Kazuo Murota

Matrices and Matroids for Systems Analysis

August 31, 1999

Springer-Verlag
Berlin Heidelberg NewYork
London Paris Tokyo
HongKong Barcelona
Budapest

Preface

Interplay between matrix theory and matroid theory is the main theme of this book, which offers a matroid-theoretic approach to linear algebra and, reciprocally, a linear-algebraic approach to matroid theory. The book serves also as the first comprehensive presentation of the theory and application of mixed matrices and mixed polynomial matrices.

A matroid is an abstract mathematical structure that captures combinatorial properties of matrices, and combinatorial properties of matrices, in turn, can be stated and analyzed successfully with the aid of matroid theory. The most important result in matroid theory, deepest in mathematical content and most useful in application, is the intersection theorem, a duality theorem for a pair of matroids. Similarly, combinatorial properties of polynomial matrices can be formulated in the language of valuated matroids, and moreover, the intersection theorem can be generalized for a pair of valuated matroids.

The concept of a mixed matrix was formulated in the early eighties as a mathematical tool for systems analysis by means of matroid-theoretic combinatorial methods. A matrix is called a mixed matrix if it is expressed as the sum of a "constant" matrix and a "generic" matrix having algebraically independent nonzero entries. This concept is motivated by the physical observation that two different kinds of numbers, fixed constants and system parameters, are to be distinguished in the description of engineering systems. Mathematical analysis of a mixed matrix can be streamlined by the intersection theorem applied to the pair of matroids associated with the "constant" and "generic" matrices. This approach can be extended further to a mixed polynomial matrix on the basis of the intersection theorem for valuated matroids.

The present volume grew out of an attempted revision of my previous monograph, "Systems Analysis by Graphs and Matroids — Structural Solvability and Controllability" (Algorithms and Combinatorics, Vol. 3, Springer-Verlag, Berlin, 1987), which was an improved presentation of my doctoral thesis written in 1983. It was realized, however, that the progress made in the last decade was so remarkable that even a major revision was inadequate. The present volume, sharing the same approach initiated in the above mono-

graph, offers more advanced results obtained since then. For developments in the neighboring areas the reader is encouraged to consult:

- A. Recski: "Matroid Theory and Its Applications in Electric Network Theory and in Statics" (Algorithms and Combinatorics, Vol. 6, Springer-Verlag, Berlin, 1989),
- R. A. Brualdi and H. J. Ryser: "Combinatorial Matrix Theory" (Encyclopedia of Mathematics and Its Applications, Vol. 39, Cambridge University Press, London, 1991).
- H. Narayanan: "Submodular Functions and Electrical Networks" (Annals of Discrete Mathematics, Vol. 54, Elsevier, Amsterdam, 1997).

The present book is intended to be read profitably by graduate students in engineering, mathematics, and computer science, and also by mathematics-oriented engineers and application-oriented mathematicians. Self-contained presentation is envisaged. In particular, no familiarity with matroid theory is assumed. Instead, the book is written in the hope that the reader will acquire familiarity with matroids through matrices, which should certainly be more familiar to the majority of the readers. Abstract theory is always accompanied by small examples of concrete matrices.

Chapter 1 is a brief introduction to the central ideas of our combinatorial method for the structural analysis of engineering systems. Emphasis is laid on relevant physical observations that are crucial to successful mathematical modeling for structural analysis.

Chapter 2 explains fundamental facts about matrices, graphs, and matroids. A decomposition principle based on submodularity is described and the Dulmage–Mendelsohn decomposition is derived as its application.

Chapter 3 discusses the physical motivation of the concepts of mixed matrix and mixed polynomial matrix. The dual viewpoint from structural analysis and dimensional analysis is explained by way of examples.

Chapter 4 develops the theory of mixed matrices. Particular emphasis is put on the combinatorial canonical form (CCF) of layered mixed matrices and related decompositions, which generalize the Dulmage–Mendelsohn decomposition. Applications to the structural solvability of systems of equations are also discussed.

Chapter 5 is mostly devoted to an exposition of the theory of valuated matroids, preceded by a concise account of canonical forms of polynomial/rational matrices.

Chapter 6 investigates mathematical properties of mixed polynomial matrices using the CCF and valuated matroids as main tools of analysis. Control theoretic problems are treated by means of mixed polynomial matrices.

Chapter 7 presents three supplementary topics: the combinatorial relaxation algorithm, combinatorial system theory, and mixed skew-symmetric matrices.

Expressions are referred to by their numbers; for example, (2.1) designates the expression (2.1), which is the first numbered expression in Chap. 2.

Similarly for figures and tables. Major symbols used in this book are listed in Notation Table.

The ideas and results presented in this book have been developed with the help, guidance, encouragement, support, and criticisms offered by many people. My deepest gratitude is expressed to Professor Masao Iri, who introduced me to the field of mathematical engineering and guided me as the thesis supervisor. I appreciate the generous hospitality of Professor Bernhardt Korte during my repeated stays at the University of Bonn, where a considerable part of the theoretical development was done. I benefited substantially from discussions and collaborations with Pawel Bujakiewicz, François Cellier, Andreas Dress, Jim Geelen, András Frank, Hisashi Ito, Satoru Iwata, András Recski, Mark Scharbrodt, András Sebő, Masaaki Sugihara, and Jacob van der Woude. Several friends helped me in writing this book. Most notable among these were Akiyoshi Shioura and Akihisa Tamura who went through all the text and provided comments. I am also indebted to Daisuke Furihata, Koichi Kubota, Tomomi Matsui, and Reiko Tanaka. Finally, I thank the editors of Springer-Verlag, Joachim Heinze and Martin Peters, for their support in the production of this book, and Erich Goldstein for English editing.

Kyoto, June 1999 *Kazuo Murota*

Contents

Preface V

1. Introduction to Structural Approach — Overview of the Book 1
1.1 Structural Approach to Index of DAE 1
1.1.1 Index of Differential-algebraic Equations 1
1.1.2 Graph-theoretic Structural Approach 3
1.1.3 An Embarrassing Phenomenon 7
1.2 What Is Combinatorial Structure? 10
1.2.1 Two Kinds of Numbers 11
1.2.2 Descriptor Form Rather than Standard Form 15
1.2.3 Dimensional Analysis 17
1.3 Mathematics on Mixed Polynomial Matrices 20
1.3.1 Formal Definitions 20
1.3.2 Resolution of the Index Problem 21
1.3.3 Block-triangular Decomposition 26

2. Matrix, Graph, and Matroid 31
2.1 Matrix 31
2.1.1 Polynomial and Algebraic Independence 31
2.1.2 Determinant 33
2.1.3 Rank, Term-rank and Generic-rank 36
2.1.4 Block-triangular Forms 40
2.2 Graph 43
2.2.1 Directed Graph and Bipartite Graph 43
2.2.2 Jordan–Hölder-type Theorem for Submodular Functions 48
2.2.3 Dulmage–Mendelsohn Decomposition 55
2.2.4 Maximum Flow and Menger-type Linking 65
2.2.5 Minimum Cost Flow and Weighted Matching 67
2.3 Matroid 71
2.3.1 From Matrix to Matroid 71
2.3.2 Basic Concepts 73
2.3.3 Examples 77
2.3.4 Basis Exchange Properties 78

2.3.5 Independent Matching Problem 84
2.3.6 Union 93
2.3.7 Bimatroid (Linking System) 97

3. Physical Observations for Mixed Matrix Formulation 107
3.1 Mixed Matrix for Modeling Two Kinds of Numbers 107
3.1.1 Two Kinds of Numbers 107
3.1.2 Mixed Matrix and Mixed Polynomial Matrix 116
3.2 Algebraic Implication of Dimensional Consistency 120
3.2.1 Introductory Comments 120
3.2.2 Dimensioned Matrix 121
3.2.3 Total Unimodularity of a Dimensioned Matrix 123
3.3 Physical Matrix 126
3.3.1 Physical Matrix 126
3.3.2 Physical Matrices in a Dynamical System 128

4. Theory and Application of Mixed Matrices 131
4.1 Mixed Matrix and Layered Mixed Matrix 131
4.2 Rank of Mixed Matrices 134
4.2.1 Rank Identities for LM-matrices 135
4.2.2 Rank Identities for Mixed Matrices 139
4.2.3 Reduction to Independent Matching Problems 142
4.2.4 Algorithms for the Rank 145
4.3 Structural Solvability of Systems of Equations 153
4.3.1 Formulation of Structural Solvability 153
4.3.2 Graphical Conditions for Structural Solvability 156
4.3.3 Matroidal Conditions for Structural Solvability 160
4.4 Combinatorial Canonical Form of LM-matrices 167
4.4.1 LM-equivalence 167
4.4.2 Theorem of CCF 172
4.4.3 Construction of CCF 175
4.4.4 Algorithm for CCF 181
4.4.5 Decomposition of Systems of Equations by CCF 187
4.4.6 Application of CCF 191
4.4.7 CCF over Rings 199
4.5 Irreducibility of LM-matrices 202
4.5.1 Theorems on LM-irreducibility 202
4.5.2 Proof of the Irreducibility of Determinant 205
4.6 Decomposition of Mixed Matrices 211
4.6.1 LU-decomposition of Invertible Mixed Matrices 212
4.6.2 Block-triangularization of General Mixed Matrices 215
4.7 Related Decompositions 221
4.7.1 Decomposition as Matroid Union 221
4.7.2 Multilayered Matrix 225
4.7.3 Electrical Network with Admittance Expression 228

