JOHANNES KEPLER
UNIVERSITY LINZ

Author / Eingereicht von
Felix Dreßler
k12105003

Submission / Angefertigt am
Institute / **Institut**

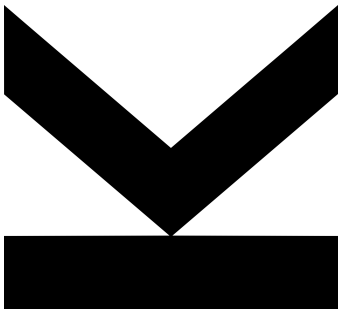
Thesis Supervisor / First Supervisor / BeurteilerIn / ErstbeurteilerIn / ErstbetreuerIn
DI Dr. **Name**

Second Supervisor / ZweitbeurteilerIn / ZweitbetreuerIn
Name

Assistant Thesis Supervisor / Mitbetreuung
Name

Month 2024 / Monat 2024

Circuit Modelling



Subtitle / Untertitel

JOHANNES KEPLER
UNIVERSITY LINZ
Altenbergerstraße 69
4040 Linz, Austria
www.jku.at
DVR 0093696

Kurzfassung

Kurzfassung auf Deutsch.

Abstract

Abstract in English.

Contents

List of Figures

List of Tables

1 Introduction

This chapter should include information about what circuit modelling wants to achieve as well as giving an overview of what this bachelor-thesis is about.

What is this thesis about?
Modelling and numerically solving systems that arrise from electrical circuits with RLC elements. Furthermore it will briefly discuss on expanding this baseline with more complicated electrical components. What is the goal of this thesis?
The goal of this thesis is to give insight into industrial standards concerning circuit modelling. It aims to elaborate on the underlying concepts of MNA as well as on the most commonly used numerical methods.

2 Formulating a Mathematical Model

based on DAE lecture and modelling and discretiztation of elec circuit problems.

2.1 Network Topology

An electrical circuit is usually considered as a graph (N, E) where $N = (n_0, n_1, n_2, ...)$ are the Nodes and $E = (e_{ij})_{ij}$ are the edges, where for some i and some j we have that $e_{ij} = (n_j, n_k)$ is the edge from node j to node k. We can store this information in an *incidence matrix* $\tilde{A} = (\tilde{a}_{ij})_{ij}$ which is defined by

$$\tilde{a}_{ij} = \begin{cases} 1 & \text{edge } j \text{ starts at node } i, \\ -1 & \text{edge } j \text{ ends at node } i, \\ 0 & \text{else.} \end{cases}$$

We call $u = (u_0, u_1, u_2, ...)$ the corresponding potentials to the nodes N. The difference of these potentials is the voltage at the associated edge. To fix the absolute values of these potentials we have to set one node to a fixed potential. We will do that by "grounding" the node n_0, this means we set the potential $u_0 := 0$. This grounding of a node allows us to remove the corresponding row from the incidence matrix to get the *reduced incidence matrix* A. The vector $v = (v_{ij})_{ij}$ represents the voltages at the edges. For some i and some j the voltage at edge ij is $v_{ij} = u_i - u_j$.

We will later see, that the components of an electrical crircuit, which will be installed along the edges, describe a relationship between the edges current and its voltage. Thus a current vector $i = (i_1, i_2, i_3, ...)$ containing the currents along the edges is required.

2.2 Energy Conservation Laws

To fully fix all the variables that arrise in the model of an electrical circuit we will need some *conservation laws*:

- **Kirchhoff's voltage law (KVL):**
  The sum of voltages along each loop of the network must equal to zero. Using the incidence matrix A this law can be formulated as

$$A^\top * u = v. \tag{2.1}$$

- **Kirchhoff's current law (KCL):**
  For any node, the sum of currents flowing into the node is equal to the sum of currents flowing out of the node. Using the incidence matrix A again, this law can be formulated as

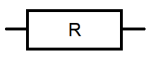
$$A * i = 0. \tag{2.2}$$

2.3 Electrical Components and their relations

Electrical components are described by equations relating their edge voltage v to their edge current i. We will mainly focus on so called RLC-networks which consist of resistors, capacitors, inductances, voltage sources and current sources. Diodes and Transistors as well as other electrical components can be described in a similar way, although these would lead to a more difficult analysis of the system.

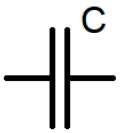
- **Resistor**
  Resistors "resist" the flow of current, which causes voltage to drop. This behaviour is described by the *resistance* R which is given in *Ohm* (Ω) and its reciprocal, the *conductance* G, which is given in *Simens* ($S = \frac{1}{\Omega}$).

$$v = R * i \quad \text{or} \quad i = G * u.$$



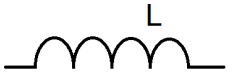
- **Capacitor**
  Capacitors "store" electrical energy by accumulating electrical charge. Their characteristic equations can be described directly using the stored charge Q or indirectly using the change in charge, which is nothing other than the current I. The *capacitance* C is given in *Farads* (F).

  $Q = C * v$ and by derivation in t $I = C * \dot{v}$.



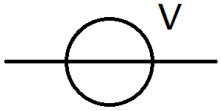
- **Inductor (Coil)**
  An electric current flowing through a conductor generates a magnetic field Φ surrounding it. This magnetic field causes a voltage drop dependant on the change in current. The *inductance* L is given in *Henry* (H).

  $\Phi = L * i$ and by derivation in t $v = L * \dot{i}$.



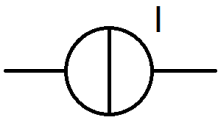
- **Voltage Source**
  A voltage source supplies the system with a voltage. It can either supply varying amounts of voltage (with the special case of alternating current AC) or a fixed amount of voltage. The unit of voltage is *Volts* (V).

  $v = v_{src}$



- **Current Source**
  A current source supplies the system with current. It can either supply varying amounts of current (with the special case of alternating current AC) or a fixed amount of current. The unit of current is *Ampere* (A).

  $$i = i_{src}$$



- Diode - to be filled with information after the rest is complete
- Transistor - unlike the other components which where all two-teminal components a transisstoor is a three-terminal component

2.4 Modified Nodal Analysis - MNA

num gew dgl steif n steif - seite 422
modelling discr circ prob - seite 19

To analyse the network further we will sort the reduced incidence matrix A such that is has the block form

$$A = (A_R A_C A_L A_V A_I)$$

where A_R, A_C, A_L, A_V and A_I all include the collumns that are related to the resistors, capacitors, coils, voltage sources and current sources.

