Math 269A HW 1 Solutions

Jeffrey Hellrung

2011 Fall

1 Pen (or Pencil) and Paper

1. Consider the problem
$$\frac{d\mathbf{u}}{dt} = \begin{pmatrix} -2 & 1 \\ \epsilon & -2 \end{pmatrix} \mathbf{u}, \quad \epsilon > 0.$$
Derive an estimate of the form
$$\|\mathbf{u}(t)\|_\infty \leq K\|\mathbf{u}(0)\|_\infty.$$
Discuss the effect of the parameter ϵ in your estimate. What is the smallest constant K?

Solution

Let
$$\mathbf{A} := \begin{pmatrix} -2 & 1 \\ \epsilon & -2 \end{pmatrix}.$$
One finds eigenvalue-eigenvector pairs of $\mathbf{A}$ to be
$$\left(\lambda^\pm, \mathbf{v}^\pm\right) := \left(-2 \pm \sqrt{\epsilon}, \begin{pmatrix} 1 \\ \pm\sqrt{\epsilon} \end{pmatrix}\right),$$
from which we derive the solution $\mathbf{u}(t)$ to be given by
$$\mathbf{u}(t) = \frac{1}{2\sqrt{\epsilon}} \begin{pmatrix} \sqrt{\epsilon}\left(e^{\lambda^+ t} + e^{\lambda^- t}\right) & e^{\lambda^+ t} - e^{\lambda^- t} \\ \epsilon\left(e^{\lambda^+ t} - e^{\lambda^- t}\right) & \sqrt{\epsilon}\left(e^{\lambda^+ t} + e^{\lambda^- t}\right) \end{pmatrix} \mathbf{u}(0).$$
Note that we must have $\epsilon < 4$ for a decaying solution.

We can bound the solution in the infinity-norm as
$$\begin{aligned} \|\mathbf{u}(t)\|_\infty &\leq \frac{1}{2\sqrt{\epsilon}} \max\left\{\sqrt{\epsilon}\left(e^{\lambda^+ t} + e^{\lambda^- t}\right) + e^{\lambda^+ t} - e^{\lambda^- t}, \epsilon\left(e^{\lambda^+ t} - e^{\lambda^- t}\right) + \sqrt{\epsilon}\left(e^{\lambda^+ t} + e^{\lambda^- t}\right)\right\} \|\mathbf{u}(0)\|_\infty \\ &= \left(\frac{1}{2}\left(e^{\lambda^+ t} + e^{\lambda^- t}\right) + \frac{1}{2}\max\left\{\sqrt{\epsilon}, \frac{1}{\sqrt{\epsilon}}\right\}\left(e^{\lambda^+ t} - e^{\lambda^- t}\right)\right) \|\mathbf{u}(0)\|_\infty. \end{aligned}$$
One finds this latter expression monotically decreases in t when $\epsilon \leq 2$ (hence one may take $K = 1$), but when $\epsilon > 2$, we find a maximum at
$$t^* = \frac{1}{2\sqrt{\epsilon}} \log \frac{\sqrt{\epsilon} + (\epsilon - 2)}{\sqrt{\epsilon} - (\epsilon - 2)}$$
which necessitates $K > 1$.

2. Consider the problem

$$\frac{d\mathbf{u}}{dt} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \mathbf{u}, \quad t > 0, \quad \mathbf{u}(0) = \mathbf{f}.$$

Prove that the leap frog method converges for this problem and provide an upper bound for the magnitude of the error in terms of $\mathbf{f}$, the truncation error of the method used in the first step, and higher derivatives of the solution.