4.8 Partitioned Matrix 230
4.8.1 Definitions 231
4.8.2 Existence of Proper Block-triangularization 235
4.8.3 Partial Order Among Blocks 238
4.8.4 Generic Partitioned Matrix 240
4.9 Principal Structures of LM-matrices 250
4.9.1 Motivations 250
4.9.2 Principal Structure of Submodular Systems 252
4.9.3 Principal Structure of Generic Matrices 254
4.9.4 Vertical Principal Structure of LM-matrices 257
4.9.5 Horizontal Principal Structure of LM-matrices 261

5. Polynomial Matrix and Valuated Matroid 271
5.1 Polynomial/Rational Matrix 271
5.1.1 Polynomial Matrix and Smith Form 271
5.1.2 Rational Matrix and Smith–McMillan Form at Infinity 272
5.1.3 Matrix Pencil and Kronecker Form 275
5.2 Valuated Matroid 280
5.2.1 Introduction 280
5.2.2 Examples 281
5.2.3 Basic Operations 282
5.2.4 Greedy Algorithms 285
5.2.5 Valuated Bimatroid 287
5.2.6 Induction Through Bipartite Graphs 290
5.2.7 Characterizations 295
5.2.8 Further Exchange Properties 300
5.2.9 Valuated Independent Assignment Problem 306
5.2.10 Optimality Criteria 308
5.2.11 Application to Triple Matrix Product 316
5.2.12 Cycle-canceling Algorithms 317
5.2.13 Augmenting Algorithms 325

6. Theory and Application of Mixed Polynomial Matrices ... 331
6.1 Descriptions of Dynamical Systems 331
6.1.1 Mixed Polynomial Matrix Descriptions 331
6.1.2 Relationship to Other Descriptions 332
6.2 Degree of Determinant of Mixed Polynomial Matrices 335
6.2.1 Introduction 335
6.2.2 Graph-theoretic Method 336
6.2.3 Basic Identities 337
6.2.4 Reduction to Valuated Independent Assignment 340
6.2.5 Duality Theorems 343
6.2.6 Algorithm 348
6.3 Smith Form of Mixed Polynomial Matrices 355
6.3.1 Expression of Invariant Factors 355

6.3.2 Proofs 363
6.4 Controllability of Dynamical Systems 364
6.4.1 Controllability 364
6.4.2 Structural Controllability 365
6.4.3 Mixed Polynomial Matrix Formulation 372
6.4.4 Algorithm 375
6.4.5 Examples 379
6.5 Fixed Modes of Decentralized Systems 384
6.5.1 Fixed Modes 384
6.5.2 Structurally Fixed Modes 387
6.5.3 Mixed Polynomial Matrix Formulation 390
6.5.4 Algorithm 395
6.5.5 Examples 398

7. Further Topics 403
7.1 Combinatorial Relaxation Algorithm 403
7.1.1 Outline of the Algorithm 403
7.1.2 Test for Upper-tightness 407
7.1.3 Transformation Towards Upper-tightness 413
7.1.4 Algorithm Description 417
7.2 Combinatorial System Theory 418
7.2.1 Definition of Combinatorial Dynamical Systems 419
7.2.2 Power Products 420
7.2.3 Eigensets and Recurrent Sets 422
7.2.4 Controllability of Combinatorial Dynamical Systems 426
7.3 Mixed Skew-symmetric Matrix 431
7.3.1 Introduction 431
7.3.2 Skew-symmetric Matrix 433
7.3.3 Delta-matroid 438
7.3.4 Rank of Mixed Skew-symmetric Matrices 444
7.3.5 Electrical Network Containing Gyrators 446

References 453

Notation Table 469

Index 479

1. Introduction to Structural Approach — Overview of the Book

This chapter is a brief introduction to the central ideas of the combinatorial method of this book for the structural analysis of engineering systems. We explain the motivations and the general framework by referring, as a specific example, to the problem of computing the index of a system of differential-algebraic equations (DAEs). In this approach, engineering systems are described by mixed polynomial matrices. A kind of dimensional analysis is also invoked. It is emphasized that relevant physical observations are crucial to successful mathematical modeling for structural analysis. Though the DAE-index problem is considered as an example, the methodology introduced here is more general in scope and is applied to other problems in subsequent chapters.

1.1 Structural Approach to Index of DAE

1.1.1 Index of Differential-algebraic Equations

Let us start with a simple electrical network[1] of Fig. 1.1 to introduce the concept of an index of a system of *differential-algebraic equations* (*DAE*s) and to explain a graph-theoretic method.

The network consists of a voltage source V (branch 1), two ohmic resistors R_1 and R_2 (branch 2 and branch 3), an inductor L (branch 4), and a capacitor C (branch 5). A state of this network is described by a 10 dimensional vector $\boldsymbol{x} = (\xi^1, \cdots, \xi^5, \eta_1, \cdots, \eta_5)^{\mathrm{T}}$ representing currents ξ^i in and the voltage η_i across branch i $(i = 1, \cdots, 5)$ with reference to the directions indicated in Fig. 1.1. The governing equations in the frequency domain are given by a system of equations $A^{(1)}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{b} = (0, 0, 0, 0, 0; V, 0, 0, 0, 0)^{\mathrm{T}}$ is another 10 dimensional vector representing the source, and $A^{(1)}$ is a 10×10 matrix defined by

[1] This example, described in Cellier [28, §3.7], was communicated to the author by P. Bujakiewicz, F. Cellier, and R. Huber.

$$A^{(1)} = \begin{array}{c|ccccc|ccccc}
 & \xi^1 & \xi^2 & \xi^3 & \xi^4 & \xi^5 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline
 & 1 & -1 & 0 & 0 & -1 & & & & & \\
 & -1 & 0 & 1 & 1 & 1 & & & & & \\ \hline
 & & & & & & -1 & 0 & 0 & 0 & -1 \\
 & & & & & & 0 & 1 & 1 & 0 & -1 \\
 & & & & & & 0 & 0 & -1 & 1 & 0 \\ \hline
 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
 & 0 & R_1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\
 & 0 & 0 & R_2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
 & 0 & 0 & 0 & sL & 0 & 0 & 0 & 0 & -1 & 0 \\
 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & sC
\end{array} . \tag{1.1}$$

As usual, s is the variable for the Laplace transformation that corresponds to $\mathrm{d}/\mathrm{d}t$, the differentiation with respect to time (see Remark 1.1.1 for the Laplace transformation). The first two equations, corresponding to the 1st and 2nd rows of $A^{(1)}$, represent *Kirchhoff's current law* (*KCL*), while the following three equations *Kirchhoff's voltage law* (*KVL*). The last five equations express the element characteristics (*constitutive equations*). The system of equations, $A^{(1)}\boldsymbol{x} = \boldsymbol{b}$, represents a mixture of differential equations and algebraic equations (i.e., a linear time-invariant DAE), since the coefficient matrix $A^{(1)}$ contains the variable s.

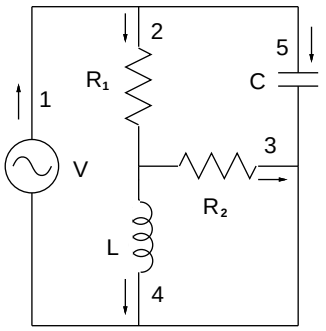


Fig. 1.1. An electrical network

For a linear time-invariant DAE in general, say $A\boldsymbol{x} = \boldsymbol{b}$ with $A = A(s)$ being a nonsingular polynomial matrix in s, the *index* is defined (see Remark 1.1.2) by

$$\nu(A) = \max_{i,j} \deg_s (A^{-1})_{ji} + 1. \tag{1.2}$$

Here it should be clear that each entry $(A^{-1})_{ji}$ of A^{-1} is a rational function in s and the degree of a rational function p/q (with p and q being polynomials) is defined by $\deg_s(p/q) = \deg_s p - \deg_s q$. An alternative expression for $\nu(A)$ is

$$\nu(A) = \max_{i,j} \deg_s((i,j)\text{-cofactor of } A) - \deg_s \det A + 1. \tag{1.3}$$

For the matrix $A^{(1)}$ of (1.1), we see

$$\max_{i,j} \deg_s((i,j)\text{-cofactor of } A^{(1)}) = \deg_s((6,5)\text{-cofactor of } A^{(1)}) = 2,$$
$$\det A^{(1)} = R_1 R_2 + sL \cdot R_1 + sL \cdot R_2 \tag{1.4}$$

by direct calculation and therefore $\nu(A^{(1)}) = 2 - 1 + 1 = 2$ by the formula (1.3).