To mathematically describe the circuit we will use *modified nodal analysis* (or short MNA). MNA uses the node voltages as well as the currents of the coils and the voltage sources as unknowns and is based on the conservation laws

$$Ai(t) = 0 \tag{2.3}$$

$$v = A^\top u(t) \tag{2.4}$$

as well as on the voltage-current relations of the electrical components. By replacing all edge-currents with their respective voltage-current relation and ell egde-voltages with their node-potentials we obtain the MNA-equations

$$\begin{aligned} A_C C A_C^\top \dot{u} + A_R G A_R^\top u + A_L i_L + A_V i_V + A_I i_{src} &= 0 \\ L\dot{i_L} - A_L^\top u &= 0 \\ -A_V^\top + v_{src} &= 0. \end{aligned}$$

In matrix form these read as

$$\begin{pmatrix} A_C C A_C^\top & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} \dot{u} \\ \dot{i_L} \\ \dot{i_V} \end{pmatrix} + \begin{pmatrix} A_R G A_R^\top & A_L & A_V \\ -A_L^\top & 0 & 0 \\ -A_V^\top & 0 & 0 \end{pmatrix} * \begin{pmatrix} u \\ i_L \\ i_V \end{pmatrix} = \begin{pmatrix} -A_I i_{src} \\ 0 \\ -v_{src} \end{pmatrix}, \tag{2.5}$$

where the diagonal matrices C, G and L contain the capacities, conductivities and inductivities.

The resulting systems are *stiff* systems. This means that for their numerical solution special care has to be put into which methods are suitable for a solving these sytsems in a stable manner.

2.5 Energy bilanz??????????

may also not be relevant - leave out?

2.6 Charge/Flux oriented formulation of MNA

may noot be too importatn - leave out?

Again using KCL 2.3 and the component equations we formulate a system of equations. (**more here**) What is flux and charge - explain either here or witht the components

$$A_C\dot{q} + A_R r(A_R^\top u, t) + A_L i_L + A_V i_V + A_I i(A^\top u, \dot{q}, i_L, i_V, t) = 0 \tag{2.6}$$
$$\dot{\phi} - A_L^\top u = 0 \tag{2.7}$$
$$v(A^\top u, \dot{q}, i_L, i_V, t) - A_V^\top u = 0 \tag{2.8}$$
$$q - q_C(A_C^\top u) = 0 \tag{2.9}$$
$$\phi - \phi_L(i_L) = 0 \tag{2.10}$$

Using

node potentials u,

branch currents through voltage and flux controlled elements i_V and i_L,

charges and fluxes q and ϕ,

voltage dependent resistors r,

voltage and current dependent charge and flux sources q_C and ϕ_L,

controlled current and voltage sources i_{src} and v_{src}.

Charge/flux oriented or conventional MNA? On which formulation — charge/flux oriented or conventional — should the numerical discretization be based, if MNA is used for the automatic generation of network equations? From a structural aspect, the conventional MNA formulation yields a standard form of numerical integration problems, while the charge/flux oriented formulation does not. There are however several reasons, not to transform (4.1) into (4.2) before applying numerical discretization schemes, although they are analytically equivalent:

Structure. (4.1) is of linear-implicit nonlinear form, while (4.2) is of nonlinear-implicit nonlinear form. This may have an impact on the choice of a suitable integrator.

Numerics. Information on the charge/flux level is lost in the conventional approach, and charge conservation may only be maintained approximately in numerical integration schemes.

Implementation. Implicit numerical integration schemes for the conventional MNA equations (4.2) require second partial derivatives of q_C und ϕ_L. These derivative informations, however, are not available in circuit simulation packages, may even not exist because of the lack of smoothness in transistor models.

Figure 2.1:

so it makes sense to use the conventional approach of MNA - compare with phd thesis

3 Differential Algebraic Equations

von num gew dgl niochtsteife steife und dae

3.1 Abstract Problem

$$F(t, y(t), y'(t)) = 0 \qquad \forall t \in I \tag{3.1}$$

with $F : \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^N$ sufficiently smooth. We will focus on linear time invaraiant systems of the form

$$Eu'(t) = Au(t) + f(t).$$

With $E, A \in \mathbb{R}^{n \times n}$. If E is regular, then this system is just an ordinary differential equation, thus we assume E to be singular to obtain a "true" DAE.

In the following chapters we will discuss important analytical properties of this such sytems. We will discuss the solvability and the index of these systems.

3.2 Types of DAEs

- **Linear systems with constant coeffiecients**
  are systems of the form

$$Ay'(t) + By(t) = f(t) \tag{3.2}$$

with $A, B \in \mathbb{R}^{n \times n}$, A singular and $f(t)$ a function.

- **linear time dependent systems**

  $$A(t)y'(t) + B(t)y(t) = f(t)$$

  with $A(t), B(t), f(t)$ functions.

- **structured (non-linear) systems**
  are semi-explicit systems of the form

  $$y'(t) = f(t, y(t), z(t)), 0 = g(t, y(t), z(t))$$

  with f and g functions.

For our analysis of electrical networks we will focus on linear systems with constant coefficients. (where do the other systems arrise? what are they about?)

3.2.1 Weierstraß-Kronecker Normalform

kapitel aus buch seite 399 num gew dgl steif, kapitel 13.2.2

To determine the solvability of a linear system with constant coefficients 3.2 we first need to introduce a Normalform for the system, the *Weierstraß-Kronecker Normalform*. This Normalform is dependant on the family $\{A, B\} := \{\mu A + B | \mu \in \mathbb{R}\}$, which is called the *matrix pencil* of the DAE.

Definition 1. *The matrix pencil $\{A, B\}$ is called* regular *if there exists some $c \in \mathbb{R}$, such that $(cA + B)$ is regular ($det(cA + B) \neq 0$), otherwise it is called singular.*

Theorem 1 (Jordan Normalform). *For every matrix $Q \in \mathbb{R}^{\times}$ there exists a regular matrix $T \in \mathbb{C}^{n\times n}$, such that*

$$T^{-1}QT = J = diag(J_1, ..., J_r) \quad with \quad J_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ 0 & \lambda_i & \ddots & \vdots \\ & \ddots & \ddots & 1 \\ 0 & \dots & 0 & \lambda_i \end{pmatrix} \in \mathbb{C}^{m_i \times m_i}$$

and $n = m_1 + ... + m_r$.

quote that from somewhere.

translate that:

The matrix J is called Jordan Normalform of Q, the J_i are called Jordan Blocks, where λ_i are the eigenvalues of Q. The matrix J is uniquely determined by Q except for the arrangement of the diagonal blocks. If Q posesses only real eigenvalues, then T can also be choosen from the reals.
A transformation from A and B in 3.2 enables a seperation into differential and algebraic variables.

aus buch seite 401

Theorem 2. *Let $\{A, B\}$ be a regular matrix pencil. There exist $P, Q \in \mathbb{C}^{n\times n}$ such that*

$$PAQ = \begin{pmatrix} I_d & 0 \\ 0 & N \end{pmatrix}, \quad PBQ = \begin{pmatrix} R & 0 \\ 0 & I_{n-d} \end{pmatrix}$$

where

$$N = diag(N_1, ..., N_r) \quad with \quad N_i = \begin{pmatrix} 0 & 1 & & 0 & \\ & \ddots & \ddots & & \\ & & & 0 & 1 \\ 0 & & & 0 & \end{pmatrix} \in \mathbb{R}^{n_i \times n_i}$$

and R has Jordan Normalform.

