Numerical Analysis – Lecture 24[1]

Discussion 5.27 *(The reason for* $e^{t(B+C)} \neq e^{tB}e^{tC}$*)* Comparing the Taylor expansions

$$e^{t(B+C)} = I + t(B+C) + \tfrac{1}{2}t^2(B^2 + BC + CB + C^2) + \mathcal{O}(t^3)$$

with

$$\begin{aligned} e^{tB}e^{tC} &= \left[I + tB + \tfrac{1}{2}t^2B^2 + \mathcal{O}(t^3)\right] \times \left[I + tC + \tfrac{1}{2}t^2C^2 + \mathcal{O}(t^3)\right] \\ &= I + t(B+C) + \tfrac{1}{2}t^2(B^2 + 2BC + C^2) + \mathcal{O}(t^3), \end{aligned}$$

we obtain

$$e^{tB}e^{tC} = e^{t(B+C)} + \tfrac{1}{2}t^2(BC - CB) + \mathcal{O}(t^3). \tag{5.16}$$

Thus, $e^{tB}e^{tC} = e^{t(B+C)}\ \forall t \geq 0$ only if B and C commute. *[Note: It is trivial to prove that commutativity is sufficient, not just necessary.]* The good news is, however, that approximating $e^{\Delta t(B+C)}$ with $e^{\Delta tB}e^{\Delta tC}$ incurs an error of $\mathcal{O}((\Delta t)^2)$ which, for 'our' diffusion equation, is $\mathcal{O}((\Delta x)^4)$. Further, suppose that r is a rational function such that $r(z) = e^z + \mathcal{O}(z^2)$. It follows from our analysis that

$$\boldsymbol{u}^{n+1} = r(\Delta tB)r(\Delta tC)\boldsymbol{u}^n$$

produces an error of $\mathcal{O}((\Delta t)^2)$. The choice $r(z) = (1+\frac{1}{2}z)/(1-\frac{1}{2}z)$ results in a *split Crank–Nicolson* scheme, whose implementation reduces to a solution of tridiagonal algebraic linear systems. The error is consistent with the error of semidiscretization (since $\Delta t = \mu(\Delta x)^2$). As far as stability is concerned, we observe that both B and C are symmetric, therefore so are $r(\Delta tB)$ and $r(\Delta tC)$. But symmetry $\Rightarrow$ normalcy $\Rightarrow$ the 2-norm equals the spectral radius. Therefore,

$$\|\boldsymbol{u}^{n+1}\| \leq \|r(\Delta tB)\| \cdot \|r(\Delta tC)\| \cdot \|\boldsymbol{u}^n\| = \rho(r(\Delta tB)) \cdot \rho(r(\Delta tC)) \cdot \|\boldsymbol{u}^n\|.$$

It is left as an exercise to verify that the eigenvalues of the (symmetric) matrices B and C are nonpositive and hence to deduce that, provided r is A-stable (therefore $|r(z)| < 1$ for $z \in \mathbb{C}$, $\mathrm{Re}\, z < 0$), it is true that $\rho(r(\Delta tB)), \rho(r(\Delta tC)) \leq 1$. This proves $\|\boldsymbol{u}^{n+1}\| \leq \|\boldsymbol{u}^n\| \leq \cdots \|\boldsymbol{u}^0\|$, hence stability.

Improvement 5.28 It follows at once from (5.16), using symmetry, that

$$\tfrac{1}{2}\left(e^{tB}e^{tC} + e^{tC}e^{tB}\right) = e^{t(B+C)} + \mathcal{O}((\Delta t)^3).$$

Moreover, an easy exercise in Taylor expansions verifies that the error in the *Strang splitting*

$$e^{\Delta t(B+C)} \approx e^{\frac{1}{2}\Delta tB}e^{\Delta tC}e^{\frac{1}{2}\Delta tB}$$

is also $\mathcal{O}((\Delta t)^3)$. Thus, as long as $r(z) = e^z + \mathcal{O}(z^3)$, the time-stepping formula

$$\boldsymbol{u}^{n+1} = r(\tfrac{1}{2}\Delta tB)r(\Delta tC)r(\tfrac{1}{2}\Delta tB)\boldsymbol{u}^n$$

carries a local error of $\mathcal{O}((\Delta t)^3)$.