The solution to $A\boldsymbol{x} = \boldsymbol{b}$ is of course given by $\boldsymbol{x} = A^{-1}\boldsymbol{b}$, and therefore $\nu(A) - 1$ equals the highest order of the derivatives of the input $\boldsymbol{b}$ that can possibly appear in the solution $\boldsymbol{x}$. As such, a high index indicates difficulty in the numerical solution of the DAE, and sometimes even inadequacy in the mathematical modeling. Note that the index is equal to one for a system of purely algebraic equations (where $A(s)$ is free from s), and to zero for a system of ordinary differential equations in the normal form ($\mathrm{d}\boldsymbol{x}/\mathrm{d}t = A_0\boldsymbol{x}$ with a constant matrix A_0, represented by $A(s) = sI - A_0$).

Remark 1.1.1. For a function $x(t)$, $t \in [0, \infty)$, the *Laplace transform* is defined by $\hat{x}(s) = \int_0^\infty x(t)\mathrm{e}^{-st}\mathrm{d}t$, $s \in \mathbf{C}$. The Laplace transform of $\mathrm{d}x(t)/\mathrm{d}t$ is given by $s\hat{x}(s)$ if $x(0) = 0$. See Doetsch [49] and Widder [341] for precise mathematical accounts and Chen [33], Kailath [152] and Zadeh–Desoer [350] for system theoretic aspects of the Laplace transformation. □

Remark 1.1.2. The definition of the index given in (1.2) applies only to linear time-invariant DAE systems. An index can be defined for more general systems and two kinds are distinguished in the literature, a differential index and a perturbation index, which coincide with each other for linear time-invariant DAE systems. See Brenan–Campbell–Petzold [21], Hairer–Lubich–Roche [100], and Hairer–Wanner [101] for details. □

Remark 1.1.3. Extensive study has been made recently on the DAE index in the literature of numerical computation and system modeling. See, e.g., Brenan–Campbell–Petzold [21], Bujakiewicz [26], Bujakiewicz–van den Bosch [27], Cellier–Elmqvist [29], Duff–Gear [60], Elmqvist–Otter–Cellier [72], Gani–Cameron [86], Gear [88, 89], Günther–Feldmann [98], Günther–Rentrop [99], Hairer–Wanner [101], Mattsson–Söderlind [188], Pantelides [264], Ponton–Gawthrop [272], and Ungar–Kröner–Marquardt [324]. □

1.1.2 Graph-theoretic Structural Approach

Structural considerations turn out to be useful in computing the index of DAE. This section describes the basic idea of the graph-theoretic structural methods.

In the graph-theoretic structural approach we extract the information about the degree of the entries of the matrix, ignoring the numerical values

of the coefficients. Associated with the matrix $A^{(1)}$ of (1.1), for example, we consider

$$A^{(1)}_{\mathrm{str}} = \begin{array}{c|ccccc|ccccc} & \xi^1 & \xi^2 & \xi^3 & \xi^4 & \xi^5 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline & t_1 & t_2 & 0 & 0 & t_3 & & & & & \\ & t_4 & 0 & t_5 & t_6 & t_7 & & & & & \\ \hline & & & & & & t_8 & 0 & 0 & 0 & t_9 \\ & & & & & & 0 & t_{10} & t_{11} & 0 & t_{12} \\ & & & & & & 0 & 0 & t_{13} & t_{14} & 0 \\ \hline & 0 & 0 & 0 & 0 & 0 & t_{15} & 0 & 0 & 0 & 0 \\ & 0 & t_{16} & 0 & 0 & 0 & 0 & t_{17} & 0 & 0 & 0 \\ & 0 & 0 & t_{18} & 0 & 0 & 0 & 0 & t_{19} & 0 & 0 \\ & 0 & 0 & 0 & s\,t_{20} & 0 & 0 & 0 & 0 & t_{21} & 0 \\ & 0 & 0 & 0 & 0 & t_{22} & 0 & 0 & 0 & 0 & s\,t_{23} \end{array}$$

where $t_1, \cdots, t_{23}$ are assumed to be independent parameters.

For a polynomial matrix $A = A(s) = (A_{ij})$ in general, we consider a matrix $A_{\mathrm{str}} = A_{\mathrm{str}}(s)$, called the *structured matrix* associated with A, in a similar manner. For a nonzero entry A_{ij}, let $\alpha_{ij} s^{w_{ij}}$ be its leading term, where $\alpha_{ij} \in \mathbf{R} \setminus \{0\}$ and $w_{ij} = \deg_s A_{ij}$. Then $(A_{\mathrm{str}})_{ij}$ is defined to be equal to $s^{w_{ij}}$ multiplied by an independent parameter t_{ij}. Note that the numerical information about the leading coefficient α_{ij} is discarded with the replacement by t_{ij}. Namely, we define the (i,j) entry of A_{str} by

$$(A_{\mathrm{str}})_{ij} = \begin{cases} t_{ij} s^{\deg_s A_{ij}} & (\text{if } A_{ij} \neq 0) \\ 0 & (\text{if } A_{ij} = 0) \end{cases} \tag{1.5}$$

where t_{ij} is an independent parameter. We refer to the index of A_{str} in the sense of (1.2) or (1.3) as the *structural index* of A and denote it by $\nu_{\mathrm{str}}(A)$, namely,

$$\nu_{\mathrm{str}}(A) = \nu(A_{\mathrm{str}}). \tag{1.6}$$

Two different matrices, say A and A', are associated with the same structured matrix, $A_{\mathrm{str}} = A'_{\mathrm{str}}$, if $\deg_s A_{ij} = \deg_s A'_{ij}$ for all (i,j). In other words, a structured matrix is associated with a family of matrices that have a common structure with respect to the degrees of the entries. Though there is no guarantee that the structural index $\nu_{\mathrm{str}}(A)$ coincides with the true index $\nu(A)$ for a particular (numerically specified) matrix A, it is true that $\nu_{\mathrm{str}}(A') = \nu(A')$ for "almost all" matrices A' that have the same structure as A in the sense of $A'_{\mathrm{str}} = A_{\mathrm{str}}$. That is, the equality $\nu_{\mathrm{str}}(A') = \nu(A')$ holds true for "almost all" values of t_{ij}'s, or, in mathematical terms, "generically" with respect to the parameter set $\{t_{ij} \mid A_{ij} \neq 0\}$. (The precise definition of "generically" is given in §2.1.)

The structural index has the advantage that it can be computed by an efficient combinatorial algorithm free from numerical difficulties. This is based on a close relationship between subdeterminants of a structured matrix and matchings in a bipartite graph.

Specifically, we consider a bipartite graph $G(A) = (\mathrm{Row}(A), \mathrm{Col}(A); E)$ with the left vertex set corresponding to the row set $\mathrm{Row}(A)$ of the matrix A, the right vertex set corresponding to the column set $\mathrm{Col}(A)$, and the edge set corresponding to the set of nonzero entries of $A = (A_{ij})$, i.e.,

$$E = \{(i,j) \mid i \in \mathrm{Row}(A), j \in \mathrm{Col}(A), A_{ij} \neq 0\}.$$

Each edge $(i,j) \in E$ is given a weight $w_{ij} = \deg_s A_{ij}$.

For instance, the bipartite graph $G(A^{(1)})$ associated with our example matrix $A^{(1)}$ of (1.1) is given in Fig. 1.2(a). The thin lines indicate edges (i,j) of weight $w_{ij} = 0$ and the thick lines designate two edges, $(i,j) = (9,4), (10,10)$, of weight $w_{ij} = 1$.

A matching M in $G(A)$ is, by definition, a set of edges (i.e. $M \subseteq E$) such that no two members of M have an end-vertex in common. The weight of M, denoted $w(M)$, is defined by

$$w(M) = \sum_{(i,j) \in M} w_{ij},$$

while the size of M means $|M|$, the number of edges contained in M. We denote by $\mathcal{M}_k$ the family of all the matchings of size k in $G(A)$ for $k = 1, 2, \cdots$, and by $\mathcal{M}$ the family of all the matchings of any size (i.e., $\mathcal{M} = \cup_k \mathcal{M}_k$).

For example, the thick lines in Fig. 1.2(b) show a matching M of weight $w(M) = 1$ and of size $|M| = 10$, and $M' = (M \setminus \{(3,10),(10,5)\}) \cup \{(10,10)\}$ is a matching of weight $w(M') = 2$ and of size $|M'| = 9$.

Assuming that A_{str} is an $n \times n$ matrix, we consider the defining expansion of its determinant:

$$\det A_{\mathrm{str}} = \sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}\, \pi \cdot \prod_{i=1}^{n} (A_{\mathrm{str}})_{i\pi(i)} = \sum_{\pi \in \mathcal{S}_n} \mathrm{sgn}\, \pi \cdot \prod_{i=1}^{n} t_{i\pi(i)} \cdot s^{\sum_{i=1}^{n} w_{i\pi(i)}},$$

where $\mathcal{S}_n$ denotes the set of all the permutations of order n, and $\mathrm{sgn}\, \pi = \pm 1$ is the signature of a permutation π. We observe the following facts:

1. Nonzero terms in this expansion correspond to matchings of size n in $G(A)$;
2. There is no cancellation among different nonzero terms in this expansion by virtue of the independence among t_{ij}'s.

