Practical Course: Modeling, Simulation, Optimization

Week 1

Daniël Veldman

Chair in Dynamics, Control, and Numerics, Friedrich-Alexander-University Erlangen-Nürnberg

Contents

1.A Organization

Practical course: modeling, simulation, optimization

Lecturer:

- Dr. Daniël Veldman (e-mail: daniel.veldman@math.fau.de)

Main objective

Gain *practical* experience in modeling, simulation, and optimization for physical systems governed by partial differential equations.

Main topics:

- Modelling, analysis, simulation and/or optimization of problems in engineering or the natural sciences
- Numerical algorithms for partial differential equation models (finite differences, finite elements, etc)
- Continuous optimization and optimal control

Structure

12 lectures divided over 3 parts:

Part I: Finite differences (FD)

- Lecture 1: 1-D Poisson equation
- Lecture 2: 2-D Poisson equation
- Lecture 3: Time-discretization
- Lecture 4: Advection-Diffusion equations

Part II: Finite elements (FE)

- Lecture 5: Finite elements in 1-D
- Lecture 6: Finite elements in 2-D
- Lecture 7: 2-D elasticity
- Lecture 8: Beam models

Part III: Optimization

- Lecture 9: Static optimization and gradient descent
- Lecture 10: Hessians and step size selection
- Lecture 11: Dynamic optimal control
- Lecture 12: Control of neural ODEs

For each lecture there is a corresponding MATLAB exercise.
There are also 3 bonus MATLAB questions for each part covering some extensions.

1.B 1-D Conservation laws

Continuum mechanics

Do not model individual molecules but instead consider averaged quantities.

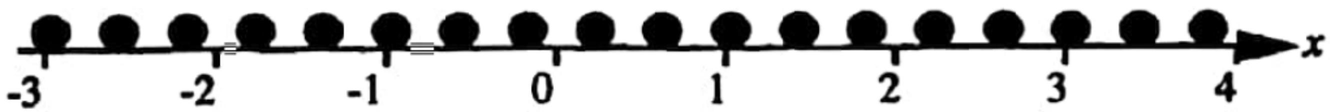

Figure: The coordinate system for a 1-D continuum with an example distribution of molecules [Roberts, 1994]

Examples:

- mass density $\rho(\mathbf{x})$ [kg/m^3] instead of positions of individual molecules
- temperature $T(\mathbf{x})$ [K] instead of kinetic energy of individual particles.

Note: The continuum assumption is only justified when the *representative physical length scale* L [m] is much larger than the *mean free path length of molecules* λ [m], i.e. when

$$Kn = \frac{\lambda}{L} \ll 1,$$

where Kn is the Knudsen number.

Conservation of mass

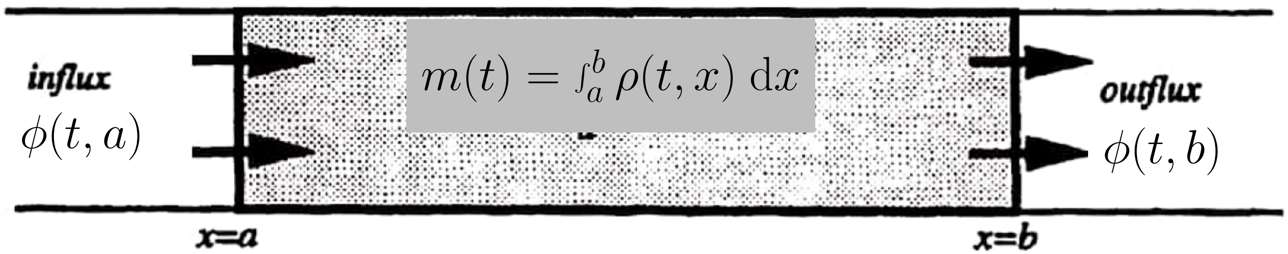

Figure: A slice of a 1-D continuum to investigate the conservation of mass (from [Roberts, 1994] but edited)