Proof. Becuase $\{A, B\}$ is a regular matrix pencil, there exists $c \in \mathbb{R}$ such that $(cA + B)$ is regular. Set

$$\hat{A} := (cA + B)^{-1}A, \quad \hat{B} := (cA + B)^{-1}B.$$

Considering

$$(cA + B)^{-1}(cA + B) = I \implies (cA + B)^{-1}B + c(cA + B)^{-1}A = I,$$

we get that

$$\hat{B} = I - c\hat{A}.$$

Let $J_{\hat{A}}$ be the Jordan Normalform of $\hat{A}$, this means that there exists a regular matrix T_1 such that

$$T_1^{-1}AT_1 = J_{\hat{A}} = \begin{pmatrix} W & 0 \\ 0 & \tilde{N} \end{pmatrix}.$$

The matrix W contains the Jordanblocks with Eigenvalues which are nonzero, the matrix $\tilde{N}$ contains the Jordan blocks with Eigenvalues equal to zero, thus $\tilde{N}$ is *nilpotent*. The Jordan Normalform $J_{\hat{B}}$ of $\hat{B}$ is given by

$$T_1^{-1}\hat{B}T_1 = J_{\hat{B}} = \begin{pmatrix} I - cW & 0 \\ 0 & I - c\tilde{N} \end{pmatrix}.$$

The following two transformations will allow us to get the desired structure. First we will transform $J_{\hat{A}}$ with

$$T_2 := \begin{pmatrix} W & 0 \\ 0 & I - c\tilde{N} \end{pmatrix}$$

in

$$T_2^{-1}J_{\hat{A}} = \begin{pmatrix} I & 0 \\ 0 & (I - c\tilde{N})^{-1}\tilde{N} \end{pmatrix}$$

and $J_{\hat{B}}$ in

$$T_2^{-1} J_{\hat{B}} = \begin{pmatrix} W^{-1} - cI & 0 \\ 0 & I \end{pmatrix}.$$

Let now R be the Jordan Normalform of $(W^{-1} - cI)$ and N be the Normalform of $(I - c\tilde{N})^{-1}\tilde{N}$, this means

$$T_W^{-1}(W^{-1} - cI)T_W = R \quad \text{and} \quad T_{\tilde{N}}^{-1}(I - c\tilde{N})^{-1}\tilde{N}T_{\tilde{N}} = N$$

Considering this definition together with the Neumann-series of $(I - c\tilde{N})^{-1}$ we obtain

$$\tilde{N}(I - c\tilde{N})^{-1} = \tilde{N}(c\sum_{i=0}^{\infty}\tilde{N}^i) = \tilde{N}(c\sum_{i=0}^{k-1}\tilde{N}^i) = c\sum_{i=0}^{k-1}\tilde{N}^{i+1} = (c\sum_{i=0}^{k-1}\tilde{N}^i)\tilde{N}$$

where we used that $\tilde{N}$ is nilpotent with nilpotency index k. This shows that $\tilde{N}$ and $(I - c\tilde{N})^{-1}$ commute.

From this we can conclude that

$$N^k = [T_{\tilde{N}}^{-1}(I - c\tilde{N})^{-1}\tilde{N}T_{\tilde{N}}]^k = T_{\tilde{N}}^{-1}[(I - c\tilde{N})^{-1}\tilde{N}]^k T_{\tilde{N}} = T_{\tilde{N}}^{-1}(I - c\tilde{N})^{-k}\underbrace{\tilde{N}^k}_{=0}T_{\tilde{N}} = 0$$

We used the commutativity in the third step here. The nilpotent matrix N thus also has the nilpotency index k. A transformation with

$$T_3 := \begin{pmatrix} T_W & 0 \\ 0 & T_{\tilde{N}} \end{pmatrix}$$

transforms $T_2^{-1} J_{\hat{A}}$ into the Jordan Normalform

$$J_{\tilde{A}} := T_3^{-1}T_2^{-1}J_{\hat{A}}T_3 = T_3^{-1}T_2^{-1}T_1^{-1}\hat{A}T_1T_3 = \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}$$

and $T_2^{-1} J_{\hat{B}}$ into

$$J_{\tilde{B}} := T_3^{-1} T_2^{-1} J_{\hat{B}} T_3 = T_3^{-1} T_2^{-1} T_1^{-1} \hat{B} T_1 T_3 = \begin{pmatrix} R & 0 \\ 0 & I \end{pmatrix}.$$

Now set

$$P := T_3^{-1} T_2^{-1} T_1^{-1} (cA + B)^{-1} \quad \text{and} \quad Q = T_1 T_3$$

to get the statement. □

Definition 2. *The nilpotency index k from the Weierstraß-Kronecker Normalform of a matrix pencil $\{A, B\}$ with A singular is called the* Kronecker-Index *of $\{A, B\}$. We write $ind\{A, B\}$. For A regular we set $ind\{A, B\} = 0$.*

next lemma needs quotation, it is lemma 13.2.1

Lemma 1. *The Kronecker-Index $ind\{A, B\}$ is independant of the choice of the matrices P and Q.*

Lemma 13.2.2 maybe

Using the findings above we are able to Transform the initial DAE 3.2 using the matrix P from 2. By multiplying P from the left we obtain

$$PAy'(t) + PBy(t) = Pf(t).$$

Setting

$$y = Q \begin{pmatrix} u \\ v \end{pmatrix}, \quad Pf(t) = \begin{pmatrix} s(t) \\ q(t) \end{pmatrix} \quad \text{with} \quad u, s \in \mathbb{R}^d,$$

we get a system of the form

$$\begin{aligned} u'(t) + Ru(t) &= s(t) \\ Nv'(t) + v(t) &= q(t) \end{aligned} \tag{3.3}$$

The first equation is an ordinary differential equation of first order and posesses a unique solution $u(t)$ in $[t_0, t_l]$ for any starting values $u_0 \in \mathbb{R}^d$. Additionally setting $q(t) \in C^{k-1}([t_0, t_l])$ then differentiating the second equation in 3.3 gives

$$\begin{aligned} v(t) &= q(t) - Nv'(t) = q(t) - N(\underbrace{q(t) - Nv'(t)}_{=v(t)})' = q - Nq' + N^2v'' \\ &= q - Nq' + N^2(q - Nv')'' = q - Nq' + N^2q'' - N^3v''' \\ &\vdots \\ &= q - Nq' + ... + (-1)^{k-1}N^{k-1}q^{(k-1)} + (-1)\underbrace{N^kv^{(k)}}_{=0} \end{aligned}$$

$$= \sum_{i=0}^{k-1}(-1)^iN^iq^{(i)}(t) \tag{3.4}$$

where k is the nilpotency index of N. This expression gives an explicit solution for $v(t)$ in $[t_0, t_l]$ with $v(t) \in \mathbb{R}^{d-1}$. It shows the depencdency of the solution and its derivatives. The higher the Kronecker index k gets, the more differentiations of $q(t)$ have to be performed.

