1 + \nu \right]^{-1} = \frac{2\tau^2}{\bar{\tau}_G} - 2\tau. \quad (6.5)$$

Our optimal  $\tau_K$  is then, letting  $\nu \rightarrow 0$ ,

$$\tau_K = \frac{\bar{\tau}_G}{2} \left( 1 + \sqrt{1 + \frac{4}{\bar{\tau}_G \lambda_1 (1 - \cos(\pi/(2(n+2))))}} \right).$$

Following similar steps from our derivation for  $\tau_H$  we expand and find

$$\tau_K \approx \left( \frac{1}{2} + \frac{2(n+2)\sqrt{1+\chi^2}}{\pi} + \frac{\pi}{16(n+2)\sqrt{1+\chi^2}} \right) \bar{\tau}_G, \quad (6.6)$$

where  $\chi = |\lambda_S|/\lambda_P$ .

The steps for the  $m > 1$  case for **K** follow from the  $m = 1$  and  $m > 1$  cases of the **T** and **H** schemes.

When  $\lambda_P = 0$ ,

$$\mathbf{K} = \prod_{k=1}^m \prod_{j=1}^n (\mathbf{I} - \tau_k \mathbf{S}_j).$$

and

$$\rho(\mathbf{K}) \leq \prod_{i=1}^m \left| [(1 - \tau_k^2 |\lambda_S|^2)^2 + \tau_k^2 |\lambda_S|^2 + O(\lambda^3)] \right|. \quad (6.7)$$

For strong stability we set  $|\lambda_S|^2 = \lambda$  and

$$\left| \prod_{k=1}^m (1 - \tau_k^2 \lambda + \tau_k^4 \lambda^2 + O(\lambda^3)) \right| \leq L_{mn} \quad \forall \lambda \in [\mu, \lambda_1] \quad (6.8)$$

So we can define

$$R_K = \sum_{k=1}^m \sum_{j=0}^n (-1)^{j \cdot k} \gamma_{k,j} \lambda^{k \cdot j},$$

where  $\lambda = |\lambda_S|^2$ ,  $\sum_{k=1}^m \gamma_{k,0} = 1$  and  $\gamma_{k,1} = \tau_k^2$ . The problem is optimized when

$$\frac{2}{\lambda w_k} = \gamma_{k,1} \quad \text{where} \quad w_k = 1 - \cos\left(\frac{(2k-1)\pi}{2mn}\right).$$

This results in the inequality,

$$\sum_{k=1}^m \tau_k < \frac{2mn\sqrt{2}}{|\lambda_S|\pi} \sum_{k=1}^m \frac{1}{2k-1}.$$

For  $m = 1$ , we see that we have

$$\tau_K < \frac{2n\sqrt{2}}{|\lambda_S|\pi}.$$

For the  $m > 1$  case we substitute in for  $\sum_{k=1}^m \frac{1}{2k-1}$  from section 3, to find

$$\tau_K < \frac{2mn\sqrt{2}}{\pi|\lambda_S|} \left[ \frac{-1}{4m(m-1)} + \ln \left( \frac{2m-1}{\sqrt{m-1}} \right) + \frac{\gamma}{2} \right].$$

□

**7. Conclusion.** We have introduced a new explicit numerical method for the solution of a system differential equations dominated by a skew-symmetric component. Our results presented include a stability proof for this new method as well as providing an analytical stability proof for the stability condition for the super-time-stepping method. We have provided an analytical argument that our new method allows a larger time step for non-symmetric problems.

A numerical demonstration comparing the **H** scheme with  $m = 1$  to super-time-stepping scheme **T** applied to the skew-dominated Hall and ambipolar diffusion problem in magnetohydrodynamics may be found in [17]. It clearly shows that the time necessary to compute the problem with the **T** scheme grows extremely large as the skew-symmetric component of the problem increases with respect to the symmetric component while the **H** scheme becomes more efficient in comparison.

We have also presented a variant of our method that we have found works well in problems in magnetohydrodynamics for matrices of size  $7 \times 7$ .

**8. Acknowledgements.** The authors wish to thank G.W. Stewart for helpful discussions.

#### REFERENCES

- [1] V. Alexiades, G. Amiez, and P. Gremaud, Super-time-stepping acceleration of explicit schemes for parabolic problems, *Comm. Num. Meth. Eng.*, 12 (1996) pp. 31-42.
- [2] M. Anum-Addo, R. Darst, and K.F. Gurski, unpublished report.
- [3] M. Anum-Addo, *Matrix Madness: A Stability Study of Five Numerical Schemes for a Skew-dominated Discretized System*, undergraduate thesis, 2008.
- [4] D. Barash, T. Schlick, M. Israeli, and R. Kimmel, Multiplicative operator splittings in nonlinear diffusion: From spatial splitting to multiple time-steps, *J Math. Imag. Vision* 19 (2003) no. 1, pp. 33-48.
- [5] R. Bellman, *Introduction to Matrix Analysis*, McGraw-Hill, New York, 1960.
- [6] I. Bendixson, Sur les racines d'une equation fondamentale, *Acta Mathematica*. 3 (1902) pp.359-366.
- [7] P.L.I. Brian, An infinite difference method of high order of accuracy for the solution of the three-dimensional heat conduction problems, *A.I.Ch.E.J.* 7 (1961) pp.367-370.
- [8] R. Courant, K. Friedrichs and H. Lewy, On the partial difference equations of mathematical physics, *IBM Journal* (1967) pp. 215-234, (English translation of the 1928 German original).
- [9] Havi, J. Gamma: Exploring Euler's Constant, Princeton Univ. Press, Princeton, N.J., 2003.
- [10] A. Hirsch, Sur les racines d'une equation fondamentale (Extrait d'une lettre de M.A. Hirsh á M.I. Bendixson), *Acta Mathematica*. 25 (1902) pp.367-370.
- [11] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, 1985.
- [12] J. Jones, On the Lyapunov Stability Criteria, *J. Soc. Indust. Appl. Math.* 13 (1965) 4, 1965 pp.941-945.
- [13] K.-A. Lie, A dimensional splitting method for nonlinear equations with variable coefficients, *BIT* 39 (1999) 4, pp. 683700.

- [14] A.M. Lyapunov, *Problème général de la stabilité du mouvement*, Annals of Mathematics Study No. 17, Princeton University Press, Princeton, 1947.
- [15] W. Markoff, *Über Polynome, die einem gegebenen Intervall möglichst wenig von Null abweichen*, Ann. 77 (1892) pp 213-285 (translation and condensation by J. Grossman of Russian article published in 1892).
- [16] G.W. Stewart and Jiguang Sun, *Matrix Perturbation Theory*, Academic Press, San Diego, 1990.
- [17] S. O'Sullivan and T.P. Downes, *A Three-dimensional Method for Modelling Weakly Ionized Plasmas*, Mon. Not. R. Astron. Soc. 376 (2007) pp.1648-1658.
- [18] D.A. Tilley and D.S. Balsara, *A Two-fluid Method for Ambipolar Diffusion*, Monthly Notices of the Royal Astronomical Society, 389 (2008), pp. 1058-1073.
- [19] O. Toeplitz, *Das algebraische Analogon zu einem Satze von Fejér*, Mathematische Zeitschrift. 2 (1918) pp.187-197.
- [20] J.G. Verwer, *Explicit Runge-Kutta methods for parabolic partial differential equations*, Appl. Num. Math 22 (1996), pp. 359-379.
- [21] N.N. Yanenko, *Method of Fractional Steps*, Springer-Verlag, NY, 1971.
- [22] Young, R.M., *Math. Gaz.* 75 (1991), pp.187-190.