Solution

The leap frog method is given by

$$\frac{\mathbf{u}^{n+1} - \mathbf{u}^{n-1}}{2\Delta t} = \mathbf{A}\mathbf{u}^n, \quad \mathbf{A} := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Since

$$\begin{aligned} \frac{\mathbf{u}\left(t^{n+1}\right) - \mathbf{u}\left(t^{n-1}\right)}{2\Delta t} &= \frac{d\mathbf{u}}{dt}\left(t^n\right) + \frac{1}{6}\mathbf{u}^{(3)}\left(t^n\right)\Delta t^2 + \mathcal{O}\left(\Delta t^4\right) \\ &= \mathbf{A}\mathbf{u}\left(t^n\right) + \frac{1}{6}\mathbf{u}^{(3)}\left(t^n\right)\Delta t^2 + \mathcal{O}\left(\Delta t^4\right) \end{aligned}$$

we obtain the following difference equation for the error $\mathbf{e}^n := \mathbf{u}\left(t^n\right) - \mathbf{u}^n$:

$$\mathbf{e}^{n+1} = \mathbf{e}^{n-1} + 2\Delta t \mathbf{A}\mathbf{e}^n + \frac{1}{3}\mathbf{u}^{(3)}\left(t^n\right)\Delta t^3 + \mathcal{O}\left(\Delta t^5\right).$$

To solve the above difference equation, we first reformulate it as a first order difference equation:

$$\mathbf{E}^n = \mathcal{A}\mathbf{E}^{n-1} + \mathbf{C}^n$$

where

$$\mathbf{E}^n := \begin{pmatrix} \mathbf{e}^{n+1} \\ \mathbf{e}^n \end{pmatrix}, \quad \mathcal{A} := \begin{pmatrix} 2\Delta t\mathbf{A} & \mathbf{I}_{2\times 2} \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} \end{pmatrix}, \quad \mathbf{C}^n := \begin{pmatrix} \frac{1}{3}\mathbf{u}^{(3)}\left(t^n\right)\Delta t^3 + \mathcal{O}\left(\Delta t^5\right) \\ \mathbf{0}_2 \end{pmatrix}.$$

One can easily verify then that

$$\mathbf{E}^n = \mathcal{A}^n\mathbf{E}^0 + \sum_{k=0}^{n-1} \mathcal{A}^k \mathbf{C}^{n-k}.$$

It follows that

$$\begin{aligned} \left\|\mathbf{e}^{n+1}\right\| &\le \left\|\mathbf{E}^n\right\| \\ &\le \left\|\mathcal{A}^n\right\| \left\|\mathbf{E}^0\right\| + \left(\max_n \left\|\mathbf{C}^n\right\|\right) \sum_{k=0}^{n-1} \left\|\mathcal{A}^k\right\| \\ &= \left\|\mathcal{A}^n\right\| \left\|\mathbf{e}^1\right\| + \left(\frac{1}{3}\left\|\mathbf{u}^{(3)}\right\|\Delta t^3 + \mathcal{O}\left(\Delta t^5\right)\right) \sum_{k=0}^{n-1} \left\|\mathcal{A}^k\right\|. \end{aligned}$$

It now remains to bound $\left\|\mathcal{A}^k\right\|$. For this, it is sufficient to consider, specifically, $\|\mathcal{A}\|_\infty = 1 + 2\Delta t$. This gives

$$\left\|\mathcal{A}^n\right\|_\infty \le \|\mathcal{A}\|_\infty^n \le e^{2\Delta t n}$$

and

$$\sum_{k=0}^{n-1} \left\|\mathcal{A}^k\right\|_\infty \le \sum_{k=0}^{n-1} \|\mathcal{A}\|_\infty^k \le n e^{2\Delta t n}.$$

This finally gives the bound

$$\left\|\mathbf{u}\left(t^{n+1}\right) - \mathbf{u}^{n+1}\right\| \le e^{2\Delta t n}\left\|\mathbf{e}^1\right\|_\infty + n e^{2\Delta t n}\left(\frac{1}{3}\left\|\mathbf{u}^{(3)}\right\|_\infty \Delta t^3 + \mathcal{O}\left(\Delta t^5\right)\right).$$

3. Derive a one-step formula for RK3 in the form

$$v^{n+1} = P(z)v^n,$$

i.e., specify $P(z)$.