These two facts imply the following:

1. The structured matrix A_{str} is nonsingular (i.e., $\det A_{\mathrm{str}} \neq 0$) if and only if there exists a matching of size n in $G(A)$;
2. In the case of a nonsingular A_{str}, it holds that

$$\deg_s \det A_{\mathrm{str}} = \max_{M_n \in \mathcal{M}_n} w(M_n). \tag{1.7}$$

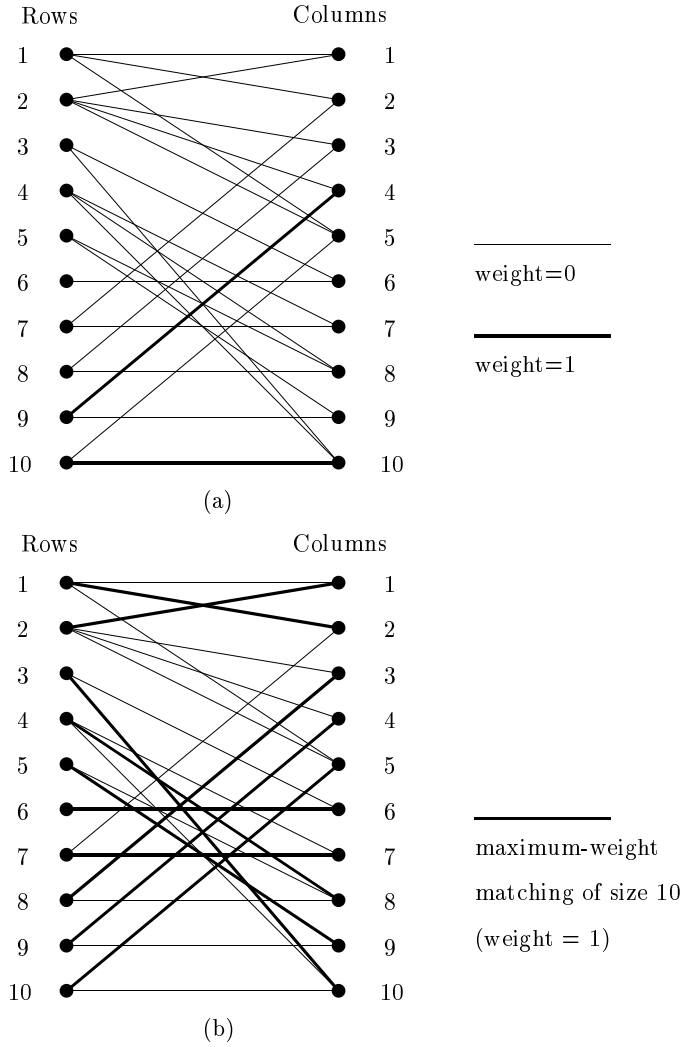


Fig. 1.2. Graph $G(A^{(1)})$ and the maximum-weight matching

A similar argument applied to submatrices of $A_{\rm str}$ leads to more general formulas:

$$\begin{aligned} \operatorname{rank} A_{\rm str} &= \max_{M\in\mathcal{M}} |M|, \\ \max_{|I|=|J|=k} \deg_s \det A_{\rm str}[I,J] &= \max_{M_k\in\mathcal{M}_k} w(M_k) \quad (k=1,\cdots,r_{\rm str}), \end{aligned} \tag{1.8}$$

where $A_{\rm str}[I,J]$ means the submatrix of $A_{\rm str}$ having row set I and column set J, and $r_{\rm str} = \operatorname{rank} A_{\rm str}$. It should be clear that the left-hand side of (1.8) designates the maximum degree of a minor (subdeterminant) of order k.

A combination of the formulas (1.3) and (1.8) yields

$$\nu_{\rm str}(A) = \max_{M_{n-1}\in\mathcal{M}_{n-1}} w(M_{n-1}) - \max_{M_n\in\mathcal{M}_n} w(M_n) + 1 \tag{1.9}$$

for a nonsingular $n \times n$ polynomial matrix A. Thus we have arrived at a combinatorial expression of the structural index.

For the matrix $A^{(1)}$ we have (cf. Fig. 1.2)

$$\max_{M^{(1)}_{n-1}\in\mathcal{M}^{(1)}_{n-1}} w(M^{(1)}_{n-1}) = 2, \qquad \max_{M^{(1)}_{n}\in\mathcal{M}^{(1)}_{n}} w(M^{(1)}_{n}) = 1$$

and therefore $\nu_{\rm str}(A^{(1)}) = 2 - 1 + 1 = 2$, in agreement with $\nu(A^{(1)}) = 2$.

It is important from the computational point of view that efficient combinatorial algorithms are available for checking the existence of a matching of a specified size and also for finding a maximum-weight matching of a specified size. Thus the structural index $\nu_{\rm str}$, with the expression (1.9), can be computed efficiently by solving weighted bipartite matching problems utilizing those efficient combinatorial algorithms.

A number of graph-theoretic techniques (which may be considered variants of the above idea) have been proposed as "structural algorithms" (Bujakiewicz [26], Bujakiewicz–van den Bosch [27], Duff–Gear [60], Pantelides [264], Ungar–Kröner–Marquardt [324]). It is accepted that structural considerations should be useful and effective in practice for the DAE-index problem and that the generic values computed by graph-theoretic "structural algorithms" have practical significance.

1.1.3 An Embarrassing Phenomenon

While the structural approach is accepted fairly favorably, its limitation has also been realized in the literature. A graph-theoretic structural algorithm, ignoring numerical data, may well fail to render the correct answer if numerical cancellations do occur for some reason or other. So the failure of a graph-theoretic algorithm itself should not be a surprise. The aim of this section is to demonstrate a further embarrassing phenomenon that the structural index of our electrical network varies with how KVL is described.

Recall first that the 3rd row of the matrix $A^{(1)}$ represents the conservation of voltage along the loop 1–5 (V–C). In place of this we now take another loop 1–2–4 (V–R_1–L) to obtain a second description of the same electrical network. The coefficient matrix of the second description is given by

$$A^{(2)} = \begin{array}{c|ccccc|ccccc} \xi^1 & \xi^2 & \xi^3 & \xi^4 & \xi^5 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline 1 & -1 & 0 & 0 & -1 & & & & & \\ -1 & 0 & 1 & 1 & 1 & & & & & \\ \hline & & & & & -1 & -1 & 0 & -1 & 0 \\ & & & & & 0 & 1 & 1 & 0 & -1 \\ & & & & & 0 & 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & R_1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & R_2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & sL & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & sC \end{array}, \tag{1.10}$$

which differs from $A^{(1)}$ in the 3rd row. The associated structured matrix $A^{(2)}_{\rm str}$ differs from $A^{(1)}_{\rm str}$ also in the 3rd row, and is given by

$$A^{(2)}_{\rm str} = \begin{array}{c|ccccc|ccccc} \xi^1 & \xi^2 & \xi^3 & \xi^4 & \xi^5 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline t_1 & t_2 & 0 & 0 & t_3 & & & & & \\ t_4 & 0 & t_5 & t_6 & t_7 & & & & & \\ \hline & & & & & t_{24} & t_{25} & 0 & t_{26} & 0 \\ & & & & & 0 & t_{10} & t_{11} & 0 & t_{12} \\ & & & & & 0 & 0 & t_{13} & t_{14} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & t_{15} & 0 & 0 & 0 & 0 \\ 0 & t_{16} & 0 & 0 & 0 & 0 & t_{17} & 0 & 0 & 0 \\ 0 & 0 & t_{18} & 0 & 0 & 0 & 0 & t_{19} & 0 & 0 \\ 0 & 0 & 0 & s\,t_{20} & 0 & 0 & 0 & 0 & t_{21} & 0 \\ 0 & 0 & 0 & 0 & t_{22} & 0 & 0 & 0 & 0 & s\,t_{23} \end{array},$$

where $\{t_i \mid i = 1, \cdots, 7, 10, \cdots, 26\}$ is the set of independent parameters.

Naturally, the index should remain invariant against this trivial change in the description of KVL, and in fact we have

$$\nu(A^{(1)}) = \nu(A^{(2)}) = 2.$$

It turns out, however, that the structural index does change, namely,

$$\nu_{\rm str}(A^{(1)}) = 2, \qquad \nu_{\rm str}(A^{(2)}) = 1,$$

where the latter is computed from the graph $G(A^{(2)})$ in Fig. 1.3; we have

$$\max_{M^{(2)}_{n-1} \in \mathcal{M}^{(2)}_{n-1}} w(M^{(2)}_{n-1}) = 2, \qquad \max_{M^{(2)}_{n} \in \mathcal{M}^{(2)}_{n}} w(M^{(2)}_{n}) = 2$$

and therefore

$$\nu_{\rm str}(A^{(2)}) = \nu(A^{(2)}_{\rm str}) = 2 - 2 + 1 = 1$$

according to the expression (1.9).

