Analysis and waveform relaxation for a differential-algebraic electrical circuit model

DISSERTATION

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Jonas Pade

Gutachter/innen: 1. Prof. Dr. Caren Tischendorf
2. Prof. Dr. Sebastian Schöps
3. Prof. Dr. Ricardo Riaza

Tag der mündlichen Prüfung: 22.03.2021

Kurzzusammenfassung

Die Hauptthemen dieser Arbeit sind einerseits eine tiefgehende Analyse von nichtlinearen differential-algebraischen Gleichungen (DAEs) vom Index 2, die aus der modifizierten Knotenanalyse (MNA) von elektrischen Schaltkreisen hervorgehen, und andererseits auf der Basis dieser Analyse die Entwicklung von Konvergenzkriterien für Waveform Relaxationsmethoden zum Lösen gekoppelter Probleme. Ein Schwerpunkt in beiden genannten Themen sind die Beziehungen zwischen der Topologie eines Schaltkreises und mathematischen Eigenschaften der zugehörigen DAE.

Der Analyse-Teil umfasst eine detaillierte Beschreibung einer Normalform für Schaltkreis DAEs vom Index 2 und Folgerungen, die sich für die Sensitivität des Schaltkreises bezüglich seiner Input-Quellen ergeben. Wir präsentieren Abschätzungen, die Aufschluss darüber geben, wie stark sich eine Änderung in den Input-Quellen auf andere Größen im Schaltkreis auswirkt. Entscheidende Konstanten dieser Abschätzungen werden angegeben und mit der topologischen Position der jeweiligen Input-Quelle im Schaltkreis in Beziehung gesetzt.

Technologische Geräte basieren auf zunehmend komplexen Schaltkreisen, für deren Modellierung sich oftmals eine Betrachtung als gekoppeltes System empfiehlt. Waveform relaxation (WR) ist ein geeigneter Ansatz zur Lösung solch gekoppelter Probleme, da sie das Verwenden von auf die Subprobleme angepassten Lösungsmethoden und Schrittweiten ermöglicht. Es ist bekannt, dass WR zwar bei Anwendung auf gewöhnliche Differentialgleichungen konvergiert, falls diese eine Lipschitz-Bedingung erfüllen, selbiges jedoch bei DAEs nicht ohne Hinzunahme eines Kontraktivitätskriteriums sichergestellt werden kann. Wir beschreiben allgemeine Konvergenzkriterien für WR auf DAEs vom Index 2. Auf der Basis der Ergebnisse aus dem Analyse-Teil leiten wir außerdem topologische Konvergenzkriterien für gekoppelte Probleme zweier Schaltkreise und solche, bei denen ein elektromagnetisches Feld mit einem Schaltkreis gekoppelt ist, her. Anhand von Beispielen wird veranschaulicht, wie überprüft werden kann, ob ein hinreichendes Konvergenzkriterium erfüllt ist. Weiterhin werden die Konvergenzraten des Jacobi WR Verfahrens und des Gauss-Seidel WR Verfahrens angegeben und verglichen. Simulationen von einfachen gekoppelten Beispiel-Systemen zeigen drastische Unterschiede des Konvergenzverhaltens von WR Methoden, abhängig davon, ob die topologischen Konvergenzbedingungen erfüllt sind oder nicht.

Abstract

The main topics of this thesis are firstly a thorough analysis of nonlinear differential-algebraic equations (DAEs) of index 2 which arise from the modified nodal analysis (MNA) for electrical circuits, and secondly, based on this analysis, the derivation of convergence criteria for waveform relaxation (WR) methods on coupled problems. In both topics, a particular focus is put on the relations between a circuit's topology and the mathematical properties of the corresponding DAE.

The analysis encompasses a detailed description of a normal form for circuit DAEs of index 2 and consequences for the sensitivity of the circuit with respect to its input source terms. More precisely, we provide bounds which describe how strongly changes in the input sources of the circuit affect its behaviour. Crucial constants in these bounds are determined in terms of the topological position of the input sources in the circuit.

The increasingly complex electrical circuits in technological devices often call for coupled systems modelling. Allowing for each subsystem to be solved by dedicated numerical solvers and time scales, WR is an adequate method in this setting. It is well-known that while WR converges on ordinary differential equations if a Lipschitz condition is satisfied, an additional convergence criterion is required to guarantee convergence on DAEs. We present general convergence criteria for WR on higher index DAEs. Furthermore, based on our results of the analysis part, we derive topological convergence criteria for coupled circuit/circuit problems and field/circuit problems. Examples illustrate how to practically check if the criteria are satisfied. If a sufficient convergence criterion holds, we specify at which rate of convergence the Jacobi and Gauss-Seidel WR methods converge. Simulations of simple benchmark systems illustrate the drastically different convergence behaviour of WR depending on whether or not the circuit topological convergence conditions are satisfied.

Preface

First of all I would like to thank my supervisor Caren Tischendorf who gave me the opportunity to work on this thesis. Her steady support as well as her scientific guidance were invaluable to me. Furthermore, I am grateful to Ricardo Riaza and Sebastian Schöps for co-examining this work. Collaborating with Sebastian Schöps and Idoia Cortes Garcia was always a pleasure and their knowledge and ideas more than once opened my eyes for new aspects of my research topic.

This thesis was written during my time in the working group of Caren Tischendorf, where I worked in the Federal Ministry for Economic Affairs funded projects KoSMos (05M13PXA) and MathEnergy (0324019E). I had the chance of meeting many kind and inspiring people during the years. Tom Streubel, Maximilian Schade, Christian Strohm, Malak Diab, Christoph Huck, Dennis Groh and Henning Sauter created a pleasant working atmosphere and made lunch breaks most enjoyable.

Finally, I want to thank the fabulous proofreaders Jan Philipp Pade, Tamina Pade, Marie-Pascale Hardy and Jochen Pade mostly for everything beyond proofreading.

Contents

Notation 1

1. Introduction 2

I. Analysis of electrical circuits 5

2. Basics: A circuit model, differential-algebraic equations and graphs 6
- 2.1. A lumped circuit model: Modified nodal analysis . . . 6
  - 2.1.1. The element equations . . . 7
  - 2.1.2. Modified nodal analysis: linear and nonlinear circuit model . . . 8
- 2.2. Differential-algebraic equations (DAEs) . . . 12
- 2.3. Graphs and matrices . . . 15
  - 2.3.1. Incidence and Laplacian matrices . . . 16
  - 2.3.2. The r-distance in graphs . . . 18
  - 2.3.3. Edge contraction and corresponding algebraic operation . . . 19
- 2.4. Conclusion . . . 22

3. Structural analysis of electrical circuits 23
- 3.1. Preliminaries from linear algebra . . . 24
- 3.2. Linear decoupling . . . 27
- 3.3. Preliminaries for nonlinear circuits . . . 38
- 3.4. Nonlinear decoupling . . . 43
- 3.5. Transient perturbation analysis . . . 54
- 3.6. Conclusion . . . 69

II. Waveform relaxation 70

4. Convergence of waveform relaxation on coupled DAEs and coupled circuits 71
- 4.1. WR on coupled DAEs . . . 72
  - 4.1.1. The WR algorithm . . . 73
  - 4.1.2. Convergence . . . 75
- 4.2. WR on coupled circuits . . . 80
  - 4.2.1. Partitioned model and WR algorithm . . . 80

4.2.2. Convergence . . . 83
4.3. WR on coupled field/circuit models . . . 92
4.3.1. Coupled field/circuit model and WR algorithm . . . 92
4.3.2. Analysis of the space-discrete field DAE . . . 94
4.3.3. Convergence . . . 96
4.4. Numerical tests . . . 98
4.4.1. Circuit/Circuit . . . 98
4.4.2. Field/Circuit . . . 99
4.5. Conclusion . . . 100

5. Summary and Outlook **103**

Notation

Abbreviations

MNA	modified nodal analysis
ODE	ordinary differential equation
DAE	differential-algebraic equation
WR	waveform relaxation
IV	initial value
IVP	initial value problem
GS	Gauss-Seidel
KCL	Kirchhoff current laws
KVL	Kirchhoff voltage laws
C	capacitance
G	conductance
R	resistance
L	inductance
V	voltage source
I	current source

symbols

$\mathcal{A}$	incidence matrix
$\mathcal{L}$	Laplacian matrix or generalized Laplacian matrix
I	identity matrix
$A^\top$	transposed of A
$\lvert \cdot \rvert$	vector norm in $\mathbb{R}^n$
$\lvert \cdot \rvert_*$	induced matrix norm
$\lVert \cdot \rVert$	maximum norm on continuous function space
x^k	superscript k denotes iteration counter
$\dot{f}$	used only in Chapter 4: time derivative of f
f'	derivative of f
$G = (V, E)$	graph
V	set of nodes
E	set of edges
$\lvert V \rvert$	cardinality of finite set V

1. Introduction

Over the past decades until today, the research in the field of circuit simulation is regularly confronted with new challenges arising from the symbiotic relationship with industrial development of electronic devices, which relies on increasingly complex electrical circuits. The evolution of differential-algebraic equations (DAEs) and their analysis in mathematics was significantly stimulated by this relationship. DAEs can be viewed as dynamical systems involving algebraic constraints. Compared to ordinary differential equations, the presence of constraints in DAEs bears certain particularities which usually pose additional difficulties for the numerical treatment, even more when they are only given implicitly. In electrical circuit models, the Kirchhoff laws are an instance for algebraic constraints.

In the 1980s, the notion of an *index* of DAEs was of growing importance. This concept is not easy to grasp since there exist many definitions of it, designed to suit specific classes of DAEs in various fields of applications. However, all the index definitions attempt to measure the level of difficulty imposed by the algebraic constraints, which is closely related to the number of differentiations of an input function required to solve the DAE. This classification of DAEs gave rise to the quest for certain *DAE normal forms*, where the dynamic parts of the DAE and the diffferent algebraic constraints are separated and can be identified directly. As with the index, different mathematical communities proposed different normal forms. One important index concept, along with a related normal form, is the *tractability index* [GM86, LMT13]. For circuit simulation, usually based on models like differential-algebraic models as for instance modified nodal analysis (MNA), the index of a circuit's describing DAE is one key for an adequate choice of solvers. However, the tractability index, like others, is generally hard to compute for DAEs of complex circuits. Therefore, a result from Estevez Schwarz and Tischendorf [ET00], which presented criteria by which the index of a circuit DAE can be determined only by means of the circuit's topology, was widely noticed. Furthermore, they found that the index of MNA DAEs (that is, circuit DAEs arising from MNA) can not exceed two if only standard elements are considered. Another important question was answered on the basis of the tractability index by [Tis04], where it was shown that MNA DAEs of index one and two are locally uniquely solvable. This result was extended to arbitrary bounded time intervals for MNA DAEs of index one [Mat12] and of higher index [Jan15]. The latter also presented a new DAE index concept, the *dissection index*, which is strongly inspired by the tractability index.

In the first part of the thesis, we exploit the principal idea of the dissection index to propose a detailed normal form for MNA DAEs. We propose assumptions on the MNA DAEs which guarantee that the corresponding normal form right hand side is locally Lipschitz continuous. It is shown that the normal form implies an alternative way of proof for the global existence and uniqueness result in [Jan15] for MNA DAEs of index two. Building on the normal form, we then study systematically the transient perturbations caused by perturbations of a circuit's independent sources, providing error bounds for the solution variables in dependence of the magnitude of a perturbation function. Like the perturbation index [HLR89], we classify the type of bound by means of the highest involved derivative of the perturbation, so that we have index-zero, index-one and index-two type bounds. However, in contrast to the perturbation index which considers a broader class of perturbations, we only admit certain structured perturbations which can interpreted as source perturbations.

We find that the sensitivity of a circuit to a source term is determined by the source's topological position in the circuit. Refinements of the results are presented adressing the effect of a current source perturbation on specifically the voltage over that current source. One part of the results is in accordance with the topological index results in [ET00] and could therefore be somewhat anticipated, another part seems to reveal new insight.

The second part of the thesis is devoted to waveform relaxation (WR), which is an iterative method for the numerical solution of dynamical systems. It is particularly well-established in circuit simulation, and additionally, it is naturally suitable to treat coupled and multiphysical systems. WR is known to converge if applied to ordinary differential equations if the vector field satisfies a Lipschitz condition [Bur95]. However, on DAEs, convergence can not be guaranteed unless the DAE satisfies and additional contractivity condition. Since the seminal works of Lelarasmee, Ruehli and Sangiovanni-Vincentelli [Lel82, LRS82], this fact was adressed by a great number of articles which formulated WR convergence criteria for different classes of DAEs, cf. [WOSR85, MN87, Sch91, CI94, JK96, Mie00, JW01, AG01, SZF06, Ebe08, SGB10, Schö11, WV12, BBGS13] to name only some of them.

This work offers a WR convergence criterion for a novel class of nonlinear implicit DAEs of higher index. It is worth noticing, however, that the criteria therein are difficult to check in general. This is a common problem shared with the proposed criteria of the just cited articles. Roughly speaking, the criteria are usually either formulated directly for DAEs in certain normal forms, or, as in our case, a normal form of the DAE seems indispensable to check if the required criteria are satisfied.

For that reason, in the case of coupled circuit DAEs, we exploit the perturbation analysis of the first part of the thesis to express the WR convergence criteria in topological terms, reformulating them as topological conditions on the position of certain coupling elements. These topological conditions are easier to check than the original abstract criteria. Notably, we can view the problems in each step of the iterative WR as circuits with perturbed inputs, where the input perturbation changes in each iteration step. This allows to apply the circuit topological results from the perturbation analysis to WR convergence analysis. The systematically topological approach to the WR convergence analysis and the resulting topological WR convergence criteria for coupled circuits are, to the knowledge of the author, a new contribution to the WR field. While it should be mentioned that topological coupling conditions which guarantee fast WR convergence were formulated in [Lel82, WOSR85, BBGS13], these results rely on considerably more restrictive coupling conditions and did not aim for a systematic topological study.

Apart from circuit/circuit couplings, topological convergence criteria are also provided for a coupled field/circuit model, where the electromagnetic field model is a DAE after space discretization.

This work is organized as follows:

Chapter 2 establishes the basics for what follows. Notably, it derives the MNA circuit model in Section 2.1, briefly introduces DAEs in Section 2.2 and provides selected results from graph and matrix theory in Section 2.3.

The analysis of electrical circuits is contained in Chapter 3. Section 3.1 is mainly concerned with certain transformation matrices which are used in Section 3.2 to derive a DAE normal form for MNA DAEs in the linear case. This linear result can be seen as a preparation for the nonlinear case, since the main principles of the linear case remain valid in the nonlinear case. Of course, considering nonlinear circuits also poses some new problems. Section 3.3 provides the tools to tackle them.

The main results of the chapter are then formulated in Sections 3.4 and 3.5. First, a detailed DAE normal form for nonlinear MNA DAEs of index two is offered in Section 3.4. On the basis of the normal form, the results of the transient perturbation analysis are presented in Section 3.5.

In chapter 4, we turn our focus to a convergence analysis of WR on coupled DAEs and coupled circuits in particular. The main result of Section 4.1 is a WR convergence theorem for coupled nonlinear higher index DAEs of a general form. Section 4.2 offers a WR convergence theorem for coupled circuits with topologically formulated convergence criteria. The result is illustrated by prototypical examples. Section 4.3 formulates topological WR convergence criteria for a coupled field/circuit model. The theoretical convergence results are tested and confirmed by numerical simulations of toy examples in Section 4.4

Finally, we briefly summarize the results and formulate open questions in Chapter 5.

Part I.

Analysis of electrical circuits

2. Basics: A circuit model, differential-algebraic equations and graphs

This thesis is devoted to the analysis of the modified nodal analyis (MNA), which is a widely used electrical circuit model. Two important characteristics set the frame for this work:

- The MNA model is a differential-algebraic equation (DAE).
- Electrical circuits can be naturally represented by graphs.

After setting up the MNA model in Section 2, we therefore collect some mainly well-known basics of DAEs (Section 2.2) and graph theory (Section 2.3). Trying to be concise, we stick only to the results which shall prove to be useful in later chapters. For more detail and a comprehensive introduction, we refer to the textbooks [CDK87, Rec89] for circuit theory, [HLR89, Ria08, LMT13] for DAEs and [GR01, Bap10, Mol12, Bol13] for graph theory.

2.1. A lumped circuit model: Modified nodal analysis

The basic mathematical system of equations investigated in this thesis describes a lumped (electrical) circuit model. This system recurs prominently in each chapter, being examined from different angles.

Lumped circuit models are based on the assumption of instantaneous electromagnetic wave propagation and the absence of interference between the elements. This assumption implies that the *current through* and the *voltage across* an element are well-defined and independent of space. Therefore, the resulting model is a system of (differential) equations only in time and not in space, which greatly reduces the mathematical complexity of the problem at hand. From an industrial point of view, this is crucial since it allows for lower computing time and costs.

However, a suitable model for industrial circuit simulation has to reconcile two contrary ambitions: On the one hand, a low computing time. On the other hand, the model should of course reflect the physical behaviour of the circuit adequatly. Whenever neglecting the effects of wave propagation is not reasonable and leads to inaccuracies of unacceptable magnitude, spatially distributed models resulting in partial differential equations should be combined with lumped circuit models, cf. [Tis04, Schö11, Bau12, Cor20, CPST20]. The results of [CPST20] are presented in Sections 4.3 and 4.4.

Either way, decades of successful implementation in industrial circuit simulation are a strong evidence advocating the use of lumped circuit models.

In a lumped circuit model, the physical elements are replaced by one or several interconnected lumped (idealized) elements which describe the essential behavior of this element and neglect other effects. Combining the model equations of the lumped elements with the Kirchhoff laws enables us to set up a reasonable system of model equations which describes the behavior of the whole network. The unknowns in these systems are the current through and the voltages across some or all elements of the circuit. They are determined by a differential-algebraic equation, which is, loosely speaking, a differential equation with algebraic constraints. One difficulty of these systems is that the constraints are usually given implicitly. Depending on which of the Kirchhoff laws are used and how they are combined with the element equations, we can obtain different circuit models, cf. [CDK87]. Here, we introduce the well-established MNA. For a detailed and thorough introduction and derivation of the model, we recommend [Rei14].

We refer to the contact points or terminals of an element as its *nodes*, and our main focus is on elements with two nodes, also known as two-terminals. Furthermore, leaving aside controlled sources and more complex mixed elements, we restrict ourselves to circuits consisting of nonlinear resistors, inductors, capacitors and independent current and voltage sources. However, we shall provide brief remarks and references subsequent to the major theorems on whether or not the required assumptions hold for circuits with generalized elements.

2.1.1. The element equations

In the following, we present how the elements of a nonlinear circuit are modeled in a lumped circuit and we introduce the Kirchhoff laws.

A nonlinear *resistor* is modeled by the equation

$$v(t) = r(i(t)) \qquad \text{or} \qquad i(t) = g(v(t)), \tag{2.1}$$

where $i(t)$ and $v(t)$ are scalars representing the current flow through the element and the voltage between the two nodes of the element, and r and g are nonlinear functions describing the resistance and the conductance of the resistor.

A *voltage source* and a *current source* provide a voltage v_s and a current i_s,

$$v(t) = v_s(t), \qquad i(t) = i_s(t). \tag{2.2}$$

The functions v_s and i_s can be nonlinear.

A *capacitor* stores the electric charge q. The lumped element is described by

$$q(t) = q_c(v(t)), \tag{2.3}$$

where q_c is a nonlinear function. Considering that it holds $i(t) = q'(t)$ at all times t, we obtain

$$i(t) = \frac{\mathrm{d}}{\mathrm{d}t} q_c(v(t)). \tag{2.4}$$

Inductances are lumped models of coils. They store energy in their magnetic fields and they are described by

$$\phi(t) = \phi_l(i(t)), \tag{2.5}$$

whith ϕ the magnetic flux. At all times t, it holds $v(t) = \phi'(t)$ and accordingly

$$v(t) = \frac{\mathrm{d}}{\mathrm{d}t} \phi_l(i(t)). \tag{2.6}$$

2.1.2. Modified nodal analysis: linear and nonlinear circuit model

Before presenting the circuit model, we briefly introduce some basics. First, we introduce the *Kirchhoff laws*, which are fundamental for circuit theory as they define the laws which must be respected when interconnecting elements.

After that, *reduced incidence matrices* are defined. They allow for an elegant and compact representation of the circuit equations.

Kirchhoff laws They consist of the *Kirchhoff current law* (KCL) and the *Kirchhoff voltage law* (KVL). The KCL states that the sum of all currents entering any node of an arbitrary lumped circuit is equal to the sum of all currents leaving the node. The KVL states that the sum of voltages in any loop of an arbitrary circuit is zero. A *loop* is a subcircuit with at least two elements where each node connects precisely two elements.

Incidence matrix Two elements are called *incident*, if they are interconnected, that is, they share a common node. A node and an element are called incident, if the node is a terminal of the element.

We associate an arbitrary orientation with each element, see Figure 2.1. That way, we set a flow direction, that is, a sign for the current flow through each element. Furthermore, we number the elements and the nodes of the network arbitrarily. For a network with n nodes and b elements, the incidence matrix $\bar{\mathcal{A}} \in \mathbb{R}^{n \times b}$ is defined by

$$\bar{a}_{ij} = \begin{cases} +1 & \text{if element } b_j \text{ leaves node } n_i, \\ -1 & \text{if element } b_j \text{ enters node } n_i, \\ 0 & \text{if } b_j \text{ and } n_i \text{ are not incident.} \end{cases} \tag{2.7}$$

For convenience, we sort the elements by type, that is, $\bar{\mathcal{A}} = \begin{pmatrix} \bar{\mathcal{A}}_c & \bar{\mathcal{A}}_r & \bar{\mathcal{A}}_l & \bar{\mathcal{A}}_v & \bar{\mathcal{A}}_i \end{pmatrix}$, where the subscripts indicate the element type.

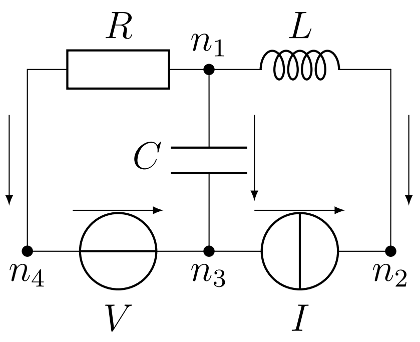


Figure 2.1.: A circuit with the nodes n_1, n_2, n_3, n_4. An arbitrary orientation is assigned to each element.

Example 2.1.1 *For the circuit in Figure 2.1, we obtain the incidence matrix*

$$\bar{\mathcal{A}} = \begin{pmatrix} \bar{\mathcal{A}}_c & \bar{\mathcal{A}}_r & \bar{\mathcal{A}}_l & \bar{\mathcal{A}}_v & \bar{\mathcal{A}}_i \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 \\ -1 & 0 & 0 & -1 & 1 \\ 0 & -1 & 0 & 1 & 0 \end{pmatrix}.$$

Remark. In the example, the incidence matrices $\bar{\mathcal{A}}_\star$ of the respective elements are column vectors. Not that this is due to the minimal nature of the example. Usually, a circuit has several elements of each type, and accordingly the matrices $\bar{\mathcal{A}}_\star$ have more than one column.

Reference node and reduced incidence matrix Incidence matrices do not have full row rank n, but if the corresponding network is connected, any choice of $n-1$ rows is linearly independent, see Chapter 2.3. This is the mathematical motivation for defining a reference node in the circuit. The corresponding incidence matrix row is deleted, resulting in a *reduced incidence matrix* $\mathcal{A}$. That way, the KCL of the reference node is not taken into account; its node potential is set to zero. The reference node is often chosen such that it corresponds to the mass node of the physical system.

Example 2.1.2 *Choosing n_4 as the reference node, the reduced incidence matrix of the circuit in Figure 2.1 becomes*

$$\mathcal{A} = \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_r & \mathcal{A}_l & \mathcal{A}_v & \mathcal{A}_i \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 \\ -1 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

The (reduced) incidence matrix allows for a compact notation of the system of nodal equations, referred to as the KCL, in a circuit:

$$\mathcal{A}i = \mathcal{A}_c i_c + \mathcal{A}_r i_r + \mathcal{A}_l i_l + \mathcal{A}_v i_v + \mathcal{A}_i i_i = 0, \qquad i = \begin{pmatrix} i_c \\ i_r \\ i_l \\ i_v \\ i_i \end{pmatrix} \tag{2.8}$$

where $i_\star$ is the vector of currents through the respective elements.

Substituting potential difference for voltage Furthermore, the KVL implies that we can replace the unknown voltage v over an element by the *potential difference* $e_i - e_j$ of its two terminal nodes n_i, n_j, that is, $v = e_i - e_j$. Since usually the number of elements in a circuit is considerably higher than the number of nodes, replacing the voltages by (differences of) node potentials reduces the number of unknowns. By means of the incidence matrix, we can denote

$$\mathcal{A}_c^\top e = v_c, \qquad \mathcal{A}_r^\top e = v_r, \qquad \mathcal{A}_l^\top e = v_l, \qquad \mathcal{A}_v^\top e = v_v, \qquad \mathcal{A}_i^\top e = v_i \tag{2.9}$$

where e the vector of node potentials (except the potential at the reference node).

Network equations If we combine the KCL (2.8) and the element equations (2.1)-(2.6) and substitute the node potentials for the voltages by using equation (2.9), we obtain

$$\begin{aligned} \mathcal{A}_c i_c + \mathcal{A}_r i_r + \mathcal{A}_l i_l + \mathcal{A}_v i_v + \mathcal{A}_i i_i &= 0, \\ i_r &= g(\mathcal{A}_r^\top e), \\ i_c &= \frac{\mathrm{d}}{\mathrm{d}t} q_c(\mathcal{A}_c^\top e), \\ \mathcal{A}_l^\top e &= \frac{\mathrm{d}}{\mathrm{d}t} \phi_l(i_l), \\ \mathcal{A}_v^\top e &= v_s, \\ i_i &= i_s. \end{aligned}$$

Note that here, the unknowns and functions are vector-valued in contrast to the single element equations (2.1)-(2.6).

Inserting i_r, i_c, i_s yields the *nonlinear MNA*

$$\mathcal{A}_c \frac{\mathrm{d}}{\mathrm{d}t} q_c(\mathcal{A}_c^\top e) + \mathcal{A}_r g(\mathcal{A}_r^\top e) + \mathcal{A}_l i_l(t) + \mathcal{A}_v i_v + \mathcal{A}_i i_s = 0, \tag{2.10a}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \phi_l(i_l) - \mathcal{A}_l^\top e = 0, \tag{2.10b}$$

$$\mathcal{A}_v^\top e - v_s = 0. \tag{2.10c}$$

If the relevant functions are differentiable, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} q_c(\mathcal{A}_c^\top e) = C(\mathcal{A}_c^\top e)\mathcal{A}_c^\top e', \quad \frac{\mathrm{d}}{\mathrm{d}t}\phi_l(i_l) = L(i_l)i_l',$$

with C, L the Jacobians of q_c, ϕ_l. Note that these Jacobian matrices are naturally diagonal by definition of q_c and ϕ_l. The given definitions are adequate to describe two-terminal elements; we remark, however, that the case of multiport elements gives rise to non-diagonal Jacobians in general.

Insertion into (2.10a)-(2.10c) leads to an equivalent formulation of the *nonlinear MNA*:

$$\mathcal{A}_c C(\mathcal{A}_c^\top e)\mathcal{A}_c^\top e' + \mathcal{A}_r g(\mathcal{A}_r^\top e) + \mathcal{A}_l i_l + \mathcal{A}_v i_v + \mathcal{A}_i i_s = 0, \tag{2.11a}$$

$$L(i_l)i_l' - \mathcal{A}_l^\top e = 0, \tag{2.11b}$$

$$\mathcal{A}_v^\top e - v_s = 0. \tag{2.11c}$$

An electrical circuit is called *linear*, if the current-voltage relations for all resistors, capacitors, and inductors are linear. That is, they are described by

$$i_r = Gv_r, \quad i_c = Cv_c', \quad v_l = Li_l',$$

where G, C, L are diagonal matrices with diagonal entries greater zero. The source functions v_s, i_s may still be nonlinear. Proceeding analoguously to the nonlinear case yields the *linear MNA*

$$\mathcal{A}_c C\mathcal{A}_c^\top e' + \mathcal{A}_r G\mathcal{A}_r^\top e + \mathcal{A}_l i_l + \mathcal{A}_v i_v + \mathcal{A}_i i_s = 0, \tag{2.12a}$$

$$Li_l' - \mathcal{A}_l^\top e = 0, \tag{2.12b}$$

$$\mathcal{A}_v^\top e - v_s = 0. \tag{2.12c}$$

Each diagonal entry of G, C and L correspond to a one element and is called its *conductance*, *capacitance* and *inductance*, respectively. The reciprocals of conductance and capacitance are called *resistance* and *elastance*.

Assumption 2.1.3 (MNA) *The linear MNA* (2.12) *and the nonlinear MNA* (2.11) *meet the following properties:*

(i) $\mathcal{A}_v$ *has full column rank, and* $\begin{pmatrix}\mathcal{A}_c & \mathcal{A}_v & \mathcal{A}_r & \mathcal{A}_l\end{pmatrix}$ *has full row rank.*

(ii) The input functions i_s, v_s *are continuously differentiable.*

The Kirchhoff laws require (i), as the two conditions are equivalent to the topological conditions “the circuit contains no V-loop” and “the circuit contains no I-cutset”, as we shall see in Section 2.3.1. Together, the assumptions provide well-posedness of the MNA, where (ii) ensures sufficient smoothness of the input functions and (i) is necessary to guarantee existence and uniqueness of solutions.

Remark 2.1.4 *The basic MNA model introduced here leads to diagonal matrices $C(\cdot)$, $L(\cdot)$ and (if exists) $g'(\cdot)$. However, incorporation of for instance controlled sources or the MOSFET model may require a model which leads to non-diagonal incremental matrices $C(\cdot), L(\cdot)$ and $g'(\cdot)$ [ET00, Bod07]. In order to cover such generalizations, we shall* not *assume that $L(\cdot)$, $C(\cdot)$ and $g'(\cdot)$ are diagonal unless explicitly stated.*

2.2. Differential-algebraic equations (DAEs)

DAEs are differential equations. Like ODEs, they do not involve partial derivatives. But the term ODE is commonly used for either *explicit* ODEs or, at least, for (implicit) ODEs which can be equivalently transformed into an explicit ODE. DAEs, however, are differential equations which can *not* be equivalently transformed into an explicit ODE; instead, parts of the solution are given by possibly implicit algebraic constraints. Formally, a DAE reads

$$F : \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{I} \subset \mathbb{R}, \qquad (y, x, t) \mapsto F(y, x, t), \qquad F(x', x, t) = 0, \tag{2.13}$$

where F is continuous and continuously differentiable in y. Then, Equation (2.13) is a DAE, if the partial derivative $\frac{\partial F}{\partial y}(y, x, t)$ is singular for all $(y, x, t) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{I}$. If the defining singularity condition is violated on any subdomain, then by the implicit function theorem we can, at least locally, solve for x' and thereby transform the system into an explicit ODE.

This DAE definition is concise and easy to comprehend. The appearance of x' in Equation (2.13) seems to indicate that we seek solutions in C^1. However, this is not necessarily the case. If F depends in fact only on parts of x', then we require only these parts to be differentiable, see Example 2.2.1 for g not differentiable. This is not reflected in Equation (2.13) in order to keep notation compact. The following examples reveal essential features of DAEs.

Example 2.2.1 *For $f : \mathbb{R}^n \times \mathcal{I} \subset \mathbb{R} \to \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathcal{I} \to \mathbb{R}^m$ continuous, we consider*

$$y' = f(y, t), \tag{2.14a}$$
$$z = g(y, t). \tag{2.14b}$$

In this DAE the constraint g for the algebraically fixed variable z is explicitly given, and the differentiated variable y is determined by an ODE independently of z. If a DAE is given in this form, it is said to be *fully decoupled*.

The presence of algebraic constraints gives rise to some problems which are characteristic for DAEs:

(i) Not for all initial values $x_0 \in \mathbb{R}^n$ is a DAE solvable. A *consistent initial value* of a DAE is one for which there exists a solution.

(ii) Solvability of a DAE requires sufficient smoothness of inhomegeinities, also called input functions.

(iii) Solving a DAE (numerically) possibly requires not only integrations, but also differentiations.

In the example DAE 2.14, we can choose initial values for y. For each initial value $y_0 := y(t_0)$, there exists only one initial value $z_0 := z(t_0)$ for which the DAE has a solution, namely $z_0 = g(y_0, t_0)$.

Example 2.2.2 *For some input function r, we consider the linear DAE*

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} x' + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} x = r(t). \tag{2.15}$$

With a nonsingular matrix T and new coordinates $v := T^{-1}x$, we equivalently transform the system as follows:

$$T^\top ATv' + T^\top BTv = T^\top r(t).$$

We choose T so that

$$T := \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}, \qquad T^\top AT = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad T^\top BT = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Writing out the equations with $v = (v_1, v_2)^\top$ and $T = (T_1\ T_2)$, that means

$$v_1' + v_1 = T_1^\top r, \tag{2.16a}$$
$$v_1 + v_2 = T_2^\top r. \tag{2.16b}$$

We notice that the DAE (2.16) is of the decoupled form of the DAE (2.14) with

$$f(t, y) = -y + T_1^\top r, \qquad g(t, y) = -y + T_2^\top r.$$

Clearly, $v = T^{-1}x$ solves the decoupled DAE (2.16) if and only if x solves (2.15). In that case, we say we have *decoupled* the DAE (2.15), and Equation (2.16a) is its *inherent ODE.*

Example 2.2.3 (Index 2 DAE in normal form)

$$y' = f(y, z_1, z_2, z_3, t) \tag{2.17a}$$
$$z_1 = g_1(y, z_2, z_3, z_3', t) \tag{2.17b}$$
$$z_2 = g_2(y, z_3, t) \tag{2.17c}$$
$$z_3 = g_3(t). \tag{2.17d}$$

This DAE is given in a somewhat "half-decoupled" form, since the variables are given explicitly, but to fully decouple the equations, we still need to insert and thereby eliminate the variables z_i in the dynamic equation. In contrast to DAE (2.14), solving the system requires a differentiation of $z_3 = g_3(t)$.

The index of a DAE

In DAE literature, the so-called *index* of a DAE is ubiquitous. An adequate choice of a numerical solver for a DAE should take its index into account, the determination of which is therefore of high relevance.

There are many definitions of a DAE index. However, they all attempt to quantify the level of difficulty of a DAE for numerical analysis and simulation. Since a major problem for simulation is the ill-posedness of numerical differentiation, the index of a DAE is closely related to the number of differentiations needed in order to solve it, as can be seen well in the *perturbation index* [HLR89].

Definition 2.2.4 (Perturbation index) *If x is a solution of (2.13) on $[t_0, T]$ with initial value x_0, and x^δ is a solution of a perturbed system*

$$F((x^\delta)', x^\delta, t) = \delta$$

with initial value x_0^δ, then the DAE (2.13) has perturbation index k *along x if k is the smallest number such that there exists $c > 0$ and an estimate of the form*

$$\|x - x^\delta\| \le c\left(|x_0 - x_0^\delta| + \|\delta\| + \cdots + \|\delta^{(k-1)}\|\right)$$

for all (sufficiently smooth) perturbations δ. Here, $\delta^{(k)}$ denotes the k-th derivative, and $\|\cdot\|$ and $|\cdot|$ are the maximum norm and a vector norm.

Among the most important index concepts are the *Kronecker index* [Gan05], *tractability index* [GM86, LMT13] and the *differentiation index* [CG95, Gea88]. Except for the Kronecker index, which is from the nineteenth century, most index concepts were developped in the eighties or nineties of the twentieth century. The mentioned indices coincide on classes of DAEs which meet certain linearity requirements. Notably, it was shown in [Tis04] that the differentiation index, the tractability index and the perturbation index coincide on the class of MNA equations (2.11) independently of the topology of the described circuit at hand.

The different index concepts mainly differ in two aspects. First, whether or not they come with a method to obtain a certain normal form or DAE decoupling, which normally goes hand in hand with determining the index. Secondly, the indices apply to different ranges of (generalized) DAE classes. A good introduction to DAEs and survey on different index concepts, including some important ones not mentioned here, is [Meh12].

The perturbation index is very directly defined via the number of differentiations impacting the solution of a perturbed system. Considering that differentiations pose major challenges in numerical analysis, this index is specifically apt for numerical convergence analysis or perturbation analysis. However, it does not offer a DAE decoupling method.

2.3. Graphs and matrices

We have seen that electrical circuits can be modeled with a DAE for which (reduced) incidence matrices play an important role. Many graph theoretical results adress the close relation of incidence matrices and the corresponding network. In this section, we "cherry-pick" only those which are relevant in the subsequent chapters. For a comprehensive introduction, we refer to standard textbooks, e.g. [CDK87, GR01, Bap10, Bol13].

Basic graph theoretical definitions and terminology A *graph* $G = (V, E)$ is a non-empty set of points $V(G)$ called *nodes* along with a (possibly empty) set of line segments $E(G)$ called *edges* joining pairs of nodes. The number of nodes and edges in a graph is denoted by $|V(G)|$ and $|E(G)|$. If it is clear from the context which graph they belong to we write just V and E. Both the edges and the nodes are numbered from 1 to $|V|$ and from 1 to $|E|$, respectively, and are referred to by these numbers. We admit *multiple edges*, that is, more than one edge joining the same pair of nodes. However, edges joining a node with itself, called *selfloops*, are excluded. Our graphs are *weighted*, which means that each edge k is associated with a weight $w_k > 0$.

If two nodes are joined by an edge, they are *adjacent*. The edge is *incident* with the two nodes it joins and vice versa. Also, two edges are *incident* if they have an incident node in common. Given two nodes i, j and two edges k, l, we write $i \sim j$ if the nodes are adjacent, and $i \sim k$ and $k \sim l$ for the respective node to edge and edge to edge incidences. The *degree* $\deg(i)$ of a node i is the number of edges it is incident with. The *weighted degree* $\deg_w(i)$ of a node i is the sum of weights of its incident edges, that is, $\deg_w(i) = \sum_{k\sim i} w_k$. The *condensed weight* $\omega_{ij} = \sum_{k\sim i \wedge k\sim j} w(k)$ between adjacent nodes i and j is the sum of weights of edges which join the nodes.

A *subgraph* of $G = (V, E)$ is a graph $\tilde{G} = (\tilde{V}, \tilde{E})$ such that $\tilde{V} \subseteq V$ and $\tilde{E} \subseteq E$. We write $\tilde{G} \subseteq G$ and call G the *supergraph* of $\tilde{G}$ in this case. If $\tilde{G} = (V, \tilde{E})$, that is, $\tilde{G}$ contains all nodes of the supergraph, it is a *spanning subgraph*. A *path* is a graph such that precisely two nodes have degree one, all other nodes have degree two, and each edge is incident with at least one other edge. Two nodes of a graph G are connected, if there exists a path $P \subseteq G$ which contains them. A graph is *connected*, all of its pairs of nodes are connected. A *connected component* of G is a maximal connected subgraph, that is, a connected subgraph $C \subseteq G$ such that none of its nodes is adjacent to any node not belonging to C. A *loop* is a connected graph such that each node has degree 2. A *forest* is a graph which does not contain loops, and a *tree* is a connected forest. In a connected graph $G = (V, E)$, a *cutset* is a subset $E_{cut} \subseteq E$ such that $G - E_{cut}$ is not connected, and adding any one edge of E_{cut} to $G - E_{cut}$ yields a connected graph.

The edges are given an arbitrary but fixed *orientation*, that is, one of its incident nodes is designated its *start node* and the other one is designated the *end node*. This has no meaning apart from fixing a sign of flow, that is, a positive and a negative direction.

Remark. Due to the historical development of graph theory in different scientific communities, the terminology is not unanimous in the literature.

2.3.1. Incidence and Laplacian matrices

We introduced the incidence matrix of a circuit in Subsection 2.1.2, Equation (2.7). The definition can naturally be applied to graphs if we replace elements by edges. The following relations between incidence and Laplacian matrix on the one hand and the corresponding graph on the other hand are well-known in graph theory and circuit theory. They refer to the non-reduced incidence and Laplacian matrix if not stated otherwise. The bar in $\bar{\mathcal{A}}, \bar{\mathcal{L}}$ indicates this.