The Kronecker index k shows, that k differentiations are required to receive an ordinary differential equation.

3.3 Index of a Differential Algebraic Equation

The Index of a DAE gives us insight about it's numerical properties and in general about the solvability. In general, the higher the index, the harder it is, to solve the system.

We will consider two types of index concepts, the differentiation index and the perturbation index.

Definition 3 (differentiation index)**.** *Consider the differential algebraic equation 3.1 to be uniquely locally solvable and F suffieciently smooth differentiable. For a given $m \in \mathbb{N}$ consider*

$$\begin{aligned} F(t, y, y') &= 0, \\ \frac{\mathrm{d}F(t, y, y')}{\mathrm{d}t} &= 0, \\ &\vdots \\ \frac{\mathrm{d}^m F(t, y, y')}{\mathrm{d}t^m} &= 0. \end{aligned}$$

The smallest natural number m for which the above System results in an explicit system of the form

$$y' = \phi(t, y)$$

from which y can be determined is called ***differentiation index****.*

In the previous chapter we have already discussed, that for a DAE with constant coefficients 3.2 and a regular matrix pencil $\{A, B\}$ we need $k = ind\{A, B\}$ differentiations to receive an ordinary differential equation. This means that the Kronecker index k is equal to the differentiation index in the case of a DAE with constant coefficients.

Definition 4 (perturbation index)**.** *Let $y(t)$ be the exact solution to 3.1. This problem has the* ***perturbation index*** *$k \in \mathbb{N}$ along $y(t), t_0 \leq t \leq T$ if for all $\tilde{y}(t)$ with $F(t, \tilde{y}, \tilde{y}') = \delta(t)$ the inequality*

$$||y(t) - \tilde{y}(t)|| \leq C \left(||y(t_0) - \tilde{y}(t_0)|| + \sum_{j=0}^{k} \max_{t_0 \leq \xi \leq T} \left\| \int_{t_0}^{\xi} \frac{\mathrm{d}^j \delta}{\mathrm{d}\tau^j}(\tau) d\tau \right\| \right)$$

for the smallest number k.

for lin const nilpot ind = ind + proof

Bemerkung 13.3.5. Im Fall linearer differential-algebraischer Systeme (13.2.1) mit regulärem Matrixbüschel $\{A, B\}$ kann die DAE

$$A(y'(t) - \widetilde{y}'(t)) + B(y(t) - \widetilde{y}(t)) = \delta(t)$$

nach Folgerung 13.2.1 transformiert werden in

$$\begin{aligned} u'(t) - \widetilde{u}'(t) + \widehat{R}(u(t) - \widetilde{u}(t)) &= \delta_1(t) \\ N(v'(t) - \widetilde{v}'(t)) &= \delta_2(t). \end{aligned} \tag{13.3.29}$$

Aus der Lösungsdarstellung von (13.3.29), vgl. (13.2.11), folgt, dass der Störungsindex pi gleich dem Kronecker-Index k ist. Mit Bemerkung 13.3.1 gilt damit $k = di = pi$. □

Bezüglich des Störungsindex von (13.3.8) gilt der

Figure 3.1:

4 Index Analysis of the MNA

der ane satz do + proof hopefully

seite 22 kap 7 netw top and dae ind for rlc - modelling and discretization circ prob

In the previous chapter we have seen two different kinds of Index concepts for differential algebraic equations. We have also seen that these two, even thoough they describe rather different structural aspects of the equation, are the same for our use-cases. This of course leads to the question : ´´What are the indices we can usually expect for MNA?".

This chapter aims to answer that question for the linear RLC case. Which means that the RLC components are described by linear functions with positive capacitances, inductances and resistances. Thus the matrices

$$C := \frac{\partial q_C(w)}{\partial w}, \quad L := \frac{\partial \phi_L(w)}{\partial w}, \quad G := \frac{\partial r(w)}{\partial w}$$

are positive definite and symmetric.

The generalization of these results to the nonlinear case still relies on positive definiteness.

Recall that we consider the equations resulting from the analysis above. These equations are of the form 3.2

$$Ay'(t) + By(t) = f(t).$$

Specifically the obtained equations from the Modified Nodal Analysis 2.5 are

$$\begin{pmatrix} A_C C A_C^\top & 0 & 0 \\ 0 & L & 0 \\ 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} \dot{u} \\ \dot{i_L} \\ \dot{i_V} \end{pmatrix} + \begin{pmatrix} A_R G A_R^\top & A_L & A_V \\ -A_L^\top & 0 & 0 \\ -A_V^\top & 0 & 0 \end{pmatrix} * \begin{pmatrix} u \\ i_L \\ i_V \end{pmatrix} = \begin{pmatrix} -A_I i_{src} \\ 0 \\ -v_{src} \end{pmatrix}.$$

4.1 General Index analysis

Assuming the system only contains linear elements or is linearized at an operating point in order to investigate the system behaviour then the corresponding network equation represent a DAE with constant coefficients 3.2. We will denote $x = (u, i_L, i_V)^\top$. The structure of the system is reliant on the matrix B, thus we consider

- **ODE-case**:
  The matrix B is regular in 3.2. This is the case iff the circuit contains no voltage sources and there are no nodes which have no path to ground via capacitors. Then the system represents a linear-implicit system of ODEs and can be transformed into the explicit ODE sytstem

  $$\dot{x} = B^{-1}(-Ax + f(t)).$$

  Thus we obtain an index of 0.

- **DAE-case**: The matrix B is singular in 3.2. This is the interesting case which we will analyse further.