The discrepancy between the structural index $\nu_{\rm str}(A^{(2)})$ and the true index $\nu(A^{(2)})$ is ascribed to the discrepancy between $\deg_s \det A^{(2)}_{\rm str} = 2$ and

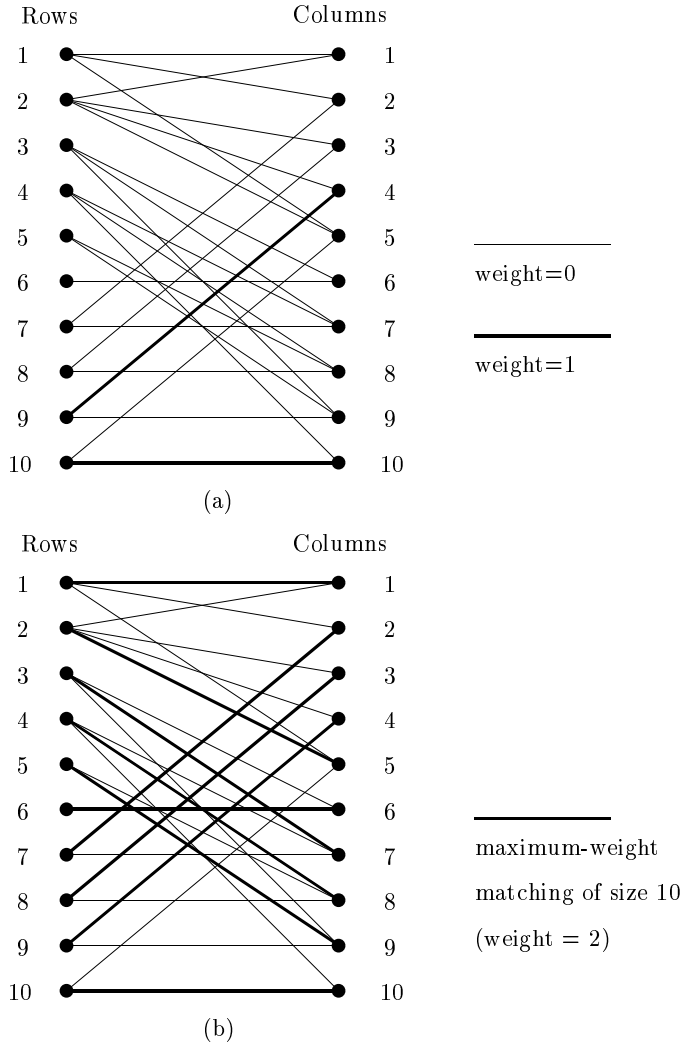


Fig. 1.3. Graph $G(A^{(2)})$ and the maximum-weight matching

$\deg_s \det A^{(2)} = 1$, which in turn is caused by a numerical cancellation in the expansion of $\det A^{(2)}$. A closer look at this phenomenon reveals that this cancellation is *not an accidental cancellation, but a cancellation with good reason* which could be better called *structural cancellation*. In fact, we can identify a 2×2 singular submatrix of the coefficient matrix for the KCL and a 3×3 singular submatrix of the coefficient matrix for the KVL:

$$\begin{array}{c|cc} & \xi^1 & \xi^5 \\ \hline & 1 & -1 \\ & -1 & 1 \end{array}, \qquad \begin{array}{c|ccc} & \eta_2 & \eta_3 & \eta_4 \\ \hline & -1 & 0 & -1 \\ & 1 & 1 & 0 \\ & 0 & -1 & 1 \end{array}$$

as the reason for this cancellation. More specifically, the expansion of $\det A^{(2)}_{\rm str}$ contains four "spurious" quadratic terms

$$\underline{t_1 \cdot t_7} \cdot \underline{t_{25} \cdot t_{11} \cdot t_{14}} \cdot t_{15} \cdot t_{16} \cdot t_{18} \cdot st_{20} \cdot st_{23}, \tag{1.11}$$

$$\underline{t_1 \cdot t_7} \cdot \underline{t_{26} \cdot t_{10} \cdot t_{13}} \cdot t_{15} \cdot t_{16} \cdot t_{18} \cdot st_{20} \cdot st_{23}, \tag{1.12}$$

$$\underline{t_3 \cdot t_4} \cdot \underline{t_{25} \cdot t_{11} \cdot t_{14}} \cdot t_{15} \cdot t_{16} \cdot t_{18} \cdot st_{20} \cdot st_{23}, \tag{1.13}$$

$$\underline{t_3 \cdot t_4} \cdot \underline{t_{26} \cdot t_{10} \cdot t_{13}} \cdot t_{15} \cdot t_{16} \cdot t_{18} \cdot st_{20} \cdot st_{23}, \tag{1.14}$$

which cancel one another when the numerical values as well as the system parameters are given to t_{ij}'s ($t_1 = t_7 = t_{10} = t_{11} = t_{14} = 1$, $t_3 = t_4 = t_{13} = t_{15} = t_{25} = t_{26} = -1$, $t_{16} = R_1$, $t_{18} = R_2$, $t_{20} = L$, $t_{23} = C$). In fact, $\det A^{(2)}$, which is equal to $\det A^{(1)} = R_1R_2 + sL \cdot R_1 + sL \cdot R_2$ given in (1.4), does not contain those terms. Note that the term (1.11) corresponds to the matching in Fig. 1.3(b), and recall that the system parameters R_1, R_2, L, C are treated as mutually independent parameters, which cannot be cancelled out among themselves.

This example demonstrates that the structural index is not determined uniquely by a physical/engineering system, but it depends on its mathematical description. It is emphasized that both

$$A^{(1)}: \begin{array}{|ccccc|} \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline -1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ \hline \end{array} \quad \text{and} \quad A^{(2)}: \begin{array}{|ccccc|} \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline -1 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ \hline \end{array}$$

are equally a legitimate description of KVL and there is nothing inherent to distinguish between the two. In this way the structural index is vulnerable to our innocent choice. This makes us reconsider the meaning of the structural index, which will be discussed in the next section.

Remark 1.1.4. The limitation of the graph-theoretic structural approach, as explained above, is now widely understood. Already Pantelides [264] recognized this phenomenon and more recently Ungar–Kröner–Marquardt [324] expounded this point with reference to an example problem arising from an analysis of distillation columns in chemical engineering. □

1.2 What Is Combinatorial Structure?

In view of the "embarrassing phenomenon" above we have to question the physical relevance of the structural index (1.6) and reconsider how we should

recognize the combinatorial structure of physical systems. The objective of this section is to discuss this issue and to introduce an advanced framework of structural analysis that uses mixed (polynomial) matrices as the main mathematical tool. The framework realizes a reasonable balance between physical faith and mathematical convenience in mathematical modeling of physical/engineering systems. As for physical faith, it is based on two different observations; the one is the distinction between "accurate" numbers (fixed constants) and "inaccurate" numbers (independent system parameters), and the other is the consistency with respect to physical dimensions. As for mathematical convenience, the analysis of mixed (polynomial) matrices and the design of efficient algorithms for them can be done successfully by means of matroid theory. Hence the name of "matroid-theoretic approach" for the advanced framework based on mixed matrices, as opposed to the conventional graph-theoretic approach to structural analysis.

1.2.1 Two Kinds of Numbers

Let us continue with our electrical network. The matrix $A^{(2)}$ of (1.10) can be written as

$$A^{(2)}(s) = A_0^{(2)} + sA_1^{(2)}$$

with

$$A_0^{(2)} = \left[\begin{array}{ccccc|ccccc} 1 & -1 & 0 & 0 & -1 & & & & & \\ -1 & 0 & 1 & 1 & 1 & & & & & \\ \hline & & & & & -1 & -1 & 0 & -1 & 0 \\ & & & & & 0 & 1 & 1 & 0 & -1 \\ & & & & & 0 & 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & R_1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & R_2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad A_1^{(2)} = \left[\begin{array}{ccccc|ccccc} 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & & & & & \\ \hline & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C \end{array}\right]. \tag{1.15}$$

We observe here that the nonzero entries of the coefficient matrices $A_k^{(2)}$ $(k = 0, 1)$ are classified into two groups: one group of fixed constants (± 1) and the other group of system parameters R_1, R_2, L and C. Accordingly, we can split $A_k^{(2)}$ $(k = 0, 1)$ into two parts:

$$A_k^{(2)} = Q_k^{(2)} + T_k^{(2)} \qquad (k = 0, 1)$$

with

$$Q_0^{(2)} = \left[\begin{array}{ccccc|ccccc} 1 & -1 & 0 & 0 & -1 & & & & & \\ -1 & 0 & 1 & 1 & 1 & & & & & \\ \hline & & & & & -1 & -1 & 0 & -1 & 0 \\ & & & & & 0 & 1 & 1 & 0 & -1 \\ & & & & & 0 & 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad T_0^{(2)} = \left[\begin{array}{ccccc|ccccc} 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & & & & & \\ \hline & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & R_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & R_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

$$Q_1^{(2)} = \left[\begin{array}{ccccc|ccccc} 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & & & & & \\ \hline & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right], \quad T_1^{(2)} = \left[\begin{array}{ccccc|ccccc} 0 & 0 & 0 & 0 & 0 & & & & & \\ 0 & 0 & 0 & 0 & 0 & & & & & \\ \hline & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ & & & & & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & C \end{array}\right].$$

It is assumed that the system parameters, R_1, R_2, L, C, are independent parameters. Even when concrete numbers are given to R_1, R_2, L, C, those numbers are not expected to be exactly equal to their nominal values, but they lie in certain intervals of real numbers of engineering tolerance. Even in the extreme case where both R_1 and R_2 are specified to be 1Ω, for example, their actual values will be something like $R_1 = 1.02\Omega$ and $R_2 = 0.99\Omega$.