Lemma 2.3.1 *Let $G = (V, E)$ be a connected graph with incidence matrix $\bar{\mathcal{A}} \in \mathbb{R}^{|V| \times |E|}$. Then, any $|V| - 1$ rows of $\bar{\mathcal{A}}$ are linearly independent.*

Lemma 2.3.2 *Let $G = (V, E)$ be a graph with $|C|$ connected components and incidence matrix $\bar{\mathcal{A}} \in \mathbb{R}^{|V| \times |E|}$. Then, rank $\bar{\mathcal{A}} = |V| - |C|$.*

Lemma 2.3.3 *Let G be a graph with incidence matrix $\bar{\mathcal{A}}$. Then, G is a forest if and only if $\bar{\mathcal{A}}$ has full column rank.*

Remark. Considering the equality of row rank and column rank in a matrix and Lemma 2.3.1, it follows immediately that Lemma 2.3.3 holds for reduced incidence matrices as well.

Proofs can be found for instance in [CDK87, pp. 25-26] for Lemma 2.3.1, and [Bap10, Theorem 2.3] and [Bap10, Lemma 2.5] for Lemma 2.3.2 and 2.3.3.

Definition 2.3.4 (Laplacian matrix) *Let $G = (V, E)$ be a graph. The Laplacian matrix $\bar{\mathcal{L}}(G) \in \mathbb{R}^{|V| \times |V|}$ is defined by the node to node relations*

$$\bar{\mathcal{L}}_{ij} = \begin{cases} \deg_w(i), & \textit{if } i = j, \\ 0, & \textit{if } i \neq j \textit{ and the nodes } i \textit{ and } j \textit{ are not adjacent}, \\ -\omega_{ij}, & \textit{if } i \neq j \textit{ and the nodes } i \textit{ and } j \textit{ are adjacent}, \end{cases} \tag{2.18}$$

where $\deg_w(i)$ is the weighted degree and ω_{ij} is the condensed weight between adjacent nodes i and j.

Remark. By construction, the Laplacian matrix is symmetric and all row sums (and column sums) are zero.

The Laplacian matrix appears also by the names graph Laplacian, Kirchhoff matrix or admittance matrix in the literature. Note that the Laplacian matrix fully describes the topology of the graph except for the fact that it makes no difference between multiple and simple edges. More precisely, replacing a simple edge of weight w by multiple edges whose weight sum amounts to w results in the same Laplacian matrix.

Incidence and Laplacian matrices are closely related as we see in the next lemma, cf [Mol12, pp. 95-96].

Lemma 2.3.5 (Laplacian and incidence matrix) *Let $G = (V, E)$ be a graph with incidence matrix $\bar{\mathcal{A}}$ and Laplacian matrix $\bar{\mathcal{L}}$. Let furthermore $W \in \mathbb{R}^{|E|\times|E|}$ be a diagonal matrix which stores the edge weights $w_1, \dots, w_{|E|}$. It holds $\bar{\mathcal{A}} W \bar{\mathcal{A}}^\top = \bar{\mathcal{L}}$.*

Remark. It follows immediately that the relation also holds for reduced matrices, that is,

$$\mathcal{A} W \mathcal{A}^\top = \mathcal{L} \in \mathbb{R}^{(|V|-1)\times(|V|-1)},$$

where $\mathcal{A}$ is the reduced incidence matrix (for a given reference node) and the $\mathcal{L}$ is the *reduced Laplacian matrix*, that is, the principal submatrix of $\bar{\mathcal{L}}$ which arises from deletion of the row and column corresponding to the reference node.

Remark. Revisiting the linear MNA Equations (2.12), we can replace the terms $\mathcal{A}_c C \mathcal{A}_c^\top$ and $\mathcal{A}_r G \mathcal{A}_r^\top$ by $\mathcal{L}_c$ and $\mathcal{L}_r$, where the latter are reduced Laplacian matrices of the respective spanning subgraphs. The capacitances and conductances of the elements then take the role of edge weights in the circuit representing graph.

Definition 2.3.6 (Generalized inverses) *Let an arbitrary matrix $A \in \mathbb{R}^{n\times n}$ be given. Then, A^g is called a* g-inverse *of A, if*

$$AA^gA = A.$$

If A^g moreover satisfies $A^gAA^g = A^g$, then it is called a reflexive g-inverse.

Furthermore, $A^\dagger$ is called Moore-Penrose inverse *of A, if it is a reflexive g-inverse and satisfies additionally*

$$A^\dagger A = (A^\dagger A)^\top, \quad AA^\dagger = (AA^\dagger)^\top.$$

Remark. The Moore-Penrose-inverse, also called $\{1, 2, 3, 4\}$-inverse, exists and is unique for any real square matrix, cf [Pen55].
The following result is mentioned without a proof in [Bap10, p.133] as "easy-to-check".

Lemma 2.3.7 *Let G be a connected graph with fixed reference node r, Laplacian matrix $\bar{\mathcal{L}}$ and reduced Laplacian $\mathcal{L}$. Furthermore, let $\bar{\mathcal{L}}^g$ be the matrix which is obtained from $\mathcal{L}^{-1}$ when a zero row and column is "squeezed into" the r-th row and column, that is, the r-th row and column are zero and the former $(r+k)$-th row and column are now row and column $r+k+1$ for $k \geq 0$. Then, $\bar{\mathcal{L}}^g$ is a g-inverse of $\bar{\mathcal{L}}$.*

Proof: Denoting the number of nodes of G by n, we assume for notational simplicity that the reference node is the n-th node so that

$$\bar{\mathcal{L}}^g = \begin{pmatrix} \mathcal{L}^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Lemma 2.3.1, 2.3.2 and 2.3.5 and the connectedness of G assure the existence of $\mathcal{L}^{-1}$. Denoting $(\bar{\mathcal{L}})_{ij} = \bar{l}_{ij}$ and $\bar{x}_n^\top := \begin{pmatrix} \bar{l}_{n1} & \cdots & \bar{l}_{n(n-1)} \end{pmatrix}$, it follows

$$\bar{\mathcal{L}}\bar{\mathcal{L}}^g\bar{\mathcal{L}} = \begin{pmatrix} \mathcal{L} & \bar{x}_n \\ \bar{x}_n^\top & \bar{l}_{nn} \end{pmatrix} \begin{pmatrix} \mathcal{L}^{-1} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathcal{L} & \bar{x}_n \\ \bar{x}_n^\top & \bar{l}_{nn} \end{pmatrix} = \begin{pmatrix} \mathcal{L} & \bar{x}_n \\ \bar{x}_n^\top & \bar{x}_n^\top \mathcal{L}^{-1}\bar{x}_n \end{pmatrix}$$

It remains to show that $\bar{x}_n^\top \mathcal{L}^{-1}\bar{x}_n = \bar{l}_{nn}$. We note that due to the zero row sum property, the vector of all ones, which we denote by $\mathbb{1}$, lies in the kernel of $\bar{\mathcal{L}}$. It follows that $\mathbb{1} \in \ker \bar{\mathcal{L}}\bar{\mathcal{L}}^g\bar{\mathcal{L}}$, which implies that all row sums are zero and notably $\bar{x}_n^\top \mathcal{L}^{-1}\bar{x}_n = \bar{l}_{nn}$. □

Lemma 2.3.8 *Let G be a graph with Laplacian matrix $\bar{\mathcal{L}}$, and let A and $\bar{A}$ be two g-inverses of $\bar{\mathcal{L}}$ with coefficients a_{ij} and $\bar{a}_{ij}$. It holds*

$$a_{ii} - a_{ij} - a_{ji} + a_{jj} = \bar{a}_{ii} - \bar{a}_{ij} - \bar{a}_{ji} + \bar{a}_{jj} \qquad \forall i, j.$$

The proof can be found in [Bap10, Lemma 9.10].

2.3.2. The r-distance in graphs

While *effective resistance* between two nodes in a resistive network is known and used in electrical circuit theory for much longer, [KR93] in a classical article formulated a mathematical framework and identified the effective resistance as a distance measure in graphs when the conductances are understood as edge weights. They named this newly identified graph distance *resistance distance*. It has proved to be the more adequate distance concept than the standard shortest path distance in many application areas. Among them are molecular graphs and the Kirchhoff index in chemistry [KR93, BBLK94, GM96, PL18], control theory [BH06], communication theory [TYN18, PYZ18], random walks [Che18, YK19] and power networks [DB10, DB12].

Remark. The common names *resistance distance* and *effective resistance* for the subsequently defined distance could be misleading in our case since we shall use it in capacitive and inductive subgraphs. Therefore, we decided to use the neutral name "r-distance" instead.

Definition 2.3.9 (The r-distance in a graph ("resistance distance")) *Let G be a connected graph with (weighted) Laplacian matrix $\bar{\mathcal{L}}$. The* r-distance *between two nodes i, j of a graph is defined by*

$$d_r^G(i,j) = (\hat{e}_i - \hat{e}_j)^\top \bar{\mathcal{L}}^\dagger (\hat{e}_i - \hat{e}_j) = \bar{\mathcal{L}}_{ii}^\dagger + \bar{\mathcal{L}}_{jj}^\dagger - 2\bar{\mathcal{L}}_{ij}^\dagger.$$

where $\hat{e}_i$ is the i-th unit vector and $\bar{\mathcal{L}}^\dagger$ is the Moore-Penrose inverse of $\bar{\mathcal{L}}$.

Due to Lemma 2.3.8, r-distance can be defined equivalently with any other g-inverse.

Corollary 2.3.10 *Let G be a connected graph with fixed reference node and Laplacian matrix $\bar{\mathcal{L}}$, and let $\bar{\mathcal{L}}^g$ be a generalized inverse of $\bar{\mathcal{L}}$. Then, it holds*

$$d_r^G(i,j) = \bar{\mathcal{L}}_{ii}^g + \bar{\mathcal{L}}_{jj}^g - \bar{\mathcal{L}}_{ij}^g - \bar{\mathcal{L}}_{ji}^g.$$

If node j is the reference node, then $d_r^G(i,j) = \bar{\mathcal{L}}_{ii}^g$.

Definition 2.3.11 (Distances in graphs) *Let $G = (V, E)$ be a graph and $i, j, k \in V$ arbitrary nodes. A function $d^G : V \times V \to \mathbb{R}$ is a* distance function *on G, if it satisfies the following three axioms.*

1. $d^G(i,j) \geq 0$ *and* $d^G(i,j) = 0 \iff i = j$
2. $d^G(i,j) = d^G(j,i)$
3. $d^G(i,k) + d^G(k,j) \geq d^G(i,j)$.

Lemma 2.3.12 ("resistance is distance", [KR93]) *The r-distance $d_r^G(i,j)$ defines a distance function in a graph.*

In a tree, or more precisely, if the path between two nodes is unique, the r-distance coincides with the standard shortest-path distance. One of the useful properties of the effective resistance, however, is that it that takes into account all paths between two nodes, not just the shortest path; the existence of alternative paths strictly decreases the distance. This is expressed by the following two lemmata, cf. [KR93].

Lemma 2.3.13 *Let $G = (V, E)$ be a graph and $i,\ j\ \in V$. The r-distance $d_r^G(i,j)$ is a nonincreasing function of the edge weights. This function is constant only for those edges not lying on any path between i and j.*

Lemma 2.3.14 *Let $G = (V, E)$ and $G_+ = (V, E_+)$ with $E \subseteq E^+$ be two graphs, and let $i,\ j\ \in V$. Then, $d_r^{G_+}(i,j) \leq d_r^G(i,j)$. Equality holds if and only if each edge lying on a path between i and j is element of E.*

2.3.3. Edge contraction and corresponding algebraic operation

In this section, after shortly introducing the well-known edge contraction on graphs, we present corresponding algebraic processes for the incidence and the Laplacian matrix.

Edge contraction Given an edge i of a graph $G = (E, V)$, the graph G/i is obtained from G by contracting the edge i; that is, G/i arises from G when the incident nodes of edge i are identified and all resulting self-loops are removed.

If $H \subseteq G$ is a subgraph, we write G/H for the graph obtained by contracting each connected component of H to one node, that is, successively contracting each edge of H. The contracted graph G/H is well-defined since it is independent of the order of edge contractions.

More details can be found in most graph theoretical textbooks, e.g. [GR01, Bol13].

Remark. Note that the bigger part of graph theoretical literature defines the contraction process so that arising parallel edges are removed, while we allow for parallel edges in contracted graphs. This is due the definition of a "graph" as a simple graph in the bigger part of the literature. However, in a circuit context it is natural to consider parallel edges in a (contracted) graph.

Corresponding algebraic operation The following matrix shall allow us to define an operation on the incidence and Laplacian matrix which can be seen as algebraic counterpart of graph contraction.

Definition 2.3.15 ($(0, 1)$ node component matrix) *Let $G = (V, E)$ be a graph with connected components $C^1, \dots, C^{|C|}$. We define the* $(0, 1)$ node component matrix $\bar{Q} \in \mathbb{R}^{|V| \times |C|}$ *of G by*

$$(\bar{Q})_{ij} = \begin{cases} 1, & \text{if node } i \in C^j, \\ 0, & \text{else.} \end{cases}$$

Remark. Note that isolated nodes count for connected components, too.

Remark. Similarly to incidence and Laplacian matrices, we define for a given reference node the *reduced* $(0, 1)$ *node component matrix* $Q \in \mathbb{R}^{(|V|-1)\times(|C|-1)}$ so that it arises from $\bar{Q}$ if the row corresponding to the reference node and the column corresponding to the connected component which contains the reference node are deleted.

Remark. Note that for a given graph the columns of $\bar{Q}$ form a basis of $\ker \bar{\mathcal{A}}^\top$.

Lemma 2.3.16 (Contraction of a transposed incidence matrix) *Let $G = (V, E)$ be a graph with spanning subgraphs $G_x = (V, E_x)$, $G_{x_t} = (V, E_{x_t})$ and $G_y = (V, E_y)$, where E_x, E_y is a partition of E and $E_{x_t} \subseteq E_x$ is the subset of X-edges whose start and end node belong to distinct connected components of G_y. Let furthermore $\bar{Q}_y \in \mathbb{R}^{|V|\times|C_y|}$ be the $(0, 1)$ node component matrix of G_y as in Definition 2.3.15. It holds:*

1. *The incidence matrix of a contracted graph is given by*

$$\bar{\mathcal{A}}(G/G_y) = \bar{\mathcal{A}}((G_y \cup G_{x_t})/G_y) = \bar{Q}_y^\top \bar{\mathcal{A}}(G_{x_t}) \in \mathbb{R}^{|C_y|\times|E_{x_t}|}.$$

2. *Denoting by $a^i_{x/y}$ and $\tilde{a}^i_{x/y}$ the columns of $\bar{\mathcal{A}}(G/G_y)$ and $\bar{Q}_y^\top \bar{\mathcal{A}}(G_x)$ which correspond to edge i, it holds $a^i_{x/y} = \tilde{a}^i_{x/y}$ if $i \in E_{x_t}$, and $\tilde{a}^i_{x/y} = 0$ if $i \in E_x \setminus E_{x_t}$.*

Remark. By definition, the edge set E_{x_t} consists of all edges which are part of an X-cutset in G; here, X-cutset means a cutset being a subset of E_x.

Proof: 1. We notice that $G/G_y = (G_y \cup G_{x_t})/G_y$ since any edge $i \notin (E_y \cup E_{x_t})$ becomes a selfloop and is removed in the contraction process. This explains the first equality. We write $\bar{\mathcal{A}}_\star$ for $\bar{\mathcal{A}}(G_\star)$ in the following. Considering that no edge $i \in E_{x_t}$ can enter and leave the same connected component C_y^j of G_y by definition, we get

$$(\bar{\mathcal{A}}_{x_t}^\top \bar{Q}_y)_{ij} = \begin{cases} 1, & \text{if edge } i \in E_{x_t} \text{ leaves } C_y^j \\ -1, & \text{if edge } i \in E_{x_t} \text{ enters } C_y^j \\ 0, & \text{if edge } i \in E_{x_t} \text{ is not incident with } C_y^j, \end{cases}$$

where C_y^j is the j-th connected component of G_y. Again by definition, any edge $i \in E_{x_t}$ enters and leaves precisely one component C_y^j, respectively. Considering furthermore that the nodes in $(G_y \cup G_{x_t})/G_y$ are the connected components of G_y, we obtain that $\bar{\mathcal{A}}_{x_t}^\top \bar{Q}_y = \bar{\mathcal{A}}^\top((G_y \cup G_{x_t})/G_y)$, and hence the second equality of the first statement follows.

2. The equality $a^i_{x/y} = \tilde{a}^i_{x/y}$ if edge $i \in E_{x_t}$ is a direct consequence from the first statement, and $\tilde{a}^i_{x/y} = 0$ if $i \in E_x \setminus E_{x_t}$ follows immediately from the structure of $\bar{Q}_y$ and $\bar{\mathcal{A}}_x$.

□

Corollary 2.3.17 (Contraction of a Laplacian matrix) *Let $G = (V, E)$ be a graph with spanning subgraphs $G_x = (V, E_x)$ and $G_y = (V, E_y)$, where E_x, E_y is a partition of E. Let furthermore $\bar{\mathcal{L}}_{x/y}$ and $\bar{\mathcal{L}}_x$ be the Laplacian matrices of G/G_y and G_x, and $\bar{Q}_y$ the $(0,1)$ node component matrix as in Definition 2.3.15. It holds*

$$\bar{\mathcal{L}}_{x/y} = \bar{Q}_y^\top \bar{\mathcal{L}}_x \bar{Q}_y \in \mathbb{R}^{|C_y| \times |C_y|},$$

where $|C_y|$ is the number of connected components of G_y.

Remark. We formulated Lemma 2.3.16 and Corollary 2.3.17 in the non-reduced version to avoid circumstantial case distinctions. However, in the following we shall use them in a *reduced version*. That is, given a reference node in the node set V, we replace all occurring incidence and Laplacian matrices and and the $(0,1)$ node component matrix by their reduced versions. It is easy to check that the results remain valid.

2.4. Conclusion

In this chapter, we presented a differential-algebraic circuit model, namely the MNA, and mostly well-known selected basics of DAE and graph theory. With the exception of the algebraic contraction process in Subsection 2.3.3, the given content is covered in textbooks on circuit theory [CDK87, Rec89], DAEs [HLR89, Ria08, LMT13] or graph theory [GR01, Bap10, Mol12, Bol13].

Lemma 2.3.16 and especially Corollary 2.3.17 in Subsection 2.3.3 can be seen as an algebraic analogue to the operation of (successive edge) contraction graphs. In the next chapter, they are essential for a topological understanding of relevant parts of the MNA equations *after* a DAE decoupling, that is, after transforming the DAE to some normal form. In contrast to the other results of the chapter, the author could not find 2.3.16 and 2.3.17 in this form in the literature.

3. Structural analysis of electrical circuits

We have seen in the previous chapter that it is desirable for the analysis to transform an implicit DAE into a normal form where an inherent ODE and algebraic constraints are decoupled from each other. Such a decoupled normal form is indispensable for the analysis of both dynamical properties and questions regarding the numerical treatment of the DAE. A well-known difficulty concerning network describing DAE models is the physical and network topological interpretation of the decoupled DAE. There are several approaches to transform a DAE into a normal form, but none of them comes with a satisfactory overall topological interpretation. In other words: The decoupled DAE can be analysed well as an abstract mathematical object, but it is hard to keep track of some physical meaning of the decoupled system's parameters and variables. Relating the decoupled DAE with the network topology is therefore usually laborious, but fruitful.

This chapters' main results are presented in Sections 3.4 and 3.5, to which the preceding sections are paving the way. The chapter is organized as follows:

Section 3.1 establishes the preliminaries for a decoupling of the linear circuit DAE model. Essentially, it introduces transformation matrices, which we shall call *kernel splitting pairs*, and which play a central role for the decoupling in the linear case as well as in the nonlinear case.

The decoupling for linear circuit DAEs is then presented in Section 3.2. The decoupling involves a high number of transformation matrices and notational definitions. Since most of the required notation for the linear case is still valid in the nonlinear case, this chapter can be seen is an intermediate point on our way to the nonlinear decoupling result, which allows to become familiar with the notation before tackling the problems posed by nonlinearity.

Section 3.3 then establishes nonlinear preliminaries, mainly related with Lipschitz continuity and strong monotonicity.

A decoupled normal form for nonlinear circuit DAEs is presented in Section 3.4, along with an existence and uniqueness result on bounded time intervals. The section is concluded by a discussion of the results in the context of comparable literature.

Finally, Section 3.5 exploits the normal form of the preceding section and offers a systematic study of the sensitivity of a circuit to (perturbations of) its independent voltage and current sources. The results show that the sensitivity to a source is determined by the topological position of the sources in the circuit. A discussion of the section's results follows.

3.1. Preliminaries from linear algebra

In this section, we collect simple results which are regularly exploited in the proof of the circuit DAE decoupling in Section 3. Transforming the circuit DAE by means of the subsequently defined *kernel splitting pairs* is the basic idea of our decoupling approach. This idea stems from [Jan15].

Lemma 3.1.1 *Let $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{m\times k}$. It holds*

$$\text{im } B = \ker A^\top \quad \Longleftrightarrow \quad \ker B^\top = \text{im } A.$$

Proof: Obviously,

$$\text{im } B \subseteq \ker A^\top \quad \Longleftrightarrow \quad A^\top B = 0 \quad \Longleftrightarrow \quad B^\top A = 0 \quad \Longleftrightarrow \quad \text{im } A \subseteq \ker B^\top.$$

Considering the rank nullity theorem, we obtain

$$\text{rank } B = \text{def } A^\top \quad \Longleftrightarrow \quad m - \text{rank } B^\top = m - \text{def } A^\top \quad \Longleftrightarrow \quad \text{def } B^\top = \text{rank } A.$$

Combining the two observations yields

$$\begin{aligned} \text{im } B = \ker A^\top \quad &\Longleftrightarrow \quad \text{im } B \subseteq \ker A^\top \quad \wedge \quad \text{rank } B = \text{def } A^\top \\ &\Longleftrightarrow \quad \text{im } A \subseteq \ker B^\top \quad \wedge \quad \text{def } B^\top = \text{rank } A \\ &\Longleftrightarrow \quad \ker B^\top = \text{im } A. \end{aligned}$$

□

Lemma 3.1.2 *Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times k}$ be such that $\begin{pmatrix} A & B \end{pmatrix}$ is invertible. Let furthermore $R \in \mathbb{R}^{m\times m}$ and $S \in \mathbb{R}^{k\times k}$ be invertible. Then, $\begin{pmatrix} AR & BS \end{pmatrix}$ is invertible.*

Proof: Factorizing yields $\begin{pmatrix} AR & BS \end{pmatrix} = \begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} R & 0 \\ 0 & S \end{pmatrix}$ and both factors are invertible. □

Lemma 3.1.3 *Let $W \in \mathbb{R}^{m\times m}$ be positive definite and $A \in \mathbb{R}^{k\times m}$. Then,*

(i) $\ker AWA^\top = \ker A^\top$,

(ii) $AWA^\top$ *is positive definite if* $\ker A^\top = \{0\}$.

Proof: Concerning (i), the inclusion $\ker A^\top \subseteq AWA^\top$ is obvious, so we only show the converse inclusion. Due to positive definiteness of W it holds

$$AWA^\top x = 0 \implies x^\top AWA^\top x = 0 \implies A^\top x = 0,$$

hence $\ker AWA^\top \subseteq \ker A^\top$.

Let now $x \neq 0$ be arbitrary. If $\ker A^\top = \{0\}$, then $y := A^\top x \neq 0$. It follows

$$x^\top AWA^\top x = y^\top W y > 0$$

since W is positive definite. This proves (ii). □

Definition 3.1.4 (kernel matrix and kernel splitting pair) *For any matrix $A \neq 0$ with nontrivial kernel, we say Q is a* ***kernel matrix*** *of A, if the columns of Q form a basis of $\ker A$.*

If Q is a kernel matrix of A, we say that $\{P, Q\}$ is a ***kernel splitting pair*** *of A, if the combined matrix $\begin{pmatrix} P & Q \end{pmatrix}$ is nonsingular.*

Example 3.1.5 (kernel splitting pairs) *Let*

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Two possible choices for kernel splitting pairs are $\{P_1, Q\}$ and $\{P_2, Q\}$, defined as

$$Q = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The example illustrates that while the image of Q and the rank of P are fixed for a given A, the image of P depends on the choice of P. The following lemma summarizes some simple results.

Lemma 3.1.6 (kernel splitting pairs) *Let $A \neq 0$ be a matrix with nontrivial kernel, and let $\{P, Q\}$ be an arbitrary orthogonal kernel pair of A. Then the following properties are satisfied.*

1. *P and Q have full column rank.*
2. *im $Q = \ker A$ and $AQ = 0$.*
3. *$\ker P = \ker AP = \{0\}$ and $\ker A^\top = \ker P^\top A^\top$.*
4. *rank P = rank A.*
5. *If A has full row rank, then $\{A^\top, Q\}$ is a kernel splitting pair of A.*

Proof: 1. and 2. follow trivially from the definition.

3. The full column rank of P implies $\ker P = \{0\}$. We consider $x \in \ker AP$. It follows $Px \in \ker A = \text{im } Q$. That is, there exists y such that $Px = Qy$, which is equivalent to $\begin{pmatrix} P & Q \end{pmatrix} \begin{pmatrix} x \\ -y \end{pmatrix} = 0$. Since $\begin{pmatrix} P & Q \end{pmatrix}$ is nonsingular, this implies $x = 0$.

For the second statement, we notice

$$\ker A^\top = \ker \begin{pmatrix} P^\top \\ Q^\top \end{pmatrix} A^\top = \ker \begin{pmatrix} P^\top A^\top \\ 0 \end{pmatrix} = \ker P^\top A^\top .$$

4. If we denote by m the number of columns of A, then P and Q have m rows, and since $\begin{pmatrix} P & Q \end{pmatrix}$ is square, it holds

$$m = \text{rank } P + \text{rank } Q.$$

Furthermore, the rank nullity theorem yields

$$m = \text{rank } A + \dim \ker A = \text{rank } A + \text{rank } Q.$$

From the two equations, it follows immediately that rank $A =$ rank P.

5. $A^\top$ has full column rank since A has full row rank. Furthermore, considering the full column rank of Q and $AQ = 0$, it follows that $\begin{pmatrix} A^\top & Q \end{pmatrix}$ has full columnn rank. It is left to show that $\begin{pmatrix} A^\top & Q \end{pmatrix}$ is square. Considering the previous result, 4., this is the case if rank $A^\top =$ rank A, which holds unconditionally. □

Remark. For a matrix with full column rank, a kernel matrix Q does not exist, while for a zero matrix there is no corresponding P. However, each matrix $M \neq 0 \in \mathbb{R}^{m \times n}$ with a nontrivial kernel has a kernel splitting pair $\{P, Q\}$. To name one, we only have to choose a basis of $\ker M$ and collect the basis vectors as columns of a matrix Q. To construct P, we choose any matrix such that $\begin{pmatrix} P & Q \end{pmatrix}$ is nonsingular.

Lemma 3.1.7 *Let $G = (V, E)$ be a graph with spanning subgraphs $G_x = (V, E_x)$ and $G_y = (V, E_y)$, where E_x, E_y is a partition of E. Let Q_x be a kernel matrix of $\mathcal{A}_x^\top$. Then,*

$$\ker Q_x^\top \mathcal{A}_y = \{0\} \qquad \Longleftrightarrow \qquad (L \subseteq G \textit{ is a loop } \implies L \subseteq G_x)$$

Proof: "$\implies$" First, we note that full column rank of $Q_x^\top \mathcal{A}_y$ implies full column rank of $\mathcal{A}_y$. Hence, Lemma 2.3.3 implies that G_y contains no loops. Furthermore, we obtain

$$\ker Q_x^\top \mathcal{A}_y = \{0\} \implies \text{im } \mathcal{A}_y \cap \ker Q_x^\top = \{0\} \overset{3.1.1}{\implies} \text{im } \mathcal{A}_y \cap \text{im } \mathcal{A}_x = \{0\}.$$

Hence, no column of $\mathcal{A}_y$ can be represented as a linear combination of columns of $\mathcal{A}_x$. Lemma 2.3.3 now implies that any subgraph of G which is a loop can not contain edges

of both E_x and E_y, which yields the desired result.
" $\Longleftarrow$ " First, we notice that from the precondition of this case and Lemma 2.3.3, it follows that $\mathcal{A}_y$ has full column rank. Let now $F_x = (V, E_{F_x})$, $F_x \subseteq G_x$ be a spanning forest, that is, F_x consists of spanning trees of the connected components of G_x. Then, Lemma 2.3.3 implies that im $\mathcal{A}_{F_x} =$ im $\mathcal{A}_x$. Furthermore, $(V, (E_{F_x} \cup E_y)) \subseteq G$ is a forest due to the precondition. Thus, the incidence matrix $(\mathcal{A}_{F_x}\ \mathcal{A}_y)$ has full column rank with Lemma 2.3.3. We obtain

$$\text{im } \mathcal{A}_{F_x} \cap \text{im } \mathcal{A}_y = \{0\} \implies \text{im } \mathcal{A}_x \cap \text{im } \mathcal{A}_y = \{0\} \overset{3.1.1}{\implies} \ker Q_x^\top \cap \text{ im } \mathcal{A}_y = \{0\}.$$

Together with the full column rank of $\mathcal{A}_y$, this yields the desired result. □

3.2. Linear decoupling

In this section, we examine the linear MNA Equations (2.12).

Assumption 3.2.1 (Passivity) *The matrices C, L and G from the linear MNA* (2.12) *are positive definite.*

This passivity assumption is usually employed in literature, cf. [ET00, Tis04, Bar04, Bod07]. These works examine the nonlinear case, but the passivity assumption therein coincides with the above assumption in the linear case.

Kernel splitting pairs for Theorem 3.2.6 Next, we present two tables with kernel splitting pairs, which are listed in the bottom row. The top row displays the matrices which the pairs correspond to. For transposed incidence matrices and corresponding kernel matrices, we employ the general notation $\mathcal{A}_{\star/\bullet}^\top := \mathcal{A}_\star^\top Q_\bullet$.

This notation is motivated by Lemma 2.3.16: If we choose $Q_\bullet$ as the $(0,1)$ node component matrix from Definition 2.3.15, the nonzero columns of $\mathcal{A}_{\star/\bullet}$ form the incidence matrix of the contracted graph $G_\star/G_\bullet$. However, this specific choice of kernel matrices is only motivating the notation; it is not necessary for what follows.

matrix	$\mathcal{A}_c^\top$	$\mathcal{A}_{v/c}^\top := \mathcal{A}_v^\top Q_c$	$(\mathcal{A}_c\ \ \mathcal{A}_v)^\top$	$\mathcal{A}_{r/cv}^\top := \mathcal{A}_r^\top Q_{cv}$	$(\mathcal{A}_c\ \ \mathcal{A}_v\ \ \mathcal{A}_r)^\top$
kernel spl. pair	$\{P_c, Q_c\}$	$\{P_v, Q_v\}$	$\{P_{cv}, Q_{cv}\}$	$\{P_r, Q_r\}$	$\{P_{cvr}, Q_{cvr}\}$

(3.1)

matrix	$\breve{\mathcal{A}}_{v/c} := P_v^\top \mathcal{A}_{v/c} = P_v^\top Q_c^\top \mathcal{A}_v$	$\mathcal{A}_{l/cvr} := Q_{cvr}^\top \mathcal{A}_l$	$\mathring{\mathcal{A}}_v^\top := \breve{Q}_v^\top \mathcal{A}_v^\top P_c$
kernel spl. pair	$\{\breve{\mathcal{A}}_{v/c}^\top, \breve{Q}_v\}$	$\{\mathcal{A}_{l/cvr}^\top, \breve{Q}_l\}$	$\{\mathring{\mathcal{A}}_v, \mathring{Q}_v\}$

(3.2)

We notice that in table (3.2), we made a specific choice for the $P_\star$ matrices. To see that this choice actually yields kernel splitting pairs $\{P_\star, Q_\star\}$, we first observe that $\breve{\mathcal{A}}_{v/c}$, $\mathcal{A}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top$ have full row rank. This is ensured by construction of P_v in the case of $\breve{\mathcal{A}}_{v/c}$ and by the following Lemma 3.2.2 in the case of $\mathcal{A}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top$. The full row rank allows to apply Lemma 3.1.6(5), which states that the choice $P_\star := A_\star^\top$ leads to kernel splitting pairs indeed.

While the specific choice of these matrices is not necessary to obtain the subsequent decoupling results, it proves convenient since it brings forth certain symmetries in the decoupled equations.

Lemma 3.2.2 *Let Assumption 2.1.3(i) hold. Then, the matrices $\mathcal{A}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top$ defined in Table (3.2) have full row rank.*

Proof: We show that the transposed matrices have full column rank.

We obtain for any $y \in \ker \mathcal{A}_{l/cvr}^\top$ and $x := Q_{cvr} y$

$$\mathcal{A}_l^\top Q_{cvr} y = 0 \implies \mathcal{A}_l^\top x = 0 \ \wedge \ \begin{pmatrix} \mathcal{A}_c^\top \\ \mathcal{A}_v^\top \\ \mathcal{A}_r^\top \end{pmatrix} x = 0 \implies \begin{pmatrix} \mathcal{A}_c^\top \\ \mathcal{A}_v^\top \\ \mathcal{A}_r^\top \\ \mathcal{A}_l^\top \end{pmatrix} x = 0 \implies x = 0.$$

Since Q_{cvr} has full column rank, it follows $y = 0$. Hence, $\mathcal{A}_{l/cvr}^\top = \mathcal{A}_l^\top Q_{cvr}$ has only the trivial kernel and therewith full column rank.

The matrix $\mathring{\mathcal{A}}_v$ has full column rank since

$$\ker \mathring{\mathcal{A}}_v = \ker P_c^\top \mathcal{A}_v \breve{Q}_v = \ker \begin{pmatrix} P_c^\top \mathcal{A}_v \breve{Q}_v \\ 0 \end{pmatrix} = \ker \begin{pmatrix} P_c^\top \\ Q_c^\top \end{pmatrix} \mathcal{A}_v \breve{Q}_v = \ker \mathcal{A}_v \breve{Q}_v = \{0\}.$$

□

Assumption 3.2.3 (Index 2 rank conditions) *The row rank of $\begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v & \mathcal{A}_r \end{pmatrix}$ is not full, and / or $\mathcal{A}_{v/c}$ has a nontrivial kernel.*

Remark. Considering Lemma 2.3.1 and Lemma 3.1.7, these rank conditions are equivalent to the topological conditions "the circuit (graph) contains an LI-cutset" and / or "the circuit (graph) contains a CV-loop with at least one V-edge". If either of the two conditions hold, the DAE of the linear MNA (2.12) has index two; otherwise, it has index zero or one, cf. [ET00].

Lemma 3.2.4 *It holds im Q_{cv} = im $Q_c Q_v$ and im Q_{cvr} = im $Q_c Q_v Q_r$.*

Proof: We have to show that

$$\text{im } Q_c Q_v = \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v \end{pmatrix}^\top, \qquad \text{im } Q_c Q_v Q_r = \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v & \mathcal{A}_r \end{pmatrix}^\top.$$

The inclusion "$\subseteq$" follows directly from the definition in both cases. The converse inclusions are left to show.

"$\text{im } Q_c Q_v \supseteq \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v \end{pmatrix}^\top$": Let $x \in \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v \end{pmatrix}^\top$, which is equivalent to $x \in \ker \mathcal{A}_c^\top$ and $x \in \ker \mathcal{A}_v^\top$. We recall that $\ker \mathcal{A}_c^\top = \text{ im } Q_c$ and $\ker \mathcal{A}_v^\top Q_c = \text{ im } Q_v$ by definition. Hence, there exists y such that

$$Q_c y = x \implies y \in \ker \mathcal{A}_v^\top Q_c = \text{ im } Q_v \implies \exists z : Q_v z = y.$$

Consequently, $x = Q_c y = Q_c Q_v z \in \text{im } Q_c Q_v$.

"$\text{im } Q_c Q_v Q_r \supseteq \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v & \mathcal{A}_r \end{pmatrix}^\top$": Analogously to the previous case, let $x \in \ker \begin{pmatrix} \mathcal{A}_c & \mathcal{A}_v & \mathcal{A}_r \end{pmatrix}^\top$, then there exists y such that

$$Q_c y = x \implies y \in \text{ im } Q_v \implies \exists z : Q_v z = y \implies \ker \mathcal{A}_r^\top Q_c Q_v z = \mathcal{A}_r^\top Q_c y = \mathcal{A}_r^\top x = 0,$$

from which the desired result follows immediately. □

The following Lemma shall be used to prove the subsequent decoupling Theorem 3.2.6.

Lemma 3.2.5 (Nonsingularity of transformation matrix T_e) *The matrix T_e, defined by*

$$T_e := \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_{cv} P_r & Q_{cvr} \end{pmatrix},$$

is nonsingular.

Proof: We define $R_1 = P_c \begin{pmatrix} \mathring{\mathcal{A}}_v & \mathring{Q}_v \end{pmatrix}$ and $R_2 := Q_c \begin{pmatrix} P_v & Q_v \end{pmatrix}$. By Lemma 3.1.2 and the nonsingularity of kernel splitting pairs, we obtain the nonsingularity of

$$\begin{pmatrix} P_c \begin{pmatrix} \mathring{\mathcal{A}}_v & \mathring{Q}_v \end{pmatrix} & Q_c \begin{pmatrix} P_v & Q_v \end{pmatrix} \end{pmatrix} = \begin{pmatrix} R_1 & R_2 \end{pmatrix}.$$

We apply Lemma 3.1.2 once more to obtain the nonsingularity of

$$\begin{pmatrix} R_1 & R_2 \begin{pmatrix} P_r & Q_r \end{pmatrix} \end{pmatrix} = \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_c Q_v P_r & Q_c Q_v Q_r \end{pmatrix}.$$

Lemma 3.2.4 furthermore implies

$$\text{im} \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_c Q_v P_r & Q_c Q_v Q_r \end{pmatrix} = \text{ im} \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_{cv} P_r & Q_{cvr} \end{pmatrix},$$

which, together with the nonsingularity of the former matrix, yields the nonsingularity of the latter one. □

In order to avoid too lengthy formulas in the following decoupling Theorem 3.2.6, we introduce a few shorthands.

Linear MNA decoupling Glossary In addition to the definitions from the Tables (3.1), (3.2), we define

$$
\begin{aligned}
T_e &= \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & P_c\mathring{Q}_v & Q_cP_v & Q_{cv}P_r & Q_{cvr} \end{pmatrix},\\
T_l &= \begin{pmatrix} \mathcal{A}_{l/cvr}^\top & \breve{Q}_l \end{pmatrix},\\
T_v &= \begin{pmatrix} \breve{\mathcal{A}}_{v/c}^\top & \breve{Q}_v \end{pmatrix},
\end{aligned}
$$

$$
\begin{aligned}
T_e^- &= \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & P_c\mathring{Q}_v & Q_cP_v & Q_{cv}P_r \end{pmatrix},\\
\bar{T}_e^- &= \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & P_c\mathring{Q}_v & Q_cP_v \end{pmatrix},
\end{aligned}
$$

$$
\bar{e} = \begin{pmatrix} \bar{e}_1 \\ \vdots \\ \bar{e}_5 \end{pmatrix}, \quad \bar{e}_{1234} = \begin{pmatrix} \bar{e}_1 \\ \vdots \\ \bar{e}_4 \end{pmatrix}, \quad \bar{e}_{123} = \begin{pmatrix} \bar{e}_1 \\ \bar{e}_2 \\ \bar{e}_3 \end{pmatrix}, \quad \bar{i}_l = \begin{pmatrix} \bar{i}_{l_1} \\ \bar{i}_{l_2} \end{pmatrix}, \quad \bar{i}_v = \begin{pmatrix} \bar{i}_{v_1} \\ \bar{i}_{v_2} \end{pmatrix},
$$

$$
\begin{aligned}
\breve{\mathcal{A}}_{v/c} &= P_v^\top \mathcal{A}_{v/c},\\
\breve{\mathcal{L}}_{v/c} &= \breve{\mathcal{A}}_{v/c}\breve{\mathcal{A}}_{v/c}^\top,\\
\mathcal{L}_r &= \mathcal{A}_r G \mathcal{A}_r^\top,\\
\breve{\mathcal{L}}_{r/cv} &= P_r^\top Q_{cv}^\top \mathcal{L}_r Q_{cv} P_r,\\
\breve{\mathcal{L}}_c &= P_c^\top \mathcal{A}_c C \mathcal{A}_c^\top P_c,\\
\mathcal{L}_{l/cvr} &= \mathcal{A}_{l/cvr} L^{-1} \mathcal{A}_{l/cvr}^\top,\\
M^\dagger &= (M^\top M)^{-1} M^\top \quad \text{for any matrix } M \text{ with full column rank.}
\end{aligned} \tag{3.3}
$$

Remark. For full column rank matrices, $M^\dagger$ defines a so-called *Moore-Penrose inverse.*

Remark. The notation for incidence and related matrices follows the general pattern

- $\mathcal{A}_\star^\top, \mathcal{A}_\bullet^\top$ with kernel splitting pairs $\{P_\star, Q_\star\}$ and $\{P_\bullet, Q_\bullet\}$, where in the basic MNA model $A_\star$ is the incidence matrix of element type $\star$. Note, however, that in order to allow for generalized models, we do not require $A_\star$ to be an incidence matrix in what follows, cf. Remark 2.1.4.
- $\mathcal{A}_{\star/\bullet}^\top := \mathcal{A}_\star^\top Q_\bullet$ with kernel splitting pair $\{P_{\star/\bullet}, Q_{\star/\bullet}\}$.
- $\breve{\mathcal{A}}_\star^\top := \mathcal{A}_\star P_\star$ and $\breve{\mathcal{A}}_{\star/\bullet}^\top := \mathcal{A}_{\star/\bullet}^\top P_{\star/\bullet}$. These matrices have full column rank with Lemma 3.1.6(3).
- Kernel splitting pairs of (non-transposed) matrices $\mathcal{A}_{\star/\bullet}$ or $\breve{\mathcal{A}}_{\star/\bullet}$ are denoted by $\{\breve{P}_{\star/\bullet}, \breve{Q}_{\star/\bullet}\}$.