Method 5.29 *(Splitting of inhomogeneous systems)* Recall our goal, namely fast methods for the two-dimensional diffusion equation. Our exposition so far has been contrived, because of the assumption that the boundary conditions are zero. In general, the linear ODE system is of the form

$$\boldsymbol{u}' = (B+C)\boldsymbol{u} + \boldsymbol{b}, \qquad \boldsymbol{u}(0) = \boldsymbol{u}^0, \tag{5.17}$$

[1] Please email all corrections and suggestions to these notes to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html`.

where $\mathbf{b}$ originates in boundary conditions (and possibly in a forcing term in the original PDE). Note that our analysis should accommodate $\boldsymbol{b} = \boldsymbol{b}(t)$, since boundary conditions might vary in time! The *exact* solution of (5.17) is provided by the *variation of constants* formula

$$\boldsymbol{u}(t) = \mathrm{e}^{t(B+C)}\boldsymbol{u}(0) + \int_0^t \mathrm{e}^{(t-\tau)(B+C)}\boldsymbol{b}(\tau)\,\mathrm{d}\tau, \qquad t \geq 0,$$

(verify), therefore

$$\boldsymbol{u}((n+1)\Delta t) = \mathrm{e}^{\Delta t(B+C)}\boldsymbol{u}(n\Delta t) + \int_{n\Delta t}^{(n+1)\Delta t} \mathrm{e}^{[(n+1)\Delta t-\tau](B+C)}\boldsymbol{b}(\tau)\,\mathrm{d}\tau, \quad n = 0, 1, \ldots.$$

The integral can be frequently evaluated explicitly, e.g. when $\boldsymbol{b}$ is a linear combination of polynomial and exponential terms. For example, $\boldsymbol{b}(t) \equiv$ const yields

$$\boldsymbol{u}((n+1)\Delta t) = \mathrm{e}^{\Delta t(B+C)}\boldsymbol{u}(n\Delta t) + (B+C)^{-1}\left(\mathrm{e}^{\Delta t(B+C)} - I\right)\boldsymbol{b}.$$

This, unfortunately, is not a helpful observation, since, even if we split the exponential, how are we supposed to 'split' $(B+C)^{-1}$? The remedy is not to evaluate the integral explicitly but, instead, to use quadrature. For example, the *trapezoidal rule* for integrals,

$$\int_0^h g(\tau)\,\mathrm{d}\tau = \tfrac{1}{2}h[g(0) + g(h)] + \mathcal{O}\big(h^3\big)$$

gives

$$\boldsymbol{u}((n+1)\Delta t) \approx \mathrm{e}^{\Delta t(B+C)}\boldsymbol{u}(n\Delta t) + \tfrac{1}{2}\Delta t[\mathrm{e}^{\Delta t(B+C)}\boldsymbol{b}(n\Delta t) + \boldsymbol{b}((n+1)\Delta t)],$$

with a local error of $\mathcal{O}\big((\Delta t)^3\big)$. We can now replace exponentials with their splittings. For example, Strang's splitting results in

$$\boldsymbol{u}^{n+1} = r(\tfrac{1}{2}\Delta t B)r(\Delta t C)r(\tfrac{1}{2}\Delta t B)[\boldsymbol{u}^n + \tfrac{1}{2}\Delta t\boldsymbol{b}^n] + \tfrac{1}{2}\Delta t\boldsymbol{b}^{n+1}.$$

As before, everything reduces to (inexpensive) solution of tridiagonal systems!

Technique 5.30 *(Splitting for nonlinear equations)* We may use splitting to resolve nonlinearities (*operatorial splitting,* as distinct from *dimensional splitting*). For example, consider the *reaction–diffusion* equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \alpha u(1-u)$$

with (for simplicity) zero boundary conditions at $x = 0$ and $x = 1$. Having discretized the space derivatives, we obtain an ODE system of the form $\boldsymbol{u}' = \boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{g}(\boldsymbol{u})$, where $\boldsymbol{f}(\boldsymbol{u}) = A\boldsymbol{u}$ originates in semidiscretization and $\boldsymbol{g}_m(\boldsymbol{u}) = \alpha u_m(1-u_m)$ in the nonlinear term. The main idea is to compose the solution advancing $\frac{1}{2}\Delta t$ with $\boldsymbol{y}' = \boldsymbol{f}(\boldsymbol{y}) + \boldsymbol{g}(\boldsymbol{y}_n)$ and $\frac{1}{2}\Delta t$ with $\boldsymbol{y}' = \boldsymbol{f}(\boldsymbol{y}_n) + \boldsymbol{g}(\boldsymbol{y})$, which gives an $\mathcal{O}\big((\Delta t)^2\big)$ method. Note that the first equation is linear, while the second consists of scalar Riccati equations, which can be solved explicitly. Alternatively, we may employ a technique similar to the Strang splitting to increase accuracy to $\mathcal{O}\big((\Delta t)^3\big)$.