

## Assignment 5

### TTK4130 Modeling and Simulation

**Problem 1 (Tank and valve-models, mass balances, linearization. 24 %)**

*NB: This is a computer exercise, and can therefore be solved in groups of 2 students. If you do so, please write down the name of your group partner in your answer.*

Consider the system of two open tanks shown in Figure 1. Tank 1 has area  $A_1$  and level  $h_1$ , while tank 2 has area  $A_2$  and level  $h_2$ . Moreover, the volume flow input  $q_i$  enters tank 1, the volume flow  $q_{12}$  runs through the valve that connects the tanks, and the volume flow  $q_o$  runs from tank 2 to the outside.

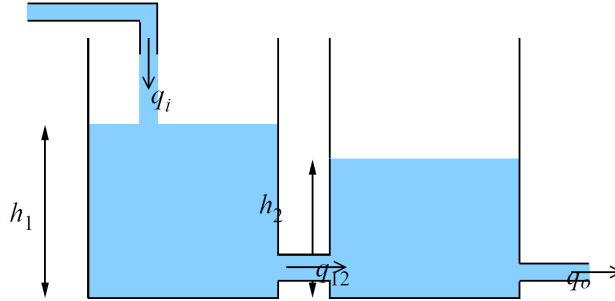


Figure 1: The two-tank system.

Assume that the valve is located at level  $h_1 = h_2 = 0$ , and that  $h_1 \geq h_2$ . Hence, the volume flow through the valves can be modeled as

$$q_{12} = C_{12} \sqrt{p_1 - p_2} \quad (1a)$$

$$q_o = C_o \sqrt{p_2 - p_o}, \quad (1b)$$

where  $p_1$  and  $p_2$  are the pressures at the bottom of tank 1 and 2, respectively,  $p_o$  is the atmospheric pressure, and  $C_{12}$  and  $C_o$  are valve constants. Furthermore, assume that the fluid that runs through the tank system has constant density  $\rho$ .

(a) By using a mass balance for each tank, show that the level of each tank is given by

$$\dot{h}_1 = \frac{1}{A_1} (q_i - C_{12} \sqrt{\rho g} \sqrt{h_1 - h_2}) \quad (2a)$$

$$\dot{h}_2 = \frac{\sqrt{\rho g}}{A_2} (C_{12} \sqrt{h_1 - h_2} - C_o \sqrt{h_2}). \quad (2b)$$

**Solution:** The mass balance for each tank is

$$\frac{d}{dt}(\rho A_1 h_1) = \rho q_i - \rho q_{12}$$

$$\frac{d}{dt}(\rho A_2 h_2) = \rho q_{12} - \rho q_o$$

Since the tanks are open tanks, the pressures at the bottom of them are given by

$$p_1 = p_o + \rho g h_1$$

$$p_2 = p_o + \rho g h_2.$$

Hence, (2) follows from the equations above and (1).

- (b) For a constant input  $q_i = q_i^*$ , find the equilibrium point  $[h_1, h_2]^T = [h_1^*, h_2^*]^T$ , and express it as a function of  $q_i^*$ ,  $C_{12}$ ,  $C_o$ ,  $\rho$ , and  $g$ .

Linearize the system around  $[h_1, h_2]^T = [h_1^*, h_2^*]^T$  and  $q_i = q_i^*$ , and write the linearized system in state-space form. Furthermore, express the state and the input matrix as a function of  $q_i^*$ ,  $C_{12}$ ,  $C_o$ ,  $\rho$ ,  $g$ ,  $A_1$  and  $A_2$ .

What happens with the eigenvalues of the linearized system when  $h_1^*, h_2^* \rightarrow 0$  or  $q_i^* \rightarrow 0$ ?

**Solution:** The equilibrium point  $[h_1, h_2]^T = [h_1^*, h_2^*]^T$  solves the equations

$$0 = q_i^* - C_{12}\sqrt{\rho g}\sqrt{h_1^* - h_2^*}$$

$$0 = C_{12}\sqrt{h_1^* - h_2^*} - C_o\sqrt{h_2^*}.$$

Hence,

$$h_1^* = \left( \frac{1}{C_{12}^2} + \frac{1}{C_o^2} \right) \frac{q_i^{*2}}{\rho g}$$

$$h_2^* = \frac{1}{C_o^2} \frac{q_i^{*2}}{\rho g}.$$

The linearized system has the form  $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u$ , where  $\mathbf{x} = [\Delta h_1, \Delta h_2]^T$ ,  $u = \Delta q_i$ ,

$$\mathbf{A} = \begin{bmatrix} -\frac{C_{12}^2 \rho g}{2A_1 q_i^*} & \frac{C_{12}^2 \rho g}{2A_1 q_i^*} \\ \frac{C_{12}^2 \rho g}{2A_2 q_i^*} & -\frac{(C_{12}^2 + C_o^2) \rho g}{2A_2 q_i^*} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \frac{1}{A_1} \\ 0 \end{bmatrix}.$$