- $\mathcal{L}_\star := \mathcal{A}_\star W_\star \mathcal{A}_\star^\top$ and $\mathcal{L}_{\star/\bullet} := \mathcal{A}_{\star/\bullet} W_\star \mathcal{A}_{\star/\bullet}^\top$ with a symmetric and positive definite matrix $W_\star$.
- $\check{\mathcal{L}}_\star := \check{\mathcal{A}}_\star W_\star \check{\mathcal{A}}_\star^\top$ and $\check{\mathcal{L}}_{\star/\bullet} := \check{\mathcal{A}}_{\star/\bullet} W_\star \check{\mathcal{A}}_{\star/\bullet}^\top$. These matrices are nonsingular with Lemma 3.1.3.

In Tables (3.1) and (3.2), only $\mathring{\mathcal{A}}_v^\top$ and its kernel splitting pair is an exception as it does not match the previously described notational pattern.

Another notational irregularity regards $W_\star$ as it can represent the weight matrix of the edge type $\star$ (in the case of C), the inverse weight matrix (in the cases of R, L) or the identity (in the case of V).

The general notation is motivated by Lemma 2.3.16 and Corollary 2.3.17, which state that if $\mathcal{A}_\star, \mathcal{A}_\bullet$ are incidence matrices $Q_\bullet$ is a particularly chosen kernel matrix of $\mathcal{A}_\bullet^\top$, then $\mathcal{A}_{\star/\bullet}$ is, up to zero columns, the incidence matrix of the contracted graph $G_\star/G_\bullet$.

Theorem 3.2.6 (Index two MNA decoupling) *Let the Assumptions 2.1.3 and 3.2.3 be satisfied. Then, $\bar{e}$, $\bar{i}_l$, $\bar{i}_v$ solve the DAE*

$$\bar{e}_2' = -\mathring{Q}_v^\dagger \check{\mathcal{L}}_c^{-1} P_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v T_v \bar{i}_v + \mathcal{A}_i i_s\right) \tag{3.4}$$

$$\bar{i}_{l_2}' = \check{Q}_l^\dagger L^{-1} \mathcal{A}_l^\top T_e \bar{e} \tag{3.5}$$

$$\bar{e}_5 = \mathcal{L}_{l/cvr}^{-1} \mathcal{A}_{l/cvr} \left(\mathcal{A}_{l/cvr}^\top \bar{i}_{l_1}' - L^{-1} \mathcal{A}_l^\top T_e^- \bar{e}_{1234}\right) \tag{3.6}$$

$$\bar{i}_{v_2} = -[\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v]^{-1} \mathring{\mathcal{A}}_v^\top \left(\mathring{\mathcal{A}}_v \bar{e}_1' + \check{\mathcal{L}}_c^{-1} P_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \mathcal{A}_i i_s\right)\right) \tag{3.7}$$

$$\bar{e}_3 = (\check{\mathcal{A}}_{v/c}^\top)^\dagger \left(v_s - \mathcal{A}_v^\top P_c (\mathring{\mathcal{A}}_v \bar{e}_1 + \mathring{Q}_v \bar{e}_2)\right) \tag{3.8}$$

$$\bar{e}_4 = -\check{\mathcal{L}}_{r/cv}^{-1} P_r^\top Q_{cv}^\top \left(\mathcal{L}_r \bar{T}_e^- \bar{e}_{123} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s\right) \tag{3.9}$$

$$\bar{i}_{v_1} = -\check{\mathcal{L}}_{v/c}^{-1} P_v^\top Q_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s\right) \tag{3.10}$$

$$\bar{e}_1 = (\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v)^{-1} \check{Q}_v^\top v_s, \tag{3.11}$$

$$\bar{i}_{l_1} = -(\mathcal{A}_{l/cvr} \mathcal{A}_{l/cvr}^\top)^{-1} \mathcal{A}_{i/cvr} i_s. \tag{3.12}$$

if and only if $e = T_e \bar{e}$, $i_l = T_l \bar{i}_l$, $i_v = T_v \bar{i}_v$ solve the linear MNA (2.12).

Proof: We introduce the new coordinates $T_e \bar{e} = e$, $T_l \bar{i}_l = i_l$, $T_v \bar{i}_v = i_v$ and we consider the transformed system

$$\tilde{T}_e^\top \left[\mathcal{L}_c T_e \bar{e}' + \mathcal{L}_r T_e \bar{e} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v T_v \bar{i}_v + \mathcal{A}_i i_s\right] = 0 \tag{3.13}$$

$$\tilde{T}_l^\top \left[L T_l \bar{i}_l' - \mathcal{A}_l^\top T_e \bar{e}\right] = 0 \tag{3.14}$$

$$T_v^\top \left[\mathcal{A}_v^\top T_e \bar{e} - v_s\right] = 0, \tag{3.15}$$

where

$$\tilde{T}_e := \begin{pmatrix} P_c(\check{\mathcal{L}}_c^{-1})^\top \mathring{\mathcal{A}}_v & P_c(\check{\mathcal{L}}_c^{-1})^\top \mathring{Q}_v & Q_c P_v & Q_{cv} P_r & Q_{cvr} \end{pmatrix},$$
$$\tilde{T}_l := \Big((L^{-1})^\top \mathcal{A}_{l/cvr}^\top \quad (L^{-1})^\top \breve{Q}_l \Big).$$

The proof consists of three steps:

1. The transformations performed to obtain System (3.13)-(3.15) from the linear MNA (2.12) are nonsingular equivalence transformations.

2. System (3.13)-(3.15) written down blockwise by the blocks of $\tilde{T}_e^\top$, $\tilde{T}_l^\top$ and $T_v^\top$ reads (3.16)-(3.24).

3. System (3.4)-(3.12) is obtained by equivalence transformations from the blockwise System (3.16)-(3.24) so that Table (3.25) holds.

1. Note that the transformations matrices $T_e, T_l, T_v, \tilde{T}_e$ and $\tilde{T}_l$ are all nonsingular: This is guaranteed by Lemma 3.2.5 for T_e and by Lemma 3.1.6(5) for T_l and T_v, whereas Lemma 3.1.2 yields the nonsingularity of $\tilde{T}_l$. Concerning $\tilde{T}_e$, we first notice that $\begin{pmatrix} P_c(\check{\mathcal{L}}_c^{-1})^\top & Q_c \end{pmatrix}$ is nonsingular due to Lemma 3.1.2; Consequently, nonsingularity of $\tilde{T}_e$ is obtained analogously to the lines of proof of Lemma 3.2.5 when substituting $\tilde{R}_1 := P_c(\check{\mathcal{L}}_c^{-1})^\top \begin{pmatrix} \mathring{\mathcal{A}}_v & \mathring{Q}_v \end{pmatrix}$ for R_1.

Hence, $\bar{e}$, $\bar{i}_l$, $\bar{i}_v$ solve the transformed DAE (3.13)-(3.15) if and only if $e = T_e\bar{e}$, $i_l = T_l\bar{i}_l$, $i_v = T_v\bar{i}_v$ solves linear MNA (2.12).

2. The equivalently transformed system (3.13)-(3.15) written down blockwise by the blocks of $\tilde{T}_e^\top$, $\tilde{T}_l^\top$ and $T_v^\top$ reads

$$\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v \bar{e}_1' + \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v \underline{\bar{i}_{v_2}} + \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \mathcal{A}_i i_s \right) = 0, \quad (3.16)$$
$$\mathring{Q}_v^\top \mathring{Q}_v \underline{\bar{e}_2'} + \mathring{Q}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v T_v \bar{i}_v + \mathcal{A}_i i_s \right) = 0, \quad (3.17)$$
$$\check{\mathcal{L}}_{v/c} \underline{\bar{i}_{v_1}} + P_v^\top Q_c^\top \left(\mathcal{L}_r T_e^- \bar{e}_{1234} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s \right) = 0, \quad (3.18)$$
$$\check{\mathcal{L}}_{r/cv} \underline{\bar{e}_4} + P_r^\top Q_{cv}^\top \left(\mathcal{L}_r \bar{T}_e^- \bar{e}_{123} + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s \right) = 0, \quad (3.19)$$
$$\mathcal{A}_{l/cvr} \mathcal{A}_{l/cvr}^\top \underline{\bar{i}_{l_1}} + \mathcal{A}_{i/cvr} i_s = 0, \quad (3.20)$$
$$-\mathcal{L}_{l/cvr} \underline{\bar{e}_5} + \mathcal{A}_{l/cvr} \left(\mathcal{A}_{l/cvr}^\top \bar{i}_{l_1}' - L^{-1} \mathcal{A}_l^\top \bar{T}_e^- \bar{e}_{1234} \right) = 0, \quad (3.21)$$
$$\breve{Q}_l^\top \breve{Q}_l \underline{\bar{i}_{l_2}'} - \breve{Q}_l^\top L^{-1} \mathcal{A}_l^\top T_e \bar{e} = 0, \quad (3.22)$$
$$\check{\mathcal{A}}_{v/c} \mathcal{A}_v^\top P_c (\mathring{\mathcal{A}}_v \bar{e}_1 + \mathring{Q}_v \bar{e}_2) + \check{\mathcal{L}}_{v/c} \underline{\bar{e}_3} - \check{\mathcal{A}}_{v/c} v_s = 0, \quad (3.23)$$
$$\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v \underline{\bar{e}_1} - \breve{Q}_v^\top v_s = 0. \quad (3.24)$$

The derivation of this blockwise representation from the former compact one is straightforward. The difficulty is, however, to become familiar with the bulk of notation and matrices introduced for this theorem. Therefore, we provide brief clarifying hints for each equation of the block-wise representation (3.16)-(3.24).

Multiplying Eq. (3.13) by $\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top$ yields Eq. (3.16):
First, we notice that

$$\begin{aligned}
&\mathcal{L}_c T_e = \mathcal{A}_c C \mathcal{A}_c^\top \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_{cv} P_r & Q_{cvr} \end{pmatrix} = \mathcal{L}_c \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & 0 & 0 & 0 \end{pmatrix}, \\
&\mathcal{L}_r T_e = \mathcal{L}_r \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_{cv} P_r & 0 \end{pmatrix}
\end{aligned}$$

by construction of Q_c and Q_{cvr}. Moreover,

$$\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \mathcal{L}_c T_e \bar{e}' = \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \mathcal{L}_c (P_c \mathring{\mathcal{A}}_v \bar{e}_1' + P_c \mathring{Q}_v \bar{e}_2') = \mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v \bar{e}_1'$$

by construction of $\mathring{Q}_v$ and since $\check{\mathcal{L}}_c = P_c^\top \mathcal{L}_c P_c$ by definition. Furthermore,

$$\begin{aligned}
\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \mathcal{A}_v T_v \bar{i}_v &= \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \mathcal{A}_v (\check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \check{Q}_v \bar{i}_{v_2}) \\
&= \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v \bar{i}_{v_2}
\end{aligned}$$

by definition of $\mathring{\mathcal{A}}_v$. Then, Equation (3.16) follows readily when Equation (3.13) is left multiplied by the first block of $\tilde{T}_e^\top$, namely $\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top$.

Multiplying Eq. (3.13) by $\mathring{Q}_v^\top \check{\mathcal{L}}_c^{-1} P_c^\top$ yields Eq. (3.17):
This equation is obtained by analogous considerations as for the previous equation.

Multiplying Eq. (3.13) by $P_v^\top Q_c^\top$ yields Eq. (3.18):
We notice that $P_v^\top Q_c^\top \mathcal{L}_c = 0$ and

$$P_v^\top Q_c^\top \mathcal{A}_v T_v \bar{i}_v = \check{\mathcal{A}}_{v/c} (\check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \check{Q}_v \bar{i}_{v_2}) = \check{\mathcal{L}}_{v/c} \bar{i}_{v_1}.$$

Multiplying Eq. (3.13) by $P_r^\top Q_{cv}^\top$ yields Eq. (3.19):
This follows from

$$P_r^\top Q_{cv}^\top \mathcal{L}_r T_e \bar{e} = P_r^\top Q_{cv}^\top \mathcal{L}_r (\bar{T}_e^- \bar{e}_{123} + Q_{cv} P_r \bar{e}_4) = P_r^\top Q_{cv}^\top \mathcal{L}_r \bar{T}_e^- \bar{e}_{123} + \check{\mathcal{L}}_{r/cv} \bar{e}_4.$$

Multiplying Eq. (3.13) by $Q_{cvr}^\top$ yields Eq. (3.20):
By construction of Q_{cvr}, it is $Q_{cvr}^\top \mathcal{L}_c = Q_{cvr}^\top \mathcal{L}_r = 0$ and $Q_{cvr}^\top \mathcal{A}_v = 0$. Furthermore,

$$Q_{cvr}^\top \mathcal{A}_l T_l \bar{i}_l = \mathcal{A}_{l/cvr} (\mathcal{A}_{l/cvr}^\top \bar{i}_{l_1} + \check{Q}_l \bar{i}_{l_2}) = \mathcal{A}_{l/cvr} \mathcal{A}_{l/cvr}^\top \bar{i}_{l_1}.$$

<u>Multiplying Eq. (3.14) by $\mathcal{A}_{l/cvr}L^{-1}$ yields Eq. (3.21):</u>
This follows from

$$
\begin{aligned}
\mathcal{A}_{l/cvr}L^{-1}LT_l\bar{i}'_l &= \mathcal{A}_{l/cvr}\mathcal{A}_{l/cvr}^\top\bar{i}'_{l_1}, \\
\mathcal{A}_{l/cvr}L^{-1}\mathcal{A}_l^\top T_e\bar{e} &= \mathcal{A}_{l/cvr}L^{-1}\mathcal{A}_l^\top T_e^-\bar{e}_{1234} + \mathcal{A}_{l/cvr}L^{-1}\mathcal{A}_l^\top Q_{cvr}\bar{e}_5 \\
&= \mathcal{A}_{l/cvr}L^{-1}\mathcal{A}_l^\top T_e^-\bar{e}_{1234} + \mathcal{L}_{l/cvr}\bar{e}_5.
\end{aligned}
$$

<u>Multiplying Eq. (3.14) by $\breve{Q}_l^\top L^{-1}$ yields Eq. (3.22):</u>
This is obvious considering that $\breve{Q}_l^\top\mathcal{A}_{l/cvr}^\top = 0$ by construction of $\breve{Q}_l^\top$.

<u>Multiplying Eq. (3.15) by $\breve{\mathcal{A}}_{v/c}$ yields Eq. (3.23):</u>
First, we note that

$$
\mathcal{A}_v^\top T_e = \mathcal{A}_v^\top \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & P_c\mathring{Q}_v & Q_cP_v & 0 & 0 \end{pmatrix}.
$$

Furthermore,

$$
\begin{aligned}
\breve{\mathcal{A}}_{v/c}\mathcal{A}_v^\top T_e\bar{e} &= \breve{\mathcal{A}}_{v/c}\mathcal{A}_v^\top(P_c\mathring{\mathcal{A}}_v\bar{e}_1 + P_c\mathring{Q}_v\bar{e}_2 + Q_cP_v\bar{e}_3) \\
&= \breve{\mathcal{A}}_{v/c}\mathcal{A}_v^\top P_c(\mathring{\mathcal{A}}_v\bar{e}_1 + \mathring{Q}_v\bar{e}_2) + \breve{\mathcal{A}}_{v/c}\breve{\mathcal{A}}_{v/c}^\top\bar{e}_3.
\end{aligned}
$$

<u>Multiplying Eq. (3.15) by $\breve{Q}_v^\top$ yields Eq. (3.24):</u>
We observe that

$$
\begin{aligned}
\breve{Q}_v^\top\mathcal{A}_v^\top T_e &= \breve{Q}_v^\top\mathcal{A}_v^\top \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & P_c\mathring{Q}_v & Q_cP_v & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \mathring{\mathcal{A}}_v^\top\mathring{\mathcal{A}}_v & \mathring{\mathcal{A}}_v^\top\mathring{Q}_v & \breve{Q}_v^\top\breve{\mathcal{A}}_{v/c}^\top & 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} \mathring{\mathcal{A}}_v^\top\mathring{\mathcal{A}}_v & 0 & 0 & 0 & 0 \end{pmatrix}.
\end{aligned}
$$

3. Next, we show that Equations (3.16)-(3.24) are equivalent to Equations (3.4)-(3.12), or more precisely that each equation of one system is equivalent to one equation of the other, as displayed in following table.

Equation	(3.16)	(3.17)	(3.18)	(3.19)	(3.20)	(3.21)	(3.22)	(3.23)	(3.24)
equiv. transformed to Eq.	(3.7)	(3.4)	(3.10)	(3.9)	(3.12)	(3.6)	(3.5)	(3.8)	(3.11)

(3.25)

For the table to hold, we need to confirm that the matrices operating on the underlined variables in Equations (3.16)-(3.24) are invertible.

We start by noticing that the matrices $\breve{\mathcal{A}}_{v/c}^\top = \mathcal{A}_v^\top Q_c P_v$, $\mathcal{A}_r^\top Q_{cv} P_r$ and $\mathcal{A}_c^\top P_c$ have full column rank by construction of P_v, P_r and P_c and by Lemma 3.1.6(3). Hence, Lemma 3.1.3 yields the invertibility (and positive definiteness) of $\breve{\mathcal{L}}_{v/c}$, $\breve{\mathcal{L}}_{r/cv}$ and $\breve{\mathcal{L}}_c$. The first two, $\breve{\mathcal{L}}_{v/c}$ and $\breve{\mathcal{L}}_{r/cv}$, are the relevant matrices for matrices for Equations (3.18), (3.19) and (3.23).

Furthermore, $\breve{\mathcal{L}}_c^{-1}$ is positive definite since $\breve{\mathcal{L}}_c$ is. Considering additionally that $\mathring{\mathcal{A}}_v$ has full column rank with Lemma 3.2.2, this implies the invertibility of $\mathring{\mathcal{A}}_v^\top \breve{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v$, which is the relevant matrix for Equation (3.16).

Lemma 3.2.2 also provides that $\mathcal{A}_{l/cvr}^\top$ has full column rank, which yields the invertibility of $\mathcal{L}_{l/cvr}$ and $\mathcal{A}_{l/cvr}\mathcal{A}_{l/cvr}^\top$ and therewith the solvability of Equations (3.21) and (3.20) as desired.

The remaining matrices in question belong to the Equations (3.17), (3.22) and (3.24); their invertibility follows immediately from the full column rank of $\mathring{Q}_v$, $\breve{Q}_l$ and $\mathring{\mathcal{A}}_v$. □

Corollary 3.2.7 (Index two MNA full decoupling) *Let the Assumptions 2.1.3 and 3.2.3 for the linear MNA (2.12) be satisfied. Then, there exists a nonsingular matrix $T = \begin{pmatrix} T_0 & T_1 & T_2 & T_3 \end{pmatrix}$ and matrices $F_\star$, $G_\star$ such that the variables y, z_1, z_2, z_3 solve the fully decoupled DAE*

$$y' = Fy + F_s s + F_{s'} s' \tag{3.26a}$$
$$z_1 = G_1 y + G_{1,s} s + G_{s'} s' \tag{3.26b}$$
$$z_2 = G_2 y + G_{2,s} s, \tag{3.26c}$$
$$z_3 = G_{3,s} s \tag{3.26d}$$

where $s := \left(\begin{smallmatrix} i_s \\ v_s \end{smallmatrix}\right)$, if and only if $\left(\begin{smallmatrix} e \\ i_l \\ i_v \end{smallmatrix}\right) = x = T_0 y + T_1 z_1 + T_2 z_2 + T_3 z_3$ solves the linear MNA (2.12).

More precisely, the following relations with the normal form (3.4)-(3.12) from the decoupling Theorem 3.2.6 hold with $M_{cv} := [\mathring{\mathcal{A}}_v^\top \breve{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v]^{-1}$:

$$y = \begin{pmatrix} \bar{e}_2 \\ \bar{i}_{l_2} \end{pmatrix}, \quad z_1 = \begin{pmatrix} \bar{e}_5 \\ \bar{i}_{v_2} \end{pmatrix}, \quad z_2 = \begin{pmatrix} \bar{e}_3 \\ \bar{e}_4 \\ \bar{i}_{v_1} \end{pmatrix}, \quad z_3 = \begin{pmatrix} \bar{e}_1 \\ \bar{i}_{l_1} \end{pmatrix}, \quad G_{s'} = \begin{pmatrix} -\mathcal{L}_{l/cvr}^{-1}\mathcal{A}_{i/cvr} & 0 \\ 0 & -M_{cv}^{-1}\breve{Q}_v^\top \end{pmatrix}$$

and $T = \begin{pmatrix} T_0 & T_1 & T_2 & T_3 \end{pmatrix}$ arises from $\left(\begin{smallmatrix} T_e & 0 & 0 \\ 0 & T_l & 0 \\ 0 & 0 & T_v \end{smallmatrix}\right)$ if the columns are adequatly permuted, that is,

$$T_0 = \begin{pmatrix} P_c\breve{Q}_v & 0 \\ 0 & \breve{Q}_l \\ 0 & 0 \end{pmatrix}, \quad T_1 = \begin{pmatrix} Q_{cvr} & 0 \\ 0 & 0 \\ 0 & \breve{Q}_v \end{pmatrix}, \quad T_2 = \begin{pmatrix} Q_c P_v & Q_{cv} P_r & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \breve{\mathcal{A}}_{v/c}^\top \end{pmatrix}, \quad T_3 = \begin{pmatrix} P_c\mathring{\mathcal{A}}_v & 0 \\ 0 & \mathcal{A}_{l/cvr}^\top \\ 0 & 0 \end{pmatrix}.$$

Proof: The desired result follows readily from the decoupling Theorem if we eliminate the variables contained in z_1 and z_2 from Equations (3.4)-(3.12) by insertion in the right hand side and resort the equations according to the order proposed by y, z_1, z_2, z_3. The matrix $G_{1,s'}$ arises from insertion of $\bar{i}'_{l_1} = -(\mathcal{A}_{l/cvr}\mathcal{A}_{l/cvr}^\top)^{-1}\mathcal{A}_{i/cvr}i'_s$ (differentiation of Equation (3.12)) and $\bar{e}'_1 = (\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v)^{-1}\breve{Q}_v^\top v'_s$ (differentiation of Equation (3.11)) into the right hand sides of Equations (3.6) and (3.7). □

The index 1 case and other special cases For the decoupled Representation (3.4)-(3.12) in Theorem 3.2.6 and the resulting Corollary 3.2.7, we implicitly assumed that all the relevant matrices for which we define kernel splitting pairs in Tables (3.1) and (3.2) are nonzero matrices with a nontrivial kernel, since otherwise the corresponding kernel splitting pairs are not well-defined. While this implicit assumption helps us to establish a consistent notation and to avoid many case distinctions, its violation poses no mathematical problem.

In the following, we provide a brief and rather informal recipe to find the normal form in these cases, and notably in the index-1 case. For that, it is useful to introduce matrices with zero rows and matrices with zero columns, that is, $A \in \mathbb{R}^{0\times n}$ and $B \in \mathbb{R}^{n\times 0}$. For a matrix $A \in \mathbb{R}^{0\times n}$ consisting of zero rows, we define $\ker A := \mathbb{R}^n$, and for the multiplication with a matrix $M \in \mathbb{R}^{n\times n}$ we define $AM \in \mathbb{R}^{0\times n}$ and $MB \in \mathbb{R}^{n\times 0}$. Finally, we define the kernel splitting pair for any matrix $R \in \mathbb{R}^{n\times m}$, $\ker R = \{0\}$ as $\{P_R = I_m, Q \in \mathbb{R}^{m\times 0}\}$, and for the zero matrix $0 = S \in \mathbb{R}^{n\times m}$ as $\{P_S \in \mathbb{R}^{m\times 0}, Q_S = I_m\}$. Here, I_m denotes the $(m \times m)$ identity matrix. To avoid confusion, we shall denote $I_m =: I$ for the remainder of this paragraph.

For an example, we first consider the case $\ker P_v^\top Q_c^\top \mathcal{A}_v \overset{3.1.6}{=} \ker Q_c^\top \mathcal{A}_v = \{0\}$, which implies that the circuit has no CV-loop with at least one voltage source. Then, we obtain $\{\breve{P}_v = I, \breve{Q}_v \in \mathbb{R}^{n\times 0}\}$ as the corresponding kernel splitting pair. It follows that $\mathring{\mathcal{A}}_v^\top \in \mathbb{R}^{0\times m}$ and the corresponding splitting pair becomes $\{\mathring{P}_v \in \mathbb{R}^{m\times 0}, \mathring{Q}_v = I\}$. Hence

$$T_v = I, \quad T_e = \begin{pmatrix} P_c & Q_c P_v & Q_{cv} P_r & Q_{cvr} \end{pmatrix}.$$

Consequently, the variables $\bar{i}_{v_2}$ and $\bar{e}_1$ vanish along with Equations (3.7) and (3.11) in the normal form of decoupling Theorem 3.2.6.

Similarly, if $\ker \mathcal{A}_r^\top Q_{cv} = \{0\}$, which means that there exists no LI-cutset in the circuit, it follows that $Q_r \in \mathbb{R}^{k\times 0}$ and $\breve{Q}_l = I$, so that

$$T_l = I, \quad T_e = \begin{pmatrix} P_c \mathring{\mathcal{A}}_v & P_c \mathring{Q}_v & Q_c P_v & Q_{cv} \end{pmatrix}$$

and the variables $\bar{e}_5$ and $\bar{i}_{l_1}$ and Equations (3.6) and (3.12) vanish.

If the circuit is of index one, that is, $\ker Q_c^\top \mathcal{A}_v$ *and* $\ker \mathcal{A}_{r/cv}^\top$ have a trivial kernel, then in the resulting normal form the variables $\bar{e}_5, \bar{i}_{v_2}, \bar{e}_1, \bar{i}_{l_1}$ vanish as well as Equations (3.6),(3.7),(3.11),(3.12). In the compact representation of Corollary 3.2.7, the variables z_1 and z_3 and Equations (3.26b) and (3.26d) vanish and $F_{s'} = 0$.

To sum it up, if the index-two Assumption 3.2.3 does not hold, or more generally any of the matrices in Tables (3.1), (3.2) is the zero matrix or has a trivial kernel, then this actually facilitates the DAE decoupling since one has to perform less equation and / or variable splittings. However, the difficulty lies in developping a notational framework which encompasses each possible case. Since we did not tackle this challenge, this paragraph provides an informal and pragmatic guide of how to find the simplified normal form in the index 1 case or other cases where one or several matrices in Tables (3.1), (3.2) have a nontrivial kernel and / or are zero matrices.

3.3. Preliminaries for nonlinear circuits

Here we present results which enable us to achieve nonlinear decoupling results similar to the linear case and exploit them for a perturbation analysis in the subsequent section.

Definition 3.3.1 (Lipschitz continuity) *Let $D_n \subseteq \mathbb{R}^n$ and $D_l \subseteq \mathbb{R}^l$. A function $f : D_n \to \mathbb{R}^m$ is* Lipschitz continuous, *if*

$$\exists L > 0 : \ |f(x_1) - f(x_2)| \leq L|x_1 - x_2| \quad \forall x_1, x_2 \in D_n,$$

and f is locally Lipschitz continuous, *if $f_{|B_n} : B_n \to \mathbb{R}^m$ is Lipschitz continuous for any compact subset $B_n \subseteq D_n$.*

A function $g : D_l \times D_n \to \mathbb{R}^m$ is Lipschitz continuous in the second argument y, *if*

$$\exists L > 0 : \ |g(x, y_1) - g(x, y_2)| \leq L|y_1 - y_2| \quad \forall x \in D_l, y_1, y_2 \in D_n,$$

and it is locally Lipschitz continuous in the second argument y, *if $g_{|(D_l \times B_n)}$ is Lipschitz continuous in the second argument for any compact subset $B_n \subseteq D_n$.*

Remark. Note that the Lipschitz constant L for g is independent of x, but the local Lipschitz constants for f and g depend on the subset B_n in contrast to the global Lipschitz constants. The given definition of local Lipschitz continuity, cf. [Tes12], is equivalent to the following more common definition

$$f : D \to \mathbb{R}^n \text{ is locally Lipsch. cont.} \iff \forall x \in D \ \exists\, U_x \subseteq D : \ f_{|U_x} \text{ is Lipsch. cont.}$$

Definition 3.3.2 (Strong monotonicity, uniform definiteness, boundedness) *Consider the functions $f : \mathbb{R}^n \to \mathbb{R}^n$, $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ and $M : D_n \to \mathbb{R}^{m \times m}$, $D_n \subseteq \mathbb{R}^n$.*

- *f is* strongly monotone *if there exists a constant $\mu_f > 0$ such that*

$$(x_2 - x_1)^\top (f(x_2) - f(x_1)) \geq \mu_f |x_2 - x_1|^2, \ \forall x_1, x_2 \in \mathbb{R}^n, \tag{3.27}$$

- *g is* strongly monotone in the second argument *if there exists a constant $\mu_g > 0$ such that*

$$(y_2 - y_1)^\top (g(x, y_2) - g(x, y_1)) \geq \mu_g |y_2 - y_1|^2, \ \forall x \in \mathbb{R}^n, y_1, y_2 \in \mathbb{R}^m, \tag{3.28}$$

- *M is* uniformly positive definite *if there exists a constant μ_M such that*

$$y^\top M(x) y \geq \mu_M |y|^2, \ \forall x \in D_n, y \in \mathbb{R}^m, \tag{3.29}$$

- *M is* bounded*, if there exists $C > 0$ such that*

$$|M(x)|_* \leq C \quad \forall x \in D_n,$$

where $|\cdot|_*$ is the induced matrix norm.

Remark. Due to equivalence of norms in $\mathbb{R}^n$, the above concepts are well-defined independently of the chosen norm. Note furthermore that in general, positive definiteness of $M(x)$ for all x does not imply uniform positive definiteness of M.

A proof for the following lemma and theorem can be found in [OR70, Theorems 5.4.3. and 6.4.4].

Lemma 3.3.3 *Let $D \in \mathbb{R}^n$ and $f : \mathbb{R}^m \to \mathbb{R}^m$ be continuously differentiable with Jacobian matrix $f'(x)$. Then, f is strongly monotone if and only if f' is uniformly positive definite.*

Theorem 3.3.4 (Browder Minty) *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and strongly monotone. Then, the equation*

$$f(x) = y$$

has a unique solution $x \in \mathbb{R}^n$ for each $y \in \mathbb{R}^n$. Furthermore, the inverse function $f^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz continuous.

The previous two results can be extended to parameter-dependent equations, cf. [Mat12, Lemma 3.4] and [JMT15, Lemma 2.6].

Lemma 3.3.5 *Let $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be continuous and continuously differentiable w.r.t. $z \in \mathbb{R}^m$ with Jacobian $\frac{\partial f}{\partial z}(x,z)$ at the point (x,z). Then, f is strongly monotone w.r.t. z if and only if $\frac{\partial f}{\partial z}$ is uniformly positive definite, which means here*

$$\exists \mu > 0 : \; y^\top \frac{\partial f}{\partial z}(x,z) y \geq \mu y^\top y \quad \forall x \in \mathbb{R}^n, z \in \mathbb{R}^m, y \in \mathbb{R}^m.$$

Remark. Notice that the constant μ is independent of the parameters x, z.

Lemma 3.3.6 (Parameter-dependent Browder Minty) *Let $\mathcal{I} \subset \mathbb{R}$ be an interval and $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{I} \to \mathbb{R}^m$ be a continuous function. Then, for all $(x,t) \in \mathbb{R}^n \times \mathcal{I}$, the equation*

$$f(x,y,t) = 0 \tag{3.30}$$

has a unique solution $y \in \mathbb{R}^m$ if f is strongly monotone w.r.t. y and globally Lipschitz continuous w.r.t. x. The solution depends on (x,t) and we write $y = \Psi(x,t)$ with the continuous function $\Psi : \mathbb{R}^n \times \mathcal{I} \to \mathbb{R}^m$ which is globally Lipschitz continuous w.r.t. x.

The following Lemma's purpose for us is only to simplify the proof of the Lemma 3.3.8 thereafter.

Lemma 3.3.7 *Let $M : D_n \to \mathbb{R}^{m\times m}$ be uniformly positive definite and let $|\cdot|$ be a vector norm. Then, there exists $\mu_M > 0$ such that*

$$|M(x)y| \geq \mu_M |y| \quad \forall x \in D_n, y \in \mathbb{R}^m.$$

Proof: There exists $\tilde{\mu}_M > 0$ such that

$$|y|_2 |M(x)y|_2 \geq y^\top M(x)y \geq \tilde{\mu}_M |y|_2^2 \quad \forall x \in D_n, y \in \mathbb{R}^m$$

for the euclidean vector norm $|\cdot|_2$. The first inequality is Cauchy-Schwarz, and the second one holds since M is uniformly positive definite. Hence, it holds

$$|M(x)y|_2 \geq \tilde{\mu}_M |y|_2 \quad \forall x \in D_n, y \in \mathbb{R}^m.$$

Due to equivalence of norms, we can conclude that for an arbitrary vector norm $|\cdot|$ there exists μ_M such that

$$|M(x)y| \geq \mu_M |y| \quad \forall x \in D_n, y \in \mathbb{R}^m.$$

□

Lemma 3.3.8 *Let $M : D_n \to \mathbb{R}^{m\times m}$, $D_n \subseteq \mathbb{R}^n$ be Lipschitz continuous and uniformly positive definite. Then, the function*

$$M^{-1} : D_n \to \mathbb{R}^{m\times m}, \qquad x \mapsto [M(x)]^{-1}$$

is Lipschitz continuous.

Proof: $M(x)$ is invertible for all $x \in \mathbb{R}^n$ since $M(x)$ is positive definite for all x.

We use the induced matrix norm $|A|_* := \sup_{z\neq 0} \frac{|Az|}{|z|}$, for which it holds the submultiplicativity property $|AB| \leq |A_*||B|_*$ $\forall A, B \in \mathbb{R}^{m\times m}$. Furthermore, we exploit Lemma 3.3.7 and the fact that $z \mapsto M(x)z$ is surjective for arbitrary x. For $x, y \in D_n$ arbitrary, it holds

$$\begin{aligned}
|[M(x)]^{-1} - [M(y)]^{-1}|_* &= |[M(y)]^{-1}M(y)[M(x)]^{-1} - [M(y)]^{-1}M(x)[M(x)]^{-1}|_* \\
&\leq |[M(y)]^{-1}|_* |M(y) - M(x)|_* |[M(x)]^{-1}|_* \\
&\leq \sup_{u\neq 0} \frac{|[M(x)]^{-1}u|}{|u|} \sup_{v\neq 0} \frac{|[M(y)]^{-1}v|}{|v|} L|x-y| \\
&= \sup_{\tilde{u}\neq 0} \frac{|[M(x)]^{-1}M(x)\tilde{u}|}{|M(x)\tilde{u}|} \sup_{\tilde{v}\neq 0} \frac{|[M(y)]^{-1}M(y)\tilde{v}|}{|M(y)\tilde{v}|} L|x-y| \\
&= \sup_{\tilde{u}\neq 0} \frac{|\tilde{u}|}{|M(x)\tilde{u}|} \sup_{\tilde{v}\neq 0} \frac{|\tilde{v}|}{|M(y)\tilde{v}|} L|x-y| \\
&\overset{3.3.7}{\leq} \sup_{\tilde{u}\neq 0} \frac{|\tilde{u}|}{\mu|\tilde{u}|} \sup_{\tilde{v}\neq 0} \frac{|\tilde{v}|}{\mu|\tilde{v}|} L|x-y| \\
&= \frac{L}{\mu^2}|x-y|.
\end{aligned}$$

□

Lemma 3.3.9 *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be strongly monotone. Let furthermore $B \in \mathbb{R}^{n\times k}$ be an arbitrary matrix and $A \in \mathbb{R}^{m\times n}$ be a matrix with full row rank. Then, the function*

$$f_A : \mathbb{R}^k \times \mathbb{R}^m \to \mathbb{R}^m, \quad (x, z) \mapsto Af(Bx + A^\top z)$$

is strongly monotone w.r.t. z.

Proof: Since $A^\top$ has full column rank, the function $|\cdot|_A$, defined by $|z|_A := |A^\top z|$ defines a norm for any vector norm $|\cdot|$. Strong monotonicity of f guarantees the existence of a $\mu > 0$ such that

$$[x_1 - x_2]^\top [f(x_1) - f(x_2)] \geq \mu |x_1 - x_2|^2 \; \forall x_1, x_2 \in \mathbb{R}^n.$$

Defining $y_i := Bx + A^\top z_i$ for arbitrary $x \in \mathbb{R}^k$ and considering the equivalence of norms in $\mathbb{R}^n$ yields the existence of $\tilde{\mu} > 0$ such that

$$\begin{aligned} [z_1 - z_2]^\top [f_A(x, z_1) - f_A(x, z_2)] &= [y_1 - y_2]^\top [f(y_1) - f(y_2)] \\ &\geq \mu |y_1 - y_2|^2 \\ &= \mu |z_1 - z_2|_A^2 \\ &\geq \tilde{\mu} |z_1 - z_2|^2 \end{aligned}$$

for all $z_1, z_2 \in \mathbb{R}^m$. □

Corollary 3.3.10 *Let $M : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ be uniformly positive definite and let A have full row rank. Then, $x \mapsto AM(x)A^\top$ is uniformly positive definite.*

Proof: Since $A^\top$ has full column rank, it holds $A^\top y \neq 0$ for any $y \neq 0$, and $|\cdot|_{A^\top} := |A^\top \cdot|$ defines a norm for any vector norm $|\cdot|$. Due to uniform positive definiteness of M, there exists $\tilde{\mu}, \mu > 0$ such that for any x, y it holds

$$y^\top AM(x)A^\top y \geq \tilde{\mu} |A^\top y|^2 = \tilde{\mu} |y|_{A^\top}^2 \geq \mu |y|^2,$$

where the last estimate holds due to equivalence of norms in $\mathbb{R}^n$. □

Lemma 3.3.11 *Let $M : \mathbb{R}^n \to \mathbb{R}^{m\times m}$ be a continuous function. Then, it holds*

(i) If M is uniformly positive definite, then $x \mapsto [M(x)]^{-1}$ is bounded.

If $M(x)$ is furthermore symmetric and positive definite for all $x \in \mathbb{R}^n$, it holds

(ii) M is uniformly positive definite if and only if $x \mapsto [M(x)]^{-1}$ is bounded.

Proof: (i): Let M be uniformly positive definite. By Lemma 3.3.7, there exists $\mu > 0$ for any norm $|\cdot|$ such that $|M(x)y| \geq \mu |y| \ \forall x, y$.