The mass $m(t)$ [kg] in $[a,b]$ changes only because of the mass fluxes $\phi(\cdot,a)$ and $\phi(\cdot,b)$ [kg/s]:

$$\frac{\partial m}{\partial t} = \phi(t,a) - \phi(t,b).$$

Conservation of mass

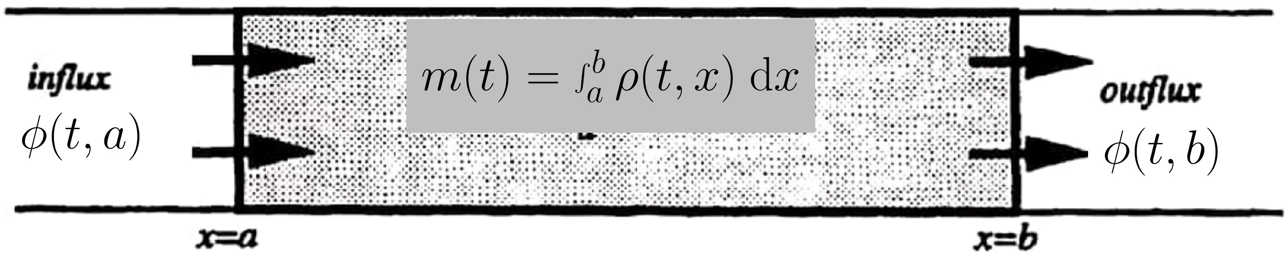

Figure: A slice of a 1-D continuum to investigate the conservation of mass (from [Roberts, 1994] but edited)

The mass $m(t)$ [kg] in $[a, b]$ changes only because of the mass fluxes $\phi(\cdot, a)$ and $\phi(\cdot, b)$ [kg/s]:

$$\frac{\partial m}{\partial t} = \phi(t,a) - \phi(t,b).$$

Using the fundamental theorem of calculus, we find

$$\int_a^b \frac{\partial \rho}{\partial t}(t,x)\,\mathrm{d}x = -\int_a^b \frac{\partial \phi}{\partial x}(t,x)\,\mathrm{d}x.$$

Conservation of mass

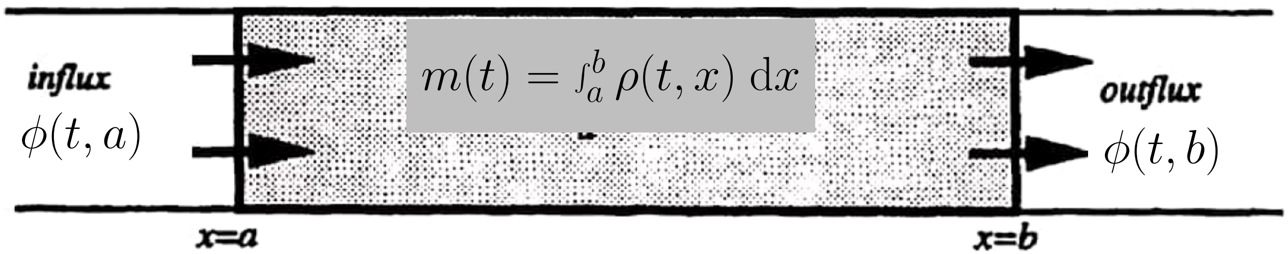

Figure: A slice of a 1-D continuum to investigate the conservation of mass (from [Roberts, 1994] but edited)

The mass $m(t)$ [kg] in $[a, b]$ changes only because of the mass fluxes $\phi(\cdot, a)$ and $\phi(\cdot, b)$ [kg/s]:

$$\frac{\partial m}{\partial t} = \phi(t, a) - \phi(t, b).$$

Using the fundamental theorem of calculus, we find

$$\int_a^b \frac{\partial \rho}{\partial t}(t, x)\,\mathrm{d}x = -\int_a^b \frac{\partial \phi}{\partial x}(t, x)\,\mathrm{d}x.$$

Because this holds for any interval $[a, b]$ in the domain $\Omega = [0, L]$:

Conservation of mass in 1-D

$$\frac{\partial \rho}{\partial t}(t, x) = -\frac{\partial \phi}{\partial x}(t, x).$$

Completing the model

$$\frac{\partial \rho}{\partial t}(t,x) = -\frac{\partial \phi}{\partial x}(t,x).$$

To complete the model, we need a *constitutive relation*
that relates the mass flux $\phi(t,x)$ [kg/s] to the mass density $\rho(t,x)$ [kg/m].

