MAT 228A Theory Homework 3

Ivan Cherkashin

Problem 1

THEOREM. *A linear finite-difference method*

$$(Lu^n)_i = \sum_s c_s u^n_{i+s}$$

is consistent $\iff \sum_s c_s = 1 \ \wedge \ \sum_s c_s s = -\sigma$

PROOF. The order of the truncation error is $O(\frac{e^{-\sigma\beta}-\lambda(\beta)}{\beta})$: therefore, consistency requires and implies that the order of the truncation error is at least $O(\beta)$ $\iff$ the first two terms of the Taylor series of the symbol $\lambda(\beta) = \sum_s c_s e^{is\beta}$ are equal to the first two terms of the Taylor series of $e^{-i\sigma\beta}$, i.e.

$$(0.1)\ \ 1 - i\sigma\beta = \lambda(0) + \beta\lambda_\beta(0) = \sum_s c_s + \beta i \sum_s c_s s \Longrightarrow \sum_s c_s = 1 \ \wedge \ \sum_s c_s s = -\sigma$$

Observation. In case all $c_s \geqslant 0$, the first condition means that c_s form a discrete probability distribution, and the second means that the mathematical expectation (i.e. average) of the index offsets s relative to i must be equal $-\sigma$. One possible physical interpretation of this requirement is the fact that a closed system that relies on information from the "future" (i.e. the case $\sum_s c_s s \geqslant 0$) is non-physical, or at least not observable. Indeed, real observable physical processes are stable: otherwise, instability leads to a qualitative change of the system until it reaches equilibrium, but by then the previous identity and information about the system will have disappeared, hence it is impossible to observe it (unless the power of measurement devices is able to resolve time and space to sufficiently small scales, but that means the system cannot be considered closed anymore, since such precise measurements supply significant amount of energy to the system).

Also, entropy is higher behind the wave than in front, since real transport processes are irreversible. Hence, when more information is used from the front of the wave than from the back, entropy can decrease, which cannot happen in a

real system. Therefore, the information (entropy) balance must be, even slightly, in favor of the information behind the wave, hence the negative sign of the mathematical expectation of the offsets. That means that the information (entropy) flux from negative offsets (cells behind the wave) must dominate over the flux from the positive offsets (front cells).

There is a little connection with probability theory and convex geometry. From this perspective, the discreet evolution operator L can be seen as a doubly stochastic matrix (Proof: Each column, as well as each row, contains all of c_s, which add up to 1: this is due to multidiagonal structure of L). That means that consistent finite difference methods form a convex set, because doubly stochastic matrixes form one. Moreover, the fact that every discreet evolution operator is represented as a linear combination of shift operators is a consequence of Caratheodory's theorem for a simplex: every element of a simplex can be uniquely represented as a convex combination of its exterior points. In our case, the exterior points are permutation matrices (shift operators).

Finally, a stochastic matrix can be seen as a matrix of transition probabilities of a stationary finite-state Markov chain. Then the components of the numerical solution can be interpreted as mathematical expectations of the values at the cells of the stencil at the previous time step. Since the entropy of a Markov chain is maximized when it reaches its stationary transition probability distribution, it is now evident why stable numerical transport processes reach equilibrium: the solution tends to a constant, because in this state the entropy is maximized.

Now it is possible to prove why consistent linear numerical schemes converge to constant solution and preserve its L^1 norm (which can be mass or any other conserved quantity that is being transported).

Indeed, since the method is consistent, the vector of ones $\Lambda_1 = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenvector of the discreet evolution operator L with the corresponding eigenvalue of 1. All other eigenvalues are less than one by absolute value since $\|L\|_1 = 1$ bounds all the eigenvalues. Therefore, for any initial data vector $u_0 h$ (assuming it is physical, i.e. all of its components are nonnegative)[1] for very large times $L^n[u_0 h]$ will converge to

$$u_\infty = \langle u_0 h, \Lambda_1 \rangle \Lambda_1 = \|u_0 h\|_1 \Lambda_1 = \|u_0\|_{L^1} \Lambda_1$$