The characteristic polynomial of the state matrix is

$$\lambda^2 + \frac{(C_{12}^2(A_1 + A_2) + C_o^2 A_1) \rho g}{2A_1 A_2 q_i^*} \lambda + \frac{C_{12}^2 C_o^2 \rho^2 g^2}{4A_1 A_2 q_i^{*2}}.$$

Hence, the eigenvalues are

$$\lambda_{1,2} = -\frac{\rho g}{4A_1 A_2 q_i^*} \left( (C_{12}^2(A_1 + A_2) + C_o^2 A_1) \pm \sqrt{(C_{12}^2(A_1 + A_2) + C_o^2 A_1)^2 - 4A_1 A_2 C_{12}^2 C_o^2} \right).$$

Therefore we conclude that if  $h_1^*, h_2^* \rightarrow 0$ , i.e.  $q_i^* \rightarrow 0$ , then  $\lambda_1, \lambda_2 \rightarrow \infty$ .

- (c) Implement and simulate the model (2) in Matlab, using the ODE solver ode45. Use the initial conditions  $h_1 = 2$  m,  $h_2 = 1$  m, and the parameter values  $q_i = 0$ ,  $C_{12} = C_o = 0.15 \text{ m}^3/\text{s}\sqrt{\text{Pa}}$ ,  $A_1 = A_2 = 4.5 \text{ m}^2$ ,  $\rho = 1000 \text{ m}^3/\text{kg}$  and  $g = 9.81 \text{ m/s}^2$ . Simulate for 2 second.

Add your Matlab script and a plot with  $h_1$  and  $h_2$  as a function of time to your answer.

Furthermore, show the step lengths by using the plot-command `plot(t, h, 'o-')`.

Comment on the step lengths based on your answer to part (b).

What is wrong with the valve equations (1)?

*Hint: Read sections 4.2.2 and 4.2.3 in the book.*

**Solution:**

```
% initial conditions and parameters
q_i = 0; h_1_0 = 2; h_2_0 = 1;
```

```

C_12 = 0.15; C_o = 0.15;
A_1 = 4.5; A_2 = 4.5;
rho = 1000; g = 9.81;
Tsim = 2.0;
% ODE
K_12 = C_12*sqrt(rho*g);
K_o = C_o*sqrt(rho*g);
f = @(t,h) ([1/A_1*(q_i - K_12*sqrt(h(1)-h(2)))
            1/A_2*(K_12*sqrt(h(1)-h(2)) - K_o*sqrt(h(2)))]);
% Simulate
[t,h] = ode45(f,[0, Tsim],[h_1_0, h_2_0]);
% Plot
figure
hold on
plot(t,h(:,1),'o-'); plot(t,h(:,2),'o-')
ylabel('h [m]'); xlabel('t [s]')
hold off

```

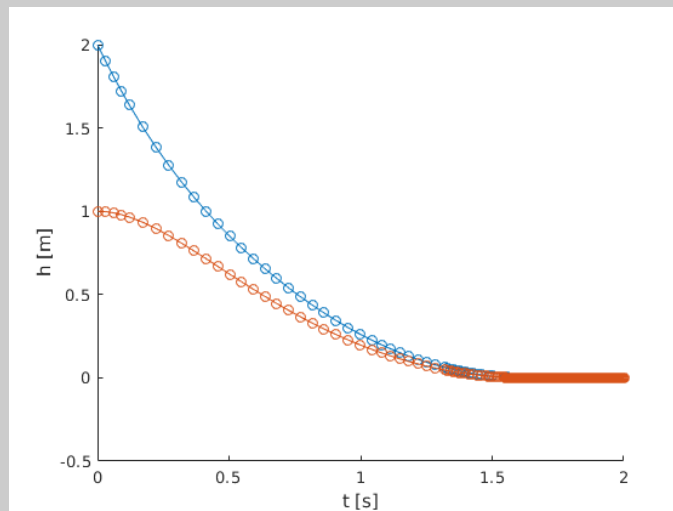


Figure 2: Two-tank system: Simulation results.

As we can see, the step lengths become very small. This is as expected since the eigenvalues become arbitrarily large when the levels of the tanks approach zero.

The valve equations in (1) are based on the assumption of turbulent flows. For low tank levels, the flows will at some point become laminar. Hence, these valve equations will no longer give a realistic model. The solution to this problem is to regularize the valve model, as explained in Section 4.2.3 in the book.

## Problem 2 (Stability functions, stability types, linear algebra. 34 %)

Consider the following Runge-Kutta methods:

1. The Lobatto IIIA of order 4, which has the Butcher array:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{5}{24} & \frac{1}{3} & -\frac{1}{24} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

2. The Lobatto IIIB of order 4, which has the Butcher array:

$$\begin{array}{c|ccc} 0 & \frac{1}{6} & -\frac{1}{6} & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} & 0 \\ 1 & \frac{1}{6} & \frac{2}{3} & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

3. The Lobatto IIIC of order 4, which has the Butcher array:

$$\begin{array}{c|ccc} 0 & \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ \frac{1}{2} & \frac{1}{6} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}$$

- (a) Find the stability function for each of these methods. Show your calculations.