Two commonly used constitutive relations:

Fick's law

$$\phi(t,x) = -\kappa(t,x)\frac{\partial \rho}{\partial x}(t,x).$$

The coefficient $\kappa(t,x)$ [m^2/s] is called the diffusivity.
'Mass flows from locations with high concentrations to locations with low concentrations'

Advective transport

$$\phi(t,x) = v(t,x)\rho(t,x).$$

The velocity field $v(t,x)$ [m/s] is given.
'Mass flows along the velocity field $v(t,x)$'

Energy conservation

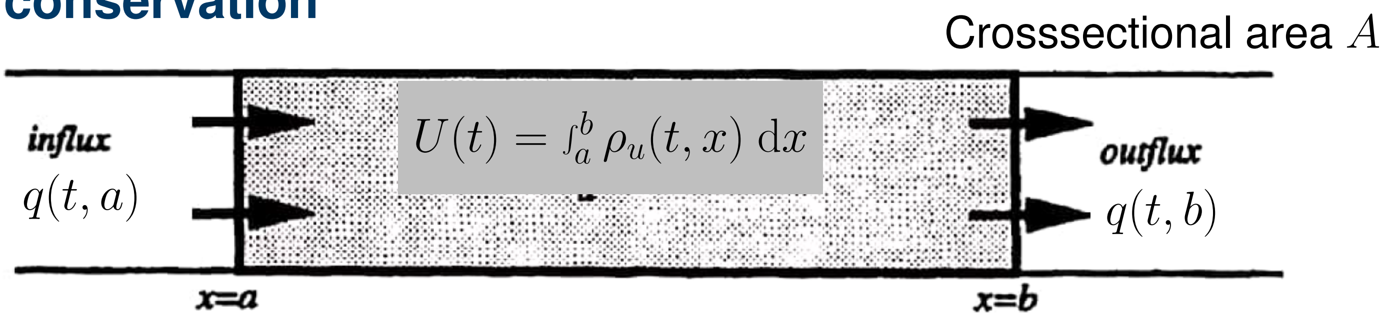

Figure: A slice of a 1-D continuum to investigate heat conduction (from [Roberts, 1994] but edited)

The internal energy $U(t)$ [J] in $[a,b]$ only changes because of the heat fluxes $q(\cdot,a)$ and $q(\cdot,b)$ [W/m^2] and the heat $\int_a^b Q(x)\,\mathrm{d}x$ [W] generated inside $[a,b]$:

$$\frac{\partial U}{\partial t} = Aq(t,a) - Aq(t,b) + \int_a^b Q(t,x)\,\mathrm{d}x.$$

Energy conservation

Crosssectional area A

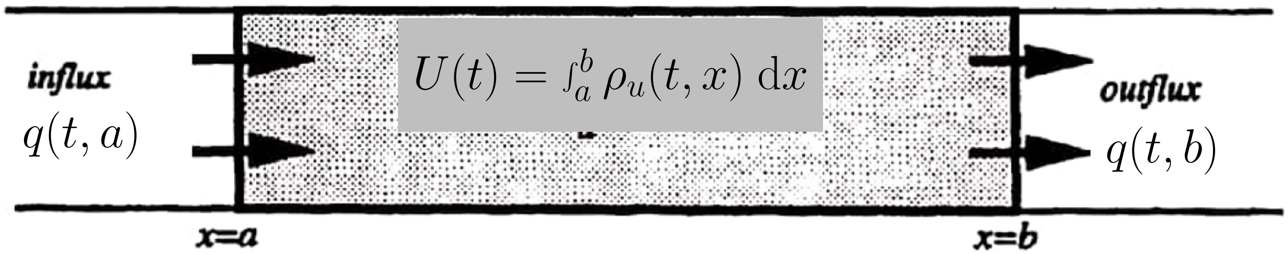

Figure: A slice of a 1-D continuum to investigate heat conduction (from [Roberts, 1994] but edited)