Furthermore, since $M(x)$ is nonsingular for all $x \in \mathbb{R}^n$, it holds

$$|M(x)^{-1}| = \sup_y \frac{|M(x)^{-1}y|}{|y|} = \sup_y \frac{|M(x)^{-1}(M(x)y)|}{|M(x)y|} = \sup_y \frac{|y|}{|M(x)y|} \leq \sup_y \frac{|y|}{\mu|y|} = \frac{1}{\mu}.$$

(ii): Since (i) proves the statement " $\Longrightarrow$ ", we only have to show " $\Longleftarrow$ ". We note that $M(x)$ is symmetric and positive definite if and only if $[M(x)]^{-1}$ is symmetric and positive definite. Furthermore, the spectra are reciprocals in that case, that is,

$$\sigma([M(x)]^{-1}) = \{\lambda_1(x), \cdots, \lambda_m(x)\}, \qquad \sigma(M(x)) = \{\frac{1}{\lambda_1(x)}, \ldots, \frac{1}{\lambda_m(x)}\},$$

where σ denotes the spectrum of a matrix. Boundedness of $[M(x)]^{-1}$ implies boundedness of its eigenvalues:

$$\exists C > 0 : \quad \lambda_{max}(x) \leq C \quad \forall x \in \mathbb{R}^n.$$

Denoting the eigenvalues of $M(x)$ by $\eta_i(x)$, it follows for the minimal eigenvalue

$$\eta_{min}(x) = \frac{1}{\lambda_{max}(x)} \geq C.$$

Since there exists a basis of eigenvectors, this implies

$$y^\top M(x) y \geq C y^\top y.$$

□

Lemma 3.3.12 *Let $D \subseteq \mathbb{R}^n$ be a compact set and the functions*

$$M : D \to \mathbb{R}^{n \times n}, \qquad G : D \to \mathbb{R}^{n \times m}$$

be Lipschitz continuous. Then, the functions K_1 and K_2, defined by

$$\begin{aligned} A_1 &: D \to \mathbb{R}^{n \times m}, & x &\mapsto M(x)G(x), \\ A_2 &: D \times D \to \mathbb{R}^{n \times m}, & (x, y) &\mapsto M(x)G(y) \end{aligned}$$

are Lipschitz continuous.

Proof: Let L_M and L_G be the Lipschitz constants of $x \mapsto M(x)$ and $x \mapsto G(x)$, and let $\max_{x \in D} |x| = C_D$. We obtain for arbitrary $x_1, x_2 \in D$, the vector norm $|\cdot|$ and induced matrix norm $|\cdot|_*$

$$\begin{aligned} |A_1(x_1) - A_1(x_2)|_* &= |M(x_1)G(x_1) - M(x_2)G(x_2)|_* \\ &= |\left(M(x_1) - M(x_2)\right)G(x_2) + M(x_1)(G(x_1) - G(x_2))|_* \\ &\leq L_M |G(x_2)|_* |x_1 - x_2| + L_G |M(x_1)|_* |x_1 - x_2| \\ &\leq (L_M C_G + L_G C_M)|x_1 - x_2|, \end{aligned}$$

where $C_G := \max_{x \in D} |G(x)|_*$ and $C_M := \max_{x \in D} |M(x)|_*$. Such bounds exist as G and M are continuous functions on a compact set. The inequality chain yields Lipschitz continuity of A_1.

For A_2, we obtain analogously

$$\begin{aligned} |A_2(x_1, y_1) - A_2(x_2, y_2)|_* &= |M(x_1)G(y_1) - M(x_2)G(y_2)|_* \\ &= |\left(M(x_1) - M(x_2)\right) G(y_2) + M(x_1)(G(y_1) - G(y_2))|_* \\ &\leq L_M C_G |x_1 - x_2| + L_G C_M |y_1 - y_2|. \end{aligned}$$

Since $|(a,b)|_+ := |a| + |b|$ defines a norm on $\mathbb{R}^n \times \mathbb{R}^n$, it follows for $L = \max\{L_M C_G, L_G C_M\}$ that

$$|A_2(x_1, y_1) - A_2(x_2, y_2)|_* \leq L|(x_1 - x_2, y_1 - y_2)|_+,$$

which yields the desired result due to equivalence of norms in $\mathbb{R}^n \times \mathbb{R}^n$. □

Lemma 3.3.13 *Let $f : [t_0, T] \subset \mathbb{R} \to \mathbb{R}^n$ be continuous. Then, it holds*

$$\left| \int_{t_0}^{T} f(t) dt \right| \leq \int_{t_0}^{T} |f(t)| \, dt.$$

A proof can be found in [OR70, Theorem 3.2.11.].

Lemma 3.3.14 (Gronwall Lemma) *Let $\mathcal{I} = [t_0, T]$ be a compact interval, $\beta > 0$ a constant and $\alpha, \Psi : \mathcal{I} \to \mathbb{R}$ continuous functions satisfying*

$$\Psi(t) \leq \alpha(t) + \beta \int_{t_0}^{t} \Psi(s) ds \quad \forall t \in \mathcal{I}.$$

If α is nondecreasing, that is, $\alpha(s) \leq \alpha(t)$ if $s \leq t$, then

$$\Psi(t) \leq \alpha(t) e^{\beta(t - t_0)} \quad \forall t \in \mathcal{I}.$$

Remark. Note that if $\alpha(t)$ is nonnegative for all $t \in \mathcal{I}$, it holds $\Psi(t) \leq c\alpha(t)$ for all $t \in \mathcal{I}$, where $c := e^{\beta(T - t_0)}$.

A proof of a slightly more general version of the Gronwall Lemma, considering a time-dependent β, can be found in [Tes12, Lemma 2.7].

3.4. Nonlinear decoupling

This section, in which we develop a decoupling of the nonlinear MNA Equations (2.11), is strongly based on the linear decoupling from Section 3.2.

Assumption 3.4.1 *The functions*

$$C : \mathbb{R}^n \to \mathbb{R}^{n\times n}, \qquad L : \mathbb{R}^m \to \mathbb{R}^{m\times m}, \qquad g : \mathbb{R}^k \to \mathbb{R}^k$$

from the nonlinear MNA (2.11) are continuous. Furthermore, $C(x)$ and $L(y)$ are positive definite for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, and g is strongly monotone.

For the nonlinear decoupling, we shall employ the same kernel matrices as defined in Tables (3.1) and (3.2) and the same decoupling Glossary (3.3), requiring only a few adaptations.

Nonlinear Extension of the linear MNA decoupling Glossary (3.3) The only necessary adaptation is that C and L are state-dependent (matrix-valued) functions now, and g is a function instead of a matrix. Hence, all the matrices defined in the Glossary (3.3) which involved L or C become matrix valued functions now. Matrices formerly involving G become vector-valued functions involving g now. We define

$$\begin{aligned}\mathcal{L}_c(\bar{e}_{12}) &:= \mathcal{A}_c C(\mathcal{A}_c^\top T_e \bar{e})\mathcal{A}_c^\top = \mathcal{A}_c C(\mathcal{A}_c^\top P_c \mathring{\mathcal{A}}_v \bar{e}_1 + \mathcal{A}_c^\top P_c \mathring{Q}_v \bar{e}_2)\mathcal{A}_c^\top \\ \bar{g}(\bar{e}_{1234}) &:= \mathcal{A}_r g(\mathcal{A}_r^\top T_e \bar{e}) = \mathcal{A}_r g(\mathcal{A}_r^\top T_e^- \bar{e}_{1234}) \\ \check{\mathcal{L}}_{r/cv}(\cdot) &:= P_r^\top Q_{cv}^\top \mathcal{A}_r g'(\cdot)\mathcal{A}_r^\top Q_{cv} P_r \\ M_{cv}(\bar{e}_{12}) &:= \mathring{\mathcal{A}}_v \check{\mathcal{L}}_c(\bar{e}_{12})^{-1} \mathring{\mathcal{A}}_v^\top \qquad , \end{aligned} \tag{3.31}$$

where g' denotes the Jacobian of g if $g \in C^1$.

Theorem 3.4.2 (Nonlinear MNA decoupling) *Consider the nonlinear MNA (2.11), and let the Assumptions 2.1.3, 3.2.3 and 3.4.1 be satisfied. Then, there exists a function Ψ_g such that $\bar{e}$, $\bar{i}_l$, $\bar{i}_v$ solve the DAE*

$$\bar{e}_2' = -\mathring{Q}_v^\dagger [\check{\mathcal{L}}_c(\bar{e}_{12})]^{-1} P_c^\top [\bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v T_v \bar{i}_v + \mathcal{A}_i i_s] \tag{3.32}$$

$$\bar{i}_{l_2}' = \breve{Q}_l^\dagger [L(T_l \bar{i}_l)]^{-1} \mathcal{A}_l^\top T_e \bar{e} \tag{3.33}$$

$$\bar{e}_5 = [\mathcal{L}_{l/cvr}(T_l \bar{i}_l)]^{-1} \mathcal{A}_{l/cvr} \left(\mathcal{A}_{l/cvr}^\top \bar{i}_{l_1}' - [L(T_l \bar{i}_l)]^{-1} \mathcal{A}_l^\top T_e^- \bar{e}_{1234}\right) \tag{3.34}$$

$$\bar{i}_{v_2} = -[M_{cv}(\bar{e}_{12})]^{-1} \mathring{\mathcal{A}}_v^\top \left(\mathring{\mathcal{A}}_v \bar{e}_1' + [\check{\mathcal{L}}_c(\bar{e}_{12})]^{-1} P_c^\top \left(\bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \bar{i}_{v_1} + \mathcal{A}_i i_s\right)\right) \tag{3.35}$$

$$\bar{e}_3 = (\check{\mathcal{A}}_{v/c}^\top)^\dagger \left(v_s - \mathcal{A}_v^\top P_c(\mathring{\mathcal{A}}_v \bar{e}_1 + \mathring{Q}_v \bar{e}_2)\right) \tag{3.36}$$

$$\bar{e}_4 = \Psi_g\left(\bar{e}_{123}, \bar{i}_l, i_s\right) \tag{3.37}$$

$$\bar{i}_{v_1} = -\check{\mathcal{L}}_{v/c}^{-1} P_v^\top Q_c^\top \left(\bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s\right) \tag{3.38}$$

$$\bar{e}_1 = (\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v)^{-1} \breve{Q}_v^\top v_s, \tag{3.39}$$

$$\bar{i}_{l_1} = -(\mathcal{A}_{l/cvr} \mathcal{A}_{l/cvr}^\top)^{-1} \mathcal{A}_{i/cvr} i_s. \tag{3.40}$$

if and only if $e = T_e\bar{e},\ i_l = T_l\bar{i}_l,\ i_v = T_v\bar{i}_v$ *solve the MNA* (2.11).

The function $\Psi_g : \mathbb{R}^n \to \mathbb{R}^m$ *is Lipschitz continuous. If the conductance function* g *from the MNA* (2.11) *is* C^1*, then for a solution of the DAE* (3.32)-(3.40) *it holds*

$$\frac{\partial \Psi_g}{\partial i_s}(\bar{e}_{123}, \bar{i}_l, i_s) = -\left[\check{\mathcal{L}}_{r/cv}(\xi)\right]^{-1} P_r^\top \mathcal{A}_{i/cv}, \tag{3.41}$$

where $\xi := \mathcal{A}_r^\top(\bar{T}_e^- \bar{e}_{123} + Q_{cv}P_r\Psi_g(\bar{e}_{123}, \bar{i}_l, i_s))$.

Proof: We closely follow the proof of the linear case. Analogously to Equations (3.13)-(3.15), we obtain

$$[\tilde{T}_e(\bar{e}_{12})]^\top [M_c(\bar{e}_{12})T_e\bar{e}' + \bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v T_v \bar{i}_v + \mathcal{A}_i i_s] = 0 \tag{3.42}$$

$$[\tilde{T}_l(T_l\bar{i}_l)]^\top \left[L(T_l\bar{i}_l)T_l\bar{i}_l' - \mathcal{A}_l^\top T_e\bar{e}\right] = 0 \tag{3.43}$$

$$T_v^\top \left[\mathcal{A}_v^\top T_e\bar{e} - v_s\right] = 0, \tag{3.44}$$

where the matrix valued functions $\tilde{T}_e(\cdot)$ and $T_l(\cdot)$ are defined by

$$\tilde{T}_e(\bar{e}_{12}) := \left(P_c[\check{\mathcal{L}}_c(\bar{e}_{12})^{-1}]^\top \mathring{\mathcal{A}}_v \quad P_c[\check{\mathcal{L}}_c(\bar{e}_{12})^{-1}]^\top \mathring{Q}_v \quad Q_cP_v \quad Q_{cv}P_r \quad Q_{cvr}\right),$$

$$\tilde{T}_l(T_l\bar{i}_l) := \left([L(i_l)^{-1}]^\top \mathcal{A}_{l/cvr}^\top \quad [L(i_l)^{-1}]^\top \check{Q}_l\right).$$

We notice that Assumption 3.4.1 and Lemma 3.3.3 imply that $C(y)$ and $L(x)$ are positive definite for all arguments. Hence we can employ the same reasoning like in the linear case for any fixed argument and we obtain nonsingularity of $\tilde{T}_e(\bar{e}_{12})$ and $\tilde{T}_l(T_l\bar{i}_l)$ for all arguments analogously.

Hence, $\bar{e}$, $\bar{i}_l$,$\bar{i}_v$ solve the transformed DAE (3.42)-(3.44) if and only if $e = T_e\bar{e},\ i_l = T_l\bar{i}_l,\ i_v = T_v\bar{i}_v$ solves the nonlinear MNA (2.11).

Writing the system by the blocks of $\tilde{T}_e(\cdot)$, $\tilde{T}_l(\cdot)$ and T_v yields a slightly modified system (3.16)-(3.24), where the modifications consist in replacing

1. the constant matrices $\check{\mathcal{L}}_c$, $\mathcal{L}_{l/cvr}$ and L by the state-dependent matrices $\check{\mathcal{L}}_c(\bar{e}_{12})$, $\mathcal{L}_{l/cvr}(T_l\bar{i}_l)$ and $L(T_l\bar{i}_l)$,
2. the term $\mathcal{L}_r T_e^- \bar{e}_{1234}$ by $\bar{g}(\bar{e}_{1234})$,
3. Equation (3.19) by

$$P_r^\top Q_{cv}^\top(\bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s) = 0. \tag{3.45}$$

Recalling that

$$P_r^\top Q_{cv}^\top \bar{g}(\bar{e}_{1234}) = P_r^\top Q_{cv}^\top \mathcal{A}_r g(\mathcal{A}_r^\top \bar{T}_e^- \bar{e}_{123} + \mathcal{A}_r^\top Q_{cv}P_r\bar{e}_4) =: \hat{g}(\bar{e}_{1234}),$$

we notice that $\hat{g}$ is strongly monotone in $\bar{e}_4$ due to Lemma 3.3.9 considering that g is strongly monotone due to Assumption 3.4.1 and Lemma 3.3.3. Then, Lemma 3.3.6 implies that we can solve Equation (3.45) for $\bar{e}_4$ yielding (3.37). Insertion of the solution function $\Psi_g(\cdot) = \bar{e}_4$ of (3.37) into (3.45) and differentiating by i_s yields

$$\begin{aligned} 0 =& \frac{d}{di_s} P_r^\top Q_{cv}^\top \left(\mathcal{A}_r^\top g(\mathcal{A}_r^\top \bar{T}_e^- \bar{e}_{123} + \mathcal{A}_r^\top Q_{cv} P_r \Psi_g\left(\bar{e}_{123}, \bar{i}_l, i_s\right)) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_i i_s \right) \\ =& \check{\mathcal{L}}_{r/cv} \left(\mathcal{A}_r^\top (\bar{T}_e^- \bar{e}_{123} + Q_{cv} P_r \Psi_g\left(\bar{e}_{123}, \bar{i}_l, i_s\right)) \right) \left(\frac{\partial \Psi_g}{\partial i_s} (\bar{e}_{123}, \bar{i}_l, i_s) \right) + P_r^\top Q_{cv}^\top \mathcal{A}_i \end{aligned}$$

which can be solved for the desired derivative of Ψ_g.

The remaining modified equations are equivalent to the according Equations (3.32)-(3.36), (3.38)-(3.40) by the same arguments like in the linear case if we consider that $C(y)$ and $L(x)$ are positive definite and hence nonsingular for all arguments. □

Assumption 3.4.3 *We consider the functions*

$$L : \mathbb{R}^n \to \mathbb{R}^{n \times n}, \qquad C : \mathbb{R}^m \to \mathbb{R}^{m \times m}, \qquad g : \mathbb{R}^k \to \mathbb{R}^k$$

from the nonlinear MNA (2.11). It holds

(i) L and C are locally Lipschitz continuous, and g is Lipschitz continuous.

(ii) L and C are uniformly positive definite, and g is strongly monotone.

Furthermore, for the functions defined by

$$\begin{aligned} L^{-1} : \mathbb{R}^n &\to \mathbb{R}^{n \times n}, & \check{\mathcal{L}}_c^{-1} : \mathbb{R}^m &\to \mathbb{R}^{m \times m} \\ x &\mapsto [L(x)]^{-1}, & z &\mapsto [P_c^\top \mathcal{A}_c C(\mathcal{A}_c^\top P_c z) \mathcal{A}_c^\top P_c]^{-1}, \end{aligned}$$

it holds

(iii) L^{-1} and $\check{\mathcal{L}}_c^{-1}$ are uniformly positive definite.

(iv) L^{-1} and $\check{\mathcal{L}}_c^{-1}$ are locally Lipschitz continuous.

Lemma 3.4.4 *Let the matrix-valued functions L and C from the nonlinear MNA (2.11) be continuous, bounded and furthermore symmetric and positive definite for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Then,*

- *Assumption 3.4.3(iii) holds.*
- *Assumptions 3.4.3(i) and 3.4.3(ii) imply 3.4.3(iv).*

Proof: Recalling that $\mathcal{A}_c^\top P_c$ has full column rank by construction of P_c, we notice that $P_c^\top \mathcal{A}_c C(P_c^\top \mathcal{A}_c z)\mathcal{A}_c^\top P_c$ inherits boundedness, symmetry and positive definiteness from $C(y)$. Hence, the preconditions for L and C yield uniform positive definiteness of L^{-1} and $\check{\mathcal{L}}_c^{-1}$ by Lemma 3.3.11(ii). Furthermore, uniform positive definiteness combined with local Lipschitz continuity of L and C yield local Lipschitz continuity of L^{-1} and $\check{\mathcal{L}}_c^{-1}$ with Lemma 3.3.8. □

Theorem 3.4.5 *Let Assumptions 2.1.3, 3.2.3 and 3.4.3 hold. Then, there exists a nonsingular matrix $T = (T_0\ T_1\ T_2\ T_3)$ and a DAE of the fully decoupled form*

$$\dot{y} = f(y, s, G_3 s') \tag{3.46a}$$
$$z_1 = g_1(y, s, G_3 s') \tag{3.46b}$$
$$z_2 = g_2(y, s) \tag{3.46c}$$
$$z_3 = M_3 G_3 s \tag{3.46d}$$

such that y, z_1, z_2, z_3 solve this DAE if and only if $x = \begin{pmatrix} e \\ i_l \\ i_v \end{pmatrix} = T_0 y + T_1 z_1 + T_2 z_2 + T_3 z_3$ solves the nonlinear MNA (2.11) with $s = \begin{pmatrix} i_s \\ v_s \end{pmatrix}$.

Furthermore, the DAE (3.46) has the following properties:

(i) It holds

$$G_3 = \begin{pmatrix} -\mathcal{A}_{i/cvr} & 0 \\ 0 & \check{Q}_v^\top \end{pmatrix}.$$

The matrices $\mathcal{A}_{i/cvr} := Q_{cvr}^\top \mathcal{A}_i$, Q_{cvr} and $\check{Q}_v$ are defined in Tables (3.1)-(3.2).

(ii) The right hand side of (3.46) is locally Lipschitz continuous, that is,

$$f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^n, \qquad g_1 : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^p, \qquad g_2 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$$

are locally Lipschitz continuous.

(iii) If a solution exists on a time interval $[t_0, T]$, then y satisfies an estimate of the form

$$\|y\|_{[t_0,t]} \le c\left(|y(t_0)| + H_t\|G_3 s'\|_{[t_0,t]} + H_t\|\theta(s)\|_{[t_0,t]}\right) e^{\beta(t-t_0)} \quad \forall t \in [t_0, T],$$

where $c > 0$ and $\beta > 0$ are constants, $H_t := t - t_0$, θ is a continuous function and $\|\cdot\|_{[t_0,t]}$ is the maximum norm on $C([t_0, t], \mathbb{R}^n)$. The constants c and β are independent of s.

(iv) For any compact time interval $[t_0, T] \subseteq \mathbb{R}$, any continuously differentiable input s and any initial value $y_0 = y(t_0) \in \mathbb{R}^n$, the corresponding IVP (3.46) is globally uniquely solvable.

Proof: The proof is strongly based on the Theorem 3.4.2 and the notation therein. The fully decoupled normal form (3.46) is obtained by collecting the variables of Equations (3.32)-(3.40) in the vectors

$$y := \begin{pmatrix} \bar{e}_2 \\ \bar{i}_{l_2} \end{pmatrix}, \quad z_1 := \begin{pmatrix} \bar{e}_5 \\ \bar{i}_{v_2} \end{pmatrix}, \quad z_2 := \begin{pmatrix} \bar{e}_3 \\ \bar{e}_4 \\ \bar{i}_{v_1} \end{pmatrix}, \quad z_3 := \begin{pmatrix} \bar{e}_1 \\ \bar{i}_{l_1} \end{pmatrix}$$

and by appropriate elimination by insertion of the respective variables in the right hand sides which is described in greater detail in the following. Clearly, as y, z_1, z_2, z_3 are only a certain resorting of the variables $\bar{e}, \bar{i}_l, \bar{i}_v$, it holds

$$x = \begin{pmatrix} \bar{e} \\ i_l \\ i_v \end{pmatrix} = \underbrace{\begin{pmatrix} T_e & 0 & 0 \\ 0 & T_l & 0 \\ 0 & 0 & T_v \end{pmatrix}}_{=;K} \begin{pmatrix} \bar{e} \\ \bar{i}_l \\ \bar{i}_v \end{pmatrix} = T \begin{pmatrix} y \\ z_1 \\ z_2 \\ z_3 \end{pmatrix}$$

where T is a permutation of K according to the resorting of the variables.

In the following, we first show from bottom to top that how the fully decoupled normal form (3.46a)-(3.46d) is obtained from the “half-way” decoupled normal form (3.32)-(3.40), along with a proof of Lipschitz continuity (on bounded sets) of each equation. After that, we prove the unique solvability of the system of equations.

Equation (3.46d) follows readily from Equations (3.39)-(3.40) with

$$M_3 = \begin{pmatrix} 0 & (\mathcal{A}_{l/cvr}\mathcal{A}_{l/cvr}^\top)^{-1} \\ (\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v)^{-1} & 0 \end{pmatrix}$$

and G_3 as desired.

Equation (3.46c): We define for shortness

$$\begin{aligned} \varphi_{\bar{i}_{l_1}}(i_s) &:= -(\mathcal{A}_{l/cvr}\mathcal{A}_{l/cvr}^\top)^{-1}\mathcal{A}_{i/cvr} i_s, \\ \varphi_{\bar{e}_1}(v_s) &:= (\mathring{\mathcal{A}}_v^\top \mathring{\mathcal{A}}_v)^{-1} \breve{Q}_v^\top v_s \end{aligned}$$

which are the right hand sides of Equations (3.39)-(3.40).

Insertion into Equations (3.36)-(3.38) yields

$$\begin{aligned} \bar{e}_3 &= (\breve{\mathcal{A}}_{v/c}^\top)^\dagger \Big(v_s - \mathcal{A}_v^\top P_c(\mathring{\mathcal{A}}_v \varphi_{\bar{e}_1}(v_s) + \mathring{Q}_v \bar{e}_2)\Big) =: \varphi_{\bar{e}_3}(y, v_s) \\ \bar{e}_4 &= \Psi_g(\varphi_{\bar{e}_1}(v_s), \bar{e}_2, \varphi_{\bar{e}_3}(v_s, y), \varphi_{\bar{i}_{l_1}}(i_s), \bar{i}_{l_2}, i_s) =: \varphi_{\bar{e}_4}(y, s) \\ \bar{i}_{v_1} &= -\breve{\mathcal{L}}_{v/c}^{-1} P_v^\top Q_c^\top \Big(\bar{g}(\varphi_{\bar{e}_1}(v_s), \bar{e}_2, \varphi_{\bar{e}_3}(y, v_s), \varphi_{\bar{e}_4}(y, s)) + \mathcal{A}_l(\mathcal{A}_{l/cvr}^\top \varphi_{\bar{e}_1}(v_s) + \breve{Q}_l \bar{i}_{l_2}) + \mathcal{A}_i i_s\Big) \\ &=: \varphi_{\bar{i}_{v_1}}(y, s), \end{aligned}$$

which is a system of the form (3.46c). This system is locally Lipschitz continuous as a composition of the locally Lipschitz continuous functions Ψ_g and $\bar{g}$ and linear functions.

Equation (3.46b) is obtained from Equations (3.34)-(3.35) after adequate variable insertions, which we will perform in greater detail in the following.

We first define

$$
\begin{aligned}
T_l \bar{i}_l &= \mathcal{A}_{l/cvr}^\top \varphi_{\bar{i}_{l_1}}(i_s) + \begin{pmatrix} 0 & \check{Q}_l \end{pmatrix} y =: \varphi_1(y, i_s) \\
\mathcal{A}_c^\top T_e \bar{e} &= \mathcal{A}_c^\top P_c(\mathring{\mathcal{A}}_v \varphi_{\bar{e}_1}(v_s) + \begin{pmatrix} \mathring{Q}_v & 0 \end{pmatrix}) y =: \varphi_2(y, v_s) \\
\mathcal{A}_l^\top T_e^- \bar{e}_{1234} &= \mathcal{A}_l^\top \Big(P_c \mathring{\mathcal{A}}_v \varphi_{\bar{e}_1}(v_s) + P_c \mathring{Q}_v \bar{e}_2 + Q_c P_v \varphi_{\bar{e}_3}(y, v_s) + Q_{cv} P_r \varphi_{\bar{e}_4}(y, s) \Big) =: \varphi_3(y, s)
\end{aligned}
$$

and

$$
\begin{aligned}
&P_c^\top (\bar{g}(\bar{e}_{1234}) + \mathcal{A}_l T_l \bar{i}_l + \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \bar{i}_{v1} - \mathcal{A}_i i_s) \\
=& P_c^\top (\bar{g}(\varphi_{\bar{e}_1}(v_s), \bar{e}_2, \varphi_{\bar{e}_3}(y, v_s), \varphi_{\bar{e}_4}(y, s)) + \varphi_1(y, i_s) + \mathcal{A}_v \check{\mathcal{A}}_{v/c}^\top \varphi_{\bar{i}_{v_1}}(y, s) - \mathcal{A}_i i_s) =: \varphi_4(y, s)
\end{aligned}
$$

with linear functions φ_1, φ_2, φ_3 and the locally Lipschitz continuous function $\varphi_4 : \mathbb{R}^n \to \mathbb{R}^k$, which is composed of linear functions and the locally Lipschitz continuous functions $\bar{g}$ and Ψ_g. Now we can rewrite Equations (3.34)-(3.35) as

$$
z_1 = \begin{pmatrix} \bar{e}_5 \\ \bar{i}_{v_2} \end{pmatrix} = K_1(y, s) G_3 s' - K_1(y, s) K_2(y, s) \begin{pmatrix} \varphi_3(y, s) \\ \varphi_4(y, s) \end{pmatrix} =: g_1(y, s, G_3 s'), \tag{3.47}
$$

with

$$
K_1(y, s) := \begin{pmatrix} [\mathcal{L}_{l/cvr}(\varphi_1(y, i_s))]^{-1} & 0 \\ 0 & -[\mathring{\mathcal{A}}_v^\top [\check{\mathcal{L}}_c(\varphi_2(y, v_s))]^{-1} \mathring{\mathcal{A}}_v]^{-1} \end{pmatrix}, \tag{3.48}
$$

$$
K_2(y, s) := \begin{pmatrix} \mathcal{A}_{l/cvr} [L(\varphi_1(y, i_s))]^{-1} & 0 \\ 0 & -\mathring{\mathcal{A}}_v^\top [\check{\mathcal{L}}_c(\varphi_2(y, v_s))]^{-1} \end{pmatrix}, \tag{3.49}
$$

which yields an equation of the form (3.46b).

K_1 and K_2 are bounded, since their diagonal blocks are bounded: L and C are uniformly positive definite by Assumption 3.4.3. Hence, Corollary 3.3.10 yields uniform positive definiteness of $\check{\mathcal{L}}_c$, and Lemma 3.3.11 implies that L^{-1} and $\check{\mathcal{L}}_c^{-1}$ are bounded. Analogously, Assumption 3.4.3 guarantees uniform positive definiteness of L^{-1} and $\check{\mathcal{L}}_c^{-1}$. Then, Corollary 3.3.10 implies that $\mathcal{L}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top [\check{\mathcal{L}}_c(\cdot)]^{-1} \mathring{\mathcal{A}}_v$ are uniformly positive definite, and their inverses are hence bounded with Lemma 3.3.11.

Furthermore, K_1 and K_2 are locally Lipschitz continuous on $\mathbb{R}^n$ (with n of appropriate size), since their diagonal blocks are locally Lipschitz continuous:

For K_2 this follows from Assumption 3.4.3(iv) and linearity of φ_1 and φ_2. Regarding K_1, Assumption 3.4.3 yields uniform positive definiteness of L^{-1}, and $\check{\mathcal{L}}_c^{-1}$, which implies with Corollary 3.3.10 that $\mathcal{L}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v$ is uniformly positive definite. Furthermore, by definition $\mathcal{L}_{l/cvr}$ and $\mathring{\mathcal{A}}_v^\top \check{\mathcal{L}}_c^{-1} \mathring{\mathcal{A}}_v$ clearly inherit local Lipschitz continuity from

L^{-1} and $\check{\mathcal{L}}_c^{-1}$, respectively. From there, Lemma 3.3.8 yields local Lipschitz continuity of both diagonal blocks of K_1.

Altogether, we obtain that the right hand side function g_1 (defined in (3.47)) is locally Lipschitz continuous. This follows from the (global) Lipschitz continuity of φ_3, φ_4, (which contain the Lipschitz continuous nonlinearities g and Ψ_g by definition) and of K_1 and K_2. Lemma 3.3.12 then yields local Lipschitz continuity of g_1.

Equation (3.46a) is obtained from Equations (3.32)-(3.33):

$$
\begin{aligned}
y' &= \begin{pmatrix} \bar{e}_2 \\ \bar{i}_{l_2} \end{pmatrix}' && (3.50)\\
&= \begin{pmatrix} \mathring{Q}_v^\dagger [\check{\mathcal{L}}_c(\varphi_2(y,v_s))]^{-1} & 0 \\ 0 & \check{Q}_l^\dagger [L(\varphi_1(y,i_s))]^{-1} \end{pmatrix} \left[\begin{pmatrix} \varphi_4(y,s) \\ \varphi_3(y,s) \end{pmatrix} + \begin{pmatrix} 0 & \mathring{\mathcal{A}}_v \\ \mathcal{A}_{l/cvr}^\top & 0 \end{pmatrix} z_1 \right] \\
&= K_3(y,s) \begin{pmatrix} \varphi_3(y,s) \\ \varphi_4(y,s) \end{pmatrix} + K_3(y,s) B \left[K_1(y,s) G_3 s' - K_1(y,s) K_2(y,s) \begin{pmatrix} \varphi_3(y,s) \\ \varphi_4(y,s) \end{pmatrix} \right] \\
&= K_4(\cdot) G_3 s' + K_5(\cdot) \begin{pmatrix} \varphi_3(\cdot) \\ \varphi_4(\cdot) \end{pmatrix} =: f(y,s,G_3 s') && (3.51)
\end{aligned}
$$

with

$$
K_3(y,s) := \begin{pmatrix} 0 & \mathring{Q}_v^\dagger [\check{\mathcal{L}}_c(\varphi_2(y,v_s))]^{-1} \\ \check{Q}_l^\dagger [L(\varphi_1(y,i_s))]^{-1} & 0 \end{pmatrix}, \qquad B := \begin{pmatrix} \mathcal{A}_{l/cvr}^\top & 0 \\ 0 & \mathring{\mathcal{A}}_v \end{pmatrix},
$$

and $K_4(\cdot) := K_3(\cdot) B K_1(\cdot)$ and $K_5(\cdot) := K_3(\cdot) - K_3(\cdot) B K_1(\cdot) K_2(\cdot)$. The function f is locally Lipschitz continuous since K_1, K_2 and the blocks of K_3 are locally Lipschitz continuous as shown above when we derived Equation (3.46b). Repeated application of Lemma 3.3.12 then yields local Lipschitz continuity of f.

(iii) and *(iv)*: The system (3.46a)-(3.46d) is globally uniquely solvable on $[t_0, T]$ for a given initial value y_0 if and only if the inherent ODE (3.46a) is globally uniquely solvable, which we shall show by means of an a-priori estimate and continuation of a local solution.

We first observe that for $y \in \mathbb{R}^n, s \in \mathbb{R}^m$ arbitrary, it holds for $\varphi_{34} := \left(\begin{smallmatrix} \varphi_3 \\ \varphi_4 \end{smallmatrix}\right)$

$$
|\varphi_{34}(y,s)| - |\varphi_{34}(0,s)| \le |\varphi_{34}(y,s) - \varphi_{34}(0,s)| \le L_{\varphi_{34}} |y|.
$$

due to Lipschitz continuity of φ_{34}. Note that $L_{\varphi_{34}}$ is independent of y and s since φ_{34} is (globally) Lipschitz continuous.

Considering furthermore the boundedness of $K_i(\cdot), i \in \{1,2,3\}$ and building the integral formulation of Equation (3.46a) yields for all $t \in [t_0, T]$

$$\begin{aligned}
|y(t)| &= \left| y(t_0) + \int_{t_0}^{t} K_3(\cdot)BK_1(\cdot)G_3 s'(\tau) - K_3(\cdot)\left(I - BK_1(\cdot)K_2(\cdot)\right)\varphi_{34}(\cdot) d\tau \right| \\
&\leq |y(t_0)| + \int_{t_0}^{t} |K_3(\cdot)BK_1(\cdot)G_3 s'(\tau)|\, d\tau + \int_{t_0}^{t} |K_3(\cdot)\left(I - BK_1(\cdot)K_2(\cdot)\right)\varphi_{34}(\cdot)|\, d\tau \\
&\leq |y(t_0)| + C_1 \int_{t_0}^{t} |G_3 s'(\tau)|\, d\tau + C_2 \int_{t_0}^{t} |\varphi_{34}(\cdot)|\, d\tau \\
&\leq |y(t_0)| + C_1 H_t \|G_3 s'\|_{[t_0,t]} + C_2 \int_{t_0}^{t} |\varphi_{34}(0, s(\tau)|\, d\tau + C_2 \int_{t_0}^{t} L_{\varphi_{34}} |y(\tau)| d\tau \\
&\leq \underbrace{|y(t_0)| + C_1 H_t \|G_3 s'\|_{[t_0,t]} + C_2 H_t \|\varphi_{34}(0, s)\|_{[t_0,t]}}_{=:\alpha(t)} + \underbrace{C_2 L_{\varphi_{34}}}_{=:\beta} \int_{t_0}^{t} |y(\tau)| d\tau
\end{aligned}$$

where $(\cdot) := (y(\tau), s(\tau))$ and $H_t := t - t_0$ for shortness. Furthermore, we defined

$$C_1 := M_3 |B|_* M_1, \quad C_2 := M_3 + M_3 |B|_* M_1 M_2, \quad M_i := \max_{x \in \mathbb{R}^n} |K_i(x)|_*$$

for the induced matrix norm $|\cdot|_*$.

At this point, we can apply the Gronwall Lemma 3.3.14 which reveals the following a-priori estimate for $t \in [t_0, T]$: *If y exists on $[t_0, T]$, then it satisfies the bound*

$$\|y\|_{[t_0,t]} \leq \alpha(t) e^{\beta(t - t_0)} \quad \forall t \in [t_0, T].$$

This confirms (iii).

Knowing that we have a local unique solution due to local Lipschitz continuity of the vector field f, this solution can furthermore be continuated within a Lipschitz rectangle by the Theorem of Picard-Lindelöf, and the a-priori estimate combined with the fact that f is Lipschitz continuous on arbitrary bounded sets guarantee that the solution does not reach the boundaries of such a rectangle until we reach the end T of the time interval $[t_0, T]$. Hence, the local unique solution is indeed a global one on $[t_0, T]$, which proves (iv). □

Corollary 3.4.6 *Let Assumptions 2.1.3, 3.2.3 and 3.4.3 hold. Then, there exists a full rank matrix $A \in \mathbb{R}^{n \times m}$, $n < m$, such that there exists a global unique solution $x \in C([t_0, T], \mathbb{R}^m)$ of the nonlinear MNA (2.11) for any bounded time interval $[t_0, T]$ and any initial value $Ax(t_0) =: x_0^{IV} \in \mathbb{R}^n$.*

Proof: By Theorem 3.4.5, the decoupled system (3.46) has a global unique solution for all initial values $y(t_0) \in \mathbb{R}^n$, and furthermore, if $\begin{pmatrix} y^\top & z_1^\top & z_2^\top & z_3^\top \end{pmatrix}^\top$ is the solution of (3.46), then

$$x = T \begin{pmatrix} y^\top & z_1^\top & z_2^\top & z_3^\top \end{pmatrix}^\top$$

solves the MNA (2.11). Since T is nonsingular, we get

$$T^{-1}x = \begin{pmatrix} y^\top & z_1^\top & z_2^\top & z_3^\top \end{pmatrix}^\top.$$

and

$$y(t_0) = Ax(t_0),$$

where A consists of the first n_y rows of T^{-1}. □

Discussion

Here we discuss Theorems 3.4.2 and 3.4.5 and Corollary 3.4.6 and put them into context, as they will be the fundamental basis of the perturbation results in the following section.

In their article [ET00], Tischendorf and Estevez Schwarz presented a DAE normal form for nonlinear MNA equations of index two (and smaller) along with an algorithm of how to achieve it, using the concept of the tractability index [GM86,LMT13]. Essentially, the idea is to transform a DAE by means of projectors onto and along (constant) kernels of the matrix-valued functions describing the DAE. The Application of this idea specifically and in great detail to the MNA model in [ET00] was the basis for several numerical and analytical results regarding the MNA. A widely noticed result offered in the same article was the topological index result, which allowed to compute the index of the MNA only by means of topological criteria, namely the existence of cutsets and loops of certain subcircuits. The same framework was exploited in [Tis04] to prove a local existence and uniqueness results for (solutions of) nonlinear index two MNA DAEs.

For MNA DAEs of index one, [Mat12, JMT15] then obtained global solvability and perturbation index results, which were then generalized to higher index circuit DAEs by [Jan15]. To achieve global existence and uniqueness results, in these works the assumptions are strengthened so that the elements are globally passive instead of only locally, which mathematically translates to uniformly positive definite matrix-valued functions L and C instead of only positive definiteness for all arguments. Additionally, the Lipschitz condition on g is strengthened from local to global.

The DAE decoupling approach in [Jan15] is slightly different from the one in [ET00, Tis04], but clearly inspired by it. Instead of transforming the MNA by means of (square) projectors, the transformation matrices in [Jan15] are rectangle and of full rank. An advantage of this approach is that it leads to a normal form of the same size as the original MNA, while the projector approach inflates the system.

The normal form in the decoupling Theorem 3.4.2 in the present work is based on such rectangle transformations P and Q as introduced in [Jan15], $\{P, Q\}$ being named *kernel splitting pairs*. A related but less detailed normal form was given in [PT19]. The one presented here provides a very high level of detail, which allows us to derive new topological results concerning perturbation sensitivity in the following section, where we relate a circuit's sensitivity to its input with the topologial position where the input enters the circuit. Furthermore, the observations of section 2.3.3 allow us to topologically interprete certain matrix products which systematically arise in the transformation to the DAE normal form: Namely, products of the form $Q_\bullet^\top \mathcal{A}_\star$ as incidence matrices and $Q_\bullet^\top \mathcal{A}_\star \mathcal{A}_\star^\top Q_\bullet$ as Laplacian matrices (of certain contracted graphs), where $\mathcal{A}_\star$ is an incidence matrix of a $\star$-subcircuit and $K_\bullet$ belongs to a kernel splitting pair of a $\bullet$-subcircuit and is chosen as the $(0,1)$ node component matrix as defined in 2.3.15. Our general notation $\mathcal{A}_{\star/\bullet}$ and $\mathcal{L}_{\star/\bullet}$ should indicate this possibility of interpretation. Theorem 3.5.5 in the following section provides a first result exploiting this insight.

However, our decoupling admits generalizations in the MNA model as required for instance for transistors or controlled sources [ET00, Bod07]. In fact, being formulated on the basis of rank assumptions, it does not even require that the matrices $\mathcal{A}_\star$ are incidence matrices. Roughly speaking, the construction of a normal form for Theorem 3.4.2 works for any system satisfying the rank (and other) Assumptions therein, but the benefits of a topological interpretation, presented in the next section, can be exploited only if $\mathcal{A}_\star$ are incidence matrices.