*Hint: Use (14.142) or (14.143) in the book.*

**Solution:**

1. For the Lobatto IIIA:

$$R(s) = \frac{1 + \frac{1}{2}s + \frac{1}{12}s^2}{1 - \frac{1}{2}s + \frac{1}{12}s^2} = P_2^2(s).$$

2. For the Lobatto IIIB:

$$R(s) = \frac{1 + \frac{1}{2}s + \frac{1}{12}s^2}{1 - \frac{1}{2}s + \frac{1}{12}s^2} = P_2^2(s).$$

3. For the Lobatto IIIC:

$$R(s) = \frac{1 + \frac{1}{4}s}{1 - \frac{3}{4}s + \frac{1}{4}s^2 - \frac{1}{24}s^3} = P_3^1(s).$$

- (b) For each of these methods, determine whether the method is A-stable, L-stable, stiffly accurate and whether it is algebraically stable. Justify your answers.

*NB: Stiffly accurate methods are explained in Section 14.6.3 in the book.*

**Solution:** The Lobatto IIIA and IIIB are A-stable because of result (14.168). Moreover, the Lobatto IIIC is L-stable due to result (14.169).

Only the  $A$  matrix of the Lobatto IIIC is non-singular. Furthermore,  $b = A^T e_\sigma$ . Hence, the Lobatto IIIC is stiffly accurate, while the Lobatto IIIA and IIIB are not.

We observe that  $b_i \geq 0$  for all methods.

Moreover, let  $M = \text{diag}(b)A + A^T \text{diag}(b) - bb^T$ . For the Lobatto III A,

$$M = \frac{1}{36} \begin{bmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{bmatrix},$$

which is not positive semidefinite. For the Lobatto III B,

$$M = \frac{1}{36} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & -1 \end{bmatrix},$$

which is not positive semidefinite. Finally, for the Lobatto III C,

$$M = \frac{1}{36} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix},$$

which is positive semidefinite.

Hence, only the Lobatto III C is algebraically stable.

- (c) Explain shortly the concepts of A- and L-stability, and give advantages and disadvantages for both types of methods, if any.

**Solution:** Read sections 14.6.1 and 14.6.2 in the book.

**Problem 3 (Intro to finite differences and finite elements. 42 %)**

NB 1: This is a computer exercise, and can therefore be solved in groups of 2 students. If you do so, please write down the name of your group partner in your answer.

NB 2: This problem is motivated by the guest lecture held by Erlend Kristiansen from COMSOL, and will serve as an introduction to the finite difference and finite elements methods, which are popular numerical methods for solving partial differential equations (PDEs). Moreover, an introduction to PDEs in the context of transmission lines can be found in Sections 4.5.1 and 4.5.2 in the book.

In problem 1 of assignment 3, your research was crucial to save humanity from a zombie apocalypse. However, recorded data of the events has shown that your simulations were too pessimistic. After reflecting about this, you conclude that the major reason for this deviation is that your models do not take into account the spatial distributions of the populations. For example, in these models, zombies could pop up anywhere. However, in reality, zombies appeared mainly at graveyards when raised from the dead, or at some medical facility as the result of a "successful" infection. These and other clear patterns were intensively exploited when quarantining and curing zombies and infected individuals, increasing the effectiveness of these measures drastically.

Therefore, in order to improve your model for the zombie apocalypse, the spatial distribution of the zombies has to be taken into account. Since these abominations move slowly and erratically, almost bumping on each other, the **diffusion equation** provides a good description of their spatial distribution over time.

For simplicity, consider only one spatial dimension. Hence, let  $Z(x, t)$  be the number of zombies at a point  $x$  at a given time  $t$ . Furthermore, assume that the zombies can only move between  $0 \leq x \leq L$ , and that at the beginning, all zombies were concentrated uniformly between  $0 \leq x \leq L_0$ , with density  $\frac{Z_0}{L_0}$ . Therefore, the corresponding partial differential equation (PDE) reads

$$\frac{\partial Z}{\partial t}(x, t) = D \frac{\partial^2 Z}{\partial x^2}(x, t) \quad (3a)$$

$$Z(x, 0) = \begin{cases} Z_0 & , \text{ for } 0 \leq x \leq L_0 \\ 0 & , \text{ for } L_0 < x \end{cases} \quad (3b)$$

$$\frac{\partial Z}{\partial x}(0, t) = 0 = \frac{\partial Z}{\partial x}(L, t), \quad (3c)$$

where the last equation means that the zombies do not leave the interval  $[0, L]$ , nor arrive at this interval from the outside.

- (a) Find the actual solution for (3).

What is the stationary value of  $Z$ ?