The internal energy $U(t)$ [J] in $[a,b]$ only changes because of the heat fluxes $q(\cdot,a)$ and $q(\cdot,b)$ [W/m^2] and the heat $\int_a^b Q(x)\,\mathrm{d}x$ [W] generated inside $[a,b]$:

$$\frac{\partial U}{\partial t} = Aq(t,a) - Aq(t,b) + \int_a^b Q(t,x)\,\mathrm{d}x.$$

Using the fundamental theorem of calculus, we find

$$\int_a^b \frac{\partial \rho_u}{\partial t}(t,x)\,\mathrm{d}x = \int_a^b \left(-A\frac{\partial q}{\partial x}(t,x) + Q(t,x)\right)\,\mathrm{d}x.$$

Because this holds for any interval $[a,b]$ in the domain $\Omega = [0,L]$:

Energy conservation in 1-D

$$\frac{\partial \rho_u}{\partial t}(t,x) = -A\frac{\partial q}{\partial x}(t,x) + Q(t,x).$$

Completing the model

$$\frac{\partial \rho_u}{\partial t}(t,x) = -A\frac{\partial q}{\partial x}(t,x) + Q(t,x).$$

We again need constitutive relations to complete the model.

Fourier's law of heat conduction

$$q(t,x) = -k\frac{\partial T}{\partial x}(t,x).$$

The coefficient k [W/m/K] is the thermal conductivity and $T(t,x)$ [K] is the temperature.
'Heat flows from locations with high temperatures to locations with low temperatures'

Internal energy

$$\rho_u(t,x) = cAT(t,x).$$

The coefficient c [J/K/m^3] heat capacity per unit length.

1.C 1-D Finite differences

Finite differences

Suppose we want to approximate the solution $u(x)$ of the boundary value problem

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + f(x) = 0, \qquad x \in (0, L),$$

$$u(0) = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}x}(L) = 0.$$

Finite differences

Suppose we want to approximate the solution $u(x)$ of the boundary value problem

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + f(x) = 0, \qquad x \in (0, L),$$
$$u(0) = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}x}(L) = 0.$$

Introduce an M-point grid in the interval $[0, L]$ with a grid spacing $\Delta x = L/(M-1)$

Also introduce $f_m = f(x_m)$ and the approximation $u_m \approx u(x_m)$.

Finite difference approximation (2nd derivative)

We want to write a system of equations in terms of f_m $(= f(x_m))$ and u_m $(\approx u(x_m))$.

Observe that (for $u \in C^4([0, L])$)

$$u(x + \Delta x) = u(x) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + \frac{(\Delta x)^2}{2} \frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + \frac{(\Delta x)^3}{6} \frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x) + O((\Delta x)^4),$$

$$u(x - \Delta x) = u(x) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + \frac{(\Delta x)^2}{2} \frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) - \frac{(\Delta x)^3}{6} \frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x) + O((\Delta x)^4).$$

Adding these two equations:

$$u(x + \Delta x) + u(x - \Delta x) = 2u(x) + (\Delta x)^2 \frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + O((\Delta x)^4).$$

Rearranging and dividing by $(\Delta x)^2$ yields

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) = \frac{u(x + \Delta x) - 2u(x) + u(x - \Delta x)}{(\Delta x)^2} + O((\Delta x)^2).$$

Finite difference approximation (for the 2nd derivative)

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x_m) \approx \frac{u_{m+1} - 2u_m + u_{m-1}}{(\Delta x)^2}.$$

Finite difference approximation (1st derivative)

We want to write a system of equations in terms of f_m $(= f(x_m))$ and u_m $(\approx u(x_m))$.

