Technische Universität München
Institut für Informatik
Lehrstuhl für numerische Programmierung

Introduction to Scientific Computing

Discrete Energy

1 Explicit time stepping and numerical energy

To obtain a numerical solution of the 1D heat equation, we have derived the following explicit scheme

$$v_j^{(m+1)} = v_j^{(m)} + \frac{\tau}{h^2}\left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)}\right). \tag{1}$$

For the resulting numerical solutions $v^{(m)}$ we introduce the discrete energy

$$E^{(m)} := h\sum_{j=1}^{n-1}\left(v_j^{(m)}\right)^2$$

for each time step m.

We would like to show that the discrete energy also decreases in time, i.e.:

$$E^{(m+1)} \leq E^{(m)} \qquad \text{for} \quad m \geq 0.$$

Thus, we compute the difference of the energy in two subsequent time steps:

$$\begin{aligned} E^{(m+1)} - E^{(m)} &= h\sum_{j=1}^{n-1}\left(\left(v_j^{(m+1)}\right)^2 - \left(v_j^{(m)}\right)^2\right) \\ &= h\sum_{j=1}^{n-1}\left(v_j^{(m+1)} + v_j^{(m)}\right)\left(v_j^{(m+1)} - v_j^{(m)}\right). \end{aligned}$$

Remembering our explicit time-stepping scheme,

$$v_j^{(m+1)} = v_j^{(m)} + \frac{\tau}{h^2}\left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)}\right), \tag{2}$$

$$\text{or} \qquad v_j^{(m+1)} - v_j^{(m)} = \frac{\tau}{h^2}\left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)}\right), \tag{3}$$

we get

$$\begin{aligned}
E^{(m+1)} - E^{(m)} &= \frac{\tau}{h} \sum_{j=1}^{n-1} \left(v_j^{(m+1)} + v_j^{(m)} \right) \left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)} \right) \\
&= \frac{\tau}{h} \Bigg\{ \underbrace{\sum_{j=1}^{n-1} v_j^{(m)} \left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)} \right)}_{=: S^{(m)}} \\
&\qquad -2 \sum_{j=1}^{n-1} v_j^{(m+1)} v_j^{(m)} + \sum_{j=1}^{n-1} v_j^{(m+1)} \left(v_{j-1}^{(m)} + v_{j+1}^{(m)} \right) \Bigg\} \\
&= \frac{\tau}{h} \left\{ S^{(m)} - 2 \sum_{j=1}^{n-1} v_j^{(m+1)} v_j^{(m)} + \sum_{j=1}^{n-1} v_j^{(m+1)} \left(v_{j-1}^{(m)} + v_{j+1}^{(m)} \right) \right\}
\end{aligned}$$

Estimates for Discrete Energy

The three sums can be examined separately.

- For the term $S^{(m)}$, we can use *summation by parts* (see appendix) to get

$$S^{(m)} = \sum_{j=1}^{n-1} v_j^{(m)} \left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)} \right) = -(v_1)^2 - \sum_{j=1}^{n-1} \left(v_{j+1}^{(m)} - v_j^{(m)} \right)^2 \leq 0.$$

  Thus, this term is always negative.

- In the second sum, we can apply the time stepping scheme (1), and obtain

$$\begin{aligned}
-2 \sum_{j=1}^{n-1} v_j^{(m+1)} v_j^{(m)} &= -2 \sum_{j=1}^{n-1} \left(v_j^{(m)} + \frac{\tau}{h^2} \left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)} \right) \right) v_j^{(m)} \\
&= \underbrace{-2 \sum_{j=1}^{n-1} (v_j^{(m)})^2}_{= \frac{1}{h} E^{(m)}} \underbrace{-2 \frac{\tau}{h^2} \sum_{j=1}^{n-1} v_j^{(m)} \left(v_{j-1}^{(m)} - 2v_j^{(m)} + v_{j+1}^{(m)} \right)}_{= S^{(m)}} \\
&= -\frac{2}{h} E^{(m)} - 2 \frac{\tau}{h^2} S^{(m)}
\end{aligned}$$

- and finally, using the inequality $ab \leq \frac{1}{2} \left(a^2 + b^2 \right)$,

$$\begin{aligned}
\sum_{j=1}^{n-1} v_j^{(m+1)} \left(v_{j-1}^{(m)} + v_{j+1}^{(m)} \right) &= \sum_{j=1}^{n-1} v_j^{(m+1)} v_{j-1}^{(m)} + \sum_{j=1}^{n-1} v_j^{(m+1)} v_{j+1}^{(m)} \\
&\leq \frac{1}{2} \sum_{j=1}^{n-1} \left[(v_j^{(m+1)})^2 + (v_{j-1}^{(m)})^2 \right] + \frac{1}{2} \sum_{j=1}^{n-1} \left[(v_j^{(m+1)})^2 + (v_{j+1}^{(m)})^2 \right]
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\sum_{j=1}^{n-1}\left(v_j^{(m+1)}\right)^2}_{=\frac{1}{h}E^{(m+1)}} + \frac{1}{2}\underbrace{\sum_{j=1}^{n-1}\left(v_{j-1}^{(m)}\right)^2}_{\leq \frac{1}{h}E^{(m)}} + \frac{1}{2}\underbrace{\sum_{j=1}^{n-1}\left(v_{j+1}^{(m)}\right)^2}_{\leq \frac{1}{h}E^{(m)}} \\
&\leq \frac{1}{h}\left(E^{(m+1)} + E^{(m)}\right)
\end{aligned}
$$