*Hint: Use the separation of variables method: 1. Find the solutions for (3a) and (3c) that have the form  $Z(x,t) = f(x)g(t)$ . Then  $f$  is a sinusoidal function, and  $g$  is an exponential function. Moreover, the frequency of  $f$  and the rate of  $g$  are related. 2. Express the actual solution as a infinite linear combination (series) of the solutions found in 1. The coefficients of this series are then chosen so that (3b) is satisfied.*

**Solution:** By replacing  $Z(x,t) = f(x)g(t)$  in (3a), we obtain

$$\frac{1}{g(t)} \frac{dg}{dt}(t) = -\alpha = \frac{D}{f(x)} \frac{d^2 f}{dx^2}(x),$$

where  $\alpha \geq 0$  is a constant since the left side of the above equation depends on  $t$ , while the right side depends on  $x$ . Hence,

$$\begin{aligned} f(x) &= A_1 \cos\left(\sqrt{\frac{\alpha}{D}}x\right) + A_2 \sin\left(\sqrt{\frac{\alpha}{D}}x\right) \\ g(t) &= A_3 e^{-\alpha t} \end{aligned}$$

The boundary conditions (3c) imply that  $A_2 A_3 \sqrt{\alpha} = 0$  and that

$$A_1 A_3 \sqrt{\alpha} \sin\left(\sqrt{\frac{\alpha}{D}}L\right) = 0$$

If one assumes that  $Z \neq 0$ , then  $A_2 = 0$  and  $\alpha = \alpha_n$  for  $n \in \mathbb{N} \cup \{0\}$ , where  $\alpha_n = D \left(\frac{n\pi}{L}\right)^2$ . Hence,

$$Z(x,t) = Z_n(x,t) = C_n \cos\left(\frac{n\pi}{L}x\right) e^{-\left(\frac{n\pi}{L}\right)^2 Dt},$$

and the actual solution to (3) can be written as

$$Z(x,t) = \sum_{n=0}^{\infty} C_n \cos\left(\frac{n\pi}{L}x\right) e^{-\left(\frac{n\pi}{L}\right)^2 Dt},$$

where

$$\begin{aligned} C_0 &= \frac{1}{L} \int_0^L Z(x,0) dx = \frac{Z_0 L_0}{L} \\ C_n &= \frac{2}{L} \int_0^L Z(x,0) \cos\left(\frac{n\pi}{L}x\right) dx = \frac{2Z_0}{n\pi} \sin\left(\frac{n\pi L_0}{L}\right), \quad n \geq 1. \end{aligned}$$

Therefore the solution to (3) is

$$Z(x,t) = \frac{Z_0 L_0}{L} + \sum_{n=1}^{\infty} \frac{2Z_0}{n\pi} \sin\left(\frac{n\pi L_0}{L}\right) \cos\left(\frac{n\pi}{L}x\right) e^{-\left(\frac{n\pi}{L}\right)^2 Dt}.$$

Furthermore,  $\lim_{t \rightarrow \infty} Z(x,t) = \frac{Z_0 L_0}{L}$ .

We will now introduce the **finite difference method**. In this method, the partial derivatives present in a PDE are approximated by expressions that depend on the solution values at several points. These expressions usually involve differences of such values. Hence, the name "finite differences".

The finite differences for (3) will be found in part (b). This will lead to a discretization of (3), which will be expressed as a set of linear equations in part (c). Finally, (3) will be solved in part (d).

- (b) Use the Taylor polynomial of  $Z$  at  $(x,t)$  of order 1 to express  $Z(x,t - \Delta t)$  as a polynomial in  $\Delta t$ . The coefficients of this polynomial depend on the values of  $Z$  and its partial derivatives respect

to time at  $(x, t)$ . Finally, find a linear combination of  $Z(x, t - \Delta t)$  and  $Z(x, t)$  that is equal to a first order approximation of  $\frac{\partial Z}{\partial t}(x, t)$ .

Analogously, by using a Taylor polynomial of  $Z$  at  $(x, t)$  of order 3, find a linear combination of  $Z(x + \Delta x, t)$ ,  $Z(x, t)$  and  $Z(x - \Delta x, t)$  that is equal to a second order approximation of  $\frac{\partial^2 Z}{\partial x^2}(x, t)$ .