Observe that (for $u \in C^4([0, L])$)

$$u(x + \Delta x) = u(x) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + \frac{(\Delta x)^2}{2}\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + \frac{(\Delta x)^3}{6}\frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x) + O((\Delta x)^4),$$
$$u(x - \Delta x) = u(x) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + \frac{(\Delta x)^2}{2}\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) - \frac{(\Delta x)^3}{6}\frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x) + O((\Delta x)^4).$$

Subtracting these two equations:

$$u(x + \Delta x) - u(x - \Delta x) = 2\Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + O((\Delta x)^3).$$

Rearranging and dividing by $2\Delta x$ yields

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x) = \frac{u(x + \Delta x) - u(x - \Delta x)}{2\Delta x} + O((\Delta x)^2).$$

(Centered) finite difference approximation (for the 1st derivative)

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x_m) \approx \frac{u_{m+1} - u_{m-1}}{2\Delta x}.$$

Intermezzo: left- and right-sided finite differences

For right-sided finite differences, observe that

$$u(x+\Delta x) = u(x) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + O((\Delta x)^2).$$

Rearranging yields

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x) = \frac{u(x+\Delta x) - u(x)}{\Delta x} + O(\Delta x).$$

Right-sided finite differences

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x_n) \approx \frac{u_{m+1} - u_m}{\Delta x}.$$

For left-sided finite differences, we do a similar derivation starting from

$$u(x-\Delta x) = u(x) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x) + O((\Delta x)^2).$$

Left-sided finite differences

$$\frac{\mathrm{d}u}{\mathrm{d}x}(x_m) \approx \frac{u_m - u_{m-1}}{\Delta x}.$$

Note: the error in left- and right-sided finite differences is $O(\Delta x)$, not $O((\Delta x)^2)$!

Boundary conditions

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x) + f(x) = 0, \qquad x \in (0, L), \qquad \frac{u_{m+1} - 2u_m + u_{m-1}}{\Delta x^2} + f_m = 0.$$

$$u(0) = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}x}(L) = 0.$$

Boundary conditions

$$\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x) + f(x) = 0, \qquad x \in (0, L), \qquad \frac{u_{m+1} - 2u_m + u_{m-1}}{\Delta x^2} + f_m = 0.$$

$$u(0) = 0, \qquad \frac{\mathrm{d}u}{\mathrm{d}x}(L) = 0.$$

Introduce two fictitious ('ghost') points x_0 and x_{M+1}.

Two additional equations for the boundary conditions:

- The boundary condition $u(0) = 0$ becomes $u_1 = 0$.
- For the boundary condition $\frac{\mathrm{d}u}{\mathrm{d}x}(L) = 0$, we use that

$$\frac{\mathrm{d}u}{\mathrm{d}x}(L) = \frac{u(x_{M+1}) - u(x_{M-1})}{2\Delta x} + O(\Delta x^2), \qquad \frac{u_{M+1} - u_{M-1}}{2\Delta x} = 0.$$

Matrix formulation (implicit formulation for the BCs)

We now have a set of $M+2$ linear equations:

$$\frac{u_{m+1} - 2u_m + u_{m-1}}{\Delta x^2} + f_m = 0, \qquad (m = 1, 2, \ldots, M),$$

$$u_1 = 0, \qquad \frac{u_{M+1} - u_{M-1}}{\Delta x} = 0.$$

Matrix formulation (implicit formulation for the BCs)

We now have a set of $M+2$ linear equations:

$$\frac{u_{m+1} - 2u_m + u_{m-1}}{\Delta x^2} + f_m = 0, \qquad (m = 1, 2, \ldots, M),$$

$$u_1 = 0, \qquad \frac{u_{M+1} - u_{M-1}}{\Delta x} = 0.$$

Write these equations in matrix form

$$\frac{1}{\Delta x^2}\begin{bmatrix} 0 & \Delta x^2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & & 0 & 0 & 0 & 0 \\ \vdots & \vdots & & & \ddots & & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{-\Delta x}{2} & 0 & \frac{\Delta x}{2} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{M-2} \\ u_{M-1} \\ u_M \\ u_{M+1} \end{bmatrix} + \begin{bmatrix} 0 \\ f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{M-2} \\ f_{M-1} \\ f_M \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Matrix formulation (towards an explicit formulation for the BCs)