Investigating the index-cases further we can again distinct into two cases:

reminder for the structure obtained in previous chapter

$$u'(t) + Ru(t) = s(t),$$
$$Nv'(t) + v(t) = q(t)$$

1. index 1 case Because N is nilpotent with nipotency index $\nu = 1$ it holds that $N^1 = 0$, thus the system transforms to

$$u'(t) + Ru(t) = s(t),$$
$$v(t) = q(t).$$

   This means that the algebraic variables are given explicitly. Thus the system can be written in ODE form

$$\dot{x} = B^{-1}(-Ax + f(t)).$$

2. index ≥ 2 case

   results in the solution we derived in 3.4. The index of the MNA is the same as $ind(A, B)$, which denotes the nilpotency index of N as we have seen in the previous analysis in chapter 3.

continue herer with modelling book from bottom of page 20-21 (did that on 26.11)

an algebraic constraint has to be fulfilled by thesolution. case of index 1 this equation is given explicitly. for index geq 2 it is given implicitly (i need to show that somewhere)

system is sensitive to perturbations - small input noise can have arbitrarily large derivatives (shows in formula above for z)

book says - no severe numerical problems with index 1. might show up for implicit algebraic constraint. Hence implicit numerical integration schemes for stiff systems are feasable.

however sever numerical problems may arrise for index geq 2. hidden algebraic constraints can make problems.

index is called algebraic index? - elaborate maybe on that

put that one chapter up?

differentiation index - since numerical differentiation is an unstable procedure this index gives a measure for the numerical problems to be expected when solving such systems

perturbation index - derivatives of perturbations enter solution

remark: tracability index

4.2 Topological Conditions

For this we consider RLC networks with independant voltage and current sources (**what does that mean?**). To obtain the perturbation index of the MNA we perturb the right-hand side of
reference to Tischendorf

basically chapter 7 of modelling book

From analyzing the MNA some conditions to the circuit topology can be obtained. We will be concidering the impact of loops containing only capacitances and voltage sources as well as cutsets containing only inductances and current sources.

In [**Tischendorf2005Topological**] they present very interesting results about the index of MNA equations. Namely the following:

Theorem 3 (Index-1 condition)**.** *[***Tischendorf2004Topological***] Let the matrices of the capacitances, inductances and resistances respectively be positive definite. If the network neither cointains* inductance-current-source cutsets *nor (controlled?)* capacitance-voltage-source loops*, then the MNA leads to an index-1 DAE.*

Theorem 4 (Index-2 condition)**.** *[***Tischendorf2004Topological***] If the Network contains* inductance-current-source cutsets *or* capacitance-voltage-source loops *except for capacitance-only loops, then the MNA leads to an index-2 DAE.*

WE will apply those results to some examples:

5 Numerical Solutions

As we are usually interested in finding real (or complex) valued solutions of our systems we also have to look into solving them numerically. This chapter focuses on the numerical solution of the mentioned systems.

We will first focus on methods used to solve a more general problem

$$y'(t) = f(t, y), \quad t \in [t_0, t_l], \tag{5.1}$$
$$y(t_0) = y_0. \tag{5.2}$$

For this we presume that the function $f(t, y)$ is continuous and Lipschitz, thus *Picard-Lindelöf* gives us, that for every y_0 it is uniquely solvable in $[t_0, t_l]$.

Numerical Methods work by discretization, this means we divide the time-intervall into

$$t_0 < t_1 < ... < t_N \leq t_l$$

and consider approximations $y_m \approx y(t_m)$ for $m = 1, ..., N$. Gitterpunkte, SChrittweite, äquidistant, nicht äquidistant.

5.1 Single-Step-Methods

The first class of numerical methods we will have a look at are single-step methods. These methods use the previous approximated value y_j and (for implicit methods) also the current approximated value to determine the current value through a *procedural function*. ... waht does implicit and explicit mean?

Definition 5. *A numerical method to approximate a differential equation 5.1 on a time-grid $t_0, ..., t_l$ with the intermediate values $y_0, ..., y_l$ is called a single-step method if it is from the form*

$$y_{j+1} = y_j + h_j \phi(t_j, y_j, y_{j+1}, h_j). \tag{5.3}$$

With the procedural function ϕ. *If ϕ is not dependent on y_{j+1} then the method is called* explicit, *otherwise it is called* implicit.

5.1.1 Consistency, Stability and Convergence

To compare different single-step methods we have to define some notions to compare their quality. This leads to the definition of the error of the mehtod, its consistency and its convergence.

Definition 6. *Let $\tilde{y}_{m+1}$ be the result of one step of 5.3 with the exact start-vector $y_m = y(t_m)$ then*

$$le_{m+1} = le(t_m + h) = y(t_{m+1}) - \tilde{y}_{m+1}, \quad m = 0, ..., N-1 \tag{5.4}$$

is called the local discretization error *of the single step method at the point t_{m+1}.*

The error encodes how far off the numerical method is from the true value of the solution. In real applications this solution is usually not known. This shows how important error bounds for numerical methods can be.

Definition 7. *A single-step method is called* consistent *if for all initial value problems 5.1*

$$\lim_{h \to 0} \frac{||le(t+h)||}{h} = 0 \quad for \quad t_0 \leq t \leq t_l \tag{5.5}$$

holds.
It is called consistent of order p*, if for a sufficiently smooth function f*

$$||le(t+h)|| \leq Ch^{p+1} \quad \textit{for all} \quad h \in (0, H] \quad \textit{and} \quad t_0 \leq t \leq t_l - h \tag{5.6}$$

holds with C not dependent on h.

Consistency aims to give insight in how similar the problem that the numerical methods solves it to the real problem that we want the solution from.

Definition 8. *A single-step method is called* convergent*, if for all initial value problems 5.1 for the* global discretization error

$$e_m = y(t_m) - y_h(t_m)$$

holds that

$$\max_m ||e_m|| \to 0 \quad \textit{for} \quad h_{max} \to 0.$$

The single-step method is called to have the convergence order *p, if*

$$\max_m ||e_m|| \leq Ch_max^p \quad \textit{for} \quad h_{max} \in (0, H] \quad \textit{with} \quad t_0 \leq t_m \leq t_l$$

with the constant C not dependent on the step size h.

As the name suggestes convergence tries to quantify how far off a numerical solution is from the real solution of a system. A very interesting result follows if we also require the Single-Step Method to be stable.

Definition 9. *A Single-Step Method is called* ()discretely) stable *if for grid-functions y_h and $\tilde{y}_h$ with*

$$y_{i+1} = y_i + h\phi(t_i, y_i), \tag{5.7}$$
$$\tilde{y}_{i+1} = \tilde{y}_i + h[\phi(t_i, \tilde{y}_i) + \theta_i], \tag{5.8}$$

and perturbations $\theta_i = \theta_h(t_i)$ of the right side as well as a bounded perturbation in the staring-values $y_0 - \tilde{y}_0$ the Error is bounded by

$$||y_h - \tilde{y}_h||_{\infty,h} \leq C(||y_0 - \tilde{y}_0||_{l^2} + ||\theta_h||_{\infty,h}$$

with a constant C which is not dependent on h.

For Single-Step Methods which are consistent and stable we obtain the following convergence theorem.