*Hint 1: The coefficients of the linear combinations depend on the corresponding steps:  $\Delta t$  and  $\Delta x$ .*

*Hint 2: Example of a second order finite difference:*

$$\frac{Z(x, t + \Delta t) - Z(x, t - \Delta t)}{2\Delta t} = \frac{\partial Z}{\partial t}(x, t) + O(\Delta t^2).$$

*Example of a fourth order finite difference:*

$$\frac{-Z(x + 2\Delta x, t) + 16Z(x + \Delta x, t) - 30Z(x, t) + 16Z(x - \Delta x, t) - Z(x - 2\Delta x, t)}{12\Delta x^2} = \frac{\partial^2 Z}{\partial x^2}(x, t) + O(\Delta x^4).$$

**Solution:** We have the Taylor expansions:

$$Z(x, t - \Delta t) = Z(x, t) - \Delta t \frac{\partial Z}{\partial t}(x, t) + O(\Delta t^2)$$

$$Z(x + \Delta x, t) = Z(x, t) + \Delta x \frac{\partial Z}{\partial x}(x, t) + \frac{\Delta x^2}{2} \frac{\partial^2 Z}{\partial x^2}(x, t) + \frac{\Delta x^3}{6} \frac{\partial^3 Z}{\partial x^3}(x, t) + O(\Delta x^4)$$

$$Z(x - \Delta x, t) = Z(x, t) - \Delta x \frac{\partial Z}{\partial x}(x, t) + \frac{\Delta x^2}{2} \frac{\partial^2 Z}{\partial x^2}(x, t) - \frac{\Delta x^3}{6} \frac{\partial^3 Z}{\partial x^3}(x, t) + O(\Delta x^4).$$

Hence,

$$\frac{Z(x, t) - Z(x, t - \Delta t)}{\Delta t} = \frac{\partial Z}{\partial t}(x, t) + O(\Delta t).$$

$$\frac{Z(x + \Delta x, t) - 2Z(x, t) + Z(x - \Delta x, t)}{\Delta x^2} = \frac{\partial^2 Z}{\partial x^2}(x, t) + O(\Delta x^2).$$

Let  $x_n = \frac{nL}{N} = n\Delta x$  and  $t_m = \frac{mT}{M} = m\Delta t$  for  $n = 0, \dots, N$  and  $m = 0, \dots, M$ . The collection of points  $(x_n, t_m)$  is known as a **grid**, where each point in the grid is called a **node**. Since the steps in each coordinate direction are constant, we have defined a **regular grid** over the region  $[0, L] \times [0, T]$ .

(c) Let  $Z_{n,m} = Z(x_n, t_m)$ . Show that the finite differences found in part (b) give the following discretization for (3a):

$$-Z_{n-1,m} + \left(2 + \frac{\Delta x^2}{D\Delta t}\right) Z_{n,m} - Z_{n+1,m} = \frac{\Delta x^2}{D\Delta t} Z_{n,m-1}. \quad (4)$$

Note that the discretization (4) requires the values of  $Z_{n,m}$  around the node  $(x_n, t_m)$ . Therefore, (4) is not well-defined on the boundary of  $[0, L] \times [0, T]$ . However, the boundary condition (3c) can be interpreted as

$$Z_{-1,m} = Z_{0,m} \quad \text{and} \quad Z_{N,m} = Z_{N+1,m}. \quad (5)$$

Hence, by adding (5) to (4), the discretization is now well-defined for all nodes.

Assume that the solution for the previous time  $t = t_{m-1}$  is known, and show that the solution for  $t = t_m$  can be calculated by solving a linear system. More precisely, show that the discretization in (4) can be written as a linear system of the form  $Ax = b$ , where  $A$  is a  $(N+1)$ -by- $(N+1)$  matrix,  $b$  is a  $(N+1)$  column vector, and

$$x = [Z_{0,m}, Z_{1,m}, \dots, Z_{N,m}]^T. \quad (6)$$

Find a general expression for  $A$  and  $b$ .

*Hint: The matrix  $A$  consists basically of the coefficients of the left side of (4), while  $b$  represents the right side of (4).*

**Solution:**

$$A = \begin{bmatrix} 1+K & -1 & & & \\ -1 & 2+K & -1 & & \\ & -1 & 2+K & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2+K & -1 \\ & & & & -1 & 1+K \end{bmatrix} \quad B = K \begin{bmatrix} Z_{0,m-1} \\ Z_{1,m-1} \\ Z_{2,m-1} \\ \vdots \\ Z_{N-1,m-1} \\ Z_{N,m-1} \end{bmatrix},$$

where  $K = \frac{\Delta x^2}{D\Delta t}$ .

- (d) Use the numerical values  $D = 1 \text{ m}^2/\text{s}$ ,  $Z_0 = 100$ ,  $L = 10 \text{ m}$ ,  $L_0 = 1 \text{ m}$  and  $T = 60 \text{ s}$ .  
Implement a Matlab script that solves (3) based on the linear systems found in part (c).

Simulate with  $N = 500$  and  $M = 3000$ .

Add your Matlab code and a plot with the numerical solution to your answer.

Comment on the results.

The plot can be a time lapse of  $Z$  as a function of  $x$  for several representative  $t$  values, or a 3D-plot (e.g. surface plot) of  $Z$  as a function of  $x$  and  $t$ .