$$\frac{1}{\Delta x^2}\begin{bmatrix} 0 & \Delta x^2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & & 0 & 0 & 0 & 0 \\ \vdots & \vdots & & & \ddots & & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & \frac{-\Delta x}{2} & 0 & \frac{\Delta x}{2} \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{M-2} \\ u_{M-1} \\ u_M \\ u_{M+1} \end{bmatrix} + \begin{bmatrix} 0 \\ f_1 \\ f_2 \\ f_3 \\ \vdots \\ f_{M-2} \\ f_{M-1} \\ f_M \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Step 1:

- The first equation states that $u_1 = 0$ (first row) $\Rightarrow$ we can eliminate u_1.
- The last line states that $u_{M+1} = u_{M-1}$ (last row) $\Rightarrow$ we can eliminate u_{M+1}.

$\Rightarrow$ Delete the first and last row.
$\Rightarrow$ Delete the second and last column.

Step 2:

- We are not interested in the value u_0, because x_0 is a fictitious (ghost) node.
- u_0 only appears in the first equation.

$\Rightarrow$ Delete the first row and column.

Matrix formulation (explicit formulation for the BCs)

(Following the steps from the previous slide)

$$\frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & & 0 & 0 & 0 & \\ 1 & -2 & & 0 & 0 & 0 & \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & & -2 & 1 & 0 & \\ 0 & 0 & & 1 & -2 & 1 & \\ 0 & 0 & \cdots & 0 & 2 & -2 & \end{bmatrix}\begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{M-2} \\ u_{M-1} \\ u_M \end{bmatrix} + \begin{bmatrix} f_2 \\ f_3 \\ \vdots \\ f_{M-2} \\ f_{M-1} \\ f_M \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Matrix formulation (explicit formulation for the BCs)

(Following the steps from the previous slide)

$$\frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & & 0 & 0 & 0 & \\ 1 & -2 & & 0 & 0 & 0 & \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & & -2 & 1 & 0 & \\ 0 & 0 & & 1 & -2 & 1 & \\ 0 & 0 & \cdots & 0 & 2 & -2 & \end{bmatrix}\begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_{M-2} \\ u_{M-1} \\ u_M \end{bmatrix} + \begin{bmatrix} f_2 \\ f_3 \\ \vdots \\ f_{M-2} \\ f_{M-1} \\ f_M \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Both the explicit and implicit formulation lead to a linear system of the form

$$\mathbf{A}\mathbf{u} + \mathbf{f} = 0.$$

This system can be solved to determine $\mathbf{u}$.

The implicit formulation for the BCs contains $M+2$ unknowns.
The explicit formulation for the BCs contains M unknowns.

In the implicit formulation, $\mathbf{u}$ contains the fictious ('ghost') values u_0 and u_{M+1}.
In the explicit formulation, $\mathbf{u}$ does not contain u_0 and u_{M+1}.

1.D Convergence analysis for 1-D finite differences

What about convergence?

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

In our example, $F(u) = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + f$.

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

In our example, $\mathbf{F}_{\Delta x}(\mathbf{u}) = \mathbf{A}\mathbf{u} + \mathbf{f}$.

What about convergence?

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

In our example, $F(u) = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + f$.

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

In our example, $\mathbf{F}_{\Delta x}(\mathbf{u}) = \mathbf{A}\mathbf{u} + \mathbf{f}$.

Let $u(\mathbf{x})$ denote the continuous solution evaluated in the grid points.

What can we say about the error $\mathbf{e} = \mathbf{u} - u(\mathbf{x})$ when $\Delta x \to 0$?

What about convergence?

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

In our example, $F(u) = \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + f$.

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

In our example, $\mathbf{F}_{\Delta x}(\mathbf{u}) = \mathbf{A}\mathbf{u} + \mathbf{f}$.