Based on the decoupling Theorem 3.4.2, we derive Theorem 3.4.5. This theorem offers various valuable results:

- A more compact representation of a normal form, however giving detail on the index two impact determined by G_3. The right hand side of the normal form is Lipschitz continuous.
- An a-priori estimate for the solution.
- A global existence and uniqueness result (on compact time intervals).

The a-priori estimate and existence and uniqueness (in a slightly generalized model) were already shown in [Jan15, Theorem 5.20] using a different approach of proof. (Local) Lipschitz continuity of the right hand side of the normal form is a property we introduced along with the necessary strengthening of the assumptions, which we want to discuss in the following. Apart from the standard MNA Assumptions 2.1.3 and the index two Assumption 3.2.3, Theorem 3.4.5 requires Assumption 3.4.3.

Local Lipschitz continuity of C, L, L^{-1} and $\check{\mathcal{L}}_c^{-1}$ as required in Conditions 3.4.3(i),(iv) is not needed to obtain the existence and uniqueness result. This can be seen from the proof, and is in accordance with [Jan15], where it is also not required. However, it is necessary in order to obtain a (locally) Lipschitz continuous right hand side, which shall allow us to derive the perturbation results in the following section.

Condition 3.4.3(ii), assuming uniform positive definiteness of C and L and strong monotonicity of g, is equivalent to Assumptions made in [Mat12, Jan15] for global existence and uniqueness results.

Condition 3.4.3(iii) requires uniform positive definiteness of L^{-1} and $\check{\mathcal{L}}_c^{-1}$ and implies boundedness of L and $\check{\mathcal{L}}_c$. Lemma 3.4.4 shows that for symmetric and positive definite matrices L and C, we can equivalently replace Condition 3.4.3(iii) by boundedness of L and C. Condition 3.4.3(iii) is used in the proof precisely to warrant boundedness of K_1 defined in Equation 3.48 and therewith boundedness of the products involving K_1 in Equation (3.51) which finally yields an a-priori bound for solutions. The proof reveals that the boundedness of these products is the necessary and sufficient condition for that. As the mentioned products are very lengthy, we required the sufficient Condition 3.4.3(iii) for reasons of a more convenient representation.

We remark, however, that no comparable condition is required in [Jan15, Theorem 5.20]. Therefore, it seems worth investigating if Condition 3.4.3(ii) implies boundedness of the prdocuts in Equation (3.51) so that Condition 3.4.3(iii) can be dropped.

The existence of a global and unique solution of the (untransformed) MNA DAE, given in Corollary 3.4.6, is a direct consequence of Theorem 3.4.5. Its main purpose is to provide a formal representation of consistent initial values for the MNA through the matrix A given in the corollary. It is of theoretical nature, since such a matrix A, or consistent initial values for higher-index DAEs in general, are usually hard to find in practice. We refer to [Est00, Bau12] for more detail on consistent initialization.

3.5. Transient perturbation analysis

In this subsection we examine sensitivity of the MNA equations with respect to perturbations of the source terms. We provide results which resemble the defining estimate of the perturbation index 2.2.4. However, the main difference is that we admit only certain *structured* perturbations, that is, perturbations of current and voltage sources.

We denote the vector of current source perturbations by i_δ, and instead of i_s, the perturbed system has the current input $i_s + P_{i_\delta} i_\delta$. Here, P_{i_δ} is the matrix which arises from the identity matrix $I_n \in \mathbb{R}^{n\times n}$ where each column represents one current source (and hence n is the number of current sources in the circuit) when columns corresponding to current sources which are *not* perturbed are deleted.

That way, if $1 \leq m \leq n$ current sources are perturbed, the vector i_δ has m components, and the vector's i-th component $(P_{i_\delta} i_\delta)_i \neq 0$ if and only if the i-th source is perturbed.

For voltage sources, P_{v_δ} and v_δ are defined analogously.

Writing $x = \begin{pmatrix} e \\ i_l \\ i_v \end{pmatrix}$, we shall compare (solutions of) the perturbed IVP

$$\mathcal{A}_c C(\mathcal{A}_c^\top e^\delta)\mathcal{A}_c^\top (e^\delta)' + \mathcal{A}_r g(\mathcal{A}_r^\top e^\delta) + \mathcal{A}_l i_l^\delta + \mathcal{A}_v i_v^\delta + \mathcal{A}_i i_s = -\mathcal{A}_i P_{i_\delta} i_\delta \tag{3.52a}$$

$$-L(i_l^\delta)(i_l^\delta)' + \mathcal{A}_l^\top e^\delta = 0 \tag{3.52b}$$

$$\mathcal{A}_v^\top e^\delta - v_s = P_{v_\delta} v_\delta \tag{3.52c}$$

$$A x^\delta(t_0) = (x_0^{IV})^\delta, \tag{3.52d}$$

to the unperturbed IVP of the nonlinear MNA (2.11), that is,

$$\mathcal{A}_c C(\mathcal{A}_c^\top e)\mathcal{A}_c^\top e' + \mathcal{A}_r g(\mathcal{A}_r^\top e) + \mathcal{A}_l i_l + \mathcal{A}_v i_v + \mathcal{A}_i i_s = 0 \quad (3.53a)$$
$$-L(i_l)i_l' + \mathcal{A}_l^\top e = 0 \quad (3.53b)$$
$$\mathcal{A}_v^\top e - v_s = 0 \quad (3.53c)$$
$$Ax(t_0) = x_0^{IV}, \quad (3.53d)$$

where A is a matrix as in Corollary 3.4.6 such that the IVPs have a unique solution on bounded time intervals (provided the assumptions of the corollary hold).

We shall examine the sensitivity of solutions of the MNA with respect to such perturbations. More precisely, we present network topological results answering the following questions:

- $|x - x^\delta|$: How can we express this difference in terms of (derivatives of) the source term perturbations $\delta = \left(\begin{smallmatrix} i_\delta \\ v_\delta \end{smallmatrix}\right)$?
- $|v_{ij} - v_{ij}^\delta|$: Considering only current source perturbations $\delta = i_\delta$, how can we express this difference in terms of (derivatives of) i_δ? Here, v_{ij} is the vector of voltages over perturbed current sources.

Remark. In the previous two sections, we required only that the matrices $\mathcal{A}_\star$ satisfy certain rank conditions. Here, however, we shall assume throughout that they are incidence matrices in order to exploit the circuit's topology.

For a given initial point t_0 and a norm $|\cdot|$, we define the balls

$$B^1(T,M) := \left\{ \delta \in C^1([t_0,T],\mathbb{R}^n) : \max_{t\in[t_0,T]} |\delta(t)| \le M \quad \wedge \quad \max_{t\in[t_0,T]} |\delta'(t)| \le M \right\} \quad (3.54)$$
$$B(T,M) := \left\{ \delta \in C([t_0,T],\mathbb{R}^n) : \max_{t\in[t_0,T]} |\delta(t)| \le M \right\} \quad (3.55)$$

Theorem 3.5.1 (Sensitivity to source perturbation) *Let $|\cdot|$ be a vector norm, $B^1(T,M)$ and $B(T,M)$ be defined as in (3.54) and (3.55), and let the MNA Assumptions 2.1.3 and 3.4.3 hold. We consider the IVPs (3.52) and (3.53), which represent the MNA of a circuit, and their solutions x^δ and x. The following two estimates hold:*

1. *For all $M > 0$ and all $T > 0$, there exists $c_2 > 0$ such that for all $\delta = \left(\begin{smallmatrix} i_\delta \\ v_\delta \end{smallmatrix}\right) \in B^1(T,M)$ and all $t \in [t_0,T]$, it holds*

$$|x(t) - x^\delta(t)| \le c_2 \left(|x_0^{IV} - (x_0^{IV})^\delta| + \int_{t_0}^t |\delta(\tau)|d\tau + \int_{t_0}^t |\delta'(\tau)|d\tau + |\delta(t)| + |\delta'(t)| \right). \quad (3.56)$$

2. If each perturbed current source in the circuit forms a loop with CVR-edges and no perturbed voltage source forms a loop with CV-edges, then for all $M > 0$ and all $T > 0$ there exists $c_1 > 0$ such that for all $\delta = \left(\begin{smallmatrix} i_\delta \\ v_\delta \end{smallmatrix}\right) \in B(T, M)$ and all $t \in [t_0, T]$, it holds

$$|x(t) - x^\delta(t)| \leq c_1 \left(|x_0^{IV} - (x_0^{IV})^\delta| + \int_{t_0}^{t} |\delta(\tau)| d\tau + |\delta(t)| \right). \tag{3.57}$$

Remark. Out terminology is such that an element forms a loop with for instance CVR-edges, if no edges other than CVR are involved in that loop. This means that a loop with (for instance) CV-edges is also a loop with CVR-edges.

Proof: We shall show the statements in order.

1. We write

$$P_\delta := \begin{pmatrix} P_{i_\delta} & 0 \\ 0 & P_{v_\delta} \end{pmatrix}, \qquad \delta := \begin{pmatrix} i_\delta \\ v_\delta \end{pmatrix},$$

and we notice that the IVPs (3.52) and (3.53) differ only in initial values and the source functions: The source term of the MNA is $s = \left(\begin{smallmatrix} i_s \\ v_s \end{smallmatrix}\right)$, and the source term of the perturbed MNA is $s + P_\delta \delta$.

Applying Theorem 3.4.5 to the perturbed and the unperturbed MNA yields equivalent equations of the form (3.46) for both systems. Building the difference of the dynamic part (3.46a) of the perturbed and unperturbed equations in integral formulation, we obtain for all $t \in [t_0, T]$ and all $\delta \in B^1(T, M)$

$$\begin{aligned}
|\Delta_y| &= |y(t_0) + \int_{t_0}^{t} f(y^\delta, s + P_\delta\delta, G_3(s + P_\delta\delta)')d\tau - y^\delta(t_0) - \int_{t_0}^{t} f(y, s, G_3 s')d\tau| \\
&\leq |\Delta_y(t_0)| + \left| \int_{t_0}^{t} f(y^\delta, s + P_\delta\delta, G_3(s + P_\delta\delta)') - f(y, s, G_3 s')d\tau \right| \\
&\leq |\Delta_y(t_0)| + \int_{t_0}^{t} |f(\underbrace{y^\delta, s + P_\delta\delta, G_3(s + P_\delta\delta)'}_{\alpha_\delta}) - f(\underbrace{y, s, G_3 s'}_{\alpha})|d\tau
\end{aligned}$$

where we dropped the arguments t and τ for better readability and denoted $\Delta_y := y - y^\delta$. The inequality between the second and the third line holds due to Lemma 3.3.13.

We notice that s and s' are bounded as fixed continuously differentiable functions on the compact time interval $[t_0, T]$, δ and δ' are bounded since $\delta \in B^1(T, M)$, and Theorem 3.4.5(*iii*) yields boundedness of y and y^δ. Note that the bound for y^δ holds uniformly for all $\delta \in B^1(T, M)$. Hence, the arguments $\alpha(t)$ and $\alpha_\delta(t)$ are contained in a compact subset $\Omega \subset \mathbb{R}^n$ at all times $t \in [t_0, T]$ and for all $\delta \in B^1(T, M)$ and f is Lipschitz continuous on Ω since f is locally Lipschitz continuous by Lemma 3.4.5.

For any Lipschitz continuous function g with Lipschitz constant L, it holds

$$\begin{aligned}|g(a,b)-g(\tilde a,\tilde b)| &= |g(a,b)-g(\tilde a,b)+g(\tilde a,b)-g(\tilde a,\tilde b)|\\ &\le |g(a,b)-g(\tilde a,b)|+|g(\tilde a,b)-g(\tilde a,\tilde b)| \le L|a-\tilde a|+L|b-\tilde b|.\end{aligned}$$

That way, we obtain

$$|\Delta_y| \le |\Delta_y(t_0)| + L_0\int_{t_0}^t |\Delta_y| d\tau + L_0|P_\delta|_*\int_{t_0}^t |\delta| d\tau + L_0|G_3P_\delta|_*\int_{t_0}^t |\delta'| d\tau, \tag{3.58}$$

where $|\cdot|_*$ is the induced matrix norm of $|\cdot|$ and L_0 is the Lipschitz constant of f on Ω. At this point we can apply the Gronwall Lemma 3.3.14, which yields

$$|\Delta_y| \le \left(|\Delta_y(t_0)| + L_0|P_\delta|_*\int_{t_0}^t |\delta| d\tau + L_0|G_3P_\delta|_*\int_{t_0}^t |\delta'| d\tau\right) e^{L_0(t-t_0)}$$

and consequently the existence of $c_y > 0$ such that

$$|\Delta_y| \le c_y\left(|\Delta_y(t_0)| + \int_{t_0}^t |\delta| d\tau + \int_{t_0}^t |\delta'| d\tau\right).$$

The algebraic equations (3.46b)-(3.46d), boundedness of sources s, perturbations δ and solutions y and y^δ on $[t_0,T]$ along with local Lipschitz continuity of g_1 and g_2 on arbitrary bounded subsets of $\mathbb{R}^n$, which is warranted by Theorem 3.4.5, yield for all $t\in[t_0,T]$ and all $\delta \in B^1(T,M)$

$$|\Delta_{z_1}| = |g_1(y^\delta, s+P_\delta\delta, G_3(s+P_\delta\delta)') - g_1(y,s,G_3s')| \le L_1(|\Delta_y| + |P_\delta|_*|\delta| + |G_3P_\delta|_*|\delta'|) \tag{3.59}$$

$$|\Delta_{z_2}| = |g_2(y^\delta, s+P_\delta\delta) - g_2(y,s)| \le L_2(|\Delta_y| + |P_\delta|_*|\delta|) \tag{3.60}$$

$$|\Delta_{z_3}| = |G_3(s+P_\delta\delta) - G_3 s| \le |G_3P_\delta|_*|\delta|, \tag{3.61}$$

where $\Delta_{z_i} := z_i - z_i^\delta$ and L_1, L_2 are the Lipschitz constants of g_1, g_2 on a bounded set which contains the relevant arguments. Dependence of t is again dropped in the notation for better readability.

Furthermore, Theorem 3.4.5 yields the existence of matrices T_i such that

$$x = T_0 y + T_1 z_1 + T_2 z_2 + T_3 z_3, \qquad x^\delta = T_0 y^\delta + T_1 z_1^\delta + T_2 z_2^\delta + T_3 z_3^\delta,$$

where x is the solution of the nonlinear MNA and x^δ is the solution of the perturbed nonlinear MNA. It follows for $\Delta_x := x - x^\delta$ and an appropriate constant $c_2 > 0$

$$\begin{aligned}|\Delta_x| &= |T_0\Delta_y + T_1\Delta_{z_1} + T_2\Delta_{z_2} + T_3\Delta_{z_3}|\\ &\le |T_0|_*|\Delta_y| + |T_1|_*|\Delta_{z_1}| + |T_2|_*|\Delta_{z_2}| + |T_3|_*|\Delta_{z_3}|\\ &= c_2\left(|\Delta_y(t_0)| + |\delta| + |\delta'| + \int_{t_0}^t |\delta| d\tau + \int_{t_0}^t |\delta'| d\tau\right)\end{aligned}$$

which proves the first statement. The crucial point here is to note for fixed T and M, c does not depend on $t \in [t_0, T]$ or $\delta \in B^1(T, M)$.

2. Revisiting Inequalities (3.58)-(3.61), we see that $|\delta'|$ only appears with a certain factor, namely $|G_3 P_\delta|_* |\delta'|$. To prove the second statement, it is therefore sufficient to show that $G_3 P_\delta = 0$ provided the preconditions of the second statement hold. We recall that this product is defined by

$$G_3 P_\delta = \begin{pmatrix} -\mathcal{A}_{i/cvr} & 0 \\ 0 & \breve{Q}_v^\top \end{pmatrix} \begin{pmatrix} P_{i_\delta} & 0 \\ 0 & P_{v_\delta} \end{pmatrix} = \begin{pmatrix} -Q_{cvr}^\top \mathcal{A}_i P_{i_\delta} & 0 \\ 0 & \breve{Q}_v^\top P_{v_\delta} \end{pmatrix}$$

and by construction of P_{i_δ}, the matrix $\mathcal{A}_i P_{i_\delta} =: \mathcal{A}_\delta$ is the incidence matrix of the perturbed current sources.

Lemma 2.3.3 implies that if all perturbed current sources build a loop with CVR-edges, then $\operatorname{im} \mathcal{A}_\delta \subseteq \operatorname{im} \mathcal{A}_{cvr}$. Furthermore, it is $\ker Q_{cvr}^\top = \operatorname{im} \mathcal{A}_{cvr}$ with Lemma 3.1.1. This proves that $Q_{cvr}^\top \mathcal{A}_i P_{i_\delta} = 0$ in that case.

Now we show that $\breve{Q}_v^\top P_{v_\delta} = 0$. We first examine $\breve{Q}_v^\top$, which is the kernel matrix of $P_v^\top Q_c^\top \mathcal{A}_v$ as defined in Table (3.2). We notice that $\ker P_v^\top Q_c^\top \mathcal{A}_v = \ker Q_c^\top \mathcal{A}_v$ by Lemma 3.1.6(3), and hence $\operatorname{im} \breve{Q}_v = \ker Q_c^\top \mathcal{A}_v$, and, by Lemma 3.1.1, $\ker \breve{Q}_v^\top = \operatorname{im} \mathcal{A}_v^\top Q_c$. Furthermore, Lemma 2.3.16 yields that $Q_c^\top \mathcal{A}_v = \begin{pmatrix} \mathcal{A}_{v/c} & 0 \end{pmatrix}$ if the V-edges are sorted accordingly, that is, those V-edges which form loops with C-edges (that is, whose incident nodes are C-connected) are represented by zero columns and the remaining V-edges are represented by $\mathcal{A}_{v/c}$. Here, $\mathcal{A}_{v/c}$ is the incidence matrix of the contracted Graph G_{cv}/G_c. We recall that this graph arises from the graph containing all C- and V-edges and all nodes of the circuit when all C-edges are contracted.

To sum it up, we have

$$\ker \breve{Q}_v^\top = \operatorname{im} \begin{pmatrix} \mathcal{A}_{v/c}^\top \\ 0 \end{pmatrix}. \tag{3.62}$$

On the other hand, the matrix P_{v_δ} consists of columns which are unit vectors denoted here by $\hat{e}_i$. More precisely, $\hat{e}_i$ is a column of P_{v_δ} if and only if V-edge i corresponds to a perturbed voltage source.

Let now edge i be such a perturbed V-edge for which the precondition holds, that is, i does not form a loop with CV-edges. In the following we show that it then holds $\hat{e}_i \in \operatorname{im} \mathcal{A}_{v/c}^\top$. Combined with the previous considerations, this implies the desired result $\breve{Q}_v^\top P_{v_\delta} = 0$ if the precondition holds.

For the remainder of this proof, given any graphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$ and any edge $j \in E_1$ with incident nodes $n_k, n_l \in V_1$, we define the operations $+$ and $-$ as $G_2 + j := (V_2 \cup \{n_k, n_l\}, E_2 \cup \{j\})$ and $G_1 - j := (V_1, E_1 \setminus \{j\})$.

Edge i remains undeleted in the contracted graph G_{cv}/G_c and it does not belong to a loop in the contracted graph. Therefore, the end node n_k and the start node n_l of

i belong to different connected components in $G_{cv}/G_c - i$. We denote the component which contains n_k by C_k^{-i} and we consider the graph $C_k^{-i} + i$.

Due to the zero-sum property of the columns of incidence matrices, it holds that the rows a non-reduced incidence matrix are linearly dependent and the sum of (all) its rows is zero.

For the non-reduced incidence matrix $\mathcal{A}_{C_k^{-i}+i}$ with rows a_j, this means that

$$\sum_{j=1,\dots,|N_k^{-i}+i|} a_j = 0,$$

where $N_k^{-i} + i$ is the node set of $C_k^{-i} + i$ and $|N_k^{-i} + i|$ its cardinality. Let a_l be the row which corresponds to n_l. We notice that $a_l = \hat{e}_i$ since n_l is a leaf (a node with degree 1) which is only incident to edge i in the graph $C_k^{-i} + i$. Therefore, it holds

$$\sum_{j=1,\dots,l-1,l+1,\dots,|N_k^{-i}+i|} a_j = -\hat{e}_i.$$

This implies that the columns of $\mathcal{A}_{v/c}^\top$ which correspond to nodes of C_k^{-i} span $\hat{e}_i$. Combined with Equation (3.62), this proves the second statement. □

Remark 3.5.2 *The estimations in Theorem 3.5.1 have a similar structure as the definition of the perturbation index given in 2.2.4. The difference is that the perturbation index admits more general perturbations, while in Theorem 3.5.1, only perturbations of the input terms are admitted.*

Therefore, it is possible that a circuit DAE has for instance perturbation index two, but satisfies an index-one type estimate of the form (3.57). The converse situation, however, is not possible.

Figure 3.1 shows an example circuit. According to Theorem 3.5.1, the voltage source V_1 and the current source I_1 are "index-one inputs", that is, a perturbation of them leads to the index-one type Estimate (3.57), and V_2 and I_2 are "index-two inputs" whose perturbation leads to the index-two type Estimate (3.56).

In the example circuit in Figure 3.2, I_1 is an "index one input" since it forms a loop with CVR-edges (in fact with CV-edges), and I_2 is an "index-two input".

For the remaining theorems of the section, we want to look at the IVP where only current sources are perturbed:

$$\mathcal{A}_c C(\mathcal{A}_c^\top e^\delta)\mathcal{A}_c^\top (e^\delta)' + \mathcal{A}_r g(\mathcal{A}_r^\top e^\delta) + \mathcal{A}_l i_l^\delta + \mathcal{A}_v i_v^\delta + \mathcal{A}_i i_s = -\mathcal{A}_i P_{i_\delta} i_\delta \tag{3.63a}$$

$$-L(i_l^\delta)(i_l^\delta)' + \mathcal{A}_l^\top e^\delta = 0 \tag{3.63b}$$

$$\mathcal{A}_v^\top e^\delta - v_s = 0 \tag{3.63c}$$

$$Ax^\delta(t_0) = (x_0^{IV})^\delta, \tag{3.63d}$$

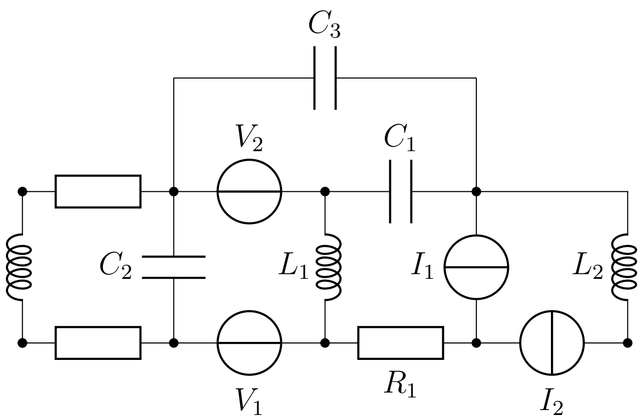


Figure 3.1.: The voltage source V_2 forms a CV-loop with C_1 and C_3, whereas any loop containing V_1 must contain either L_1 or R_1 and is hence not a CV-loop. Similarly, any loop containing the current source I_2 must contain contain also L_2 and is hence not a CVR-loop. I_1, however, forms a CVR-loop with R_1, V_1, C_2, C_3. This means the circuit is "index-one sensitive" to perturbations of I_1 and V_1, that is, a perturbation of (only) them leads to Estimate (3.57) in Theorem 3.5.1. Perturbations of V_2 and / or I_2 lead to the index-two type Estimate (3.56).

We notice that P_{i_δ} is constructed such that it deletes columns corresponding to unperturbed current sources in the product $\mathcal{A}_i P_{i_\delta}$ and leaves the columns corresponding to perturbed sources unchanged. Therefore, we define

$$\mathcal{A}_\delta := \mathcal{A}_i P_{i_\delta}, \tag{3.64}$$

which is the incidence matrix of the perturbed current sources.

Theorem 3.5.3 (Sensitivity of voltages over perturbed currents) *Let $|\cdot|$ be a vector norm, $B(T,M)$ be defined as in (3.55), and let the MNA Assumptions 2.1.3 and 3.4.3 hold. We consider the IVPs (3.53) and (3.63), which represent the MNA of a circuit, and their solutions x and x^δ.*

If each perturbed current source in the circuit forms a loop with CV-edges, then for all $T > 0$ and all $M > 0$ there exists $c > 0$ such that for all $t \in [t_0, T]$ and all $i_\delta \in B(T,M)$, it holds

$$|\mathcal{A}_\delta^\top (e(t) - e^\delta(t))| \leq c \left(|x_0^{IV} - (x_0^{IV})^\delta| + \int_{t_0}^{t} |i_\delta(\tau)| d\tau \right). \tag{3.65}$$

Remark. Note that e and e^δ are the node potentials of the circuit (the former in the perturbed case, the latter in the unperturbed case). Accordingly, $\mathcal{A}_\delta^\top e$ and $\mathcal{A}_\delta^\top e^\delta$ are the vectors describing the voltage over the current sources which are perturbed.

Proof: This proof is based on Theorem 3.4.2 and the notation therein. The idea is to rewrite both the perturbed and the unperturbed MNA in the normal form (3.32)-(3.40), build the difference of the equations which determine $\bar{e}_1, \dots, \bar{e}_5$ and exploit $e = T_e \bar{e}$. We recall that the perturbed and unperturbed MNA only differ in the current source terms.

First, we notice that by construction of P_{i_δ}, we know that $\mathcal{A}_i P_{i_\delta} =: \mathcal{A}_\delta$ is the incidence matrix of the perturbed current sources. Lemma 2.3.3 implies that if all perturbed current sources build a loop with CV-edges, then $\text{im } \mathcal{A}_\delta \subseteq \text{im } \mathcal{A}_{cv} \subseteq \text{im } \mathcal{A}_{cvr}$. Furthermore, it is $\ker Q_{cv}^\top = \text{im } \mathcal{A}_{cv}$ and $\ker Q_{cvr}^\top = \text{im } \mathcal{A}_{cvr}$ with Lemma 3.1.1. Hence $Q_{cv}^\top \mathcal{A}_\delta = 0$ and $Q_{cvr}^\top \mathcal{A}_\delta = 0$.

For the transformation matrix T_e from the decoupling Theorem 3.4.2, this means that

$$\mathcal{A}_\delta^\top T_e = \mathcal{A}_\delta^\top (P_c \mathring{\mathcal{A}}_v \ P_c \mathring{Q}_v \ Q_c P_v \ Q_{cv} P_r \ Q_{cvr}) = \mathcal{A}_\delta^\top (P_c \mathring{\mathcal{A}}_v \ P_c \mathring{Q}_v \ Q_c P_v \ 0 \ 0).$$

Furthermore, considering Equation (3.39), we have $\bar{e}_1^\delta - \bar{e}_1 = 0$. Thus, using the notation $\Delta_{\bar{e}} := \bar{e}^\delta - \bar{e}$ and $\Delta_{\bar{e}_i} := \bar{e}_i^\delta - \bar{e}_i$, we obtain

$$\mathcal{A}_\delta^\top T_e(\Delta_{\bar{e}}) = \mathcal{A}_\delta^\top (P_c \mathring{Q}_v \Delta_{\bar{e}_2} + Q_c P_v \Delta_{\bar{e}_3}) \tag{3.66}$$

Building the difference of Equation (3.36) of the perturbed and unperturbed MNA yields

$$\Delta_{\bar{e}_3} = (\breve{\mathcal{A}}_{v/c}^\top)^\dagger \mathcal{A}_v^\top P_c \mathring{Q}_v \Delta_{\bar{e}_2} =: M \Delta_{\bar{e}_2}, \tag{3.67}$$

since v_s cancels out and $\Delta_{\bar{e}_1} = 0$. Hence, we can rewrite for $\Delta_e := e^\delta - e$

$$\mathcal{A}_\delta^\top \Delta_e = \mathcal{A}_\delta^\top [P_c \mathring{Q}_v \Delta_{\bar{e}_2} + M \Delta_{\bar{e}_2}] =: \tilde{M} \Delta_{\bar{e}_2}. \tag{3.68}$$

Exploiting Theorem 3.4.5, we obtain analogously to Inequality (3.58) from the preceding proof of Theorem 3.5.1

$$\begin{aligned}
\left| \begin{pmatrix} \Delta_{\bar{e}_2} \\ \Delta_{\bar{i}_{l_2}} \end{pmatrix} \right| &= |\Delta_y| \\
&= \left| y(t_0) + \int_{t_0}^t f(y^\delta, s + P_\delta i_\delta, G_3(s + P_\delta i_\delta)') d\tau - y^\delta(t_0) - \int_{t_0}^t f(y, s, G_3 s') d\tau \right| \\
&\leq |\Delta_y(t_0)| + \left| \int_{t_0}^t f(y^\delta, s + P_\delta i_\delta, G_3 s') - f(y, s, G_3 s') d\tau \right| \\
&\leq |\Delta_y(t_0)| + L_0 \int_{t_0}^t |\Delta_y| d\tau + L_0 |P_\delta|_* \int_{t_0}^t |i_\delta| d\tau \quad \forall t \in [t_0, T].
\end{aligned}$$

where L_0 is the Lipschitz constant of f on a compact set which contains the relevant arguments (for greater detail, see derivation of (3.58)) and the second line holds since $G_3 P_\delta = 0$ in our case as shown in the proof of the second statement of Theorem 3.5.1. By $|\cdot|_*$, we denote an induced matrix norm. Note that $P_\delta = \left(\begin{smallmatrix} P_{i_\delta} \\ 0 \end{smallmatrix} \right)$ since only current

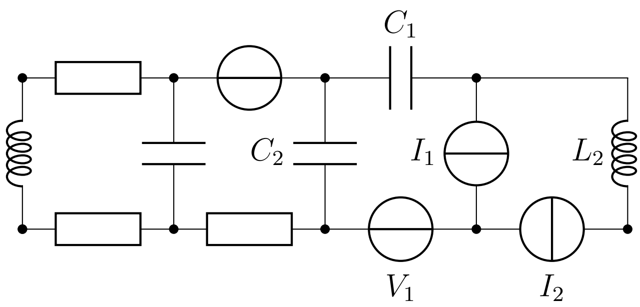


Figure 3.2.: The current source I_1 forms a CV-loop with C_1, C_2 and V_1, whereas any loop which contains I_2 must contain L_1 and is hence not a CV-loop. This means that the voltage over I_1 is "ODE-sensitive" to perturbations of I_1, that is, a perturbation of I_1 leads to Estimate (3.65) in Theorem 3.5.3, where the left hand side of the estimate describes (the norm of) the voltage over I_1 in this case. In contrast, I_2 does not satisfy the requirements for Theorem 3.5.3.

sources are perturbed. Using the Gronwall Lemma 3.3.14, we obtain a bound for $|\Delta_y|$ for an appropriate $c_y > 0$:

$$|\Delta_y| \leq c_y \left(|\Delta_y(t_0)| + \int_{t_0}^{t} |i_\delta| d\tau \right), \tag{3.69}$$

which implies the existence of a constant $c_{\bar{e}_2} > 0$ such that

$$|\Delta_{\bar{e}_2}| \leq c_{\bar{e}_2} \left(|\Delta_y(t_0)| + \int_{t_0}^{t} |i_\delta| d\tau \right). \tag{3.70}$$

For given $T > 0$, $M > 0$ the constants in these estimates are independent of $t \in [t_0, T]$ and $i_\delta \in B(T, M)$ as argued in the derivation of Inequality (3.58).

Considering Equations (3.66) and (3.67), we can conclude the desired result

$$|\mathcal{A}_\delta \Delta_e| \leq |\tilde{M}|_* |\Delta_{\bar{e}_2}|_{[t_0,t]} \leq |\tilde{M}|_* c_{\bar{e}_2} \left(|\Delta_y(t_0)| + \int_{t_0}^{t} |i_\delta| d\tau \right),$$

where $\tilde{M}$ is defined as in (3.67). □

In the example circuit in Figure 3.1, neither I_1 nor I_2 form a loop with CV-edges, hence they do not satisfy the preconditions for Theorem 3.5.3.

In Figure 3.2, I_1 forms a loop with CV-edges and Theorem 3.5.3 can be applied.

In the next theorem, we consider circuits in which only one current source is perturbed. In this case, P_{i_δ} is a column vector and $i_\delta(t) \in \mathbb{R}$ a scalar value. The incidence matrix of the perturbed current source, $\mathcal{A}_i P_{i_\delta} = \mathcal{A}_\delta =: a_\delta$ is then a column vector.

Specific choice of kernel splitting pairs We choose specific kernel splitting pairs $(P_\star, Q_\star)$ (which were defined in 3.1.4) to obtain the normal forms in the Decoupling Theorems 3.4.2 and 3.4.5. Recall that kernel splitting pairs are transformation matrices with which we achieved these normal form for circuit DAEs, and Tables (3.1) and (3.2) define precisely which kernel splitting pairs are needed for the DAE decoupling.

Let $G = (V, E)$ be a disconnected (subgraph of a) circuit graph and $C^j = (V^j, E^j)$, $j = 1, \ldots, |C|$ its connected components. For convenience, let the reference node of the circuit be contained in the component $C^{|C|}$. Furthermore, let

$$Q \in \mathbb{R}^{(|V|-1)\times(|C|-1)}$$

be the reduced $(0, 1)$ component kernel matrix of G as defined in 2.3.15 and subsequent remarks.

For $n_{C^1} \in C^1, n_{C^2} \in C^2, \ldots, n_{|C|-1} \in C^{|C|-1}$ an arbitrary selection of one node per connected component (except the component which contains the reference node), we then define

$$P \in \mathbb{R}^{(|V|-1)\times(|V|-|C|)}$$

such that it arises from the identity matrix $I_{|V|-1} \in \mathbb{R}^{(|V|-1)\times(|V|-1)}$ after deletion of the columns $n_{C_1}, \ldots, n_{|C|-1}$.

It is easy to check that this way, the pair $\{P, Q\}$ is a kernel splitting pair of the transposed incidence matrix $\mathcal{A}^\top$ of G. The benefits of this definition are that certain products with Laplacians and incidence matrices maintain a topological meaning:

On the one hand, by construction of P, a product MP arises from M after deletion of the columns $n_{C_1}, \ldots, n_{|C|-1}$. That way, in a circuit graph context, multiplying P (from the right) to a transposed incidence matrix corresponds to "setting one reference node in each connected component" in the graph.

On the other hand, we can exploit the Results 2.3.16 and 2.3.17 for a topological interpretation of products of transposed incidence matrices and Laplacians with Q. According to these results, such products correspond to certain graph contractions, which motivated our notational convention $\mathcal{A}_{\star/\bullet}$.

Definition 3.5.4 *Let a circuit graph G with numbered CV-components be given. Let the conductance function g be C^1 with a diagonal Jacobian g'. For a node i, let $p(i)$ be the CV-component which contains i. We define*

$$R_{ij} := d_r^{G_{r/cv}}(p(i), p(j)),$$

where d_r^G is the r-distance as defined in 2.3.9 and $G_{r/cv} := G_{rcv}/G_{cv}$ is a contracted graph based on edge contraction as defined in the first paragraph of Subsection 2.3.3. The nodes of $G_{r/cv}$ are the CV-components of G, and its edge weights are given by the diagonal entries of g'.

Remark. The representation of a nonlinear circuit as a graph requires state-dependent edge weights. In that case, the "resistive distance" $R_{ij} = R_{ij}(\mathcal{A}_r^\top e)$ is a function of the state.

Remark. The MNA model (2.11) gives rise to matrices C, L and g' which are diagonal if only the basic (nonlinear) RLC-elements and independent sources are to be modelled. However, elements like transistors may require to admit non-diagonal matrices C, L and g'. For that reason, we did *not* assume that C, L and g' are diagonal up to this point.

In contrast, the following result is based on the assumption that g' is diagonal. The reason behind this is that we want to exploit the structure of weighted Laplacian matrices $\mathcal{A}_* g' \mathcal{A}_*^\top$ for certain incidence matrices $\mathcal{A}_*$. By definition, such matrix products are Laplacians if the weight matrix g' is diagonal, cf. Section 2.3.1. Notably, the r-distance in graphs, which is defined by means of Laplacian matrices, plays an important role in the following theorem.

Remark. The graph $G_{r/cv}$ is generally not connected, such that $d_r^{G_{r/cv}}(p(i), p(j))$ is not necessarily a real number / well-defined. However, we apply this distance only to CVR-connected nodes, for which $d_r^{G_{r/cv}}(p(i), p(j)) \in \mathbb{R}$ is therefore well-defined.

Remark. If two nodes i and j belong to the same CV component in a graph G, then they are identified in the contracted graph $G_{r/cv}$. Hence, $R_{ij} := 0$ in that case.

Theorem 3.5.5 (Voltage over one perturbed current source) *Let $|\cdot|$ be a vector norm, $B(T, M)$ be defined as in Equations (3.54) and (3.55), and let the MNA Assumptions 2.1.3 and 3.4.3 hold. We consider the IVPs (3.53) and (3.63), which represent the MNA of a circuit, and their solutions x and x^δ.*

We assume that only one current source is perturbed so that $P_{i_\delta} \in \mathbb{R}^{n \times 1}$ and $i_\delta(t) \in \mathbb{R}$, and we denote the voltage over this current source by v_* (Equation (3.53)) and v_*^δ (Equation (3.63)).

If the conductance function g is continuously differentiable and its Jacobian matrix $g'(x)$ is diagonal for all x, and if furthermore the perturbed current source forms a loop with CVR-edges, then for all $M > 0$ and all $T > 0$ there exists $c > 0$ such that for all $i_\delta \in B(T, M)$ and all $t \in [t_0, T]$, it holds

$$|v_*(t) - v_*^\delta(t)| \leq c \left(|x_0^{IV} - (x_0^{IV})^\delta| + \int_{t_0}^{t} |i_\delta(\tau)| d\tau \right) + \sup_{e \in \mathbb{R}^n} R_{ij}(\mathcal{A}_r^\top e) |i_\delta(t)|. \tag{3.71}$$

Remark. The proof also reveals that the supremum in Equation (3.71) is finite, hence well-defined.

Remark. The supremum of R_{ij} is generally not a known parameter in a circuit in the case of a nonlinear graph. However, since it describes the r-distance between two nodes in a contracted (resistive) graph, the monotonicity properties of this distance concept (cf. Lemma 2.3.13 and 2.3.14) offer insight on which topological manipulations or changes of the conductance functions increase or decrease R_{ij}.

Proof: The proof is strongly based on the decoupling Theorem 3.4.2, and the notation therein shall be required here.

First, we notice that, with the notation $\Delta_e := e^\delta - e$ and analogous definitions for the other variables,

$$\Delta_e = T_e\Delta_{\bar{e}} = P_c\mathring{\mathcal{A}}_v\Delta_{\bar{e}_1} + P_c\mathring{Q}_v\Delta_{\bar{e}_2} + Q_cP_v\Delta_{\bar{e}_3} + Q_{cv}P_r\Delta_{\bar{e}_4} + Q_{cvr}\Delta_{\bar{e}_5} \tag{3.72}$$

and by Equation (3.39), we have

$$\Delta_{\bar{e}_1} = 0. \tag{3.73}$$

This implies by Equation (3.36)

$$\Delta_{\bar{e}_3} = (\breve{\mathcal{A}}_{v/c}^\top)^\top \mathcal{A}_v^\top P_c\mathring{Q}_v\Delta_{\bar{e}_2} =: M\Delta_{\bar{e}_2}. \tag{3.74}$$

We introduce the notation $a_\delta := \mathcal{A}_iP_{i_\delta}$. We do not use a capital letter for a_δ as a reminder that a_δ has only one column. This column represents the incidences of the perturbed current source.

It follows from Equation (3.40) that

$$\Delta_{\bar{i}_{l_1}} = 0 \tag{3.75}$$

since

$$a_{\delta/cvr} = 0 \tag{3.76}$$

as shown in the proof of the second statement of Theorem 3.5.1. It follows that Δ_y, built from Equations (3.32)-(3.33), does not involve derivatives anymore after insertion, since the differentiated quantities $\Delta_{\bar{i}_{l_1}}$ and $\Delta_{\bar{e}_1}$ are both zero now, which implies for all $t \in [t_0, T]$ and all $i_\delta \in B(T, M)$

$$|\Delta_y(t)| \leq c_y\left(|\Delta_y(t_0)| + \int_{t_0}^t |i_\delta| d\tau\right) \tag{3.77}$$

analogously to Equation (3.69) in the proof of Theorem 3.5.3. As $\bar{e}_2$ is contained in y, such a bound holds also for $\Delta_{\bar{e}_2}$.