**Solution:** The Matlab script and simulation results:

```
% parameters
N = 500; M = 3000;
L = 10; T = 60;
Z_0 = 100; L_0 = 1;
D = 1;

% numerical solution
dx = L/N; x = dx*(0:N);
dt = T/M; t = dt*(0:M);
K = dx^2/D/dt;
v1 = (2+K)*ones(N+1,1);
v1(1) = 1+K; v1(N+1) = 1+K;
v2 = ones(N,1);
A = diag(v1) - diag(v2,-1) - diag(v2,1);
A = A/K;
b = zeros((N+1),1);
index = floor(L_0/dx);
b(1:index) = Z_0;
Z = zeros(N+1,M+1);
Z(:,1) = b;
for m=2:M+1
    Zm = A\b;
    Z(:,m) = Zm;
    b = Zm;
end
```



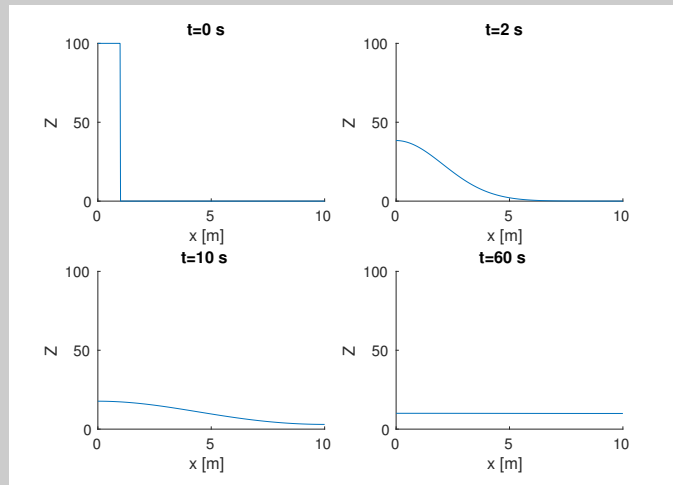


Figure 3: Time lapse for the numerical solution.

We will now introduce another popular method for solving PDEs: the **finite element method**. This method is more mathematically sophisticated than the finite difference method. However, the extra technical challenges will pay off as outstanding flexibility.

The finite element method starts by rewriting the PDE as a **variational problem**, i.e. the solution of the PDE will be reinterpreted as the solution of a wide family of integrals. The variational formulation is also known as the "weak" formulation, while the PDE is known as the "strong" formulation. For most well-posed problems, such as (3), the "weak" and "strong" formulations are equivalent. In this problem, the variational formulation of (3) will be found in part (e).

Once the variational formulation is in place, the actual solution is approximated by a linear combination of so-called **basis functions**. A basis function is usually non-zero on a small subregion, which is known as its **support**. Hence, the name "finite element". The properties of the simplest of all basis functions, the linear basis functions over a regular grid, will be studied in part (f).

If a linear combination of basis functions provides a good approximation to the actual solution, this linear combination can replace the actual solution in the variational formulation. Hence, we obtain a discretization of the variational formulation similar to the one in part (c), where the solution of a set of linear equations provides the numerical solution. In this problem, the discretization of (3) using the finite difference method will be found in part (g).

(e) **(Optional)** Use (3a) and (3c) to show that

$$\int_0^T \int_0^L \left( \frac{\partial Z}{\partial t}(x, t) v(x, t) + D \frac{\partial Z}{\partial x}(x, t) \frac{\partial v}{\partial x}(x, t) \right) dx dt = 0, \quad (7)$$

for all continuously differentiable functions  $v$  on  $[0, L] \times [0, T]$ .

This is the variational formulation of (3).

*Hint: Use partial integration.*

**Solution:** By multiplying both sides of (3a) with  $v$  and integrating over  $[0, L] \times [0, T]$ , we obtain

$$\begin{aligned} \int_0^T \int_0^L \frac{\partial Z}{\partial t}(x, t) v(x, t) dx dt &= D \int_0^T \int_0^L \frac{\partial^2 Z}{\partial x^2}(x, t) v(x, t) dx dt \\ &= D \int_0^T \left[ \frac{\partial Z}{\partial x}(x, t) v(x, t) \right]_{x=0}^L dt - D \int_0^T \int_0^L \frac{\partial Z}{\partial x}(x, t) \frac{\partial v}{\partial x}(x, t) dx dt \\ &= -D \int_0^T \int_0^L \frac{\partial Z}{\partial x}(x, t) \frac{\partial v}{\partial x}(x, t) dx dt. \end{aligned}$$

- (f) **(Optional)** The basis functions will be defined by reusing the regular grid introduced in part (c). For  $n = 0, \dots, N$  and  $m = 0, \dots, M$ , define the basis function  $v_{n,m}$  as

$$v_{n,m}(x, t) = \max \left\{ 0, 1 - \max \left\{ \frac{|x - x_n|}{\Delta x}, \frac{|t - t_m|}{\Delta t} \right\} \right\}, \quad (x, t) \in [0, L] \times [0, T]. \quad (8)$$

Make a sketch of the regular grid, and a 3D-sketch of a general basis function  $v_{n,m}$ .

What is the support of  $v_{n,m}$ ?