Generally, when a physical system is described by a polynomial matrix

$$A(s) = \sum_{k=0}^{N} s^k A_k, \tag{1.16}$$

it is often justified (see §1.2.2) to assume that the nonzero entries of the coefficient matrices A_k $(k = 0, 1, \cdots, N)$ are classified similarly into two groups. In other words, we can distinguish the following *two kinds of numbers*, together characterizing a physical system. We may refer to the numbers of the first kind as "fixed constants" and to those of the second kind as "system parameters."

Accurate numbers (fixed constants): Numbers accounting for various sorts of conservation laws such as Kirchhoff's laws which, stemming from topological incidence relations, are precise in value (often ± 1), and therefore cause no serious numerical difficulty in arithmetic operations on them.

Inaccurate numbers (system parameters): Numbers representing independent system parameters such as resistances in electrical networks and masses

in mechanical systems which, being contaminated with noise and other errors, take values independent of one another, and therefore can be modeled as algebraically independent numbers.[2]

Accurate numbers often appear in equations for conservation laws such as Kirchhoff's laws, the law of conservation of mass, energy, or momentum, and the principle of action and reaction, where the nonvanishing coefficients are either 1 or -1, representing the underlying topological incidence relations. Integer coefficients in chemical reactions (*stoichiometric coefficients*), such as "2" and "1" in $2 \cdot \mathrm{H_2O} = 2 \cdot \mathrm{H_2} + 1 \cdot \mathrm{O_2}$, are also accurate numbers. Another example of accurate numbers appears in the defining relation $\mathrm{d}x/\mathrm{d}t = 1 \cdot v$ between velocity v and position x. Typical accurate numbers are illustrated in Fig. 1.4.

The above observation leads to the assumption that the coefficient matrices A_k $(k = 0, 1, \cdots, N)$ in (1.16) are expressed as

$$A_k = Q_k + T_k \qquad (k = 0, 1, \cdots, N), \tag{1.17}$$

where

(A-Q1): Q_k $(k = 0, 1, \cdots, N)$ are matrices over $\mathbf{Q}$ (the field of rational numbers), and

(A-T): The collection $\mathcal{T}$ of nonzero entries of T_k $(k = 0, 1, \cdots, N)$ is algebraically independent over $\mathbf{Q}$.

Namely, each A_k may be assumed to be a *mixed matrix*, in the terminology to be introduced formally in §1.3. Then $A(s)$ is split accordingly into two parts:

$$A(s) = Q(s) + T(s) \tag{1.18}$$

with

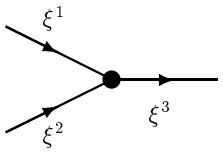
$$Q(s) = \sum_{k=0}^{N} s^k Q_k, \qquad T(s) = \sum_{k=0}^{N} s^k T_k. \tag{1.19}$$

Namely, $A(s)$ is a *mixed polynomial matrix* in the terminology of §1.3.

Our intention in the splitting (1.17) or (1.18) is to extract a more meaningful combinatorial structure from the matrix $A(s)$ by treating the Q-part numerically and the T-part symbolically. This is based on the following observations.

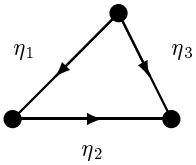
Q-part: The nonzero pattern of the Q-matrices is subject to our arbitrary choice in the mathematical description, as we have seen in our electrical network, and hence the structure of the Q-part should be treated numerically, or linear-algebraically. In fact, this is feasible in practice, since the entries of the Q-matrices are usually small integers, causing no serious numerical difficulty in arithmetic operations.

[2] Informally, "algebraically independent numbers" are tantamount to "independent parameters," whereas a rigorous definition of algebraic independence will be given in §2.1.1.



$$-1 \cdot \xi^1 - 1 \cdot \xi^2 + 1 \cdot \xi^3 = 0$$

KCL



$$-1 \cdot \eta_1 - 1 \cdot \eta_2 + 1 \cdot \eta_3 = 0$$

KVL

$$2 \cdot H_2O = 2 \cdot H_2 + 1 \cdot O_2$$

Stoichiometry

Velocity v – displacement x	$v = 1 \cdot \dot{x}\ (= s \cdot x)$
Current ξ – charge Q	$\xi = 1 \cdot \dot{Q}\ (= s \cdot Q)$

Fig. 1.4. Accurate numbers

T-part: The nonzero pattern of the T-matrices is relatively stable against our arbitrary choice in the mathematical description of constitutive equations and therefore it can be regarded as representing some aspect of the combinatorial structure of the system. It can be treated properly by graph-theoretic concepts and algorithms.

Combination: The structural information from the Q-part and the T-part can be combined properly and efficiently by virtue of the fact that each part defines a well-behaved and well-studied combinatorial structure called matroid. Mathematical and algorithmic results from matroid theory afford effective methods of system analysis.

We may summarize the above as follows:

Q-part	by	linear algebra
T-part	by	graph theory
Combination	by	matroid theory

In §1.3 we shall take a glimpse at how the DAE-index problem can be treated using mixed polynomial matrices and how the embarrassing phenomenon of §1.1.3 can be resolved properly.

1.2.2 Descriptor Form Rather than Standard Form

In introducing mixed polynomial matrices we have assumed that the nonzero entries of the coefficient matrices are either fixed constants or independent parameters. This is an assumption on a description of a physical system, and not an assumption on the system itself. For a system in question there can be many different descriptions, but some of them may satisfy the assumption and others may fail to meet it. In this section we discuss this issue by comparing the state-space equations (Kalman [153]) and the descriptor equations (Luenberger [182, 183]).

Let us consider another example, a simple mechanical system (Fig. 1.5) which consists of two masses m_1, m_2, two springs k_1, k_2, and a damper f; u is the force exerted from outside.

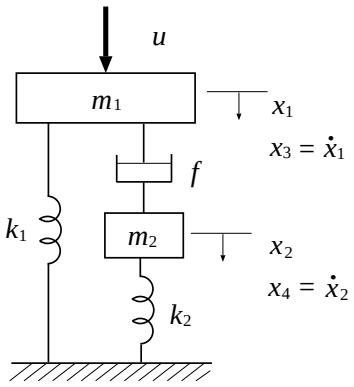


Fig. 1.5. A mechanical system

We may describe the system in the form of *state-space equations*:

$$\dot{\boldsymbol{x}}(t) = \hat{A}\boldsymbol{x}(t) + \hat{B}\boldsymbol{u}(t) \tag{1.20}$$

in terms of $\boldsymbol{x} = (x_1, x_2, x_3, x_4)$ and $\boldsymbol{u} = (u)$, where x_1 and x_2 are vertical displacements (downwards, as indicated in Fig. 1.5) of masses m_1 and m_2, respectively, and x_3 and x_4 are their velocities, and

$$\hat{A} = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 \end{matrix} \\ \left[\begin{matrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -k_1/m_1 & 0 & -f/m_1 & f/m_1 \\ 0 & -k_2/m_2 & f/m_2 & -f/m_2 \end{matrix}\right] \end{array}, \quad \hat{B} = \begin{array}{c} \begin{matrix} u \end{matrix} \\ \left[\begin{matrix} 0 \\ 0 \\ 1/m_1 \\ 0 \end{matrix}\right] \end{array}. \tag{1.21}$$

It should be clear that $\dot{\boldsymbol{x}}$ is a short-hand notation for $\mathrm{d}\boldsymbol{x}/\mathrm{d}t$, the time derivative of $\boldsymbol{x}$.

The state-space equations (1.20) have been useful for investigating analytic and algebraic properties of a dynamical system, and the structural or combinatorial analysis at the early stage[3] was based on it. It is gradually recognized, however, that the state-space equations are not very suitable for representing the combinatorial structure of a system in that the entries of matrices $\hat{A}$ and $\hat{B}$ of (1.20) are usually not independent but interrelated to one another, being subject to algebraic relations. For instance, we have $\hat{A}_{33} + \hat{A}_{34} = 0$ in (1.21), and consequently $\hat{A}$ of (1.21) does not admit a splitting into Q-part and T-part satisfying (A-Q1) and (A-T).