We notice that if the current sources are sorted conveniently, then $i_s = \begin{pmatrix} i_s^- \\ i_s^* \end{pmatrix}$, where i_s^- is the vector of current inputs from unperturbed sources and i_s^* is the (unperturbed) scalar input from the source which is perturbed. Being a little sloppy on column and row vector representation, we write

$$\Psi_g(\bar{e}_{123}, \bar{i}_l, i_s) = \Psi_g(\bar{e}_{123}, \bar{i}_l, i_s^-, i_s^*).$$

From the partial derivative $\frac{\partial \Psi_g}{\partial i_s}$ given in Equation (3.41) of the Decoupling Theorem 3.4.2, it follows

$$|a_{\delta/cv}^\top P_r \frac{\partial \Psi_g}{\partial i_s^*}(\bar{e}_{123}, \bar{i}_l, i_s^-, i_s^*)| = |a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\xi)\right]^{-1} P_r^\top a_{\delta/cv}|$$
$$\leq \sup_{\zeta \in \mathbb{R}^n} |a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top \zeta)\right]^{-1} P_r^\top a_{\delta/cv}|,$$

where $|\cdot|$ is the absolute value in $\mathbb{R}$, $\xi := \mathcal{A}_r^\top(\bar{T}_e^- \bar{e}_{123} + Q_{cv} P_r \Psi_g(\bar{e}_{123}, \bar{i}_l, i_s))$ and n is the number of rows of $\mathcal{A}_r$. Furthermore, $a_{\delta/cv} := Q_{cv}^\top a_\delta$ and we recall that a_δ is the column of $\mathcal{A}_i$ which corresponds to the perturbed input. Hence, a Lipschitz constant of $a_{\delta/cv}^\top P_r \psi_g(\cdot)$ w.r.t. i_s^* is given by

$$L_{\Psi_g}^{i_s^*} = \sup_{e \in \mathbb{R}^n} |a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top e)\right]^{-1} P_r^\top a_{\delta/cv}| \overset{\text{to show}}{=} \sup_{e \in \mathbb{R}^n} R_{ij}(\mathcal{A}_r^\top e). \tag{3.78}$$

Recalling that g is strongly monotone, it follows with Corollary 3.3.10 that $\check{\mathcal{L}}_{r/cv}$ is uniformly positive definite, and has therefore a bounded inverse with Lemma 3.3.11. Hence, the supremum in the above equation is well-defined.

Furthermore, we notice that the term whose supremum we take is scalar since $a_{\delta/cv}$ is a column vector, and nonnegative since $\check{\mathcal{L}}_{r/cv}$ is symmetric and positive definite for all arguments, implying the same properties for the inverse.

To show that the second equation of (3.78) holds, we show that

$$a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top e)\right]^{-1} P_r^\top a_{\delta/cv} = R_{ij}(\mathcal{A}_r^\top e) \quad \forall e \in \mathbb{R}^n. \tag{3.79}$$

Lemma 2.3.16 and Corollary 2.3.17 allow to interprete $a_{\delta/cv}$ as the incidence matrix of a single-edge subgraph of the contracted graph $G_{i/cv}$, and $Q_{cv}^\top \mathcal{A}_r g'(\cdot) \mathcal{A}_r^\top Q_{cv}$ as a weighted Laplacian of $G_{r/cv}$. Furthermore, P_r by construction "sets one reference node in each (yet ungrounded) component of $G_{r/cv}$", which algebraically corresponds to deleting columns of a matrix when multiplied to it from the right. It follows, with two exceptions discussed subsequently, that

$$a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top e)\right]^{-1} P_r^\top a_{\delta/cv} = (\hat{e}_{p(i)} - \hat{e}_{p(j)})^\top [\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top e)]^{-1} (\hat{e}_{p(i)} - \hat{e}_{p(j)}), \tag{3.80}$$

where $\hat{e}_i$ denotes the i-th uni vector and $p(i)$ and $p(j)$ are the incident nodes of the edge from $G_{i/cv}$ represented by $a_{\delta/cv}$. By Corollary 2.3.10, the right hand side of Equation 3.80 is precisely the r-distance between $p(i)$ and $p(j)$ in $G_{r/cv}$, and hence R_{ij} by Definition 3.5.4.

There are two special cases to be mentioned: In the first case, the edge represented by $a_{\delta/cv}$ forms a loop with CV-edges. Then, it gets deleted in the contraction process. Lemma 2.3.16 states that it follows $a_{\delta/cv} = 0$. This is in accordance with the topological definition of R_{ij}, since the nodes i and j in that case get identified in the contraction process and it holds $p(i) = p(j)$ in $G_{r/cv}$.

The second special case is that one of the incident nodes, say i, is "set as a reference node by P_r", that is, the corresponding nonzero entry in $a_{\delta/cv}$ gets deleted in the product $a_{\delta/cv}P_r^\top$. Then, Equation 3.80 becomes

$$a_{\delta/cv}^\top P_r \left[\check{\mathcal{L}}_{r/cv}(\mathcal{A}_r^\top e)\right]^{-1} P_r^\top a_{\delta/cv} = \hat{e}_{p(j)}^\top [\check{\mathcal{L}}_{r/cv}(\cdot)]^{-1}\hat{e}_{p(j)}.$$

On the basis of Lemma 2.3.7 and 2.3.8, Corollary 2.3.10 states that this expression determines the distance between the reference node and $p(j)$ in the graph $G_{r/cv}$, which confirms our desired Equality 3.79.

Having convinced ourselves that the Lipschitz constant can be expressed as in Equation (3.78), and considering that Ψ_g is globally Lipschitz continuous (in all variables), Equation (3.37) from the decoupling Theorem implies the existence of a $c > 0$ such that

$$|a_{\delta/cv}^\top P_r \Delta_{\bar{e}_4}(t)| \leq c \left(|\Delta_y(t_0)| + \int_{t_0}^t |i_\delta(t)| \right) + L_{\Psi_g}^{i_s^*} |i_\delta(t)|. \tag{3.81}$$

where we inserted Equations (3.73),(3.75),(3.74) and (3.77), (note that y consists of $\bar{e}_2$ and $\bar{i}_{l_2}$).

Inserting the gathered information about each summand of Equation (3.72), we obtain an estimate of the same form as (3.81), but with a different constant $\tilde{c}$ (since $\Delta_{\bar{e}_1}$ and $a_\delta^\top Q_{cvr}\Delta\bar{e}_5$ vanish, $\Delta_{\bar{e}_3}$ can be expressed in terms of $\Delta_{\bar{e}_2}$, and $\Delta_{\bar{e}_2}$ is contained in Δ_y, for which estimate (3.77) holds), that is:

$$|v_*(t) - v_*^\delta(t)| = |a_\delta^\top \Delta e| \leq \tilde{c}\left(|\Delta_y(t_0)| + |\Delta_y(t)|\right) + L_{\Psi_g}^{i_s^*}|i_\delta(t)|$$

□

Figures 3.3 and 3.4 illustrate Estimate (3.71) of Theorem 3.5.5. The resistive distance R_{ij} therein can be understood as the resistive distance between the CV-components which contain the incident nodes of the perturbed current source. If R_{ij} is small, than a perturbation of a current source has only a mild effect on the voltage over that current source.

R_{ij} is composed of those resistances which are not deleted in the contracted graph $G_{r/cv}$, in other words those which have incident nodes in different CV-components. The remaining resistances are irrelevant for R_{ij}. If the relevant resistances are linear, R_{ij} can theoretically be computed by only the circuit topology and the conductances.

If any of the relevant resistances are nonlinear, however, then the conductances $g(\mathcal{A}_r^\top e)$ depend on the voltage across them. Since these voltages are generally not known a-priori, $R_{ij}(\mathcal{A}_r^\top e)$ must be considered unknown. Nontheless, the estimate has some interesting consequences even in that case when we take into account the monotonicity properties of the r-distance, cf. Lemma 2.3.13 and 2.3.14. They imply that the distance R_{ij} is a decreasing function of the connectivity of $G_{r/cv}$. More precisely, given a circuit graph G, the following actions are *decreasing* R_{ij}:

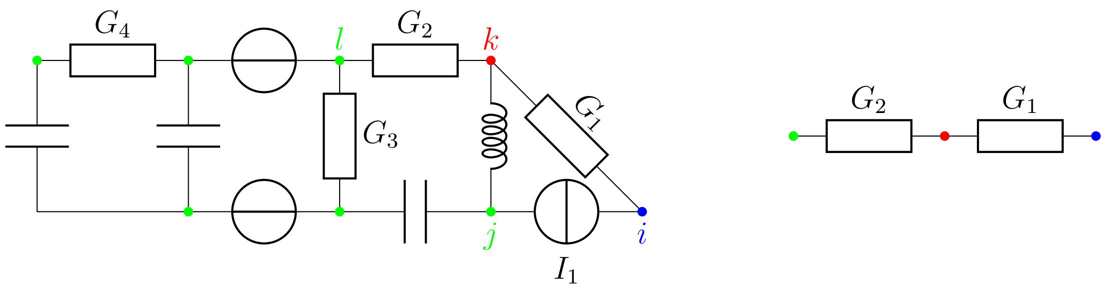


Figure 3.3.: The circuit on the left has three CV-components: The isolated nodes i and k and the component which contains all the remaining (green) nodes with j and l among them. If the current source I_1 is perturbed, the magnitude of the index-one impact of the perturbation on the voltage over I_1 is given by R_{ij} according to Theorem 3.5.5, where $R_{ij} = G_1^{-1} + G_2^{-1}$ is the resistive distance between the CV-components which contain the incident nodes i and j of I_1. On the right, we see the contracted graph $G_{r/cv}$, on which Definition 3.5.4 is based. Its edge weights are the conductances $G_i = R_i^{-1}$. The resistances G_3 and G_4 get deleted in the contraction process since they form loops with CV-elements each.

- increasing conductances (replacing g by $\tilde{g} > g$ in the case of a constant conductance),
- adding R-paths between CV-components.

Discussion

In this section, we presented three related results all aiming at quantifying the sensitivity of a circuit with respect to perturbations of its independent sources. The first result, Theorem 3.5.1, is not surprising in the light of [ET00], where the index of a circuit DAE was linked with the existence or absence of C-paths, LI-cutsets and CV-loops. In contrast to Theorem 3.5.1, which presented estimates for the effect of perturbations on the entire solution vector x, the next result, Theorem 3.5.3, offers an estimate for only the voltage over a perturbed current source. While this result seems not precisely implied in the results of [ET00], it follows the same "routine" of a thorough analysis of a normal form, in our case the normal form presented in the previous chapter, and evaluating crucial null spaces. Finally, Theorem 3.5.5 differs from the first two results in that it quantifies the "leading constant" of a perturbation estimate in terms of the circuit topology. The contracted graphs used for that appear in the circuit context in [Chen76, Est00, Ebe08]. However, the identification of certain relevant matrices of

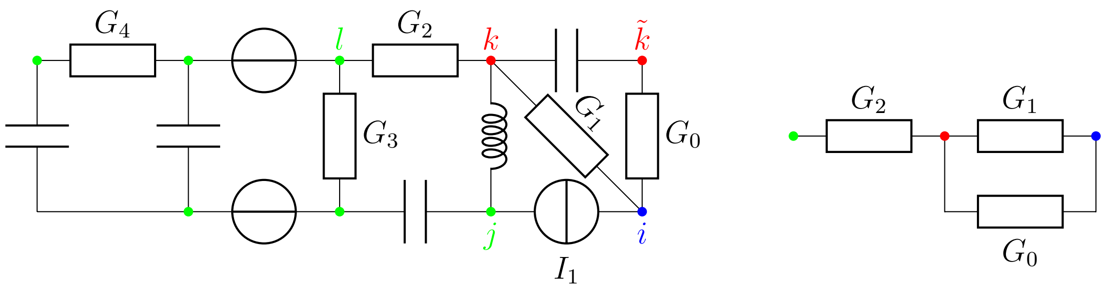


Figure 3.4.: A circuit graph on the left side and the corresponding contracted graph $G_{r/cv}$ on the right. The only change compared to Figure 3.3 is that we added the CR-path on the nodes k, $\tilde{k}$, i. Due to the monotonicity properties of the r-distance, cf. Lemma 2.3.14, it holds $R_{ij}^{+} < R_{ij}$, where R_{ij}^{+} and R_{ij} are the resistive distance between i and j in the circuit displayed here and in Figure 3.3, respectively. Precisely, we have $R_{ij}^{+} = G_2^{-1} + \frac{1}{G_0+G_1}$ in this case. Note that in the nonlinear case, the conductances G_i are state-dependent.

the circuit DAE normal form as Laplacians of such contracted graphs, and the resulting Theorem 3.5.5, is a new result.

The following chapter deals with waveform relaxation and certain *source couplings* of coupled circuits. The convergence analysis therein is based on the results of this section. In particular, the relevance of precisely the sensitivity of the voltage over a perturbed current source can be seen there.

Equally, the sensitivity of a current through a perturbed voltage source would be of interest. However, results comparable to Theorems 3.5.3 and 3.5.5, but for perturbed voltage sources instead of perturbed current sources, could not yet be achieved.

3.6. Conclusion

The main results of this chapter are presented in Sections 3.4 and 3.5.

A DAE normal form for nonlinear circuit DAEs of index two, with a locally Lipschitz continuous right hand side, is presented in Section 3.4, which furthermore provides a an a priori estimate for solutions and a global existence and uniqueness result. The transient perturbation analysis in Section 3.5 offers a systematic study of a circuit's sensitivity with respect to its independent sources. Both sections are concluded by discussions of the results.

Part II.

Waveform relaxation

4. Convergence of waveform relaxation on coupled DAEs and coupled circuits

The modeling of a large number of today's applications is increasingly complex and in many cases calls for a multiphysical approach resulting in coupled systems. *Waveform relaxation* (WR) methods (also called dynamic iteration in literature) are well-established iterative methods for coupled dynamical problems. They allow for simulator coupling, hence for each subsystem to be solved by a dedicated numerical solver taking into account the different structure and time scales of the subsystems. Another situation which requires simulator coupling is when proprietary blackbox solvers for coupled physical problems such as field/circuit problems or coupled power/gas networks are involved.

WR methods can be seen as generalization, or enhancement, of the well-known Picard iteration. They were first introduced in 1982 by [Lel82] and [LRS82] in the context of electrical circuits, where they have, under certain assumptions, proven to be particularly efficient [WV12].

For Lipschitz continuous ODE problems (on bounded time intervals), WR methods are known to be unconditionally convergent [Lel82, Bur95]. However, this is not true in general in the case of DAEs. Therefore, a number of studies were dedicated to finding convergence criteria for different classes of coupled DAEs, e.g. [Lel82, LRS82, WOSR85, MN87, Sch91, CI94, JK96, Mie00, JW01, AG01, Bar04, SZF06, Ebe08, SGB10, Schö11, WV12, BBGS13]. A discussion of these articles in comparison to the results presented here can be found at the end of Section 4.1.

Generally, for a given coupled problem, the convergence criteria of the mentioned works have in common that checking if they are satisfied can be costly; it usually requires the computation of certain Lipschitz constants in the system, which, loosely speaking, can be determined via the norm of (inverses of) certain Jacobian matrices of system functions. Furthermore, the coupled DAE is not necessarily given in the required (normal) form.

For that reason, we shall look at the question of WR convergence on coupled circuit DAEs from a novel perturbation-based point of view, strongly relying on the circuit topological analysis and perturbation results of the previous chapter, notable Section 3.5. That way, we are able to provide sufficient convergence criteria in terms of the circuit topological relations between the coupling nodes, that is, the contact points between the subsystems. The coupling conditions which guarantee convergence are less restrictive than in previous literature discussing topological WR convergence criteria, cf. e.g. [AG01, Bar04, BBGS13].

The guiding questions to be adressed throughout this chapter shall remain the following:

- What are sufficient conditions on the coupled (circuit) DAEs to guarantee WR convergence?
- At which rate does WR converge if the convergence conditions are satisfied?

The chapter is organized as follows:

Section 4.1, apart from setting up the preliminaries, introduces the WR algorithm and provides a convergence result for WR on coupled DAEs of higher index, where no circuit structure is assumed yet. Section 4.2 combines this convergence result with the circuit analysis of the previous chapter, leading to topological convergence criteria for coupled circuits. Sections 4.1 and 4.2 are both concluded by discussions of the results. In Section 4.3, we present WR convergence criteria for a coupled field/circuit model. Finally, numerical simulations of toy examples confirming our theoretical results of Sections 4.2 and 4.3 are presented in the fourth section. The simulations also show that WR can fail indeed if the convergence criteria are not satisfied.

4.1. WR on coupled DAEs

Not adressing circuit related questions yet, this section shall introduce the WR algorithm and provide a convergence result for coupled DAEs.

Preliminaries A Lipschitz continuous function $f : D_n \to \mathbb{R}^n$, where $D_n \subseteq \mathbb{R}^n$, is *contractive*, if its Lipschitz constant is smaller one, that is,

$$\exists\, 0 < L < 1 : \ |f(x_1) - f(x_2)| \leq L|x_1 - x_2| \quad \forall x_1, x_2 \in D_n.$$

A sequence $(x^k)^{k\in\mathbb{N}} \in X$ in a Banach space $(X, \|\cdot\|_X)$ has *linear rate of convergence* c, if it is convergent to x^* and

$$\exists\, 0 < c < 1 : \ \|x^k - x^*\|_X \leq c\|x^{k-1} - x^*\|_X \quad \forall\, 1 \leq k \in \mathbb{N},$$

and it has rate of convergence $\sqrt{c}$, if

$$\exists\, 0 < c < 1 : \ \|x^k - x^*\|_X \leq c\|x^{k-2} - x^*\|_X \quad \forall\, 2 \leq k \in \mathbb{N}.$$

Remark. For equivalent norms, Lipschitz continuity is norminvariant. However, contractivity is not, since the magnitude of the Lipschitz constant L is generally norm-dependent. That is, a (Lipschitz continuous) function may be contractive with respect to one norm and not contractive with respect to another (equivalent) norm. Hence, the notion of contractivity implies the choice of a fixed norm.

For practical reasons we change our notational convention for time derivatives: While we denoted the derivative of a differentiable function $x : \mathcal{I} \to \mathbb{R}^n$ by x' in in the previous chapters, we now denote

$$\frac{d}{dt}x(t) = \dot{x}(t).$$

In the previous chapters, this dot notation for the derivative would have been unsound when applied to the often needed letter i. Now, however, we need to "make space" for the superscript indicating the iteration step of WR methods, which is compatible with dot notation but would be conflicting with prime notation of derivatives.

4.1.1. The WR algorithm

Here we introduce the WR algorithm. Before we get to DAEs, we have a look at coupled ODEs first.

For an ODE $\dot{y} = f(y)$ with an initial value $y_0 = y(t_0)$ which is given in the partitioned form

$$\begin{aligned} \dot{y}_1 &= f_1(y_1, y_2), \qquad & y_{1,0} &= y_1(t_0), \\ \dot{y}_2 &= f_2(y_1, y_2), \qquad & y_{2,0} &= y_2(t_0), \end{aligned}$$

the Picard iteration reads

$$\begin{aligned} \dot{y}_1^k &= f_1(y_1^{k-1}, y_2^{k-1}), \qquad & y_{1,0} &= y_1(t_0), \\ \dot{y}_2^k &= f_2(y_1^{k-1}, y_2^{k-1}), \qquad & y_{2,0} &= y_2(t_0), \end{aligned}$$

where $k = 0, 1, 2, \ldots$ indicates the iteration step. To get the iteration started, we need initial guess functions y_1^0 and y_2^0. Note that these are not initial values, but functions over the entire time interval in question. Usually, they are chosen as constant extrapolation of the given initial values.

While the Picard iteration is convergent (for ODEs), the slow rate of convergence makes the use of this technique inefficient [Bur95]. Therefore, WR methods rely on more current iterates in the right hand side. That is, the *Jacobi WR* reads

$$\begin{aligned} \dot{y}_1^k &= f_1(y_1^k, y_2^{k-1}), \qquad & y_{1,0} &= y_1(t_0), \\ \dot{y}_2^k &= f_2(y_1^{k-1}, y_2^k), \qquad & y_{2,0} &= y_2(t_0), \end{aligned}$$

and the *Gauss-Seidel WR* (GS WR) reads

$$\dot{y}_1^k = f_1(y_1^k, y_2^{k-1}), \qquad y_{1,0} = y_1(t_0),$$
$$\dot{y}_2^k = f_2(y_1^k, y_2^k), \qquad y_{2,0} = y_2(t_0).$$

We notice that in the Gauss-Seidel WR, using the current iterate y_1^k in the second subsystem requires the solution of the first subsystem. In other words, the trade-off is parallelism against (the hope for) a faster rate of convergence here. Indeed, for coupled ODEs and certain classes of coupled DAEs, it has been shown that if both the GS WR and the Jacobi WR converge, then the GS WR has a better rate of convergence, e.g. [Bur95, JK96].

Applying the Picard iteration straightforwardly to DAEs does not work, as for a general linear DAE $E\dot{x} + Fx = g$ we obtain for in the k-th iteration step the equation

$$E\dot{x}^k + Fx^{k-1} = g(t),$$

which we have to resolve for x^k. For E singular, the problem is ill-posed since matrix pencil $\{E, 0\}$ is singular.

Hence, we focus solely on the Jacobi and GS WR, who can be used straightforwardly on DAEs if mild assumptions on the subsystems are satisfied.

Coupled DAE problem For simplicity, we shall restrict ourselves on (no more than) two subsystems in the remainder of this work. They shall be given in an input-output formulation on a compact time interval $\mathcal{I} = [t_0, T]$ as follows:

$$f_1(\dot{x}_1, x_1, s_2, t) = 0, \qquad p_1(x_1) = s_1, \qquad A_1 x_1(t_0) = x_{1,0}^{IV} \tag{4.1a}$$
$$f_2(\dot{x}_2, x_2, s_1, t) = 0, \qquad p_2(x_2) = s_2, \qquad A_2 x_2(t_0) = x_{2,0}^{IV}. \tag{4.1b}$$

Here, f_1 and f_2 represent the subsystems, $p_1(x_1) = s_1$ is the output of the first subsystem and the input of the second subsystem, and vice versa for $p(x_2) = s_2$. That way, the output equations are coupling conditions. Depending on the viewpoint, we refer to s_1 as output of the first subsystem or *coupling input* of the second subsystem. The matrices $A_1 \in \mathbb{R}^{n\times m}, n < m$ and $A_2 \in \mathbb{R}^{k\times l}$, $k < l$ are matrices which take the role of A in Corollary 3.4.6 for each subsystem. Then, $x_{1,0}^{IV}$ and $x_{2,0}^{IV}$ are those initial values which are not fixed by algebraic constraints.

Waveform relaxation iteration schemes The iteration scheme of the GS WR method applied to the coupled DAE (4.1) reads

$$f_1(\dot{x}_1^k, x_1^k, s_2^{k-1}, t) = 0, \qquad p_1(x_1^k) = s_1^k, \qquad A_1 x_1(t_0) = x_{1,0}^{IV} \tag{4.2a}$$
$$f_2(\dot{x}_2^k, x_2^k, s_1^k, t) = 0, \qquad p_2(x_2^k) = s_2^k, \qquad A_2 x_2(t_0) = x_{2,0}^{IV} \tag{4.2b}$$

where k denotes again the iteration index. The Jacobi WR iteration scheme reads

$$f_1(\dot{x}_1^k, x_1^k, s_2^{k-1}, t) = 0, \qquad p_1(x_1^k) = s_1^k, \qquad A_1 x_1(t_0) = x_{1,0}^{IV} \tag{4.3a}$$
$$f_2(\dot{x}_2^k, x_2^k, s_1^{k-1}, t) = 0, \qquad p_2(x_2^k) = s_2^k \qquad A_2 x_2(t_0) = x_{2,0}^{IV}. \tag{4.3b}$$

4.1.2. Convergence

We shall relate the convergence behaviour to the subsystems' sensitivity to perturbations of the input functions. Therefore, we examine the perturbation sensitivity of the two subproblems in (4.1) independently of one another: We consider

$$f_1(\dot{x}_1, x_1, s_2, t) = 0, \qquad p_1(x_1) = x_{1out}, \qquad A_1 x_1(t_0) = x_{1,0}^{IV} \tag{4.4}$$

and the perturbed problem

$$f_1(\dot{x}_1^\delta, x_1^\delta, s_2 + \delta_{s_2}, t) = 0, \qquad p_1(x_1^\delta) = x_{1out}^\delta, \qquad A_1 x_1^\delta(t_0) = (x_{1,0}^{IV})^\delta. \tag{4.5}$$

Analogously, we consider the second subproblem independently:

$$f_2(\dot{x}_2, x_2, s_1, t) = 0, \qquad p_2(x_2) = x_{2out}, \qquad A_2 x_2(t_0) = x_{2,0}^{IV} \tag{4.6}$$

and the perturbed problem

$$f_2(\dot{x}_2^\delta, x_2^\delta, s_1 + \delta_{s_1}, t) = 0, \qquad p_2(x_2^\delta) = x_{2out}^\delta, \qquad A_2 x_2(t_0)^\delta = (x_{2,0}^{IV})^\delta. \tag{4.7}$$

We assume that the subproblems are uniquely solvable.

Assumption 4.1.1 *Consider the IVPs (4.4) and (4.6), where the functions f_1, f_2, p_1, p_2 are the same as in the coupled DAE (4.1).*

The output functions p_1 and p_2 and the inputs s_1 and s_2 are continuous.

Furthermore, the IVP (4.4) has a unique solution $(x_1, x_{1out}) \in C([t_0, T], \mathbb{R}^n)$ for all $T > t_0$, all initial values $x_{1,0}^{IV}$ and all (continuous) inputs s_2, and analogously, the IVP (4.6) has a unique solution $(x_2, x_{2out}) \in C([t_0, T], \mathbb{R}^n)$ for all $T > t_0$, all initial values $x_{2,0}^{IV}$ and all (continuous) inputs s_1.

Note that this Assumption implies also the existence of global and unique solutions for the coupled IVP (4.1).

Assumption 4.1.2 *We consider the IVPs (4.4)-(4.7), where the functions f_1, f_2, p_1, p_2 are the same as in the coupled DAE (4.1). Let $x_1^{IV} = (x_1^{IV})^\delta$ and $x_2^{IV} = (x_2^{IV})^\delta$. Furthermore, let $|\cdot|$ be a vector norm and the ball $B(T, M)$ be defined as in Equation (3.55).*

If Assumption 4.1.1 holds, then for all $T > t_0$ and all $M > 0$, there exist constants $c > 0$ and $k_1, k_2 \geq 0$, $k_1 k_2 < 1$ such that for all $t \in [t_0, T]$ and all $\delta_{s_1}, \delta_{s_2} \in B(T, M)$, the following four conditions hold :

$$|x_1^\delta(t) - x_1(t)| \leq c \left(\int_{t_0}^t |\delta_{s_2}(\tau)| d\tau + |\delta_{s_2}(t)| \right) \tag{4.8}$$

$$|x_{1out}^\delta(t) - x_{1out}(t)| \leq c \int_{t_0}^t |\delta_{s_2}(\tau)| d\tau + k_1 |\delta_{s_2}(t)| \tag{4.9}$$

$$|x_2^\delta(t) - x_2(t)| \leq c \left(\int_{t_0}^t |\delta_{s_1}(\tau)| d\tau + |\delta_{s_1}(t)| \right) \tag{4.10}$$

$$|x_{2out}^\delta(t) - x_{2out}(t)| \leq c \int_{t_0}^t |\delta_{s_1}(\tau)| d\tau + k_2 |\delta_{s_1}(t)| \tag{4.11}$$

Remark. Note that for given T and M, the constants c, k_1, k_2 are independent of $t \in [t_0, T]$ and $\delta_{s_1}, \delta_{s_2} \in B(T, M)$.

Remark. The special cases where $k_1 = 0$ or $k_2 = 0$ shall play an important role in the following.

Using the same informal but intuitive terminology as in Section 3.5, we could rephrase the Conditions (4.8) and (4.10) as "both subsystems are index-one sensitive with respect to their coupling inputs." Note that this does not imply anything about the index of the subsystems or the coupled DAE (4.1). The sensitivity to the coupling inputs expresses precisely what determines the convergence behaviour of WR, which in each iteration step can be viewed as a problem with perturbed coupling inputs. The special case that $k_1 = 0$ or $k_2 = 0$ could be informally reformulated as "one subsystem's output is ODE sensitive to the coupling input".

Theorem 4.1.3 (Convergence of WR on coupled DAEs) *We consider the coupled problem* (4.1) *and the corresponding GS WR* (4.2) *and Jacobi WR* (4.3)*. Let $\|\cdot\|_{[t_0,t]}$ be the maximum norm on $C([t_0, T], \mathbb{R}^n)$. Furthermore, let Assumption 4.1.1 and Assumption 4.1.2 hold, and let c, k_1, k_2, $k_1 k_2 < 1$ be the constants from 4.1.2, and let $\hat{c} = c(c + k_1 + k_2)$.*

Then, for all $T \in \mathcal{I}_{con} := \left[t_0, \frac{1-k_1 k_2}{\hat{c}} + t_0\right)$ and all continuous initial guess functions (s_1^0, s_2^0), it holds:

1. *The GS sequence x^k, defined by Iteration* (4.2)*, converges in $(C([t_0, T], \mathbb{R}^n), \|\cdot\|_{[t_0,T]})$ to the solution x of* (4.1) *and has linear rate of convergence $H_T \hat{c} + k_1 k_2$, where $H_T := T - t_0$.*

2. *The Jacobi sequence x^k, defined by Iteration* (4.3)*, converges in $(C([t_0, T], \mathbb{R}^n), \|\cdot\|_{[t_0,T]})$ to the solution x of* (4.1) *and has rate of convergence $\sqrt{H_T \hat{c} + k_1 k_2}$.*

Remark. Note that $T < \frac{1-k_1 k_2}{\hat{c}} + t_0$ implies that the contraction factor $H_T \hat{c} + k_1 k_2$ is smaller one indeed.

Proof: We first show convergence of the Jacobi sequence. The convergence proof for the GS sequence is kept short since it works analogously to the Jacobi part.

Jacobi sequence: We start with the Jacobi WR. Let $T \in \mathcal{I}_{con}$ arbitrary but fixed, and let (x_1, x_2, s_1, s_2) be the solution of the coupled DAE (4.1). Furthermore, let x_1^k, s_1^k be the solution of (4.3a) for given input s_2^{k-1}, and let x_2^k, s_2^k be the solution of (4.3b) for given input s_1^{k-1}. Assumption 4.1.1 makes sure that continuity of s_1^{k-1}, s_2^{k-1} implies continuity of s_1^k, s_2^k. For fixed k, Equations (4.3a) and (4.3b) can be seen as independent problems with source perturbations $\delta_{s_2}^k := s_2^k - s_2$ and $\delta_{s_1}^k := s_1^k - s_1$. Let $k \geq 2$ be fixed. We choose M such that $\delta_{s_1}^{\tilde{k}}, \delta_{s_2}^{\tilde{k}} \in B(T, M)$ for all $\tilde{k} \in \mathbb{N}, \tilde{k} < k$. This is possible since $\delta_{s_i}^{\tilde{k}}$

are continuous functions on a compact interval. By Condition (4.11) from Assumption 4.1.2, we have for all $t \in [t_0, T]$

$$\left|s_2^k(t) - s_2(t)\right| \le c \int_{t_0}^{t} \left|s_1^{k-1}(\tau) - s_1(\tau)\right| d\tau + k_2 \left|s_1^{k-1}(t) - s_1(t)\right|,$$

and an analogous estimate for $|s_1^k(t) - s_1(t)|$ by Condition (4.9). This implies

$$\begin{aligned}\left\|s_2^k - s_2\right\| &\le c \left\| \int_{t_0}^{\cdot} \left|s_1^{k-1}(\tau) - s_1(\tau)\right| d\tau \right\| + k_2 \left\|s_1^{k-1} - s_1\right\| \\ &\le cH_T \left\|s_1^{k-1} - s_1\right\| + k_2 \left\|s_1^{k-1} - s_1\right\|. \end{aligned} \tag{4.12}$$

for the maximum norm $\|\cdot\|$ on $[t_0, T]$ which is based on the vector norm $|\cdot|$. Exploiting Condition (4.9) and inserting (4.12) then yields

$$\begin{aligned}\left\|s_1^k - s_1\right\| \le& cH_T \left\|s_2^{k-1} - s_2\right\| + k_1 \left\|s_2^{k-1} - s_2\right\| \\ \le& c^2 H_T^2 \left\|s_1^{k-2} - s_1\right\| + ck_2 H_T \left\|s_1^{k-2} - s_1\right\| + ck_1 H_T \left\|s_1^{k-2} - s_1\right\| \\ &+ k_1 k_2 \left\|s_1^{k-2} - s_1\right\| \\ =& \left(H_T c(H_T c + k_1 + k_2) + k_1 k_2\right) \left\|s_1^{k-2} - s_1\right\| \\ \le& \left(H_T \hat{c} + k_1 k_2\right) \left\|s_1^{k-2} - s_1\right\|. \end{aligned} \tag{4.13}$$

Now the precondition $T \in \mathcal{I}_{con}$ yields

$$T < t_0 + \frac{1 - k_1 k_2}{\hat{c}} \qquad \Longleftrightarrow \qquad H_T \hat{c} + k_1 k_2 < 1.$$

Since $(s_2^{k-2} - s_2) \in B(T, M)$, it follows that $(s_2^k - s_2) \in B(T, M)$. Analogously we obtain $(s_1^k - s) \in B(T, M)$.

This implies that Conditions (4.9) and (4.11) still hold in next iteration step $k+1$ (with the same constants c, k_1, k_2).

Hence, the sequences given by s_1^k and s_2^k converge to s_1 and s_2 in $(C([t_0, T], \mathbb{R}^n, \|\cdot\|)$ with rate of convergence $\sqrt{H_T \hat{c} + k_1 k_2}$. Conditions (4.8) and (4.10) then imply convergence with the same rate for x_1^k and x_2^k, since

$$|x_1^k(t) - x_1(t)| \le c \left(\int_{t_0}^{t} |s_2^k(\tau) - s_2(\tau)| d\tau + |s_2^k(t) - s_2(t)| \right)$$

implies

$$\begin{aligned}\|x_1^k - x_1\| &\le c \left(\left\| \int_{t_0}^{\cdot} |s_2^k(\tau) - s_2(\tau)| d\tau \right\| + \|s_2^k - s_2\| \right) \\ &\le cH_T \|s_2^k - s_2\| + c\|s_2^k - s_2\| = c(1 + H_T)\|s_2^k - s_2\| =: \tilde{c}\|s_2^k - s_2\|. \end{aligned}$$

GS sequence: As this proof is analogous to the Jacobi case, we keep it short. We choose again $T \in I_{con}$ arbitrary and M such that $\delta_{s_1}^{\tilde{k}}, \delta_{s_2}^{\tilde{k}} \in B(T, M)$ for all $\tilde{k} \in \mathbb{N}, \tilde{k} < k$. Conditions (4.9) and (4.11) yield

$$|s_2^{k-1} - s_2| \leq c \int_{t_0}^{t} \left|s_1^{k-1} - s_1\right| d\tau + k_2 \left|s_1^{k-1} - s_1\right|,$$

and

$$\left|s_1^k - s_1\right| \leq c \int_{t_0}^{t} \left|s_2^{k-1} - s_2\right| d\tau + k_1 \left|s_2^{k-1} - s_2\right|,$$

leading to an estimate

$$\left\|s_1^k - s_1\right\| \leq (H_T \hat{c} + k_1 k_2) \left\|s_1^{k-1} - s_1\right\|.$$

Since by precondition $H_T \hat{c} + k_1 k_2 < 1$, this makes sure that $(s_1^k - s_1) \in B(T, M)$. Hence, we obtain an analogous estimate for $\|s_2^k - s_2\|$.

This means the sequences s_1^k and s_2^k converge with linear rate of convergence $H_T \hat{c} + k_1 k_2$. Conditions (4.8) and (4.10) then imply convergence with the same rate for x_1^k and x_2^k. □

Discussion of Theorem 4.1.3 and state of the art

Here we discuss the convergence Theorem 4.1.3 and its preconditions and compare it to the state of the art in the field.

First, we notice that the considered coupled DAE form (4.1) has a high level of generality. It is implicit and nonlinear, and no assumptions concerning the index of the coupled DAE or its subsystems are made. Furthermore, (global) Lipschitz continuity of f is not required. Also, the crucial convergence Criteria (4.8)-(4.11) from Assumption 4.1.2 do not imply a restriction on the index of the coupled DAE or its subsystems. However, the fact that (4.8)-(4.11) hold on a bounded ball $B(T, M)$ of arbitrary size is usually related to vector fields f_1, f_2 which are locally Lipschitz continuous on the whole euclidean space $\mathbb{R}^n$. Compared to the literature in the field, this level of generality for the DAE class is exceptional.

Among the articles investigating WR convergence criteria for *higher index DAEs*, the DAE class is strongly restricted to linear DAEs [Mie00], pseudo-dynamic Hessenberg DAEs [SZF06] or DAEs in a semi-explicit normal form which satisfies a number of global Lipschitz and contractivity conditions [CI94]. A main challenge for the latter article here is to compute these Lipschitz constants (and the normal forms in the first place).

Furthermore, numerous relevant articles studied coupled *semi-explicit DAEs of index one*. The seminal works of Lelarasmee [Lel82, LRS82] as well as [WOSR85] examined

DAEs with global Lipschitz and contractivity conditions. Two years later, [Sch91] reformulated the convergence criteria of [Lel82] in an attempt to make the easier to check. The works of [JK96, JW01] carried out the convergence proofs in a functional analytical (operator-based) framework in order to admit generalizations concerning DAEs with delay. These mentioned works have in common that they require globally Lipschitz continuous vector fields describing the coupled DAEs, resulting in convergence results on time intervals of arbitrary size.

Semi-explicit DAEs of index one satisfying a contractivity condition and a local Lipschitz condition are considered in [Ebe08], who assumed that the DAEs can be reduced to their inherent ODEs, [AG01] with a focus on mechanical applications, [Schö11] and [BBGS13]. The local Lipschitz condition leads to convergence results on sufficiently small time intervals as is the case in the present work. Furthermore, the last two mentioned works provide applications to coupled circuits, notably field/circuit systems, and [Ebe08] offers a lot of formal detail for the (WR) partitioning of coupled circuits and adresses practical challenges for the simulation.

Finally, [AG01, Ebe08, BBGS13, Schö11] also adress the *error propagation from window to window*: As Theorem 4.1.3 in the present work as well as the results in the mentioned articles show, convergence can considerably accelerated by reducing the length of the time interval over which we integrate. This insight led to the common practice of *windowing* in WR simulation, which means the time interval $\mathcal{I}$ of interest is partitioned into the smaller intervals $\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_n$. Then, WR is applied on $\mathcal{I}_1$, and after a certain number of iterations, the computed data is used to create initial guesses on $\mathcal{I}_2$ and the WR process starts there, and so on. Naturally, this raises the question of how sensitive the WR method reacts to faulty initial data and if the convergence acceleration in a single window dominates the error resulting from faulty initial data.

The coupled DAE (4.1) along with the convergence Conditions (4.8)-(4.11), in contrast to the mentioned articles, does not (explicitly) require a restriction of the DAE index, a DAE given in a normal form, or a Lipschitz condition. We believe this formulation is therefore benefitial, since it identifies that the convergence behaviour of a WR method is exclusively determined by the sensitivity of the subsystems to the coupling input. This sensitivity is limited by the perturbation index; however, the perturbation index, admitting general perturbations, is not a sharp measure when a specific input perturbation is considered. In short, the DAE can have a high index on the one hand, and at the same time be only mildly sensitive with respect to perturbations of a specific input.

For semi-explicit index one DAEs, Conditions (4.8)-(4.11) and $k_1 k_2 < 1$ in Assumption 3.4.3 are comparable to the criterion offered for instance in [Schö11, BBGS13], since they are usually derived by means of a local Lipschitz condition, and the contractivity condition here and therein both limit the impact of non-dynamic coupling inputs.

All of the mentioned articles, including the present work until this point, have in common that for a given coupled DAE, it is generally hard to check (that is, compute) if the respective contractivity condition is satisfied. If a crucial contractivity constant is zero, this implies a better rate of WR convergence on the one hand and it is generally easier to determine than a nonzero contractivity constant. For both reasons, it is recommendable to model coupled problems in practice such that such the mentioned contractivity constant is zero, which in our case means $k_1 = 0$ or $k_2 = 0$, where k_1 and k_2 are the constants from Conditions (4.8)-(4.11).