*Hint 1: What are the level curves  $\max \left\{ \frac{|x - x_n|}{\Delta x}, \frac{|t - t_m|}{\Delta t} \right\} = c$  for  $c \in [0, 1]$ ?*

Furthermore, calculate the partial derivatives  $\frac{\partial v_{n,m}}{\partial x}(x, t)$  and  $\frac{\partial v_{n,m}}{\partial t}(x, t)$ .

*Hint 2: Note that  $v_{n,m}$  is a piecewise linear function.*

*NB: The partial derivatives are not defined everywhere. This is not a problem: The smoothing effect of the integrals in the variational formulation (7) will take care of the singularities.*

Finally, assume that a function  $v$  can be expressed as a linear combination of the basis function  $v_{n,m}$ , i.e.

$$v = \sum_{n=0}^N \sum_{m=0}^M a_{n,m} v_{n,m}. \quad (9)$$

Evaluate (9) at each node of the regular grid, and find the relation between  $a_{n,m}$  and the values of  $v(x, t)$  at the nodes.

**Solution:** The level curve

$$\max \left\{ \frac{|x - x_n|}{\Delta x}, \frac{|t - t_m|}{\Delta t} \right\} = c$$

for  $c \in [0, 1]$  is the rectangle

$$[x_n - c\Delta x, x_n + c\Delta x] \times [t_m - c\Delta t, t_m + c\Delta t].$$

Hence, the graph of  $v_{n,m}$  is a rectangular pyramid centered at  $(x_n, t_m)$  with base  $[x_{n-1}, x_{n+1}] \times [t_{m-1}, t_{m+1}]$  and height 1. Outside this base,  $v_{n,m}$  only takes the value zero. Therefore the support of  $v_{n,m}$  is  $(x_{n-1}, x_{n+1}) \times (t_{m-1}, t_{m+1})$ .

Since

$$v_{n,m}(x, t) = \begin{cases} -\frac{x-x_n}{\Delta x}, & 1 \geq \frac{x-x_n}{\Delta x} \geq \frac{|t-t_m|}{\Delta t} \\ -\frac{t-t_m}{\Delta t}, & 1 \geq \frac{t-t_m}{\Delta t} \geq \frac{|x-x_n|}{\Delta x} \\ \frac{x-x_n}{\Delta x}, & -1 \leq \frac{x-x_n}{\Delta x} \leq -\frac{|t-t_m|}{\Delta t} \\ \frac{t-t_m}{\Delta t}, & -1 \leq \frac{t-t_m}{\Delta t} \leq -\frac{|x-x_n|}{\Delta x} \\ 0, & |t - t_m| \geq \Delta t \text{ or } |x - x_n| \geq \Delta x, \end{cases}$$

then

$$\nabla v_{n,m}(x, t) = \begin{cases} [-\frac{1}{\Delta x}, 0], & 1 > \frac{x-x_n}{\Delta x} > \frac{|t-t_m|}{\Delta t} \\ [0, -\frac{1}{\Delta t}], & 1 > \frac{t-t_m}{\Delta t} > \frac{|x-x_n|}{\Delta x} \\ [\frac{1}{\Delta x}, 0], & -1 < \frac{x-x_n}{\Delta x} < -\frac{|t-t_m|}{\Delta t} \\ [0, \frac{1}{\Delta t}], & -1 < \frac{t-t_m}{\Delta t} < -\frac{|x-x_n|}{\Delta x} \\ 0, & |t-t_m| > \Delta t \text{ or } |x-x_n| > \Delta x \\ \text{undefined,} & \text{otherwise.} \end{cases}$$

Finally, note that

$$v_{n,m}(x_i, t_j) = \begin{cases} 1, & n = i \text{ and } m = j \\ 0, & n \neq i \text{ or } m \neq j. \end{cases}$$

Hence,  $a_{n,m} = v(x_n, t_m)$ .

(g) **(Optional)** Let  $Z = \sum_{i=0}^N \sum_{j=0}^M Z_{i,j} v_{i,j}$  and  $v = v_{n,m}$ . Show that (7) then becomes

$$\sum_{i=n-1}^{n+1} \sum_{j=m-1}^{m+1} Z_{i,j} \int_0^T \int_0^L \frac{\partial v_{i,j}}{\partial t} v_{n,m} dx dt + D \sum_{i=n-1}^{n+1} \sum_{j=m-1}^{m+1} Z_{i,j} \int_0^T \int_0^L \frac{\partial v_{i,j}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt = 0, \quad (10)$$

for all  $n = 0, \dots, N$  and  $m = 0, \dots, M$ .

Calculate the integrals in (10), and update this equation. Compare the obtained result to (4).

NB: The integrals only need to be calculated on the region, where the integrand is non-zero.

**Solution:** Equation (10) follows from the linearity of the integral and the fact that the supports of  $v_{i,j}$  and  $v_{n,m}$  do not overlap for  $|i-n| > 1$  or  $|j-m| > 1$ .