Solution

For any RK3 method, $P(z) = 1 + z + \frac{1}{2}z^2 + \frac{1}{6}z^3$.

4. Consider the system

$$\frac{\partial \mathbf{u}}{\partial t} = \mathbf{A}(t)\mathbf{u} + \mathbf{F}(t), \quad t > t_0, \quad \mathbf{u}(t_0) = \mathbf{u}_0 \tag{1}$$

and the homogeneous version

$$\frac{\partial \mathbf{v}}{\partial t} = \mathbf{A}(t)\mathbf{v}, \quad t > t_0, \quad \mathbf{v}(t_0) = \mathbf{v}_0. \tag{2}$$

(2) has a unique solution for each $\mathbf{v}_0$, specifically, $\mathbf{v}(t) = \mathbf{S}(t, t_0)\mathbf{v}_0$.

a. Show the following.
   (i) $\mathbf{S}(t_0, t_0) = \mathbf{I}$
   (ii) $\mathbf{S}(t_2, t_0) = \mathbf{S}(t_2, t_1)\mathbf{S}(t_1, t_0)$
   (iii) $\|\mathbf{S}(t, t_0)\|_2^2 \leq Ke^{\alpha(t-t_0)}$ for $\alpha = \max_t \rho(A^* + A)$
   (iv) $\frac{\partial \mathbf{S}}{\partial t}(t, t_0) = \mathbf{A}(t)\mathbf{S}(t, t_0)$
   (v) $\frac{\partial \mathbf{S}}{\partial t_0}(t, t_0) = -\mathbf{S}(t, t_0)\mathbf{A}(t_0)$

b. Show that the solution $\mathbf{u}(t)$ satisfies

$$\mathbf{u}(t) = \mathbf{S}(t, t_0)\mathbf{u}_0 + \int_{t_0}^{t} \mathbf{S}(t, \tau)\mathbf{F}(\tau)d\tau.$$

c. Show that

$$\|\mathbf{u}(t)\|_2 \leq \|\mathbf{u}_0\|_2 e^{\frac{\alpha}{2}(t-t_0)} + \phi^*(t, t_0) \max_{t_0 \leq \tau \leq t} \|\mathbf{F}(t)\|_2$$

where

$$\phi^*(t, t_0) = \begin{cases} \frac{2}{\alpha}\left(e^{\frac{\alpha}{2}(t-t_0)} - 1\right), & \alpha \neq 0 \\ t - t_0, & \alpha = 0 \end{cases}.$$

Solution

a. (i) Since $\mathbf{v}(0) = \mathbf{v}_0$ and $\mathbf{v}(t) = \mathbf{S}(t, t_0)\mathbf{v}_0$ for all $t \geq t_0$ (so, specifically, when $t = t_0$), we must have $\mathbf{v}_0 = \mathbf{S}(t_0, t_0)\mathbf{v}_0$ for all $\mathbf{v}_0$, hence $\mathbf{S}(t_0, t_0) = \mathbf{I}$.
   (ii) By uniqueness, $\mathbf{S}(t_2, t_0)\mathbf{v}_0 = \mathbf{v}(t_2) = \mathbf{S}(t_2, t_1)\mathbf{v}(t_1)$, and $\mathbf{v}(t_1) = \mathbf{S}(t_1, t_0)\mathbf{v}_0$. Thus, $\mathbf{S}(t_2, t_0)\mathbf{v}_0 = \mathbf{S}(t_2, t_1)\mathbf{S}(t_1, t_0)\mathbf{v}_0$ for all $\mathbf{v}_0$, hence $\mathbf{S}(t_2, t_0) = \mathbf{S}(t_2, t_1)\mathbf{S}(t_1, t_0)$.
   (iii) Observe that

$$\frac{\partial}{\partial t}\|\mathbf{v}\|_2^2 = 2\mathbf{v} \cdot \frac{\partial \mathbf{v}}{\partial t} = 2\mathbf{v} \cdot \mathbf{A}\mathbf{v} = \mathbf{v} \cdot (\mathbf{A} + \mathbf{A}^*)\mathbf{v} \leq \rho(\mathbf{A} + \mathbf{A}^*)\|\mathbf{v}\|_2^2$$