Since

$$\|\Lambda_1\|_1 = M \Longrightarrow \|\Lambda_1\|_{L^1} = h\|\Lambda_1\|_1 = 1$$

it follows, finally, that

$$\|u_\infty\|_{L^1} = \|u_0\|_{L^1} \|\Lambda_1\|_{L^1} = \|u_0\|_{L^1}$$

□

[1] $h = \frac{1}{M}$, M is the number of grid cells: it is included to be consistent with the discreet L^1 norm and regular one-norm of $\mathbb{R}^M$, i.e. $\|u_0\|_{L^1} = \|u_0 h\|_1$

Problem 2

THEOREM. *A linear finite-difference method*

$$(Lu^n)_i = \sum_s c_s u^n_{i+s}$$

is second order accurate $\sum_s c_s s^2 = \sigma^2$

PROOF. The order of the truncation error is $O(\frac{e^{-\sigma\beta}-\lambda(\beta)}{\beta})$: therefore, second-order convergence requires and implies that the order of the truncation error is at least $O(\beta^2) \implies$ the first three terms of the Taylor series of the symbol $\lambda(\beta) = \sum_s c_s e^{is\beta}$ and $e^{-i\sigma\beta}$ must be equal, i.e. for the third term:

$$\frac{-\sigma^2\beta^2}{2} = \frac{\beta^2}{2}\lambda_{\beta\beta}(0) = \frac{\beta^2}{2}\sum_s -c_s s^2 \implies \sum_s c_s s^2 = \sigma^2 \tag{0.2}$$

□

Problem 3

THEOREM. *Only even powers of* β *appear in the Taylor expansion (about 0) of* $|\lambda(\beta)|$

PROOF. $|\lambda(-\beta)| = |\overline{\lambda(\beta)}| = |\lambda(\beta)|$ since complex conjugation preserves lengths of vectors. Since $\lambda(\beta)$ is an even function infinitely differentiable at zero, its Taylor series contains only even powers of β. □

Problem 4

THEOREM. *The relationship between the order of accuracy of the truncation error* p *and the leading order term in Taylor expansion of* $|\lambda(\beta)|$ *is* $2q = p+1$ *when* p *is odd, and* $2q = p+2$ *when* p *is even.*

PROOF. The order of the truncation error is $O(\frac{|e^{-\sigma\beta}-\lambda(\beta)|}{\beta}) = O(\beta^p)$.

Due to the reverse triangle inequality,

$$1 - |\lambda(\beta)| \leqslant |e^{-\sigma\beta} - \lambda(\beta)| \iff O(\beta^{2q}) \leqslant O(\beta^{p+1})$$

If p is odd, then $p+1$ is even, and the smallest q for which the inequality is satisfied is $q = \frac{p+1}{2} \iff 2q = p+1$.

If p is even, then $p+1$ is odd, and the smallest q for which the inequality is satisfied is $q = \frac{(p+1)+1}{2} \iff 2q = p+2$ (i.e. $2q$ must be one order smaller than $p+1$ in order for $2q$ to be an even number).

□

Problem 5

THEOREM. *The convergence rate of Fromm's method is $O(h^3)$ when $\sigma = 0.5$*

PROOF. Since the amplitude error is

$$|\lambda(\beta)| - 1 = \frac{\sigma(\sigma-1)(\sigma^2-\sigma+1)}{8}\beta^4 + O(\beta^5) = O(\beta^4)$$

when $\sigma = 0.5$, and since the phase error is

$$e^{-i\delta(\sigma,\beta)} = \frac{2\sigma^2-3\sigma+1}{12}\beta^3 + O(\beta^5) = \frac{(2\sigma-1)(\sigma-1)}{12}\beta^3 + O(\beta^5) = O(\beta^5)$$

it follows that the amplitude error dominates the local truncation error, since the amplitude error is one order larger than the phase error.

Thus, the local truncation error of Fromm's method when $\sigma = 0.5$ is

$$\frac{O(\beta^4)}{\Delta t} = \frac{O(h^4)}{h} = O(h^3)$$

which explains the computed third-order convergence rate of Fromm's method with $\sigma = 0.5$ and Gaussian Pulse initial condition.

□