Theorem 5 (Lax-Richtmyer)**.** *A consistent (with order p) and discretely stable Single-Step Method is convergent (with order p). (assuming smoothness of the solution y)*

This theorem is due to Lax and Richtmyer. The converse of this statement is also true.

5.1.2 Runge-Kutta Methods

A very prevalent family of numerical single-step methods are the *Runge-Kutta* methods. ... Whats the idea behind those methods?

Definition 10. *Let $s \in \mathbb{N}$. A single-step method of the form*

$$y_{m+1} = y_m + h \sum_{i=1}^{s} b_i f(t_m + c_i h, y_{m+1}^{(i)}) \tag{5.9}$$

$$y_{m+1}^{(i)} = y_m + \sum_{j=1}^{s} a_{ij} f(t_m + c_j h, y_{m+1}^{(j)}) \tag{5.10}$$

is called a Runge-Kutta Method *with s steps.*

We usually collect the coefficients into the vectors and matrices $c = (c_1, ..., c_s)$, $A = (a_{ij})_{ij}$ and $b = (b_1, ..., b_s)$.

If A is a strictly lower triangle matrix, this means for all $j \geq i$ holds $a_{ij} = 0$ then the Runge-Kutta method is explicit, otherwise it is implicit. In general implicit Runge-Kutta methods might need more computational effort because to calculate $y_m^{(i)}$ a nonlinear system of

equations has to be solved. But in contrast those methods can also lead to very good stability characteristics.

Lemma 2. *A Runge-Kutta mehtod is consistent, if and only if*

$$\sum_{i=1}^{s} b_i = 1$$

The coefficients of a Runge-Kutta method are usually represented in the *Butchertableau*, which was introduced by John C. Butcher and has the following form.

$$\begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \dots & a_{ss} \\ \hline & b_1 & \dots & b_s \end{array} \quad \text{of in matrix form} \quad \begin{array}{c|c} c & A \\ \hline & b \end{array}$$

5.1.3 further stability properties

from numpdgl skript
In this section we consider the following simple problem

$$u' = \lambda u, \quad t > 0 \tag{5.11}$$

$$u(0) = u_0 \tag{5.12}$$

with $\lambda \in \mathbb{C}$ and u_0 fixed.

Definition 11. *1. If a single-step method can be written in the form*

$$u_{i+1} = R(z)u_i, \quad z := h * y \tag{5.13}$$

then we call $R : \mathbb{C} \to \mathbb{C}$ the stability function *of the single-step method.*

2. *The set*

$$S := z \in \mathbb{C} : |R(z)| \leq 1 \tag{5.14}$$

is called the region of stability of the method.

3. *A single-step method is called*

   - 0-stable, *if* $0 \in S$.
   - A-stable, *if* $\mathbb{C}^- \subset S$.
   - L-stable, *if* $R(z) \to 0$ *for* $Re(z) \to -\infty$.

5.2 Multistep-Methods

based on chapter 4 of book num gew dgl steif nichtsteif
Linear multistep methods use approxiamtions u_{m+l} along the gridpoints t_{m+l}, $l = 0, 1, ..., k-1$ to calculate the new approximation u_{m+k} at t_{m+k}. WE will first discuss topics related to the order of the methods depending on its parameters, stability and convergence.

Definition 12. *For given* $\alpha_0, ..., \alpha_k$ *and* $\beta_0, ..., \beta_k$ *the iteration rule*

$$\sum_{l=0}^{k} \alpha_l u_{m+l} = h \sum_{l=0}^{k} \beta_l f(t_{m+l}, u_{m+l}), \quad m = 0, 1, ..., N-k \tag{5.15}$$

is called a linear multistep method *(linear k-step method). It is always assumed that* $\alpha_k \neq 0$ *and* $|\alpha_0| + |\beta_k| > 0$. *If* $\beta_k = 0$ *holds, then the method is called explicit, otherwise implicit.*

Durch die Forderung $|\alpha_0| + |\beta_0| > 0$ ist die Schrittzahl k eindeutig festgelegt. Im Falle $\beta_k = 0$ lässt sich die Näherungsfolge $\{u_{m+k}\}$, $m = 0, 1, \ldots, N - k$, direkt berechnen. Die Verfahrensvorschrift (4.2.1) liefert demzufolge für jedes äquidistante Gitter I_h eine eindeutig bestimmte Gitterfunktion $u_h(t)$. Ist $\beta_k \neq 0$, so hat man zur Bestimmung von u_{m+k} ein i. Allg. nichtlineares Gleichungssystem der Form

$$u_{m+k} = h\frac{\beta_k}{\alpha_k} f(t_{m+k}, u_{m+k}) + v, \tag{4.2.2}$$

zu lösen, wobei der von u_{m+k} unabhängige Vektor v durch

$$v = \frac{1}{\alpha_k} \sum_{l=0}^{k-1} \Big(h\beta_l f(t_{m+l}, u_{m+l}) - \alpha_l u_{m+l} \Big)$$

gegeben ist. Zur Lösung von (4.2.2) verwendet man für nichtsteife Systeme Funktionaliteration, d. h.

$$u_{m+k}^{(\varkappa+1)} = h\frac{\beta_k}{\alpha_k} f(t_{m+k}, u_{m+k}^{(\varkappa)}) + v, \quad \varkappa = 0, 1, \ldots$$

Unter der Schrittweiteneinschränkung

$$h \left| \frac{\beta_k}{\alpha_k} \right| L < 1, \tag{4.2.3}$$

wobei L eine Lipschitz-Konstante für $f(t, y)$ darstellt, konvergiert die Folge $\{u_{m+k}^{(\varkappa)}\}$ bei beliebig vorgegebenem Startvektor $u_{m+k}^{(0)}$ gegen die eindeutige Lösung von (4.2.2). Für ein nichtsteifes Anfangswertproblem ist die Bedingung (4.2.3) an h keine wesentliche Einschränkung.

Ein lineares Mehrschrittverfahren setzt sich zusammen aus zwei Bestandteilen:

1. Der *Startphase* zur Berechnung der Näherungswerte $u_1, \ldots, u_{k-1}$ in den Gitterpunkten $t_l = t_0 + lh$, $l = 1, \ldots, k-1$, die mit einem Einschrittverfahren, z. B. mit einem expliziten Runge-Kutta-Verfahren, oder mit Mehrschrittformeln niedriger Schrittzahlen und sehr kleinen Schrittweiten bestimmt werden können.
2. Der *Laufphase*, d. h. einer Mehrschrittformel (4.2.1) zur sukzessiven Berechnung der Approximationen u_{m+k} in den Gitterpunkten t_{m+k}.

Figure 5.1:

A linear multi-step method consists of two parts:

1. In the *starting-phase* approximations $u_1, ..., u_{k-1}$ for the first $k-1$ gridpoints $t_l = t_0 + th, l = 1, ..., k-1$ are calculated using a single-step method. For example using an explicit Runge-Kutta Methodor a multi-step method with fewer steps.