Let $u(\mathbf{x})$ denote the continuous solution evaluated in the grid points.

What can we say about the error $\mathbf{e} = \mathbf{u} - u(\mathbf{x})$ when $\Delta x \to 0$?

Definition (Convergent numerical scheme)

The numerical scheme is convergent if the error $||\mathbf{e}|| \to 0$ when $\Delta x \to 0$.

Note: the problem is subtle because the number of elements in $\mathbf{e}$ grows when $\Delta x \to 0$.
Convergence in one norm does not (always) imply convergence in another norm!

Two criteria for a convergent scheme

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

Let $u(\mathbf{x})$ denote the continuous solution evaluated in the grid points.

Two criteria for a convergent scheme

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

Let $u(\mathbf{x})$ denote the continuous solution evaluated in the grid points.

Theorem (Lax)

The numerical scheme is convergent if it is both

- *consistent and*
- *stable.*

Definition (Consistent numerical scheme)

The numerical scheme is consistent iff $\mathbf{F}_{\Delta x}(u(\mathbf{x})) = O(\Delta x^p)$ for some $p > 0$.

Two criteria for a convergent scheme

Two ingredients:

1) A continuous (ODE, PDE) problem with a continuous solution $u(x)$.

$$F(u) = 0.$$

2) A numerical (FD) scheme with a discrete solution $\mathbf{u}$.

$$\mathbf{F}_{\Delta x}(\mathbf{u}) = 0.$$

Let $u(\mathbf{x})$ denote the continuous solution evaluated in the grid points.

Theorem (Lax)

The numerical scheme is convergent if it is both

- *consistent and*
- *stable.*

Definition (Consistent numerical scheme)

The numerical scheme is consistent iff $\mathbf{F}_{\Delta x}(u(\mathbf{x})) = O(\Delta x^p)$ for some $p > 0$.

Stability question: does $\mathbf{F}_{\Delta x}(\mathbf{u}) - \mathbf{F}_{\Delta x}(u(\mathbf{x})) = O(\Delta x^p)$ imply $\mathbf{u} - u(\mathbf{x}) = O(\Delta x^p)$?

Definition (Stable numerical scheme (for a static problem))

The numerical scheme is stable if there exists a constant K independent of Δx s.t.

$$|\mathbf{u} - u(\mathbf{x})| \leq K|\mathbf{F}_{\Delta x}(\mathbf{u}) - \mathbf{F}_{\Delta x}(u(\mathbf{x}))|$$

Consistency (the easy part)

By Taylor's theorem, we have

$$u(x_{m+1}) = u(x_m) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x_m) + \frac{\Delta x^3}{6}\frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4 u}{\mathrm{d}x^4}(\xi_m^+),$$
$$u(x_{m-1}) = u(x_m) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2 u}{\mathrm{d}x^2}(x_m) - \frac{\Delta x^3}{6}\frac{\mathrm{d}^3 u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4 u}{\mathrm{d}x^4}(\xi_m^-),$$

for certain $\xi_m^+ \in [x_m, x_{m+1}]$ and $\xi_m^- \in [x_{m-1}, x_m]$.

Consistency (the easy part)

By Taylor's theorem, we have

$$u(x_{m+1}) = u(x_m) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) + \frac{\Delta x^3}{6}\frac{\mathrm{d}^3u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^+),$$
$$u(x_{m-1}) = u(x_m) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) - \frac{\Delta x^3}{6}\frac{\mathrm{d}^3u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^-),$$

for certain $\xi_m^+ \in [x_m, x_{m+1}]$ and $\xi_m^- \in [x_{m-1}, x_m]$.
Adding these two equations shows that

$$u(x_{m+1}) + u(x_{m-1}) = 2u(x_m) + \Delta x^2 \frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) + \frac{\Delta x^4}{24}\left(\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^+) + \frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^-)\right).$$

Consistency (the easy part)