Putting these three results together, we get that

$$
\begin{aligned}
E^{(m+1)} - E^{(m)} &\leq \frac{\tau}{h}\left(S^{(m)} - \frac{2}{h}E^{(m)} - 2\frac{\tau}{h^2}S^{(m)} + \frac{1}{h}\left(E^{(m+1)} + E^{(m)}\right)\right) \\
&= \frac{\tau}{h}\left(1 - 2\frac{\tau}{h^2}\right)S^{(m)} + \frac{\tau}{h^2}\left(E^{(m+1)} - E^{(m)}\right),
\end{aligned}
$$

which is equivalent to

$$
\left(1 - \frac{\tau}{h^2}\right)\left(E^{(m+1)} - E^{(m)}\right) \leq \frac{\tau}{h}\left(1 - 2\frac{\tau}{h^2}\right)S^{(m)}.
$$

We already know that $S^{(m)} \leq 0$. As τ and h are positive, the right hand side will be negative if (and only if)

$$
-1 + 2\frac{\tau}{h^2} \leq 0 \qquad \Leftrightarrow \qquad \frac{\tau}{h^2} \leq \frac{1}{2}.
$$

As a consequence, the left hand side has got to be negative, too:

$$
\left(1 - \frac{\tau}{h^2}\right)\left(E^{(m+1)} - E^{(m)}\right) \leq 0.
$$

However, if $\frac{\tau}{h^2} \leq \frac{1}{2}$, then $1 - \frac{\tau}{h^2} > 0$, which means that

$$
E^{(m+1)} - E^{(m)} \leq 0.
$$

Conclusion:

As the final result, we may state that the energy function $E^{(m)}$ is decreasing, if a certain restriction of the size of the time step is enforced:

$$
E^{(m+1)} - E^{(m)} \leq 0 \qquad \text{if} \quad \frac{\tau}{h^2} \leq \frac{1}{2}.
$$

For larger time steps, energy conservation is not guarateed, and the numerical scheme might diverge.

2 Implicit time stepping and numerical energy

For the numerical solution $v^{(m)}$ in time step m, we introduced the discrete energy

$$E^{(m)} := h \sum_{j=1}^{n-1} \left(v_j^{(m)} \right)^2 .$$

Using vector notation, we can also write this as

$$E^{(m)} := h \left(v_j^{(m)} \right)^T v_j^{(m)}.$$

Again, we want to examine whether energy is decreasing with time:

$$E^{(m+1)} \leq E^{(m)}$$

so we will compute the term

$$E^{(m+1)} - E^{(m)} = h \left(\left(v^{(m+1)} \right)^T v^{(m+1)} - \left(v^{(m)} \right)^T v^{(m)} \right)$$

for the implicit scheme.

Energy estimates

According to our implicit scheme, we have

$$v^{(m+1)} = (I + rA)^{-1} v^{(m)} \qquad \text{or} \quad v^{(m+1)} = M v^{(m)},$$

using the *iteration matrix* $M := (I + rA)^{-1}$.

Then

$$\begin{aligned} E^{(m+1)} - E^{(m)} &= h \left(\left(Mv^{(m)} \right)^T Mv^{(m)} - \left(v^{(m)} \right)^T v^{(m)} \right) \\ &= h \left(\left(v^{(m)} \right)^T M^T Mv^{(m)} - \left(v^{(m)} \right)^T v^{(m)} \right) \\ &= h \left(v^{(m)} \right)^T \left(M^T M - I \right) v^{(m)} \end{aligned}$$

The respective term will be ≤ 0 for every $v^{(m)}$, if the matrix $M^T M - I$ is negative definite, i.e. if it has only negative eigenvalues.

Eigenvalues of $M^T M - I$

Let λ be an eigenvalue of $A = \text{tridiag}(-1, 2, -1)$, then

- $1 + r\lambda$ is an eigenvalue of $I + rA$
- $(1 + r\lambda)^{-1}$ is an eigenvalue of $(I + rA)^{-1} = M$

- $\left(\frac{1}{1+r\lambda}\right)^2 - 1$ is an eigenvalue of $M^T M - I$

A is a so-called *weakly diagonal dominant* matrix: for every line i of the matrix, the inequailty

$$|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$$

holds, which means that the absolute value of the diagonal element is never smaller than the sum of the absolute values of all other elements in that line (for A that's actually only two non-zero values).