Bearing in mind the practical problem of determining if the contractivity condition is satisfied, we designed the Conditions (4.8)-(4.11) to be of the same form as the perturbation results for circuits in Section 3.5. For coupled circuits (that is, coupled DAEs whose at least one subsystem is a circuit), this offers the opportunity to check if the conditions are satisfied by purely topological criteria regarding the position of the coupling inputs in the circuit. The next Section shall elaborate on that.

4.2. WR on coupled circuits

In this section, we shall combine the perturbation results of Section 3.5 with the Convergence Theorem 4.1.3 to derive topological convergence criteria for coupled circuit/circuit problems of index two. Before we come to the main results in Subsection 4.2.2, we briefly introduce the coupled circuit model.

4.2.1. Partitioned model and WR algorithm

We employ a *source coupling*, as illustrated in Figures 4.1 and 4.2. In the following, we briefly introduce the source coupling. For a detailed description of this coupling model we refer to [Ebe08, Section 4.3]. In terms of the circuit graph $G = (V, E)$, such a source coupling is based on an (arbitrary) edge partition $\bar{E} \subseteq E$, $\mathring{E} \subseteq E$, and a resulting node partition $\bar{V}, \mathring{V}, V_m$, where $\bar{V}$ is the set of nodes which are incident with edges of *only* $\bar{E}$, $\mathring{V}$ is the set of nodes which are incident with edges of *only* $\mathring{E}$, and V_m is the set of *coupling nodes* $\mathring{V}_m$, which are those that have incident edges in both $\bar{E}$ and $\mathring{E}$.

The idea of source coupling is to introduce certain dummy sources, which we shall refer to as *coupling inputs*, into the circuit (see Figures 4.1 and 4.2). That way, the coupled subsystems of the partitioned circuit can be interpreted as circuits again. Formally, we duplicate the coupling nodes to have them as independent nodes in each subsystem. This can be seen well in the Examples in Figures 4.1 and 4.2.

The resulting coupled model of the IVP then reads

$$\bar{f}(\dot{\bar{x}}, \bar{x}, t) + \bar{\mathcal{A}}_m^* \bar{\imath}_m = 0, \quad \bar{\mathcal{A}}_m^\top \bar{e} = \mathring{v}_m \qquad \bar{p}(\bar{x}, \bar{\imath}_m) = \bar{\imath}_m, \qquad \bar{A}\bar{x}(t_0) = \bar{x}_0^{IV} \tag{4.14a}$$

$$\mathring{f}(\dot{\mathring{x}}, \mathring{x}, t) = \mathring{\mathcal{A}}_m^* \bar{\imath}_m, \qquad \mathring{p}(\mathring{x}) = \mathring{\mathcal{A}}_m^\top \mathring{e} =: \mathring{v}_m, \qquad \mathring{A}\mathring{x}(t_0) = \mathring{x}_0^{IV} \tag{4.14b}$$

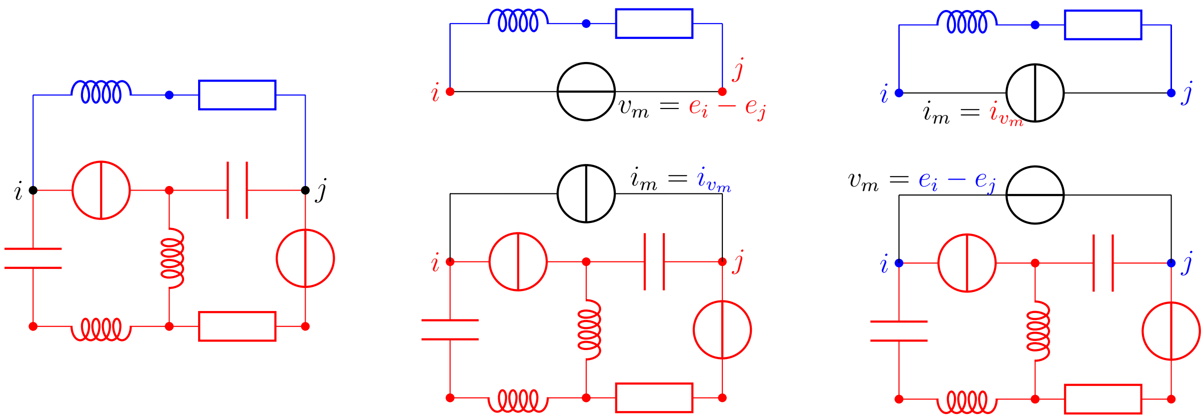


Figure 4.1.: The picture on the left shows an (arbitrary) circuit which is partitioned into a blue and a red subcircuit. The coupling nodes are i and j with node potentials e_i and e_j. The middle and the right picture depict source coupling models of the partitioned circuit. In the middle, the coupling nodes are internal unknowns of the red subsystem and externally given input variables for the blue subsystem, whereas the right picture shows the converse coupling situation.

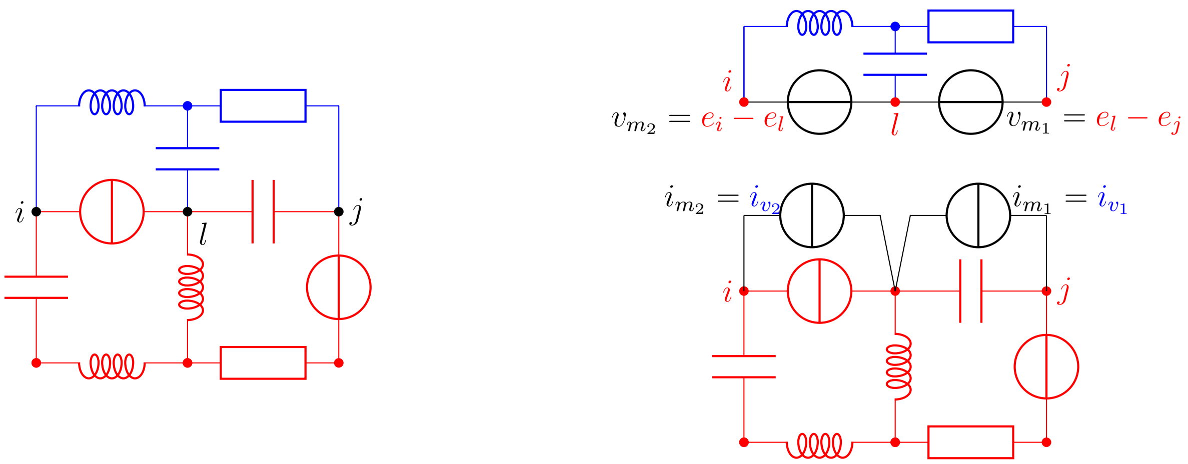


Figure 4.2.: The left picture displays an (arbitrary) circuit which is partitioned into a blue and a red subcircuit. The right picture shows a possible corresponding source coupling.

Both subsystems (4.14a) can be understood as the MNA of a circuit; however, the coupling voltage $\mathring{v}_m$ of (4.14a) and the coupling current $\bar{i}_m$ of (4.14b) are not independent, but computed as internal variable of the other subsystem; $\bar{i}_m$ is, as an internal unknown, the (vector of) current flow between two coupling nodes in Subsystem (4.14a). At the same time, it is the coupling current in Subsystem (4.14b). Analogously, $\mathring{v}_m$ is the coupling voltage in Subsystem (4.14a) and, as an internal unknown, the voltage across the coupling nodes in Subsystem (4.14b).

That way, these coupling terms $\bar{i}_m$ and $\mathring{v}_m$ describe the input that Subsystem (4.14a) receives from Subsystem (4.14b), namely $\bar{i}_m$, and conversely the input that (4.14b) receives from (4.14a), namely $\mathring{v}_m$.

The matrix $\bar{\mathcal{A}}_m^*$ is defined by $\bar{\mathcal{A}}_m^* = \left(\begin{smallmatrix}\bar{\mathcal{A}}_m \\ 0\end{smallmatrix}\right)$, where $\bar{\mathcal{A}}_m$ is the incidence matrix describing the coupling currents.The number of columns of $\bar{\mathcal{A}}_m$ is $n-1$, if n is the number of coupling nodes. Furthermore, $\bar{x} = (\bar{e}, \bar{i}_l, \bar{i}_v)$ contains the node potentials $\bar{e}$ corresponding to the (duplicated) coupling nodes $\mathring{V}_m$ and the nodes $\bar{V}$ and currents i corresponding to edges of $\bar{E}$; the corresponding KCL and element equations are described by $\bar{f} + \bar{\mathcal{A}}_m^* \bar{i}_m = 0$. Note, however, that the voltage drop $\bar{\mathcal{A}}_m^\top \bar{e} = \mathring{v}_m$ between two (duplicated) coupling nodes in (4.14a) is forced to be the same as in (4.14b) by the coupling condition. That way, we make sure that the potentials at the coupling nodes and their duplicates are the same. The variable $\mathring{x} = (\mathring{e}, \mathring{i}_l, \mathring{i}_v)$ contains the node potentials and currents corresponding to nodes of $\mathring{V}$ and $\mathring{V}_m$ and edges of $\mathring{E}$, and $\mathring{f} = \mathring{\mathcal{A}}_m^* \bar{i}_m$ describes the corresponding KCL and element equations. Here, $\mathring{\mathcal{A}}_m^* = \left(\begin{smallmatrix}\mathring{\mathcal{A}}_m \\ 0\end{smallmatrix}\right)$ describes the position of the coupling currents in Subsystem (4.14b). Finally, $\mathring{p}$ and $\bar{p}$ are the output functions, which are clearly linear, and $\bar{A}$ and $\mathring{A}$ are matrices as in Corollary 3.4.6, selecting consistent initial values.

Remark. We consider the (monolithic) MNA of a circuit with graph G, and for a given partition the MNA of the two coupled circuits built from G as in Figures 4.1 and 4.2, and we assume that the monolithic and the coupled system both have a unique solution. Let these solutions be denoted by x for the monolithic system and $y_{coupl} = (y, y_{in})$ for the coupled system, where y_{coupl} is sorted such that y_{in} are currents through or voltages across coupling inputs.

Then, the Kirchhoff laws warrant $x = y$. This justifies the coupled model (4.14).

Definition 4.2.1 (Coupling terminology) *Consider the coupled circuit model* (4.14) *and the corresponding circuit graphs. We say that Subsystem* (4.14a) *is* voltage-driven, *and Subsystem* (4.14b) current-driven. *The voltages* $\mathring{v}_m$ *are called* coupling voltages, *and the currents* $\bar{i}_m$ *are called* coupling currents. *We refer to the corresponding edges, whose incidences are described by* $\bar{\mathcal{A}}_m$ *and* $\mathring{\mathcal{A}}_m$, *as* coupling current edges *and* coupling voltage edges.

Remark. Since many of the following examples are coupled circuits with two coupling nodes, it is worth noticing that in this case we have precisely one coupling voltage and one coupling current, that is, $\bar{\mathcal{A}}_m$ and $\mathring{\mathcal{A}}_m$ are column vectors and $\bar{i}_m$ and $\mathring{v}_m$ are scalar values.

For the coupled system, the standard MNA rank assumptions should be met:

Assumption 4.2.2 *We consider a circuit which is partitioned into a voltage-driven and a current-driven subsystem. Any two coupling nodes are not V-connected in the voltage-driven subsystem, and they are CVRL-connected in the current-driven subsystem.*

This assumption guarantees that the rank assumption in 2.1.3 holds for each of the subcircuits.

The Jacobi WR iteration scheme of the coupled circuit model (4.14) reads

$$\bar{f}(\dot{\bar{x}}^k, \bar{x}^k, t) + \bar{\mathcal{A}}_m^* \bar{\imath}_m^k = 0, \quad \bar{\mathcal{A}}_m^\top \bar{e}^k = \mathring{v}_m^{k-1}, \quad \bar{p}(\bar{x}^k, \bar{\imath}_m^k) = \bar{\imath}_m^k, \qquad \bar{A}\bar{x}^k(t_0) = \bar{x}_0^{IV} \tag{4.15a}$$

$$\mathring{f}(\dot{\mathring{x}}^k, \mathring{x}^k, t) = \mathring{\mathcal{A}}_m^* \bar{\imath}_m^{k-1}, \qquad \mathring{p}(\mathring{x}^k) = \mathring{\mathcal{A}}_m^\top \mathring{e}^k =: \mathring{v}_m^k, \qquad \mathring{A}\mathring{x}^k(t_0) = \mathring{x}_0^{IV}, \tag{4.15b}$$

and the GS WR scheme reads

$$\bar{f}(\dot{\bar{x}}^k, \bar{x}^k, t) + \bar{\mathcal{A}}_m^* \bar{\imath}_m^k = 0, \quad \bar{\mathcal{A}}_m^\top \bar{e}^k = \mathring{v}_m^{k-1}, \quad \bar{p}(\bar{x}^k, \bar{\imath}_m^k) = \bar{\imath}_m^k, \qquad \bar{A}\bar{x}^k(t_0) = \bar{x}_0^{IV} \tag{4.16a}$$

$$\mathring{f}(\dot{\mathring{x}}^k, \mathring{x}^k, t) = \mathring{\mathcal{A}}_m^* \bar{\imath}_m^{k}, \qquad \mathring{p}(\mathring{x}^k) = \mathring{\mathcal{A}}_m^\top \mathring{e}^k =: \mathring{v}_m^k, \qquad \mathring{A}\mathring{x}^k(t_0) = \mathring{x}_0^{IV}. \tag{4.16b}$$

4.2.2. Convergence

We briefly review Theorems 3.5.1, 3.5.3 and 3.5.5, which adressed the perturbation sensitivity to inputs in Section 3.5, and the convergence Theorem 4.1.3 for coupled DAEs. Combined, they directly imply topological convergence criteria for coupled circuits.

First, we interprete the iterates $\mathring{v}_m^k = \mathring{v}_m + \mathring{v}_\delta^k$ and $\bar{\imath}_m^k = \bar{\imath}_m + \bar{\imath}_\delta^k$ from the WR Schemes (4.15) and (4.16) as values of perturbed sources, which implies the definitions

$$\mathring{v}_\delta^k := \mathring{v}_m^k - \mathring{v}_m \qquad \bar{\imath}_\delta^k := \bar{\imath}_m^k - \bar{\imath}_m$$

for the input perturbations.

That way, we can view the subsystems of the Jacobi scheme (and analogously for GS) as the input-perturbed subsystems of the coupled circuit.

Then, we recall that the Jacobi and GS scheme (4.15) and (4.16) of the coupled circuit (4.14) imply the assumption that the iterative schemes and the coupled circuit have the same dynamic (that is, not algebraically fixed) initial values.

Now we compare the theorems of Section 3.5 with the conditions of the convergence Theorem 4.1.3.

Second statement of Theorem 3.5.1: With the remaining two theorems of the section only offering results about the sensitivity to current sources, this is the only theorem providing results about the circuit's sensitivity to voltage sources (besides current sources, too).

If the second statement of Theorem 3.5.1 holds for one subcircuit, a direct comparison shows that it satisfies an estimate of the form of Condition (4.8) in Assumption 4.1.2. Furthermore, since the coupled circuit (4.14) has linear output functions (defined by $\bar{p}(\bar{x}, \bar{i}_m) = \bar{i}_m$ and $\mathring{p}(\mathring{x}) = \mathring{A}_m^\top \mathring{e}$, where $\mathring{x} = (\mathring{e}^\top, \mathring{i}_l^\top, \mathring{i}_v^\top)^\top$), a subcircuit for which the second statement of Theorem 3.5.1 holds equally satisfies a condition of the form (4.9) for the output with *some* constant k_1 which is not further specified.

A voltage-driven subcircuit satisfies the preconditions of the second statement of 3.5.1, if no coupling voltage forms a loop with CV-edges, and a current-driven subcircuit satisfies the preconditions, if each coupling current forms a loop with CVR-edges.

<u>Theorem 3.5.3</u>: This theorem applies only to current source perturbations, and provides a bound for their impact on the voltage across these perturbed inputs. If the current-driven subcircuit satisfies the precondition, namely "each coupling current forms a loop with CV-edges", then the output satisfies a condition of the form (4.8) with $k_1 = 0$.

These observations immediately yield the following convergence theorem for WR on coupled circuit/circuit systems.

Theorem 4.2.3 (WR convergence for coupled circuit/circuit system) *Let* (4.14) *be the nonlinear MNA of a coupled circuit whose both subcircuits satisfy Assumptions 2.1.3 and 3.4.3.*

Let the coupling inputs of the coupled circuit Equations (4.14) *satisfy the following topological properties:*

(i) No coupling voltage edge forms a loop with CV-edges in the voltage-driven subcircuit which corresponds to Subsystem (4.14a)*. Here, (other) coupling voltage edges are considered V-edges, too.*

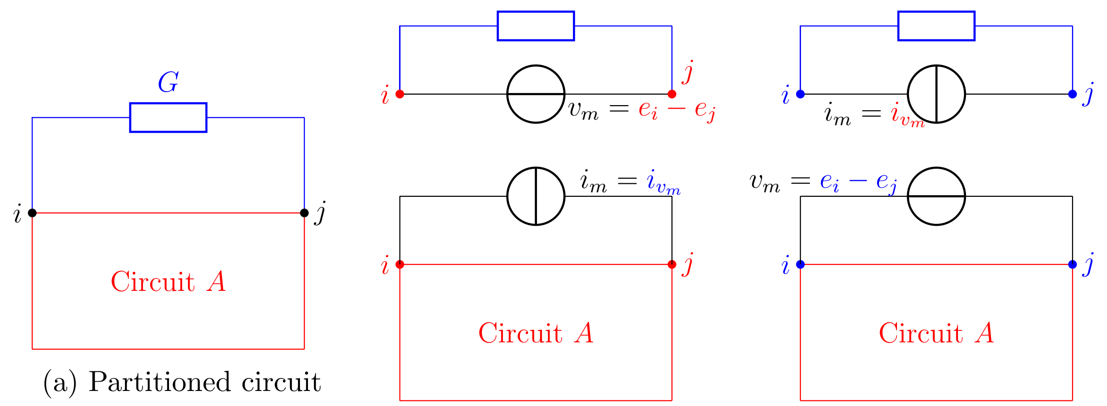
(ii) each coupling current edge forms a loop with CV-edges in the current-driven subcircuit which corresponds to Subsystem (4.14b)*.*

Then, there exists a constant $\hat{c} > 0$ such that for $H_T := T - t_0$ sufficiently small to satisfy $H_T < \frac{1}{\hat{c}}$, it holds:

1. *The GS sequence x^k, defined by Iteration* (4.16)*, converges in $(C([t_0, T], \mathbb{R}^n), \|\cdot\|_{[t_0,T]})$ to the solution x of* (4.14)*, and it has linear rate of convergence $H_T \hat{c}$.*

2. *The Jacobi sequence x^k, defined by Iteration* (4.15)*, converges in $(C([t_0, T], \mathbb{R}^n), \|\cdot\|_{[t_0,T]})$ to the solution x of* (4.14)*, and it has rate of convergence $\sqrt{H_T \hat{c}}$.*

Proof: The considerations preceding the theorem show that if (i) and (ii) are satisfied, the coupled circuit satisfies the Conditions (4.8)-(4.11) with $k_1 \in \mathbb{R}$ and $k_2 = 0$ (or $k_1 = 0$ and $k_2 \in \mathbb{R}$, the order is irrelevant). Hence, the convergence Theorem 4.1.3 applies with $k_1 k_2 = 0$. □



(a) Partitioned circuit

(b) A is voltage-driven

(c) A is current-driven

Figure 4.3.: (a) shows a partitioned circuit with coupling nodes i, j. The blue subcircuit consists only of G and the red subcircuit A is arbitrary. In (b) and (c), the two source coupling options are depicted.

If $k_1 = 0$ or $k_2 = 0$ does not hold, we have to determine both constants in order to check if $k_1 k_2 < 1$ is satisfied. Theorem 3.5.5 allows to compute such a k-constant corresponding to the output voltage of the current-driven subcircuit. This is discussed in Example 4.2.4.

Examples

Here we discuss some prototypical examples of circuit/circuit couplings, which are depicted in Figures 4.3, 4.4 and 4.5. For each of these cases, we briefly discuss convergence criteria for both options of source coupling.

Example 4.2.4 (Circuit/resistance (Figure 4.3)) *We consider the coupled circuit in Figure 4.3a.*

Source coupling leaves us a choice: The resistance can be the voltage-driven subsystem, which implies that subcircuit A is current-driven. This coupling situation is depicted in Figure 4.3b. Conversely, the resistance can be the current-driven subsystem and circuit A voltage-driven. This coupling corresponds to Figure 4.3c.

In Figure 4.3b, where the resistance subsystem is voltage-driven, the convergence Theorem 4.2.3 yields that

- *WR is convergent with rate $\sqrt{Hc}$ (Jacobi) or Hc (GS) if H is sufficiently small and the coupling nodes are CV-connected in subcircuit A*

for some generally unknown (or hard to obtain) constant c, where H denotes the time interval length. Note that according to the theorem, this still holds if the resistance subsystem is replaced by any LRI circuit, or even by any circuit as long as the coupling nodes are not CV-connected in the (sub)circuit replacing the resistance.

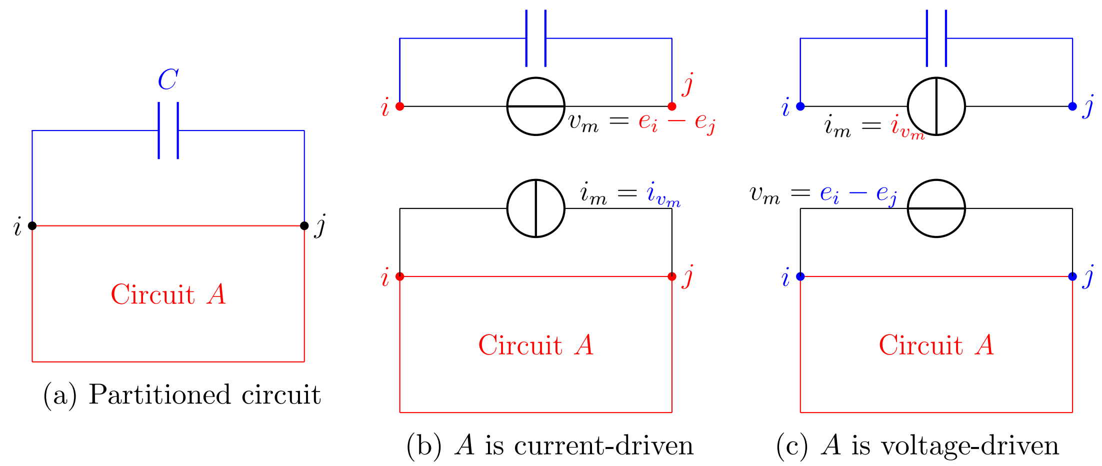


Figure 4.4.: (a) shows a partitioned circuit with coupling nodes i, j. The blue subcircuit consists only of C and the red subcircuit A is arbitrary. In (b) and (c), the two source coupling options are depicted.

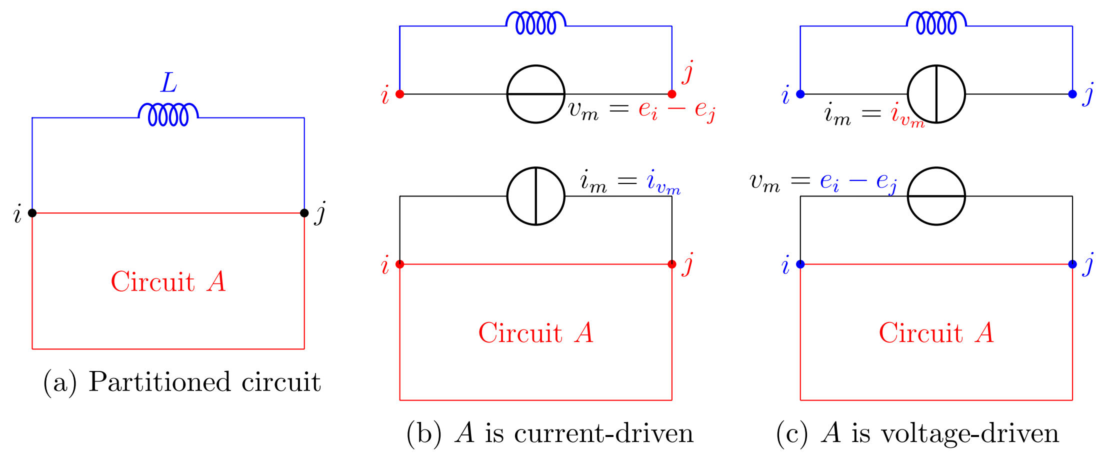


Figure 4.5.: (a) shows a partitioned circuit with coupling nodes i, j. The blue subcircuit consists only of L and the red subcircuit A is arbitrary. In (b) and (c), the two source coupling options are depicted.

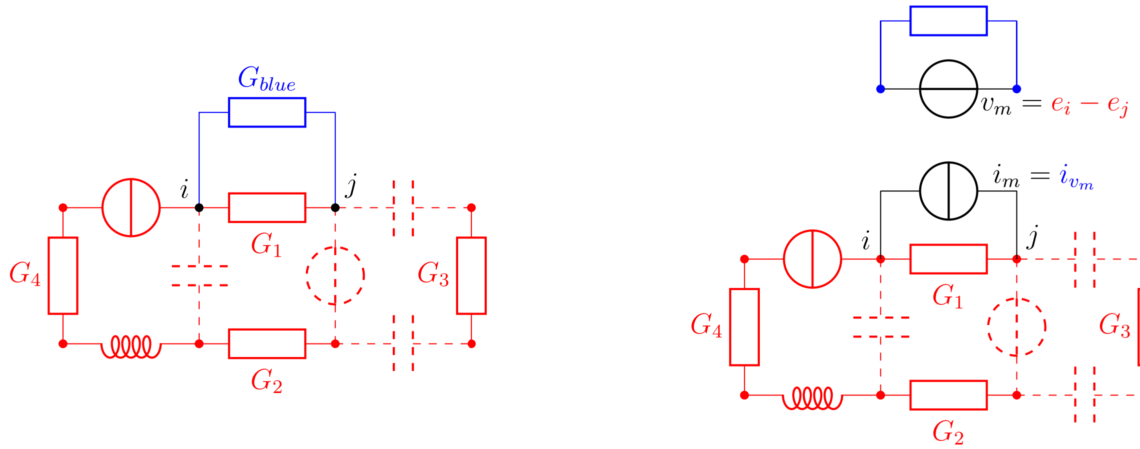


Figure 4.6.: A specific circuit discussed in Example 4.2.4. The CV-components containing the coupling nodes are dashed.

If the coupling nodes in circuit A are not CV-connected, but CVR-connected, Theorem 4.2.3 does not apply. However, we can derive a criterion exploiting Theorems 3.5.5 and 4.1.3. This scenario with a specific subcircuit A is shown in Figure 4.6, where the coupling nodes are R-connected.

For simplicity and to sketch the principal idea, we assume linear resistors in the following. The Jacobi WR scheme of the coupled system reads

$$i_{v_m}^k = G(e_i^{k-1} - e_j^{k-1})$$
$$f_A(\dot{x}^k, x^k, t) = (\hat{e}_i - \hat{e}_j) i_{v_m}^{k-1},$$

where e_i denotes a node potential and $\hat{e}_i$ the i-th unit vector. (That both are described by the letter e is a coincidence here). The functions f_A describes MNA of circuit A. In the first subsystem, i_{v_m} is both the output and the only unknown. Hence, the first subsystem satisfies Conditions (4.8) and (4.9) of the convergence Theorem 4.1.3 with $k_1 = G$. Note that this still holds for the case of resistive networks if we adjust $k_1 = G_{ij}$ and G_{ij} is the effective conductance between the coupling nodes.

The second subsystem represented by circuit A satisfies Condition (4.10) since we assumed that the coupling current forms a loop with CVR-edges, and in that case Theorem 3.5.1 yields an according estimate. Applying Theorem 3.5.5 to circuit A yields the constant k_2 : The estimate therein satisfies Condition 4.11 with $k_2 = R_{ij}^A$, where R_{ij}^A is the resistive distance as defined in 3.5.4 and the superscript A only indicates that it is measured in circuit A. R_{ij}^A measures the resistive distance between the two CV-components containing the coupling nodes i and j.

Hence, we obtain for Figure 4.3b that

- *GS WR converges with rate $Hc + \frac{G}{G_{ij}^A}$, if H is sufficiently small, the coupling nodes* *are CVR-connected in subcircuit A and*

$$GR^Aij = \frac{G}{G^A_{ij}} < 1, \qquad G^A_{ij} := (R^A_{ij})^{-1}.$$

In the generalized case, where the resistance is replaced by a resistive network, this criterion becomes $G_{ij}R^A_{ij} = \frac{G_{ij}}{G^A_{ij}} < 1$*. We could therefore reformulate the topological criterion as:* The coupling nodes are CVR-connected in A and the effective conductance G_{ij} between the coupling nodes in the resistive (voltage-driven) subsystem is smaller than the "effective conductance" G^A_{ij} between the coupling nodes in the (current-driven) subcircuit A.

In the specific example in Figure 4.6, we accordingly obtain WR convergence if $G_{blue} < G_1 + G_2$*. The conductances* G_3 *and* G_4 *are irrelevant.*

If the resistance G_2 *in Figure 4.6 is replaced by a capcitance or a voltage source, the coupling nodes become CV-connected and we obtain convergence according to Theorem 4.2.3. If* G_2 *is replaced by an inductance or voltage source instead, the convergence criterion becomes* $G_{blue} < G_1$*.*

In Figure 4.3c, the resistance is the current-driven subsystem. Our results do not provide any prediction for a possible convergence of WR. Both subcircuits satisfy the preconditions of the second statement of Theorem 3.5.1, and hence Conditions (4.8)-(4.11) *are satisfied for* $k_1 = G^{-1} = R$ *and* some k_2*, which we can not further specify.*

Example 4.2.5 (Circuit/capacitance (Figure 4.4)) *We have the same situation as in Example 4.2.4, but the resistance therein is replaced by a capacitance.*

In the coupling situation of Figure 4.4b, the capacitance is the voltage-driven subsystem. The preconditions of Theorem 4.2.3 guaranteeing WR convergence are not satisfied. According to Theorem 3.5.1, a perturbation of the coupling voltage leads to an index-two type estimate. Hence, the coupling from Figure 4.4b is not recommendable for WR.

If the capacitance is the current-driven subsystem as shown in Figure 4.4c, the situation is much better: Theorem 4.2.3 yields for some $c > 0$ *that*

- *WR converges with rate* $\sqrt{Hc}$ *(Jacobi) or* Hc *(GS), if* H *is sufficiently small and the coupling nodes* i *and* j *are* not *CVR-connected in the (voltage-driven) subcircuit* A*.*

Note that this criterion is still satisfied if the capacitance is replaced by any subcircuit with a CV-path between the coupling nodes i *and* j*.*

Example 4.2.6 (Circuit/inductance (Figure 4.5)) *Here, one subcircuit is represented by an inductance.*

In the coupling situation of Figure 4.5b, the inductance is the voltage-driven subsystem. Theorem 4.2.3 yields that the existence of a constant $c > 0$ *such that*

- *WR converges with rate* $\sqrt{Hc}$ *(Jacobi) or* Hc *(GS), if* H *is sufficiently small and the coupling nodes are CV-connected in (the current-driven) subcircuit* A*.*

However, it is easy to see that the criterion of Theorem 4.2.3 is not sharp here. Let us, for simplicity, consider a linear inductance. The Jabobi WR scheme of the coupled system then reads

$$Li_{v_m}^k = i_{v_m}^k(t_0) + \int_{t_0}^t e_i^{k-1} - e_j^{k-1} d\tau$$
$$f_A(\dot{x}^k, x^k, t) = (\hat{e}_i - \hat{e}_j) i_{v_m}^{k-1},$$

where e_i *denotes a node potential and* $\hat{e}_i$ *the* i*-th unit vector, and* f_A *describes MNA of circuit* A*. In the first subsystem,* i_{v_m} *is both the output and the only unknown. Clearly, the first subsystem satisfies Conditions* (4.8) *and* (4.9) *with* $k_1 = 0$ *of the convergence Theorem 4.1.3. Circuit* A *is the current-driven subsystem. It satisfies the preconditions for the index-one type estimate of the second statement of Theorem 3.5.1 if the coupling nodes* i *and* j *are CVR-connected in* A*. Consequently, the second subsystem satisfies Conditions* (4.10) *and* (4.11) *for* some k_2 *if the coupling nodes are CVR-connected. Since* $k_1 = 0$ *and hence* $k_1 k_2 = 0$*, the convergence Theorem 4.1.3 then yields for Figure 4.5b that*

- *WR converges with rate* $\sqrt{Hc}$ *(Jacobi) or* Hc *(GS), if* H *is sufficiently small and the coupling nodes are CVR-connected in (the current-driven) subcircuit* A*,*

which relaxes the requirement on the coupling nodes in A *from CV-connected to only CVR-connected.*

This "lack of sharpness" in the topological convergence criterion of Theorem 4.2.3 is due to an incomplete understanding of voltage source perturbations. For current source perturbations, Section 3.5 yields Theorems 3.5.3 and 3.5.5, which can be seen as refinements of Theorem 3.5.1 adressing only the voltage over a perturbed current source. In contrast, analogous results could not (yet) be achieved for voltage source perturbations.

Without elaborating further we mention here that replacing the inductance by an LR-series in Figure 4.5b, we have the same situation: the topological convergence criterion of Theorem 4.2.3 requires that the coupling nodes are CV-connected in (the current-driven) subcircuit A*. However, it is easy to check that here too the output current of the LR-subcircuit is of ODE-type, that is, the LR subsystem satisfies Conditions* (4.8) *and* (4.9) *with* $k_1 = 0$*, which means that the general convergence Theorem 4.1.3 implies that CVR-connected coupling nodes in subcircuit* A *are sufficient to guarantee WR convergence.*

The coupling in Figure 4.5c, where the inductance is the current-driven subsystem, is not recommendable. The preconditions for the index-one type estimate of the second statement of Theorem 3.5.1 are not fulfilled. Hence, we have no basis for Theorem 4.1.3 to apply.

Discussion

As a direct consequence of on the hand the perturbation analysis in Section 3.5, notably Theorems 3.5.1 and 3.5.3, and on the other hand the convergence Theorem 4.1.3 for coupled DAEs, we obtained convergence Theorem 4.2.3 for coupled circuits. It mainly provides a sufficient topological convergence criterion. If this criterion is satisfied, WR converges with rate Hc (or $\sqrt{Hc}$ in the Jacobi case) for some constant c and the time interval lenght H. This means that choosing $H \to 0$ sufficiently small, the convergence gets arbitrarily fast, which is a good situation to apply the windowing technique. The constant c, however, is generally unknown, so that the rate Hc is no guarantee for fast convergence for any fixed H.

While the focus of this work are coupled DAEs, it is worth mentioning the well-known fact that WR methods converge even superlinearly in the case of coupled ODEs, cf. e.g. [Gand15]. This seems to indicate the possibility of superlinear convergence for coupled DAEs if certain favourable dynamic coupling structures are satisfied.

One practical challenge is disregarded in Theorems 4.1.3 and 4.2.3, namely consistent initialization. Since we deal with (implicit) index-two DAEs, this is a complex matter, which we bypassed by assuming that the set of variables whose initial values are not algebraically fixed is given. (Precisely, this assumption is implied in the matrix A, which first appeared in Corollary 3.4.6 and which is generally hard to obtain.) For a systematic study of consistent initialization for DAEs and circuit DAEs, we refer to [Est00, Bau12]. In [Cor20], the problem of consistent initialization is discussed in the context of WR.

We want to point that according to the convergence Theorem 4.1.3, we have to expect WR convergence also for weaker (topological) requirements than formulated in Theorem 4.2.3, leading to a rate of convergence of $Hc + k$ (or $\sqrt{Hc + k}$ in the Jacobi case) with $k < 1$. However, for the practical application to complex problems, including such cases in a convergence criterion seems of less interest for two reasons:

Firstly, the convergence rate is obviously “worse”, meaning even for H small we can not reach a rate of convergence below $0 < k < 1$ (or $\sqrt{k}$).

Secondly, WR in that case could only be guaranteed to converge *if* $k < 1$. This means to predict convergence, we have to specifically determine k, which is usually costly. For coupled circuit problems, such a constant k generally depends on the elements resistances, inductances and / or capacitances. For nonlinear circuits, these values are state-dependent, which makes the precise computation of k a nearly hopeless project.

An Example for this case where $k \neq 0$ is discussed in 4.2.4 (Figure 4.3b).

It is worth noticing that convergence Theorem 4.2.3 has one more “blind spot” where it fails to predict WR convergence. Example 4.2.5 (Figure 4.5b) provides a scenario for this situation and discusses it. This blind spot stems from an incomplete understanding of perturbation of voltage sources: For the effect of current source perturbations on the circuit, Theorem 3.5.3 can be seen as a refinement or add-on of Theorem 3.5.1, specifying in which situations the perturbation of the voltage over a perturbed current source is of

ODE-type. A comparable result must be expected to exist for perturbed voltage sources, but could not yet be formulated. The mentioned blind spot is due to the lack of such a result.

Apart from these two special cases of a nonzero k and the blind spot, the examples mainly serve the purpose to illustrate the topological convergence criterion by simple, but enlightening prototypical cases. Even though no rigorous proof is given, the examples strongly suggest that it can be crucial for WR convergence which subsystem is chosen as current-driven and which one as voltage-driven.

To our best knowledge, a comparable result to Theorem 4.2.3, providing a sufficient topological convergence criterion for nonlinear index-two coupled circuit DAEs, does not exist in literature. In [Lel82, WOSR85, AG01, Bar04, BBGS13], instances for convergent coupling settings are given. However, they rely on the comparably restrictive coupling conditions.

4.3. WR on coupled field/circuit models

Lumped circuit models such as MNA are well-established in electrical engineering. However, they neglect the spatial dimension and therefore distributed phenomena like the skin effect. For certain devices, this may lead to inaccuracies of unacceptable magnitude in the simulation, e.g. for electric machines [Sal95] or the quench protection system of superconducting magnets in particle accelerators [Bor+17]. These cases call for field/circuit coupling [SGB10], [CSMBPAV17]. To solve such coupled systems, it is often advisable to use WR methods, since they allow for dedicated step sizes and suitable solvers for the different subsystems, and even for the use of proprietary blackbox solvers. The coupled field/circuit model considered here is a DAE in the time domain after space discretisation of the field system.

This section presents convergence criteria for coupled field/circuit models whose

- space-discrete field DAE is of index 1.
- field surrounding circuit DAE is of index 2.

As in the previous sections, the convergence criteria are formulated in terms of the circuit topology between the coupling nodes.

This whole Section 4.3 and the field/circuit simulation results in the subsequent Section 4.4 were developped in a joint work with Idoia Cortes Garcia, Sebastian Schöps and Caren Tischendorf, cf. [CPST20].

4.3.1. Coupled field/circuit model and WR algorithm

To describe the electromagnetic (EM) field part, we consider a magnetoquasistatic approximation of Maxwell's equations in a reduced magnetic vector potential formulation [ET18]. This leads to the curl-curl eddy current partial differential equation (PDE). The circuit side is formulated with the MNA. For the numerical simulation of the coupled system, the method of lines is used with a finite element (FE) discretisation. Altogether, this leads to a time-dependent coupled system of DAE initial value problems (IVPs), described by

$$M\dot{a} + K(a)a - Xi_m = 0, \quad X^\top \dot{a} = v_c, \qquad p_m(a, i_m) = i_m \tag{4.17}$$

$$f_{MNA}(\dot{x}, x, t) = \begin{pmatrix} \mathcal{A}_m i_m \\ 0 \end{pmatrix}, \qquad p_c(x) = \mathcal{A}_m^\top e =: v_c. \tag{4.18}$$

The circuit Equation (4.18) arises from the MNA. The incidence matrix $\mathcal{A}_m$ describes the position of the field device in the circuit. Its number of rows equals the number of nodes in the circuit (minus one due to the reference node). As used throughout the thesis, $x = \left(\begin{smallmatrix} e \\ i_l \\ i_v \end{smallmatrix}\right)$ collects all node potentials e and currents through inductors and voltage sources i_l and i_v. Furthermore, i_m is the current through the field device, which acts like an (externally computed) current source on the circuit.

The first Equation (4.17) represents the space-discrete field model based on the matrices

$$(M)_{ij} = \int_\Omega \sigma\omega_i \cdot \omega_j \,\mathrm{dV}, \qquad (K(a))_{ij} = \int_\Omega \nu(a)\nabla\times\omega_i \cdot \nabla\times\omega_j \,\mathrm{dV}\ , \tag{4.19}$$

which follow from the Ritz-Galerkin approach using a finite set of Nédélec basis functions ω_i [Mon03] defined on the domain Ω. Here, σ denotes the space-dependent electric conductivity and $\nu(a)$ the magnetic reluctivity that can additionally depend nonlinearly on the unknown magnetic vector potential a.