In this respect, the so-called *descriptor form*

$$\bar{F}\dot{\boldsymbol{x}}(t) = \bar{A}\boldsymbol{x}(t) + \bar{B}\boldsymbol{u}(t) \tag{1.22}$$

is more promising, having more flexibility to avoid complicated algebraic relations among entries of the coefficient matrices. Here $\boldsymbol{x}$ is called the descriptor-vector and $\boldsymbol{u}$ is the input-vector. The matrix $\bar{F}$ is not necessarily nonsingular, so that the reduction of (1.22) to the standard state-space form (1.20) is not straightforward. Even when $\bar{F}$ is nonsingular, the reduction to the standard state-space form (1.20) with $\hat{A} = \bar{F}^{-1}\bar{A}$ and $\hat{B} = \bar{F}^{-1}\bar{B}$ entailing complicated algebraic relations among the entries of $\hat{A}$ and $\hat{B}$, is not advantageous from the combinatorial point of view.

To describe our mechanical system in the descriptor form (1.22), it may be natural to introduce two additional variables x_5 (= force by the damper f) and x_6 (= relative velocity of the two masses). Additional equations (constraints) for these variables are given by[4]

$$x_5 = f x_6, \qquad x_6 = \dot{x}_1 - \dot{x}_2.$$

Then the coefficient matrices in (1.22) are given by

$$\bar{F} = \left[\begin{matrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \end{matrix}\right], \quad \bar{A} = \left[\begin{matrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ -k_1 & 0 & 0 & 0 & -1 & 0 \\ 0 & -k_2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & f \\ 0 & 0 & 0 & 0 & 0 & 1 \end{matrix}\right], \quad \bar{B} = \left[\begin{matrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{matrix}\right]. \tag{1.23}$$

[3] Structural approach in the literature of control theory was initiated by Lin [173] in the mid-seventies.

[4] We could replace the equation $x_6 = \dot{x}_1 - \dot{x}_2$ by $x_6 = x_3 - x_4$, which may be more natural. Our choice is to make the example less trivial.

The Laplace transform of the equation (1.22) gives a *frequency domain description*:

$$s\bar{F}\hat{\boldsymbol{x}}(s) = \bar{A}\hat{\boldsymbol{x}}(s) + \bar{B}\hat{\boldsymbol{u}}(s), \quad \text{or} \quad \left[\bar{A} - s\bar{F}\middle|\bar{B}\right]\begin{bmatrix}\hat{\boldsymbol{x}}(s)\\ \hat{\boldsymbol{u}}(s)\end{bmatrix} = \boldsymbol{0},$$

where $\boldsymbol{x}(0) = \boldsymbol{0}$, $\boldsymbol{u}(0) = \boldsymbol{0}$ is assumed (see Remark 1.1.1 for the Laplace transform). Then the system is described by a polynomial matrix

$$A(s) = \left[\bar{A} - s\bar{F}\middle|\bar{B}\right]. \tag{1.24}$$

For our mechanical system we have

$$A(s) = \begin{array}{c|cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & u \\ \hline & -s & 0 & 1 & 0 & 0 & 0 & 0 \\ & 0 & -s & 0 & 1 & 0 & 0 & 0 \\ & -k_1 & 0 & -sm_1 & 0 & -1 & 0 & 1 \\ & 0 & -k_2 & 0 & -sm_2 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 0 & -1 & f & 0 \\ & -s & s & 0 & 0 & 0 & 1 & 0 \end{array} \tag{1.25}$$

as the matrix of (1.24). Note that no complicated algebraic expressions are involved in this matrix, for which it is reasonable to assume (A-Q1) and (A-T) above. Consequently, $A(s)$ of (1.25) is expressed as $A(s) = Q(s) + T(s)$ with

$$Q(s) = \begin{array}{c|cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & u \\ \hline & -s & 0 & 1 & 0 & 0 & 0 & 0 \\ & 0 & -s & 0 & 1 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ & -s & s & 0 & 0 & 0 & 1 & 0 \end{array}, \quad T(s) = \begin{array}{c|cccccc|c} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & u \\ \hline & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & -k_1 & 0 & -sm_1 & 0 & 0 & 0 & 0 \\ & 0 & -k_2 & 0 & -sm_2 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & f & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}. \tag{1.26}$$

Here we have $\mathcal{T} = \{m_1, m_2, k_1, k_2, f\}$ as the set of system parameters.

It is emphasized again that the coefficient matrices $\hat{A}$ and $\hat{B}$ in the standard state-space form do not admit such natural splitting into two parts. Thus we may conclude that the descriptor form is more suitable for representing the combinatorial structure than the standard state-space form.

1.2.3 Dimensional Analysis

Here is a kind of *dimensional analysis* concerning "accurate numbers," i.e., concerning the constant part $Q(s) = \sum_{k=0}^{N} s^k Q_k$ of the matrix $A(s)$ in (1.18).

First we consider the physical dimensional consistency in the system of equations $A(s)\boldsymbol{x} = \boldsymbol{b}$, where $A(s)$ is assumed to be an $m \times n$ matrix. Since this system is to represent a physical system, relevant physical dimensions are

associated with both the variables (corresponding to the components of $\boldsymbol{x}$) and the equations (corresponding to the components of $\boldsymbol{b}$), or alternatively, with both columns and rows of the matrix $A(s)$. Also the entries of $A(s)$ have physical dimensions.

In our mechanical system, for instance, we may choose time T, length L and mass M as the *fundamental quantities* in the dimensional analysis. Then the dimensions of velocity and force are given by $T^{-1}L$ and $T^{-2}LM$, respectively. The physical dimensions associated with the equations, i.e., with the rows of $A(s)$ of (1.25), are

row 1	row 2	row 3	row 4	row 5	row 6
velocity	velocity	force	force	force	velocity
$T^{-1}L$	$T^{-1}L$	$T^{-2}LM$	$T^{-2}LM$	$T^{-2}LM$	$T^{-1}L$

(1.27)

whereas those with the variables (x_i and u), i.e., with the columns of $A(s)$, are

col 1	col 2	col 3	col 4	col 5	col 6	col 7
length	length	velocity	velocity	force	velocity	force
L	L	$T^{-1}L$	$T^{-1}L$	$T^{-2}LM$	$T^{-1}L$	$T^{-2}LM$

(1.28)

The $(3,1)$-entry "$-k_1$" of $A(s)$, for example, has a dimension of $T^{-2}M$.

The *principle of dimensional homogeneity* demands that

$$\begin{aligned}&[\text{Dimension of } i\text{th row}]\\ &= [\text{Dimension of } (i,j) \text{ entry}] \times [\text{Dimension of } j\text{th column}]\end{aligned} \tag{1.29}$$

for each (i,j) with $A_{ij} \neq 0$. For instance, this identity reads

$$T^{-2}LM = T^{-2}M \times L$$

for $(i,j) = (3,1)$ in our mechanical system.

Choosing time as one of the fundamental dimensions, we denote by $-r_i$ and $-c_j$ the exponent to the dimension of time associated respectively with the ith row and the jth column. Then the (i,j) entry of $A(s)$ should have the dimension of time with exponent $c_j - r_i$.

In our mechanical system we have

$$\begin{aligned}&r_1 = r_2 = 1,\ r_3 = r_4 = r_5 = 2,\ r_6 = 1;\\ &c_1 = c_2 = 0,\ c_3 = c_4 = 1,\ c_5 = 2,\ c_6 = 1,\ c_7 = 2\end{aligned}$$

from (1.27) and (1.28).

The "accurate numbers" usually represent topological and/or geometrical incidence coefficients (cf. Fig. 1.4), which have no physical dimensions, so that it is natural to expect that the entries of Q_k in (1.19) are dimensionless constants. On the other hand, the variable (indeterminate) "s" should have

the physical dimension of the inverse of time, since it corresponds to $\mathrm{d}/\mathrm{d}t$, the differentiation with respect to time. This implies, in particular, that each entry of the term $s^k Q_k$ has the physical dimension of time with exponent $-k$. On the other hand, the (i,j) entry of $A(s)$, and hence the (i,j) entry of $Q(s)$, should have the dimension of time with exponent $c_j - r_i$, as pointed out above.

Combining these two facts we obtain

$$r_i - c_j = k \qquad \text{if} \quad (Q_k)_{ij} \neq 0, \tag{1.30}$$

or in matrix form:

$$Q(s) = \mathrm{diag}\,[s^{r_1}, \cdots, s^{r_m}] \cdot Q(1) \cdot \mathrm{diag}\,[s^{-c_1}, \cdots, s^{-c_n}], \tag{1.31}$$

where $\mathrm{diag}\,[d_1, d_2, \cdots]$ means a diagonal matrix having diagonal entries $d_1, d_2, \cdots$. It follows from this decomposition that every nonvanishing subdeterminant of $Q(s)$ is a monomial in s over $\mathbf{Q}$, i.e., of the form αs^p with a nonvanishing rational number α and a nonnegative integer p.