Such a matrix is positiv definite, i.e. has only positive eigenvalues, if at least for one line i

$$|A_{ii}| > \sum_{j \neq i} |A_{ij}| .$$

This is true for the lines that correspond to the boundary conditions.

Hence, A is positive definite, and therefore:

- all eigenvalues of A are positive,
- all eigenvalues of $I + rA$ are therefore > 1
- all eigenvalues of M are therefore positive, and < 1
- all eigenvalues of $M^T M - I$ are therefore negative

As a direct result, we get

$$E^{(m+1)} - E^{(m)} = h \left(v^{(m)}\right)^T \left(M^T M - I\right) v^{(m)} \leq 0,$$

so for the implicit the energy is never increasing. As a cosequence,

- the size of the time step τ can be chosen independent of mesh size Δx, which means that
- the implicit scheme is always stable.

Appendix: Summation by Parts

Integration by parts

Integration by parts may be derived from the chain rule

$$(uv)' = u'v + uv'. \tag{4}$$

By integrating this equation, we get

$$[uv]_a^b = \int_a^b u'v\,dx + \int_a^b uv'\,dx,$$

which is equivalent to

$$\int_a^b u'v\,dx = [uv]_a^b - \int_a^b uv'\,dx. \tag{5}$$

Can we find a similar rule for summation?

Summation by parts

We start by observing that

$$y_{j+1}z_{j+1} - y_jz_j = \left(y_{j+1} - y_j\right) z_j + \left(z_{j+1} - z_j\right) y_{j+1},$$

where you should note the similarity to equation (4)!

By summing up for $j = 0$ to $j = n-1$, we get:

$$\begin{aligned}
&\sum_{j=0}^{n-1} \left(y_{j+1}z_{j+1} - y_jz_j\right) &&= \sum_{j=0}^{n-1} \left(y_{j+1} - y_j\right) z_j + \sum_{j=0}^{n-1} \left(z_{j+1} - z_j\right) y_{j+1} \\
\Leftrightarrow \quad &\sum_{j=0}^{n-1} y_{j+1}z_{j+1} - \sum_{j=0}^{n-1} y_jz_j &&= \sum_{j=0}^{n-1} \left(y_{j+1} - y_j\right) z_j + \sum_{j=0}^{n-1} \left(z_{j+1} - z_j\right) y_{j+1} \\
\Leftrightarrow \quad &y_nz_n - y_0z_0 &&= \sum_{j=0}^{n-1} \left(y_{j+1} - y_j\right) z_j + \sum_{j=0}^{n-1} \left(z_{j+1} - z_j\right) y_{j+1}.
\end{aligned}$$

As a result, we get that

$$\sum_{j=0}^{n-1} z_j \left(y_{j+1} - y_j\right) = y_nz_n - y_0z_0 - \sum_{j=0}^{n-1} y_{j+1} \left(z_{j+1} - z_j\right), \tag{6}$$

which is the discrete equivalent to equation (5). We will refer to this rule as *summation by parts*.

Second "derivatives"

In the continous case, we may use integration by parts to obtain

$$\int_a^b u'' v \, dx = \left[u'v\right]_a^b - \int_a^b u'v' \, dx. \tag{7}$$

As we will see, the discrete analogon also holds. We will use a slightly modified version of equation (6)

$$\sum_{j=1}^{n-1} z_j \left(y_{j+1} - y_j\right) = y_n z_n - y_1 z_1 - \sum_{j=1}^{n-1} y_{j+1} \left(z_{j+1} - z_j\right), \tag{8}$$

where we only changed the starting index. We now set $z_j = v_j$, and $y_j = u_j - u_{j-1}$, or

$$y_{j+1} - y_j = (u_{j+1} - u_j) - (u_j - u_{j-1}) = u_{j+1} - 2u_j + u_{j-1},$$

and substitute the respective terms in equation (8) to get

$$\sum_{j=1}^{n-1} v_j \left(u_{j+1} - 2u_j + u_{j-1}\right) = (u_n - u_{n-1})v_n - (u_1 - u_0)v_1 - \sum_{j=1}^{n-1} \left(u_{j+1} - u_j\right) \left(v_{j+1} - v_j\right). \tag{9}$$

Homogeneous Dirichlet boundaries

If the unknowns u_j and v_j result from a discretization of a PDE, we will often have situations where $u_0 = v_0 = 0$, or $u_n = v_n = 0$ ("homogeneous Dirichlet boundaries"). In that case, summation by parts leads to the follwoing equations:

$$\sum_{j=0}^{n-1} z_j \left(y_{j+1} - y_j\right) = -\sum_{j=0}^{n-1} y_{j+1} \left(z_{j+1} - z_j\right), \tag{10}$$

and

$$\sum_{j=1}^{n-1} v_j \left(u_{j+1} - 2u_j + u_{j-1}\right) = -u_1 v_1 - \sum_{j=1}^{n-1} \left(u_{j+1} - u_j\right) \left(v_{j+1} - v_j\right). \tag{11}$$