Most of the integrals in (10) can be calculated directly using geometrical considerations, such as symmetry, the volume of a pyramid and the area of the support of the integrand. Hence, we have that

$$\begin{aligned} \int_0^T \int_0^L \frac{\partial v_{n-1,m-1}}{\partial t} v_{n,m} dx dt &= -\frac{1}{6} \Delta x \quad (\text{Volume of pyramid}) \\ \int_0^T \int_0^L \frac{\partial v_{n,m-1}}{\partial t} v_{n,m} dx dt &= -\Delta x \int_{-1}^1 \int_{-|x|}^{|x|} 1 - \max\{|x|, |t|\} dt dx = -\frac{2}{3} \Delta x \\ \int_0^T \int_0^L \frac{\partial v_{n+1,m-1}}{\partial t} v_{n,m} dx dt &= -\frac{1}{6} \Delta x \quad (\text{Volume of pyramid}) \\ \int_0^T \int_0^L \frac{\partial v_{n-1,m}}{\partial t} v_{n,m} dx dt &= 0 \quad (\text{Symmetry}) \\ \int_0^T \int_0^L \frac{\partial v_{n,m}}{\partial t} v_{n,m} dx dt &= 0 \quad (\text{Symmetry}) \\ \int_0^T \int_0^L \frac{\partial v_{n+1,m}}{\partial t} v_{n,m} dx dt &= 0 \quad (\text{Symmetry}) \\ \int_0^T \int_0^L \frac{\partial v_{n-1,m+1}}{\partial t} v_{n,m} dx dt &= \frac{1}{6} \Delta x \quad (\text{Volume of pyramid}) \\ \int_0^T \int_0^L \frac{\partial v_{n,m+1}}{\partial t} v_{n,m} dx dt &= \Delta x \int_{-1}^1 \int_{-|x|}^{|x|} 1 - \max\{|x|, |t|\} dt dx = \frac{2}{3} \Delta x \\ \int_0^T \int_0^L \frac{\partial v_{n+1,m+1}}{\partial t} v_{n,m} dx dt &= \frac{1}{6} \Delta x \quad (\text{Volume of pyramid}) \end{aligned}$$

and that

$$\begin{aligned}
\int_0^T \int_0^L \frac{\partial v_{n-1,m-1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= 0 \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n,m-1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= \frac{1}{2} \frac{\Delta t}{\Delta x} \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n+1,m-1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= 0 \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n-1,m}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= -\frac{1}{2} \frac{\Delta t}{\Delta x} \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n,m}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= 2 \frac{\Delta t}{\Delta x} \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n+1,m}}{\partial x} \frac{\partial v_{n+1,m}}{\partial x} dx dt &= -\frac{1}{2} \frac{\Delta t}{\Delta x} \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n-1,m+1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= 0 \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n,m+1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= \frac{1}{2} \frac{\Delta t}{\Delta x} \quad (\text{Support of integrand}) \\
\int_0^T \int_0^L \frac{\partial v_{n+1,m+1}}{\partial x} \frac{\partial v_{n,m}}{\partial x} dx dt &= 0 \quad (\text{Support of integrand})
\end{aligned}$$

Hence, the final discretization is

$$\begin{aligned}
& -\frac{\Delta x^2}{12D\Delta t} Z_{n-1,m-1} + \left( \frac{1}{4} - \frac{\Delta x^2}{3D\Delta t} \right) Z_{n,m-1} - \frac{\Delta x^2}{12D\Delta t} Z_{n+1,m-1} \\
& -\frac{1}{4} Z_{n-1,m} + Z_{n,m} - \frac{1}{4} Z_{n+1,m} \\
& + \frac{\Delta x^2}{12D\Delta t} Z_{n-1,m+1} + \left( \frac{1}{4} + \frac{\Delta x^2}{3D\Delta t} \right) Z_{n,m+1} + \frac{\Delta x^2}{12D\Delta t} Z_{n+1,m+1} = 0.
\end{aligned}$$

In this problem, the finite difference and finite element schemes are very similar in form. The main reason for this is that the basis functions for the finite element scheme were defined on the regular grid used for the finite difference scheme. In other words, for regular grids the finite element method is not necessarily superior to the finite difference method.

However, if the geometry of the region where the PDE is defined, is so complicated that it does not admit a regular grid, or if one desires a finer grid on some subregions, the finite difference method cannot be used, and the flexibility of the finite element method comes in handy.

- (h) **(Optional)** Assume that we have an irregular grid, i.e. there is no obvious pattern in the positions of the nodes or the distance between them.

Explain how you would implement a finite element method for that case on a computer. In particular, explain how you would define the basis functions and how you would discretize the variational formulation.

**Solution:** For each node in the grid, a basis function is defined as 1 at that node, and as 0 at the rest of nodes. Spline interpolation or other interpolation methods can be used to find the values of these basis functions and their derivatives at other points.

The integrals between basis functions and their derivatives present in the discretization of the variational formulation, can be solved using numerical integration.