(since $\mathbf{A} + \mathbf{A}^*$ is self-adjoint). Thus, by Gronwall's inequality,

$$\|\mathbf{v}(t)\|_2^2 \leq \|\mathbf{v}_0\|_2^2 \exp\left(\int_{t_0}^{t} \rho(\mathbf{A}(s) + \mathbf{A}(s)^*)\, ds\right) \leq \|\mathbf{v}_0\|_2^2 e^{\alpha(t-t_0)}$$

where $\alpha = \max_{t_0 \leq s \leq t} \rho\left(\mathbf{A}(s) + \mathbf{A}(s)^*\right)$. Using this, we obtain

$$\left\|\mathbf{S}\left(t, t_0\right) \mathbf{v}_0\right\|_2 = \|\mathbf{v}(t)\|_2 \leq \left\|\mathbf{v}_0\right\|_2 e^{\frac{1}{2}\alpha\left(t-t_0\right)}$$

and hence

$$\left\|\mathbf{S}\left(t, t_0\right)\right\|_2 = \max\left\{\left\|\mathbf{S}\left(t, t_0\right) \mathbf{v}_0\right\|_2 : \left\|\mathbf{v}_0\right\|_2 = 1\right\} \leq e^{\frac{\alpha}{2}\left(t-t_0\right)}.$$

(iv) Starting with $\mathbf{v}(t) = \mathbf{S}(t, s)\mathbf{v}(s)$ and differentiating with respect to t yields $\frac{\partial \mathbf{v}}{\partial t}(t) = \frac{\partial \mathbf{S}}{\partial t}(t, s)\mathbf{v}(s)$. But we also have that $\frac{\partial \mathbf{v}}{\partial t}(t) = \mathbf{A}(t)\mathbf{v}(t) = \mathbf{A}(t)\mathbf{S}(t, s)\mathbf{v}(s)$. Hence, $\frac{\partial \mathbf{S}}{\partial t}(t, s)\mathbf{v}(s) = \mathbf{A}(t)\mathbf{S}(t, s)\mathbf{v}(s)$ for all $\mathbf{v}(s)$, from which we conclude that $\frac{\partial \mathbf{S}}{\partial t}(t, s) = \mathbf{A}(t)\mathbf{S}(t, s)$.

(v) Starting with $\mathbf{v}(t) = \mathbf{S}(t, s)\mathbf{v}(s)$ and differentiating with respect to s yields

$$\mathbf{0} = \frac{\partial \mathbf{S}}{\partial s}(t, s)\mathbf{v}(s) + \mathbf{S}(t, s)\frac{\partial \mathbf{v}}{\partial s}(s) = \frac{\partial \mathbf{S}}{\partial s}(t, s) + \mathbf{S}(t, s)\mathbf{A}(s)\mathbf{v}(s),$$

which must hold for all $\mathbf{v}(s)$, hence $\frac{\partial \mathbf{S}}{\partial s}(t, s) = -\mathbf{S}(t, s)\mathbf{A}(s)$.

b. Indeed, if $\mathbf{u}(t) = \mathbf{S}\left(t, t_0\right) \mathbf{u}_0 + \int_{t_0}^{t} \mathbf{S}(t, \tau)\mathbf{F}(\tau) d\tau$, then

$$\mathbf{u}\left(t_0\right) = \mathbf{S}\left(t_0, t_0\right) \mathbf{u}_0 + 0 = \mathbf{u}_0$$