2. In the *run-phase* the multi-step formula is used to determine new approximations u_{m+k} for the gridpoint t_{m+k}

For theoretical analysis of the multi-step methods we consider the generating polynomials

$$\rho(x) := \sum_{l=0}^{k} \alpha_l x^l \tag{5.16}$$

$$\sigma(x) := \sum_{l=0}^{k} \beta_l x^l \tag{5.17}$$

Eine zentrale Rolle bei der theoretischen Untersuchung linearer Mehrschrittverfahren spielen die beiden *erzeugenden Polynome*

$$\rho(\xi) := \alpha_k \xi^k + \alpha_{k-1}\xi^{k-1} + \cdots + \alpha_0$$
$$\sigma(\xi) := \beta_k \xi^k + \beta_{k-1}\xi^{k-1} + \cdots + \beta_0.$$

Sie wurden erstmals von Dahlquist [79] zur Stabilitätsuntersuchung linearer Mehrschrittverfahren verwendet. Mit den erzeugenden Polynomen lassen sich die Konsistenzbedingungen (4.2.7) in der Form

$$\rho(1) = 0 \text{ und } \rho'(1) = \sigma(1). \tag{4.2.11}$$

Figure 5.2:

5.2.1 Consistency, Stability and Convergence

local discretization error - def 4.2.2

Definition 13. *Let $\tilde{y}_{m+k}$ be the result of one step of the multi-step method 5.15 with the start-vectors $y_m = y(t_m)$. This means*

$$\alpha_k \tilde{u}_{m+k} = \sum_{l=0}^{k-1} (h\beta_l f(t_{m+l}, y(t_{m+l})) - \alpha_l y(t_{m+l})) + h\beta_k f(t_{m+k}, \tilde{u}_{m+k}).$$

Then

$$le_{m+k} = le(t_{m+k}) = y(t_{m+k}) - \tilde{u}_{m+k}, \quad m = 0, 1, ..., N-k$$

is called the local discretization error *(local error) of the linear multi-step method 5.15 at the point* t_{m+k}.

We will assign the linear difference opreator

$$L[y(t), h] = \sum_{l=0}^{k} \left(\alpha_l y(t + lh) - h\beta_l y'(t + lh) \right) \tag{5.18}$$

to the local discretization error. Using this we gain the following definition.

Definition 14. *A linear multi-step method is called* preconsistent *if for all functions* $y(t) \in C^1[t_0, t_l]$

$$\lim_{h \to 0} L[y(t), h] = 0$$

holds. It is called consistent, *if for all functions* $y(t) \in C^2[t_0, t_l]$

$$\lim_{h \to 0} \frac{1}{h} L[y(t), h] = 0$$

holds. It has the consistency order p, *if for all functions* $y(t) \in C^{p+1}[t_0, t_l]$

$$L[y(t), h] = \mathcal{O}(h^{p+1}) \quad for \quad h \to 0$$

holds.

Definition 15. *A linear multi-step method is called stable, if for solutions* u_h *and* $\tilde{u}_h$ *of*

$$\sum_{l=0}^{k} \alpha_l u_{m+l} = h \sum_{l=0}^{k} \beta_l f(t_{m+l}, u_{m+l}), \tag{5.19}$$

$$\sum_{l=0}^{k} \alpha_l \tilde{u}_{m+l} = h \sum_{l=0}^{k} \beta_l f(t_{m+l}, \tilde{u}_{m+l}) + h\theta_n \tag{5.20}$$

and bounded initial values $y_j - y_j$ for $j \in 0, ..., k$ we have that

$$\max_{t_0 \leq t_n \leq T} ||y_n - \tilde{y}_n|| \leq C \sum_{j=0}^{k-1} ||y_j - \tilde{y}_j|| + \max_{t_0 \leq t_n \leq T} ||\theta_n||.$$

Definition 16. *A linear multi-step method is called* zero-stable *if all solutions of the difference equation*

$$\sum_{l=0}^{k} \alpha_l u_{m+l} = 0$$

are bounded.

Theorem 6. *A linear multi-step method is zero-stable, if and only if the polynomial $\rho(x)$ fullfills the root-condition, this means:*

1. *All roots $\bar{x}$ of $\rho(x)$ are within the unit-circle $|\bar{x}| \leq$ in the complex plane.*
2. *All roots $\bar{x}$ with $|x| = 1$ are singular.*

Theorem 7 (DAE lecture)**.** *A linear multistep method is stable if and only if it is zero-stable.*

from circuit book below, above from modelling book

5.2.2 further stability properties

In this section we consider the following simple problem

$$u' = \lambda u, \quad t > 0 \tag{5.21}$$

$$u(0) = u_0 \tag{5.22}$$

with $\lambda \in \mathbb{C}$ and u_0 fixed.

Thus the resulting linear multistep method is of the form

$$\sum_{l=0}^{k} \alpha_l u_{n+l} = h \sum_{l=0}^{k} \beta_l \lambda u_{n+l}$$
$$\iff \sum_{l=0}^{k} [\alpha_l - h\beta_l \lambda] u_{n+l}$$

Definition 17. *1. The set*

$$S := z \in \mathbb{C} : \rho(\xi) - z\sigma(\xi) = 0 \implies \xi \textit{fullfills root criteria} \tag{5.23}$$

is called the region of stability of the method.

2. A linear multistep method is called

- 0-stable, *if* $0 \in S$.
- *stable in the point* $z \in \mathbb{C}$, *if Wurzelbedingung erfüllt bei* z, *aber doch eigentlich* z *in* S *oder?*
- $A(\alpha)$-stable, *if it is stable in all* z *that lie within the set*

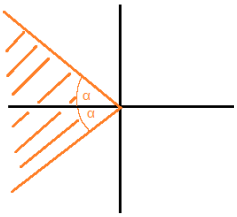


Figure 5.3:

5.3 Implicit linear multi-step formulas

These kinds of multi-step methods are conventionally used to numerically solve the systems obtained using modified nodal analysis. We will assume that the network equations arrise from networks only consisting of RLC components as well as controlled sources which keep the index between 1 and 2.