The current through the field device is described by i_m. Furthermore, the excitation matrix is computed from a winding density function χ_j modelling the j-th stranded conductor [Schö11] as

$$(X)_{ij} = \int_\Omega \chi_j \cdot \omega_i \,\mathrm{dV}\ . \tag{4.20}$$

Assumption 4.3.1 *The space-discrete field DAE* (4.17) *meets the following properties:*

(i) *M is symmetric and positive semi-definite.*

(ii) *X has full column rank.*

(iii) $X^\top M = 0$.

(iv) *The function κ defined by $\kappa(a) := K(a)a$ is strongly monotone and locally Lipschitz continuous.*

The assumptions are in agreement with previous formulation in the literature, e.g. [CGS20, Schö11]. The first Assumption 4.3.1(i) follows naturally if a Ritz-Galerkin formulation (4.19) is chosen.

The full column rank Assumption 4.3.1(ii) follows from the fact that the columns are discretisations of different coils that are located in spatially disjoint subdomains.

The orthogonality Assumption 4.3.1(iii) is a consequence of the stranded conductor model, in which eddy currents are neglected [Schö11] and thus a vanishing conductivity is assumed. On the continuous level, conductive domains are disjoint to the support of the stranded conductor winding functions. This assumption may be violated on the discrete level, e.g. if coils and conductors are in contact due to the support of the basis functions, see [Schö11, Assumption 3.5].

Finally, the monotonicity Assumption 4.3.1(iv) follows from the strong monotonicity of the underlying nonlinear material law, i.e. the BH-curve [Pec04].

4.3.2. Analysis of the space-discrete field DAE

In order to apply Theorem 4.1.3 to the WR scheme of a coupled field/circuit problem, examine the sensitivity of the field DAE (4.17) to perturbations of the WR input v_c.

Lemma 4.3.2 *Let Assumption 4.3.1 hold. Then, for a given continuous source term v_c, there exists a coordinate transformation $(w,u) = T^{-1}a$ and a system of the form*

$$\dot{u} = \theta(u) + Gv_c, \qquad w = \Psi(u), \qquad i_m = \varphi(u) \tag{4.21}$$

where θ, Ψ and φ are locally Lipschitz continuous, such that (a, i_m) solves the space-discrete field DAE

$$M\dot{a} + K(a)a - Xi_m = 0, \quad X^\top \dot{a} = v_c \tag{4.22}$$

if and only if (u, w, i_m) solves Equation (4.21).

Proof: We recall that $X^\top M = 0$ and X has full column rank by Assumption 4.3.1. Hence, the columns X_i of X can be extended to a basis B_M of $\ker M$ such that

$$B_M = \{X_1, \dots, X_n, b_1, \dots, b_l\}$$

and the vectors $X_i \perp b_j$ for any $i = 1,\dots,n$ and $j = 1,\dots,l$. Note that this way of construction guarantees that $b_j \in (\ker M \cap \ker X^\top)$ for $j = 1,\dots,l$.

By means of this basis, we define the matrices

$$Q_{MX} := \begin{pmatrix} b_1 & \cdots & b_l \end{pmatrix}, \qquad Q_M := \begin{pmatrix} Q_{MX} & X \end{pmatrix}$$

and a matrix P_M such that $T := \begin{pmatrix} Q_{MX} & X & P_M \end{pmatrix}$ is nonsingular and the columns satisfy the pairwise orthogonality conditions

$$P_M^\top X_i = 0 \ \forall i = 1,\dots,n, \qquad P_M^\top b_i = 0 \ \forall j = 1\dots,l.$$

The pair $\{P_M, Q_M\}$ is a kernel splitting pair in the sense of Definition 3.1.4.

Writing $\kappa(a) := K(a)a$, we equivalently transform the field DAE with new coordinates $T\alpha = a$:

$$\begin{aligned} T^\top MT\dot{\alpha} + T^\top\kappa(T\alpha) - T^\top Xi_m = 0,\\ X^\top T\dot{\alpha} = v_c. \end{aligned} \tag{4.23}$$

With $\alpha := \left(\begin{smallmatrix} w \\ u \end{smallmatrix}\right)$ and $u := \left(\begin{smallmatrix} u_1 \\ u_2 \end{smallmatrix}\right)$, it holds

$$T\alpha = Q_{MX}w + Xu_1 + P_M u_2$$

and the transformed DAE (4.23) has the detailed form

$$
\begin{aligned}
Q_{MX}^\top \kappa(Q_{MX} w + (X \quad P_M) u) &= 0, \\
X^\top \kappa(T\alpha) - \underline{X^\top X} i_m &= 0, \\
\underline{P_M^\top M P_M} \dot{u}_2 + P_M^\top \kappa(T\alpha) &= 0, \\
\underline{X^\top X} \dot{u}_1 &= v_c.
\end{aligned}
$$

Next we show that

(i) the first equation can be resolved for w.

(ii) the underlined matrices are nonsingular.

(i) Since κ is strongly monotone and Q_{MX} has full column rank, it follows with Lemma 3.3.9 that

$$(w, u) \mapsto Q_{MX}^\top \kappa(Q_{MX} w + (X \quad P_M) u)$$

is strongly monotone w.r.t. w. Thus, Lemma 3.3.6 yields the desired the existence of a Lipschitz continuous Ψ such that $w = \Psi(u)$.

(ii) Obviously $X^\top X$ is nonsingular since X has full column rank.

Recalling that M is symmetric and positive semi-definite, it follows that there exists a matrix L such that the factorization $M = L^\top L$ holds. Furthermore, $MP_M = L^\top L P_M$ has full column rank with Lemma 3.1.6. This implies full column rank of LP_M, and consequently

$$P_M^\top M P_M = P_M^\top L^\top L P_M$$

is nonsingular.

Inverting the nonsingular matrices then yields a system of the form

$$\dot{u} = \tilde{\theta}(u, w) + G v_c, \qquad w = \Psi(u), \qquad i_m = \tilde{\varphi}(u, w).$$

Defining $\theta(u) := \tilde{\theta}(u, \Psi(u))$ and $\varphi(u) := \tilde{\varphi}(u, \Psi(u))$ then yields the desired result, where local Lipschitz continuity of φ and θ follows from local Lipschitz continuity of κ and Ψ. □

We assume that there exists an a priori estimate for the inherent ODE $\dot{u} = \theta(u) + G v_c$ of the field DAE.

Assumption 4.3.3 *On a fixed time interval $[t_0, T]$, consider the space-discrete field DAE and the corresponding inherent ODE $\dot{u} = \theta(u) + G v_c$ as in Lemma 4.3.2 Then, any solution u of the inherent ODE satisfies an a priori bound of the form*

$$|u(t)| \leq |u(t_0)| + c(t)|v_c(t)| \quad \forall t \in [t_0, T],$$

where $c(t)$ is continuous.

4.3.3. Convergence

For given (consistent) initial values, we consider the Jacobi WR of the coupled DAE (4.17)-(4.18), which reads

$$M\dot{a}^k + K(a^k)a^k - Xi_m^k = 0, \quad X^\top \dot{a} = v_c^{k-1}, \qquad p_m(a^k, i_m^k) = i_m^k, \quad A_m a(t_0) = a_0^{IV} \tag{4.24}$$

$$f_{MNA}(\dot{x}^k, x^k, t) = \begin{pmatrix} \mathcal{A}_m i_m^{k-1} \\ 0 \end{pmatrix}, \qquad p_c(x^k) = \mathcal{A}_m^\top e^k =: v_c^k \quad A_c x(t_0) = x_0^{IV}. \tag{4.25}$$

Remark. Note that this Jacobi WR scheme implies that the circuit subsystem is current-driven; that is, it receives currents i^{k-1} as inputs in each iteration step, and the coupling voltage input v_c is internally computed. Conversely, the field subsystem is voltage-driven.

Theorem 4.3.4 *Consider the coupled field/circuit DAE* (4.17)-(4.18)*. Let the field subsystem* (4.17) *satisfy Assumptions 4.3.1 and 4.3.3, and let the circuit subsystem* (4.18) *satisfy Assumptions 2.1.3 and 3.4.3.*

Furthermore, let the coupling inputs of the coupled circuit Equations (4.14) *satisfy the following topological property:*

(i) Each coupling current edge forms a loop with CVR-edges in the circuit which corresponds to Subsystem (4.18)*.*

Then, there exists a constant $\hat{c} > 0$ such that for $H_T := T - t_0$ sufficiently small to satisfy $H_T < \frac{1}{\hat{c}}$, it holds:

The Jacobi sequence (a^k, i_m^k, x^k), defined by Iteration (4.24)-(4.25)*, converges in $(C([t_0,T],\mathbb{R}^n), \|\cdot\|_{[t_0,T]})$ to the solution (a, i_m, x) of* (4.17)-(4.18)*, and it has rate of convergence $\sqrt{H_T \hat{c}}$.*

Remark. The convergence result holds for any continuous initial guess functions v_c^0, i_m^0.

Remark. An analogous result holds for the GS WR, where we obtain linear rate of convergence $H_T\hat{c}$. This follows from Theorem 4.1.3, whose preconditions are satisfied as the subsequent proof shows.

Proof: Instead of the field DAE (4.17), we consider the equivalent normal form (4.21). From Assumption 4.3.3, it follows that for any constant $C > 0$ and any bounded input $v_c \in B(C)$, where $B(C)$ is a ball of continuous functions on $[t_0, T]$ bounded by C (cf. (3.55)), the corresponding solution $|u(v_c)(t)|$ is bounded. This implies that for any $C > 0$ and any $v_c, v_c^\delta \in B(C)$, the corresponding solutions u and u^δ stay within a domain on which θ is (globally) Lipschitz continuous with a Lipschitz constant L which is independent of v_c, v_c^δ. Hence, we can derive

$$\begin{aligned}|u(t)-u^\delta(t)| &\leq |u(t_0)-u^\delta(t_0)| + \int_{t_0}^t |\theta(u(\tau))-\theta(u^\delta(\tau))|d\tau + |A|_* \int_{t_0}^t |v_c(\tau)-v_c^\delta(\tau)|d\tau \\ &\leq |u(t_0)-u^\delta(t_0)| + L\int_{t_0}^t |(u(\tau))-(u^\delta(\tau))|d\tau + |A|_* \int_{t_0}^t |v_c(\tau)-v_c^\delta(\tau)|d\tau,\end{aligned}$$

with $|\cdot|_*$ an induced matrix norm. Then, the Gronwall Lemma 3.3.14 yields for some $c>0$

$$|u(t)-u^\delta(t)| \leq c\left(|u(t_0)-u^\delta(t_0)| + \int_{t_0}^t |v_c(\tau)-v_c^\delta(\tau)|d\tau,\right) \tag{4.26}$$

For the output $i_m = \varphi(u)$, it follows for some $\tilde{c}>0$ that

$$|i_m(t)-i_m^\delta(t)| \leq \tilde{c}\left(|u(t_0)-u^\delta(t_0)| + \int_{t_0}^t |v_c(\tau)-v_c^\delta(\tau)|d\tau,\right) \tag{4.27}$$

due to local Lipschitz continuity of φ and the a priori estimate for u.

If $u(t_0)=u^\delta(t_0)$, Equations (4.26) and (4.27) satisfy Conditions (4.8) and (4.9) with $k_1=0$.

Concerning the circuit subsystem, we note that we have only perturbed current sources and each of the corresponding edges forms a loop with CVR-edges by precondition. Hence, the second statement of Theorem 3.5.1 and the linearity of the output function imply that Conditions (4.10) and (4.11) hold for some k_2.

Since $k_1=0$, it follows that $k_1k_2=0$ regardless of k_2. Thus, the preconditions of the convergence Theorem 4.1.3 are satisfied. □

Sketch for a field model with relaxed assumptions

A careful review of the proof of the decoupling Lemma 4.3.2 hints at the following: If we drop or relax Assumption 4.3.1(*iii*), this may lead to a normal form, where instead of $i_m=\varphi(u)$ in Equation (4.21), we obtain an output i_m which depends on the input v_c, that is,

$$i_m = \tilde{\varphi}(u, v_c)$$

for some locally Lipschitz continuous function $\tilde{\varphi}$. Analogous considerations as in the proof of Theorem 4.3.4 reveal that the decoupled field DAE still satisfies conditions of the form (4.8) and (4.9) (of the convergence Theorem 4.1.3), but in contrast to the case considered therein, they hold for *some* constant $k_1 \in \mathbb{R}$; we do not obtain $k_1=0$ anymore.

To guarantee convergence, we can still strengthen the requirements on the topological position of the coupling elements in the circuit: According to Theorem 3.5.3, the circuit subsystem satisfies conditions of the form (4.10) and (4.11) with $k_2 = 0$, if the coupling current edges according to i_m form loops with CV-loops. This yields convergence of WR again according to Theorem 4.1.3.

That way, we compensate the relaxed assumptions on the field model with stronger assumptions on the coupling conditions for the circuit: Instead of requiring that the coupling current edges form loops with CVR-edges, we now require that they form loops with CV-edges.

To sum it up, we get the following result:

- Consider the case that relaxed assumptions on the field model result in a normal form (4.21) of the field DAE, where $i_m = \varphi(u)$ is replaced by an equations of the form $i_m = \tilde{\varphi}(u, v_c)$. Then, WR is convergent if each coupling current of the circuit subsystem forms a loop with CV-edges.

In Figure 4.10, the example on the left satisfies this stronger requirement for the circuit, since the coupling nodes are connected by a (dashed) CV-path. The example on the right does not satisfy it. In fact, the coupling nodes there are not even CVR-connected.

4.4. Numerical tests

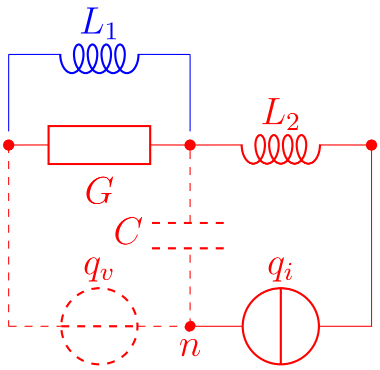
This section presents simulations of toy examples for coupled circuit/circuit (Subsection 4.4.1) and field/circuit (Subsection 4.4.2) problems. Each of the subsections features one problem on which WR is convergent and one on which WR is divergent.

4.4.1. Circuit/Circuit

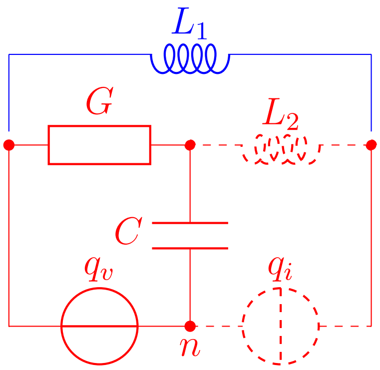
To illustrate the convergence behaviour of the GS WR scheme according to the derived criteria, we consider the toy example coupled circuits in Figures 4.7a and 4.7b with two coupling nodes. Both are described with linear MNA (2.12) and the (arbitrary) parameters $R = 1\Omega$, $L_1 = 0.1\text{H}$, $L_2 = 5\text{H}$, $C = 1\text{F}$, $i_\text{s}(t) = \sin(2t) + 5\sin(20t)$ and $v_\text{s}(t) = \sin(t) + \sin(20t)$ are set.

The GS WR algorithm is applied on the simulation time window $\mathcal{I} = [0,\ 0.8]\,\text{s}$ and the internal time integration is performed with the implicit Euler scheme with time step size $\delta t = 10^{-2}\,\text{s}$.

The coupled circuit in Figure 4.7a satisfies the conditions of the convergence Theorem 4.3.4: In the voltage-driven subcircuit, represented by the single element subcircuit L_1 in blue, the coupling voltage does not form a loop with CV-elements. In the current-driven subcircuit drawn in red, the coupling current forms a loop with CV-elements.



(a) Convergent case.



(b) Divergent case.

Figure 4.7.: Two examples of a circuit/circuit coupling. For WR, we chose the blue inductance L_1 as the voltage-driven subsystem in both cases. Hence, the coupling voltage does not form a loop with CV-elements in the voltage-driven subsystem. In (a), the coupling current forms a loop with the dashed CV-elements in the current-driven subcircuit drawn in red. In (b), the coupling current forms no loop with CV-elements, not even with CVR-elements. The LI-cutset between the coupling nodes is dashed. The node n is labeled since its potential is displayed in the plots in Figure 4.8.

The theoretical result is illustrated by the successful simulation, see Figure 4.8a, of the model shown in Figure 4.7a which satisfies the convergence criterion of Theorem 4.3.4. However, numerical simulations displayed in Figure 4.8b of the model shown in Figure 4.7b show that WR can diverge indeed if the criterion is not satisfied: The coupling current does not form a loop with CV-elements in the (red) current-driven subcircuit. In fact, it does not even form a loop with CVR-elements.

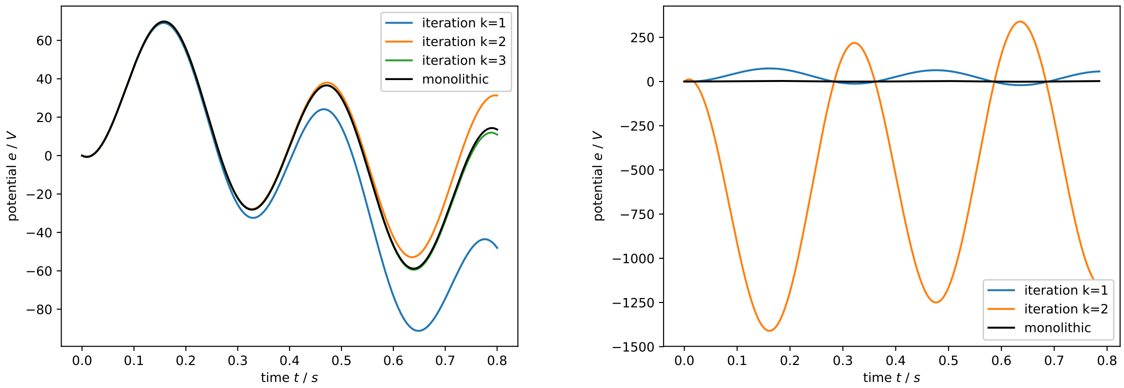
4.4.2. Field/Circuit

As in the previous subsection for the circuit/circuit case, we here present illustrative examples for the field/circuit case which confirm our theoretical convergence result 4.3.4.

The toy example circuits in Figures 4.10a and 4.10b are again described with linear MNA (2.12) and the (arbitrary) parameters $R = 1\Omega$, $L = 5\text{H}$, $C = 1\text{F}$, $i_\text{s}(t) = \sin(2t) + 5\sin(20t)$ and $v_\text{s}(t) = \sin(t) + \sin(20t)$ are set. The field model is taken from [Mee18].

The GS WR algorithm is applied on the simulation time window $\mathcal{I} = [0,\ 0.8]\,\text{s}$ and the internal time integration is performed with the implicit Euler scheme with time step size $\delta t = 10^{-2}\,\text{s}$.

The theory is illustrated by the successful simulation, see Figure 4.11a, of the model shown in Figure 4.10a which satisfies the convergence criterion of Theorem 4.3.4: The coupling current in the circuit subsystem forms a loop with CVR-elements (even with CV-elements). Numerical simulations of the model shown in Figure 4.10b show that



(a) Convergent case.

(b) Divergent case.

Figure 4.8.: Monolithic and WR solution for $k = 1, 2, 3$ iterations in (a) and $k = 1, 2$ in (b). Due to rather slow convergence in (a), we showed a third iteration. Figure (a) refers to the circuit/circuit example in 4.8a and Figure (b) refers to 4.8b.

WR can diverge indeed if the criterion is not satisfied.

4.5. Conclusion

This chapter is concerned with the convergence analysis of WR.

First, a convergence theorem for WR on coupled DAEs is presented in Section 4.1. This theorem is exploited in Section 4.2 to derive topological WR convergence criteria for coupled circuits. The topological criteria are illustrated by prototypical examples. Sections 4.1 and 4.2 are concluded by discussions of the results. Section 4.3 offers topological WR convergence criteria for a coupled field/circuit model. The theoretical results of the preceding sections are confirmed by simulations of numerical tests with toy examples, which are presented in Section 4.4.

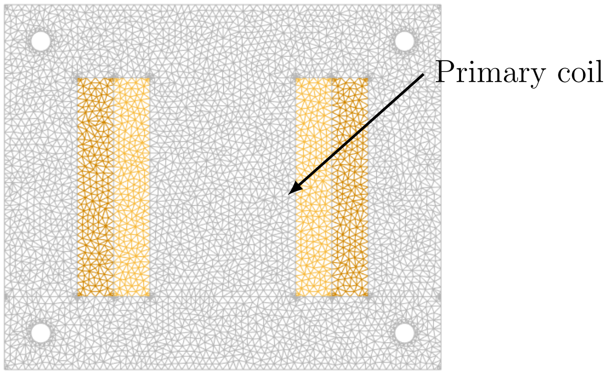
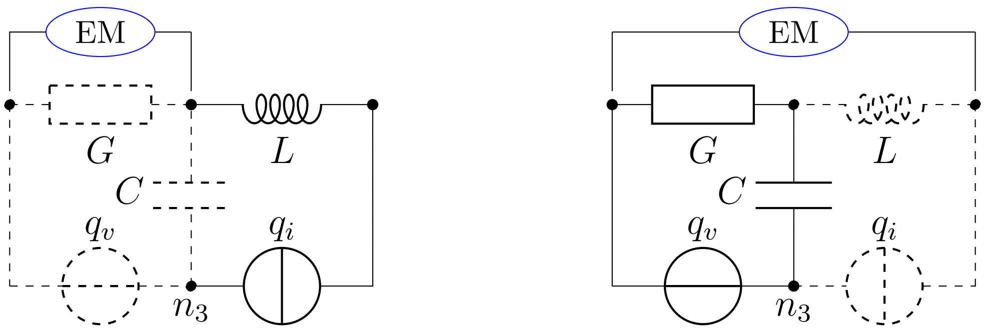


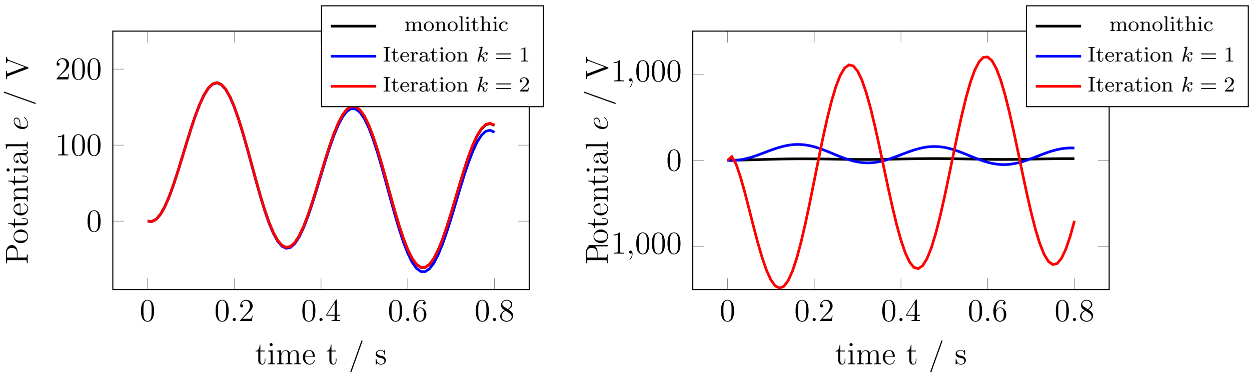
Figure 4.9.: Single phase isolation transformer ('MyTransformer'), see [Mee18].



(a) Convergent case.

(b) Divergent case.

Figure 4.10.: Field/circuit coupling with field model from Figure 4.9. For WR, we chose the field as the voltage-driven subsystem. In (a), the input element forms a loop with CVR-elements (in the circuit subsystem). The CVR-elements are dashed. More precisely, it forms two such loops: One loop with the resistance, and one loop with the CV-path. However, the existence of at least one loop is relevant. In (b), the coupling element forms no loop with CVR-elements. The LI-cutset between the coupling nodes is dashed. The node n is labeled since its potential is displayed in the plots in Figure 4.11.



(a) Convergent case of the field/circuit coupling from Figure 4.10a.

(b) Divergent case of the field/circuit coupling from Figure 4.10b

Figure 4.11.: Monolithic and WR solution for $k = 1, 2$ iterations. In (a), the graph of the monolithic solution is covered by the second iterate due to fast convergence.

5. Summary and Outlook

This thesis is devoted to the study of a differential-algebraic electrical circuit model. Specifically, the focus is two-fold: In the first part of the thesis, the class of circuit describing differential-algebraic equations (DAEs) arising from the model is analyzed. The second part examines the convergence behaviour of waveform relaxation (WR), which is an iterative method to solve dynamical systems numerically.

Regarding the analysis, we provided a detailed DAE normal form for nonlinear circuit DAEs of index two, and we formulated assumptions guaranteeing the right hand side of the normal form is locally Lipschitz continuous. Furthermore, offering an alternative proof, we reproduced an existence and uniqueness result on bounded time intervals for the circuit DAE. This result was first proved by [Jan15].

Exploiting the normal form, we established a systematic perturbation analysis which relates the circuit's sensitivity with respect to its independent sources on the one hand to the network-topological position of the sources on the other hand. That way, we can predict that the effect of an input perturbation on the circuit is comparably mild if the position of the input satisfies certain toplogical criteria. Refinements of the criteria are given if the perturbed input is a current source.

Building on the ground of the analysis, the second part of the thesis is dedicated to WR and the quest for criteria guaranteeing WR convergence. It is a well-known fact that WR can diverge on coupled DAEs, and WR convergence can not be guaranteed unless the coupled DAE satisfies certain contractivity conditions. This fact, along with sufficient contractivity conditions for different DAE classes, was broadly discussed in the literature since the seminal work of Lelarasmee [Lel82] in 1982.

We first prove a convergence result containing sufficient WR contracticity conditions on a general class of nonlinear implicit coupled DAEs of index two. However, similarly to conditions formulated previously in the literature, it is hard to check if the conditions are satisfied in general. Essentially, this requires the determination of certain Lipschitz constants of a normal form. For that reason, prepared by the previous perturbation analysis and the general WR convergence result, we offer a theorem stating comparably "easy-to-check" topological contractivity conditions for the case of coupled circuit DAEs. Furthermore, topological WR convergence criteria are given for the case of a specific field/circuit model. Finally, the theoretical results are confirmed by simulations of toy examples of circuit/circuit and field/circuit couplings.

A few questions arise naturally from this work.

Concerning the perturbation analysis, we mentioned that we presented refinements of the results for the case of perturbed current sources. We expect that similar refinements exist for the case of perturbed voltage sources. To find them, besides being of general interest for the circuit analysis, would presumably also allow to formulate topological WR convergence criteria for coupled circuit scenarios which are not yet covered by our results.

The fact that the presented WR convergence results are formulated on sufficiently small time intervals rises the question if, or under which conditions, the convergence results can be maintained on arbitrary bounded intervals. The "ingredients" for a global convergence result, namely an a-priori estimate for solutions and local Lipschitz continuity of the circuit DAE normal form, seem to be available.

On a more general note, the derivation of comparable topological WR convergence criteria in other coupled multiphysical systems such as power/gas would be a interesting challenge. The circuit DAE normal form which this work is based on is formulated under the assumption of more or less general rank conditions, which allows to apply similar principles to derive normal forms for other DAEs evolving from flow networks.

Finally, we find it intriguing to explore the opportunities of a consequent topological interpretation of normal forms for circuit DAEs. When presenting the circuit normal form, we indicated that Laplacian matrices and their inverses of certain contracted graphs appear regularly therein. A first result exploiting this insight is given by Theorem 3.5.5, giving hope that more results of this type, expressing relevant (Lipschitz) constants of the circuit DAE in terms of its topology, can be achieved.

Bibliography

[AG01] Arnold, M. and Günther, M.: Preconditioned Dynamic Iteration for Coupled Differential-Algebraic Systems. BIT Numerical Mathematics, 41, 1–25 (2001).

[Bap10] Bapat, R.: Graphs and matrices. Springer (2010).

[BH06] Barooah, P. and Hespanha, J. P.: Graph effective resistance and distributed control: Spectral properties and applications. In Proceedings of the 45th IEEE Conference on Decision and Control (pp. 3479-3485). IEEE (2006).

[Bar04] Bartel, A.: Partial Differential-Algebraic Models in Chip Design - Thermal and Semiconductor Problems. Fortschritt-Berichte VDI. Nummer 391 in Fortschritt-Berichte VDI, Reihe 20. VDI-Verlag, Düsseldorf. Dissertation (2004).

[BBGS13] Bartel, A., Brunk, M., Günther, M. and Schöps, S.: Dynamic iteration for coupled problems of electric circuits and distributed devices SIAM Journal on Scientific Computing, 35(2), B315–B335 (2013).

[Bau12] Baumanns, S.: Coupled Electromagnetic Field/Circuit Simulation: Modeling and Numerical Analysis. Dissertation, Universität zu Köln (2012).

[Big93] Biggs, N.: Algebraic graph theory (Vol. 67). Cambridge university press (1993).

[Bod07] Bodestedt, M.: Perturbation analysis of refined models in circuit simulation. Dissertation, Technische Universität Berlin (2007).

[Bol13] Bollobás, B.: Modern graph theory (Vol. 184). Springer Science & Business Media (2013)

[BBLK94] Bonchev, D., Balaban, A. T., Liu, X. and Klein, D. J.: Molecular cyclicity and centricity of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. International journal of quantum chemistry, 50(1), 1-20 (1994).

[Bor+17] Bortot, L. et. al.: STEAM: A hierarchical co-simulation framework for superconducting accelerator magnet circuits. IEEE Transactions on applied superconductivity, 28(3), 1-6 (2017).

[Bur95] Burrage, K.: Parallel and Sequential Methods for Ordinary Differential Equations. Oxford University Press (1995).

[CG95] Campbell, S. L. and Gear, C. W.: The index of general nonlinear DAEs. Numerische Mathematik, 72(2), 173-196 (1995).

[Che18] Chen, H.: Hitting times for random walks on subdivision and triangulation graphs. Linear and Multilinear Algebra, 66(1), 117-130 (2018).

[Chen76] Chen, W. K.: Applied graph theory graphs and electrical networks. North Holland series in Applied Mathematics and Mechanics (1976).

[CDK87] Chua, L., Desoer, C. and Kuh, E.: Linear and Nonlinear Circuits, McGraw-Hill (1987).

[CSMBPAV17] Garcia, I. C., Schöps, S., Maciejewski, M., Bortot, L., Prioli, M., Auchmann, B. and Verweij, A.: Optimized field/circuit coupling for the simulation of quenches in superconducting magnets. IEEE Journal on Multiscale and Multiphysics Computational Techniques, 2, 97-104 (2017).

[Cor20] Cortes Garcia, I.: Mathematical Analysis and Simulation of Field Models in Accelerator Circuits. Dissertation, Technische Universität Darmstadt (2020).

[CGS20] Garcia, I. C., De Gersem, H. and Schöps, S.: A structural analysis of field/circuit coupled problems based on a generalised circuit element. Numerical Algorithms, 83(1), 373-394 (2020).

[CPST20] Cortes Garcia, I., Pade, J., Schöps, S. and Tischendorf, C.: Waveform relaxation for low frequency coupled field/circuit differential-algebraic models of index 2. arXiv, arXiv-2005 (2020).

[CI94] Crow, M. L. and Ilic, M. D.: The waveform relaxation method for systems of differential/algebraic equations. Mathematical and computer modelling, 19(12), 67-84 (1994).

[DB10] Dörfler, F. and Bullo, F.: Spectral analysis of synchronization in a lossless structure-preserving power network model. In 2010 First IEEE International Conference on Smart Grid Communications (pp. 179-184), IEEE (2010).

[DB12] Dörfler, F. and Bullo, F.: Kron reduction of graphs with applications to electrical networks. IEEE Transactions on Circuits and Systems I: Regular Papers, 60(1), 150-163 (2012).

[Ebe08] Ebert, F.: On partitioned simulation of electrical circuits using dynamic iteration methods. Dissertation, Technische Universität Berlin (2009).

[ESMJK11] Ellens, W., Spieksma, F. M., Van Mieghem, P., Jamakovic, A. and Kooij, R. E.: Effective graph resistance. Linear algebra and its applications, 435(10), 2491-2506 (2011).

[ET18] Emson, C.R.I. and Trowbridge, C.W.: Transient 3d eddy currents using modified magnetic vector potentials and magnetic scalar potentials. IEEE Transactions on Magnetics, 24(1), 86-89 (1988).

[Est00] Estévez Schwarz, D.: Consistent initialization for index-2 differential algebraic equations and its application to circuit simulation. Dissertation, Humboldt-Universität zu Berlin (2000).

[ET00] Estévez Schwarz, D. and Tischendorf, C.: Structural analysis of electric circuits and consequences for MNA. International Journal of Circuit Theory and Applications, 28(2), 131-162 (2000).

[Fou12] Foulds, L. R.: Graph theory applications. Springer Science & Business Media (2012).

[Gand15] Gander, M. J.: 50 years of time parallel time integration. Multiple shooting and time domain decomposition methods. Springer, Cham, 69-113 (2015).

[Gan05] Gantmacher, F. R.: Applications of the Theory of Matrices. Courier Corporation (2005).

[Gea88] Gear, C. W.: Differential-algebraic equation index transformations. SIAM Journal on Scientific and Statistical Computing, 9(1), 39-47 (1988).

[GR01] Godsil, C. and Royle, G.: Algebraic graph theory. Graduate texts in mathematics 207, Springer (2001).

[GM86] Griepentrog, E. and März, R.: Differential-algebraic equations and their numerical treatment. Vol. 88. BSB Teubner (1986).

[GM96] Gutman, I. and Mohar, B.: The quasi-Wiener and the Kirchhoff indices coincide. Journal of Chemical Information and Computer Sciences, 36(5), 982-985 (1996).

[HLR89] E. Hairer, C. Lubich, and M. Roche. The Numerical Solution of Differential-Algebraic Systems by Runge-Kutta Methods. Springer-Verlag (1989).

[JK96] Jackiewicz, Z. and Kwapisz, M.: Convergence of waveform relaxation methods for differential-algebraic systems. SIAM Journal on Numerical Analysis, 33(6), 2303-2317 (1996).

[Jan15] Jansen, L.: A Dissection concept for DAEs - structural decoupling, unique solvability, convergence theory and half-explicit methods. Dissertation, Humboldt-Universität zu Berlin (2015).

[JMT15] Jansen, L., Matthes, M., Tischendorf, C.: Global unique solvability for memristive circuit DAEs of index 1. International Journal of Circuit Theory and Applications, 43(1), 73-79 (2015).

[JW01] Jiang, Y. L. and Wing, O.: A note on convergence conditions of waveform relaxation algorithms for nonlinear differential-algebraic equations. Applied numerical mathematics, 36(2-3), 281-297 (2001).

[KR93] Klein, D. J. and Randić, M.: Resistance distance. Journal of mathematical chemistry, 12(1), 81-95 (1993).

[LMT13] Lamour, R., März, R. and Tischendorf, C.: Differential-Algebraic Equations: A Projector Based Analysis. Springer (2013).

[Lel82] Lelarasmee, E.: The Waveform Relaxation Method for Time Domain Analysis of Large Scale Integrated Circuits: Theory and Applications. Dissertation, University of California, Berkeley (1982).

[LRS82] Lelarasmee, E., Ruehli, A. E. and Sangiovanni-Vincentelli, A. L.: The waveform relaxation method for time-domain analysis of large scale integrated circuits. IEEE transactions on computer-aided design of integrated circuits and systems, 1(3), 131-145 (1982).

[Mat12] Matthes, M.: Numerical Analysis of Nonlinear Partial Differential-Algebraic Equations: A Coupled and an Abstract Systems Approach. Logos Verlag Berlin GmbH (2012).

[Mee18] Meeker, D.: Finite Element Method Magnetics: pyFEMM, version 4.2. User's Manual (2018).

[Meh12] Mehrmann, V.: Index concepts for differential-algebraic equations. Encyclopedia of Applied and Computational Mathematics 1: 676-681 (2012).

[Mie00] Miekkala, U.: Dynamic iteration methods applied to linear DAE systems. International Journal of Circuit Theory and Applications, 28, 131–162 (2000).

[MN87] Miekkala, U. and Nevanlinna, O.: Convergence of dynamic iteration methods for initial value problems. SIAM Journal on Scientific and Statistical Computing, 8(4), 459-482 (1987).

[Mol12] Molitierno, J.: Applications of combinatorial matrix theory to Laplacian matrices of graphs. CRC Press (2012).

[Mon03] Monk., P.: Finite Element Methods for Maxwell's Equations. Oxford University Press (2003).

[OR70] Ortega, J. M., Rheinboldt, W. C.: Iterative solution of nonlinear equations in several variables. Academic Press, New York (1970).

[PT19] Pade, J. and Tischendorf, C.: Waveform relaxation: a convergence criterion for differential-algebraic equations. Numerical Algorithms, 81(4): 1327-1342 (2019).

[PL18] Pan, Y. and Li, J.: Kirchhoff index, multiplicative degree and spanning trees of the linear crossed hexagonal chains. International Journal of Quantum Chemistry, 118(24), e25787 (2018).

[PYZ18] Patterson, S., Yi, Y. and Zhang, Z.: A resistance-distance-based approach for optimal leader selection in noisy consensus networks. IEEE Transactions on Control of Network Systems, 6(1), 191-201 (2018).

[Pec04] Pechstein., C.: Multigrid-Newton-methods for nonlinear-magnetostatic problems. Master's thesis, Universität Linz (2004).

[Pen55] Penrose, R.: A generalized inverse for matrices. In Mathematical proceedings of the Cambridge philosophical society (Vol. 51, No. 3, pp. 406-413). Cambridge University Press (1955).

[Rec89] Recski, A.: Matroid Theory and its Applications in Electric Network Theory and Statics. Vol. 6, Springer Science & Business Media (1989).

[Rei14] Reis, T.: Mathematical modeling and analysis of nonlinear time-invariant RLC circuits. In Large-Scale Networks in Engineering and Life Sciences (pp. 125-198). Birkhäuser, Cham (2014).

[Ria08] Riaza, R. Differential-algebraic systems: analytical aspects and circuit applications. World Scientific (2008).

[Sal95] Salon, S.J.: Finite Element Analysis of Electrical Machines. Kluwer (1995).

[Sch91] Schneider, K. S.: A remark on the wave-form relaxation method. International journal of circuit theory and applications, 19(1), 101-104 (1991).

[Schö11] Schöps, S.: Multiscale Modeling and Multirate Time-Integration of Field/Circuit Coupled Problems. Dissertation, Bergische Universität Wuppertal (2011).

[SGB10] Schöps, S., De Gersem, H. and Bartel, A.: A cosimulation framework for multirate time integration of field/circuit coupled problems. IEEE Transactions on Magnetics, 46(8), 3233–3236 (2010).

[SZF06] Sun, W., Zou, J. H. and Fan, X. G.: Convergence of parallel dynamic iteration methods for nonlinear daes of index-2. In 2006 IEEE International Conference on Automation Science and Engineering (pp. 129-133) (2006).

[Tes12] Teschl, G.: Ordinary differential equations and dynamical systems (Vol. 140). American Mathematical Society (2012).

[Tis04] Tischendorf, C.: Coupled Systems of Differential Algebraic and Partial Differential Equations in Circuit and Device Simulation. Modeling and Numerical Analysis. Habilitation thesis, Institut für Mathematik, Humboldt-Universität zu Berlin (2004).

[TYN18] Thulasiraman, K., Yadav, M. and Naik, K.: Network science meets circuit theory: Resistance distance, Kirchhoff index, and Foster's theorems with generalizations and unification. IEEE Transactions on Circuits and Systems I: Regular Papers, 66(3), 1090-1103 (2018).

[WOSR85] White, J. and Odeh, F., Sangiovanni-Vincentelli, A. L., Ruehli, A.: Waveform Relaxation: Theory and Practice, EECS Department. University of California, Berkeley (1985).

[WV12] White, J. K. and Sangiovanni-Vincentelli, A. L.: Relaxation techniques for the simulation of VLSI circuits (Vol. 20). Springer Science & Business Media (2012).

[YK19] Yang, Y. and Klein, D. J.: Two-point resistances and random walks on stellated regular graphs. Journal of Physics A: Mathematical and Theoretical, 52(7), 075201 (2019).