and

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t}(t) &= \frac{\partial \mathbf{S}}{\partial t}\left(t, t_0\right) \mathbf{u}_0 + \int_{t_0}^{t} \frac{\partial \mathbf{S}}{\partial t}(t, \tau)\mathbf{F}(\tau) d\tau + \mathbf{S}(t, t)\mathbf{F}(t) \\
&= \mathbf{A}(t)\mathbf{S}\left(t, t_0\right) \mathbf{u}_0 + \int_{t_0}^{t} \mathbf{A}(t)\mathbf{S}(t, \tau)\mathbf{F}(\tau) d\tau + \mathbf{F}(t) \\
&= \mathbf{A}(t)\left(\mathbf{S}\left(t, t_0\right) \mathbf{u}_0 + \int_{t_0}^{t} \mathbf{S}(t, \tau)\mathbf{F}(\tau) d\tau\right) + \mathbf{F}(t) \\
&= \mathbf{A}(t)\mathbf{u}(t) + \mathbf{F}(t),
\end{aligned}$$

as desired.

c. This inequality follows directly from the expression for $\mathbf{u}(t)$ in b. and the bound on $\left\|\mathbf{S}\left(t, t_0\right)\right\|_2$ in a. (iii):

$$\begin{aligned}
\|\mathbf{u}(t)\|_2 &\leq \left\|\mathbf{S}\left(t, t_0\right)\right\|_2 \left\|\mathbf{u}_0\right\|_2 + \int_{t_0}^{t} \|\mathbf{S}(t, \tau)\|_2 \, d\tau \max_{t_0 \leq \tau \leq t} \|\mathbf{F}(\tau)\|_2 \\
&\leq e^{\frac{\alpha}{2}\left(t-t_0\right)} \left\|\mathbf{u}_0\right\|_2 + \phi^*\left(t, t_0\right) \max_{t_0 \leq \tau \leq t} \|\mathbf{F}(\tau)\|_2
\end{aligned}$$

as

$$\int_{t_0}^{t} e^{\frac{\alpha}{2}\left(\tau-t_0\right)} d\tau = \begin{cases} \frac{2}{\alpha}\left(e^{\frac{\alpha}{2}\left(t-t_0\right)} - 1\right), & \alpha \neq 0 \\ t - t_0, & \alpha = 0 \end{cases} =: \phi^*\left(t, t_0\right).$$

2 Programming

1. Write a program to solve

$$\frac{d\mathbf{u}}{dt} = \begin{pmatrix} -2 & 1 \\ \epsilon & -2 \end{pmatrix} \mathbf{u}, \quad \epsilon > 0$$

using the forward Euler method. Run it with initial data $\mathbf{u}(0) = (1, 3)^t$, $\Delta t = 0.01$, and enough steps to reach $T = 3.0$. Display your results for $\epsilon = 5, 4, 0.1, 0$. Discuss your numerical observations with respect to the bounds derived in the Pen and Paper problem.

Solution

We expect an exponentially increasing solution when $\epsilon = 5$, a solution which approaches a non-zero steady-state when $\epsilon = 4$, and exponentially decaying solutions when $\epsilon = 0.1$ or $\epsilon = 0$.

2. Write a program to solve
$$\frac{d\mathbf{u}}{dt} = \begin{pmatrix} a & 1 \\ -1 & a \end{pmatrix} \mathbf{u}, \quad t > 0$$
with the leap frog method. Use the forward Euler method for the first step. (a) Demonstrate the order of accuracy is 2 if $a = 0$. (b) Study the solutions for $a < 0$ and explain the numerical behavior.

   Solution

   (a) One can demonstrate the order of accuracy by computing a numerical solution for a spectrum of time step sizes Δt, plotting the logarithm of the error at some fixed final time T versus the logarithm of the time step size Δt, and verifying that the slope of a line through these plotted points is -2. (b) When $a < 0$, leap frog introduces artificial oscillations into the numerical solution that quickly become unstable.

3. Write a program that plots the boundary of the region $\{z \in \mathbb{C} : |P(z)| \leq 1\}$ for RK3.

   Solution

   One easy way to do this is to plot $z(\theta)$, where $z(\theta)$ is defined implicitly as the solution to $P(z(\theta)) = e^{i\theta}$, $\theta \in [0, 2\pi]$.