By Taylor's theorem, we have

$$\begin{aligned} u(x_{m+1}) &= u(x_m) + \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) + \frac{\Delta x^3}{6}\frac{\mathrm{d}^3u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^+), \\ u(x_{m-1}) &= u(x_m) - \Delta x \frac{\mathrm{d}u}{\mathrm{d}x}(x_m) + \frac{\Delta x^2}{2}\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) - \frac{\Delta x^3}{6}\frac{\mathrm{d}^3u}{\mathrm{d}x^3}(x_m) + \frac{\Delta x^4}{24}\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^-), \end{aligned}$$

for certain $\xi_m^+ \in [x_m, x_{m+1}]$ and $\xi_m^- \in [x_{m-1}, x_m]$.
Adding these two equations shows that

$$u(x_{m+1}) + u(x_{m-1}) = 2u(x_m) + \Delta x^2 \frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) + \frac{\Delta x^4}{24}\left(\frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^+) + \frac{\mathrm{d}^4u}{\mathrm{d}x^4}(\xi_m^-)\right).$$

Rearranging shows that

$$\frac{u(x_{m+1}) - 2u(x_m) + u(x_{m-1})}{\Delta x^2} - \frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) = O(\Delta x^2).$$

Note that $\frac{\mathrm{d}^2u}{\mathrm{d}x^2}(x_m) + f(x_m) = 0$.
Similarly, we can check that $u(x)$ also satisfies the discretized BCs up to $O(\Delta x^2)$.
Therefore,

$$\mathbf{A}u(\mathbf{x}) + \mathbf{f} = O(\Delta x^2).$$

Stability (the hard part)

By definition of $\mathbf{u}$

$$\mathbf{A}\mathbf{u} + \mathbf{f} = 0.$$

From the previous slide

$$\mathbf{A}u(\mathbf{x}) + \mathbf{f} = O(\Delta x^2).$$

Stability (the hard part)

By definition of $\mathbf{u}$

$$\mathbf{A}\mathbf{u} + \mathbf{f} = 0.$$

From the previous slide

$$\mathbf{A}u(\mathbf{x}) + \mathbf{f} = O(\Delta x^2).$$

Therefore also,

$$\mathbf{A}(\mathbf{u} - u(\mathbf{x})) = \mathbf{A}\mathbf{e} = O(\Delta x^2).$$

Stability (the hard part)

By definition of $\mathbf{u}$

$$\mathbf{A}\mathbf{u} + \mathbf{f} = 0.$$

From the previous slide

$$\mathbf{A}u(\mathbf{x}) + \mathbf{f} = O(\Delta x^2).$$

Therefore also,

$$\mathbf{A}(\mathbf{u} - u(\mathbf{x})) = \mathbf{A}\mathbf{e} = O(\Delta x^2).$$

But we need to bound $|\mathbf{e}|$, not $|\mathbf{A}\mathbf{e}|$!
Stability: there exists a constant K independent of Δx such that $|\mathbf{e}| \leq K|\mathbf{A}\mathbf{e}|$.
(in which norm $|\cdot|$?)

Stability (the hard part)

By definition of $\mathbf{u}$

$$\mathbf{A}\mathbf{u} + \mathbf{f} = 0.$$

From the previous slide

$$\mathbf{A}u(\mathbf{x}) + \mathbf{f} = O(\Delta x^2).$$

Therefore also,

$$\mathbf{A}(\mathbf{u} - u(\mathbf{x})) = \mathbf{A}\mathbf{e} = O(\Delta x^2).$$

But we need to bound $|\mathbf{e}|$, not $|\mathbf{A}\mathbf{e}|$!
Stability: there exists a constant K independent of Δx such that $|\mathbf{e}| \leq K|\mathbf{A}\mathbf{e}|$.
(in which norm $| \cdot |$?)

It can be shown that we can take the ℓ^∞-norm and $K = \frac{1}{8}$.
The argument is based on the discrete maximum principle.
(see e.g. [L. Chen, Finite difference methods for Poisson equation])

We conclude there exists a constant C such that

$$|\mathbf{u} - u(\mathbf{x})|_\infty \leq C(\Delta x)^2,$$