The conventional approach can be split into three main steps:

1. Computation of consistent initial values
2. numerical integration based on multi-step schemes
3. transformation of the DAE into a nonlinear system and its numerical solutioon by Newton's procedure (??????????????????will not be discussed further because not very specific)

Consistant initial values

Consistent initial values. The first step in the transient analysis is to compute consistent initial values (x_0, y_0) for the initial time point t_0. In the index-1 case, this can be done by performing a steady state (DC operating point) analysis, i.e. to solve

$$\mathcal{F}(0, x_0, t_0) = 0 \tag{10.2}$$

for x_0 and then set $y_0 := g(x_0)$. If there are no controlled sources, the Jacobian $\partial\mathcal{F}/\partial x$ of (10.2) with respect to x_0 reads

$$\frac{\partial \mathcal{F}}{\partial x} = \begin{pmatrix} \widetilde{G}(A_R^\top u_0, t_0) & A_L & A_V \\ -A_L^\top & 0 & 0 \\ -A_V^\top & 0 & 0 \end{pmatrix}$$

with the definition $\widetilde{G}(A_R^\top u, t) := A_R G(A_R^\top u, t) A_R^\top$ already introduced in Section 4. Since $\ker(\partial\mathcal{F}/\partial x) = \ker(A_R, A_L, A_V)^\top \times \ker(A_L, A_V)$ holds, the matrix is only regular, if there are neither loops of independent voltage sources and/or inductors, nor cutsets of independent current sources and/or capacitors. If these topological conditions are violated, no steady state solution can be computed, and so most circuit analysis programs check and refuse these circuit configurations. Additional assumptions are implied in the case of controlled sources. But note that in the nonlinear case the Jacobian matrix also may become numerically singular, e. g. due to vanishing partial derivatives or in the case of bifurcation.

An approach always feasible in the index-1 case is to extract the algebraic constraints using the projector Q_C onto $\ker A_C^\top$:

$$\begin{aligned} Q_C^\top(A_R r(A_R^\top u, t) + A_L \jmath_L + A_V \jmath_V + A_I \imath(u, \jmath_L, \jmath_V, t)) &= 0 \\ v(u, \jmath_L, \jmath_V, t) - A_V^\top u &= 0. \end{aligned}$$

If the index-1 topological conditions hold, this nonlinear system uniquely defines for $t = t_0$ the algebraic components $Q_C u_0$ and $\jmath_{V,0}$ for given (arbitrary) differential components $(I - Q_C)u_0$ and $\jmath_{L,0}$. The derivatives $\dot{y}_0$ have then to be chosen such that $A\dot{y}_0 + f(x_0, t_0) = 0$ holds.

Figure 5.4:

Nunmerical integration.

5.3.1 BDF-schemes

chapter 9.2 numerik book wikipedia

The most commonly used numerical methods for solving the systems that arrise in electrical circuits are the BDF-scheme and the trapezoidal rule.

We will not give a deeper look into their construction but will only state their properties. (specifically for BDF schemes)

The *.backward differentiation fooormula (BDF)* is a family of implicit linear multistep methods. They have the general form

$$\sum_{k=0}^{s} \alpha_k y_{n+k} = h\beta f(t_{n+s}, y_{n+s}) \tag{5.24}$$

Since we are interested in the unknown y_{n+s} which is used to evaluate f, this method is implicit. The coefficients α_k and β are chosen, so that the method achieves order s which is the maximum possible.

The BDF or BDF-k formulas for $k = 1, ...6$ have the following form

$$\begin{aligned}
k = 1 &: hf_{m+1} = u_{m+1} - u_m \\
k = 2 &: hf_{m+2} = \frac{1}{2}(3u_{m+2} - 4iu_{m+1} + u_m) \\
k = 3 &: hf_{m+3} = \frac{1}{6}(11u_{m+3} - 18u_{m+2} + 9u_{m+1} - 2u_m) \\
k = 4 &: hf_{m+4} = \frac{1}{12}(25u_{m+4} - 48u_{m+3} + 36u_{m+2} - 16u_{m+1} + 3u_m) \\
k = 5 &: hf_{m+5} = \frac{1}{260}(137u_{m+5} - 300u_{m+4} + 300u_{m+3} - 200u_{m+2} + 75u_{m+1} - 12u_m) \\
k = 6 &: hf_{m+6} = \frac{1}{60}(147u_{m+6} - 360u_{m+5} + 450u_{m+4} - 400u_{m+3} + 225u_{m+2} - 72u_{m+1} + 10u_m)
\end{aligned}$$

Methods with $s > 6$ are not zero-stable and (cannot be used?).

BDF schemes have consistency order $p = k$.

BDF schemes are methods for solving stiff equations, thus their stability is indicated by their region of absolute stability. They are not A stable but their stability region still contains a large part of the complex left half-plane. They are the most efficient linear multistep methods of this kind.

The first timestep is always performed by BDF1 (implicit Euler scheme) as a starting procedure.

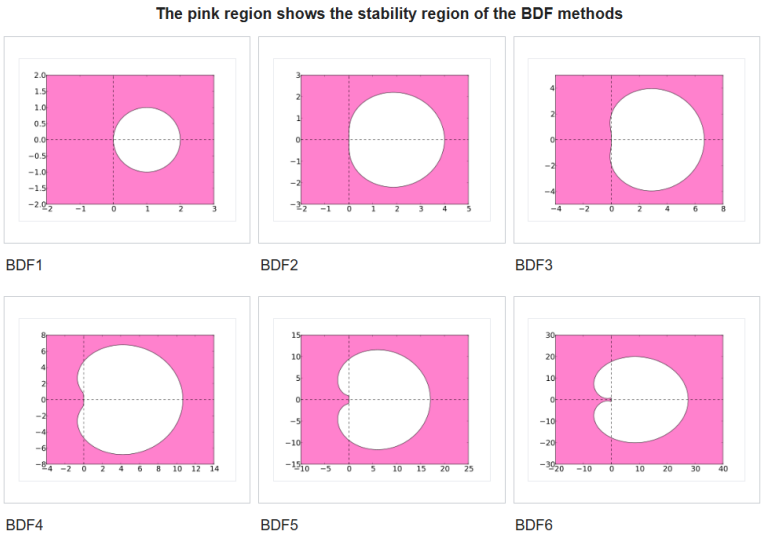


Figure 5.5:

5.3.2 trapezoidal rule

The trapezoidal rule is a somewhat natural alternative to BDF2 since it is A-stable and as a linear multistep method of order 2 the one with the smallest leading error coefficient (source? - aus modelling book)

It works by approximating the region under the graph of the function $f(x)$ as a trapezoid, hence the name. It follows that (picture)

$$\int_a^b f(x)dx \approx (b-a)\frac{1}{2}(f(a)+f(b))$$

This procedure is repeated for small subsections of the Intervall $[a,b]$. Thus we obtain the iteration formula

$$u_h(t+h) = u_h(t) + \frac{h}{2}[f(t,u_h(t)) + f(t+h,u_h(t+h))].$$

This iteration rule can also be formulated using the butcher tableau

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & \frac{1}{2} & \frac{1}{2} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

modelling book says

Because $u_h(t+h)$ appears in f again we see that this is an implicit method. The butcher tableau confirms this as well.

energy conserving (BDF methods are not)