In our mechanical system, it can be verified that $Q(s)$ of (1.26) admits an expression of the form (1.31):

$$\left[\begin{array}{cccccc|c} -s & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -s & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ -s & s & 0 & 0 & 0 & 1 & 0 \end{array}\right]$$

$$= \left[\begin{array}{cccccc|c} s & 0 & 0 & 0 & 0 & 0 & \\ 0 & s & 0 & 0 & 0 & 0 & \\ 0 & 0 & s^2 & 0 & 0 & 0 & \\ 0 & 0 & 0 & s^2 & 0 & 0 & \\ 0 & 0 & 0 & 0 & s^2 & 0 & \\ 0 & 0 & 0 & 0 & 0 & s & \end{array}\right] \cdot \left[\begin{array}{cccccc|c} -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}\right] \cdot \left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & s^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & s^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & s^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & s^{-1} & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & s^{-2} \end{array}\right].$$

Note that the diagonal entries s^{r_i} and s^{-c_j} are determined from the negative of the exponents to T (time) in (1.27) and (1.28).

We have thus arrived at a subclass of mixed polynomial matrices suitable for representing the structure of linear time-invariant dynamical systems. Namely, we are to consider the class of polynomial matrices $A(s)$ in indeterminate s with rational coefficients which are represented as

$$A(s) = Q(s) + T(s),$$

where

(A-Q2): Every nonvanishing subdeterminant of $Q(s)$ is a monomial in s over $\mathbf{Q}$, and

(A-T): The collection $\mathcal{T}$ of the nonzero coefficients of the entries of $T(s)$ is algebraically independent over $\mathbf{Q}$.

The dual viewpoint of dimensional analysis and structural analysis constitutes the physical foundation of the mathematical development explained in this book. Chapter 3 will be devoted to a full discussion about this issue.

1.3 Mathematics on Mixed Polynomial Matrices

While the previous section is devoted to physical motivations for mixed polynomial matrices, this section offers an informal introduction to their mathematical aspects through a successful treatment of the DAE-index problem left unanswered in §1.1.3.

1.3.1 Formal Definitions

The concept of a mixed matrix is defined formally as follows. Let $\boldsymbol{K}$ be a subfield of a field $\boldsymbol{F}$. A matrix $A = (A_{ij})$ over $\boldsymbol{F}$ (i.e., $A_{ij} \in \boldsymbol{F}$) is called a *mixed matrix* with respect to $(\boldsymbol{K}, \boldsymbol{F})$ if

$$A = Q + T, \tag{1.32}$$

where

(M-Q) $Q = (Q_{ij})$ is a matrix over $\boldsymbol{K}$ (i.e., $Q_{ij} \in \boldsymbol{K}$), and
(M-T) $T = (T_{ij})$ is a matrix over $\boldsymbol{F}$ (i.e., $T_{ij} \in \boldsymbol{F}$) such that the set of its nonzero entries is algebraically independent over $\boldsymbol{K}$.

For example, $A_0^{(2)}$ in (1.15) is a mixed matrix with respect to $(\boldsymbol{K}, \boldsymbol{F}) = (\mathbf{Q}, \mathbf{Q}(\mathcal{T}))$, where $\mathcal{T} = \{R_1, R_2\}$ and $\mathbf{Q}(\mathcal{T})$ is the field of rational functions in $\mathcal{T}$ with rational coefficients.

Similarly, a polynomial matrix $A(s)$ over $\boldsymbol{F}$ (i.e., $A_{ij} \in \boldsymbol{F}[s]$) is called a *mixed polynomial matrix* with respect to $(\boldsymbol{K}, \boldsymbol{F})$ if

$$A(s) = Q(s) + T(s) = \sum_{k=0}^{N} s^k Q_k + \sum_{k=0}^{N} s^k T_k \tag{1.33}$$

for some integer $N \geq 0$, where

(MP-Q1) Q_k $(k = 0, 1, \cdots, N)$ are matrices over $\boldsymbol{K}$, and
(MP-T) T_k $(k = 0, 1, \cdots, N)$ are matrices over $\boldsymbol{F}$ such that the set of their nonzero entries is algebraically independent over $\boldsymbol{K}$.

A mixed polynomial matrix with respect to $(\boldsymbol{K}, \boldsymbol{F})$ is a mixed matrix with respect to $(\boldsymbol{K}(s), \boldsymbol{F}(s))$. It should be obvious that (MP-Q1) and (MP-T) generalize (A-Q1) and (A-T), respectively. Corresponding to (A-Q2) we consider

(MP-Q2) Every nonvanishing subdeterminant of $Q(s)$ is a monomial in s over $\boldsymbol{K}$.

The major part of this book is devoted to the development of a theory of mixed (polynomial) matrices. Mixed matrices can be treated successfully by means of the standard results in the theory of matroids and submodular functions. For mixed polynomial matrices, on the other hand, a quantitative generalization of matroid theory, the theory of valuated matroids, is needed.

Matroid	(Chap. 2)	Mixed matrix	(Chap. 4)
Valuated matroid	(Chap. 5)	Mixed polynomial matrix	(Chap. 6)

Remark 1.3.1. The concept of a mixed matrix is useful also in solving a system of linear/nonlinear equations $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$. Suppose we solve this equation numerically by the *Newton method*. This amounts to solving

$$J(\boldsymbol{x})\Delta\boldsymbol{x} = -\boldsymbol{f}(\boldsymbol{x})$$

for a correction $\Delta\boldsymbol{x}$, where $J(\boldsymbol{x})$ is the Jacobian matrix of $\boldsymbol{f}(\boldsymbol{x})$. The equations may be divided into linear and nonlinear parts as

$$\boldsymbol{f}(\boldsymbol{x}) = Q\boldsymbol{x} + \boldsymbol{g}(\boldsymbol{x}),$$

where Q is a constant matrix representing the linear part and $\boldsymbol{g}(\boldsymbol{x})$ is the nonlinear part. Accordingly, the Jacobian matrix $J(\boldsymbol{x})$ takes the form of

$$J(\boldsymbol{x}) = Q + T(\boldsymbol{x}),$$

where $T(\boldsymbol{x})$ is the Jacobian matrix of the nonlinear part $\boldsymbol{g}(\boldsymbol{x})$. This expression suggests that we may treat $J(\boldsymbol{x}) = Q + T(\boldsymbol{x})$ as a mixed matrix by regarding (or modeling) the nonvanishing entries of $T(\boldsymbol{x})$ as being algebraically independent over $\boldsymbol{K} = \mathbf{R}$, where $\boldsymbol{F}$ is a certain field consisting of functions (e.g., the field of rational functions). This direction will be pursued further in §4.4.5. □

1.3.2 Resolution of the Index Problem

In this section we describe how the DAE-index problem can be treated successfully by means of mixed polynomial matrices. In so doing we intend to convey a general idea of the mathematical issues around mixed polynomial matrices without entering into technical details. Precise definitions and statements are given in subsequent chapters (§6.2 in particular).

For the DAE-index problem, we consider the degree of the determinant of an $n \times n$ mixed polynomial matrix $A(s) = Q(s) + T(s)$. The row set and the column set of $A(s)$ are denoted by R and C, respectively, and the submatrices of Q and T with row set I and column set J are written as $Q[I, J]$ and $T[I, J]$.

The following identity (cf. Theorem 6.2.4) shows how to combine the structural information from Q-part and T-part to obtain the information about A.

Theorem 1.3.2. *For a nonsingular mixed polynomial matrix* $A(s) = Q(s) + T(s)$,

$$\deg_s \det A = \max_{\substack{|I|=|J| \\ I \subseteq R, J \subseteq C}} \{\deg_s \det Q[I,J] + \deg_s \det T[R \setminus I, C \setminus J]\}. \quad (1.34)$$

(We adopt the convention that $\deg_s 0 = -\infty$*, so that the maximum in* (1.34) *is taken effectively over all* (I,J) *such that both* $Q[I,J]$ *and* $T[R \setminus I, C \setminus J]$ *are nonsingular.)* □

It is mentioned that the assumptions (MP-Q1) and (MP-T) are crucial for this formula, whereas generally the right-hand side of (1.34) is only an upper bound on $\deg_s \det A$.

The right-hand side of the identity (1.34) involves a maximization over all pairs (I,J), the number of which is almost as large as $2^{|R|+|C|}$, too large for an exhaustive search for maximization. Fortunately, however, it is possible to design an efficient algorithm to compute this maximum on the basis of the facts that each of the functions

$$f_Q(I,J) = \deg_s \det Q[I,J], \qquad f_T(I,J) = \deg_s \det T[I,J]$$

can be evaluated easily, and that the maximization ("combination of Q-part and T-part") can be done efficiently, as follows.

Q-part: The matrix $Q(s)$ is a polynomial matrix with coefficients in $\boldsymbol{K}$, and the function $f_Q(I,J)$ may be computed in many different ways (see §7.1). If the stronger condition (MP-Q2) on $Q(s)$ is satisfied with the expression (1.31), it holds that

$$f_Q(I,J) = \begin{cases} \sum_{i \in I} r_i - \sum_{j \in J} c_j & (\text{if } \det Q(1)[I,J] \neq 0) \\ -\infty & (\text{otherwise}) \end{cases}$$

where $Q(1)$ is a matrix over $\boldsymbol{K}$ and therefore $\det Q(1)[I,J]$ can be computed by means of arithmetic operations over $\boldsymbol{K}$.

T-part: The matrix $T(s)$ is a structured matrix in that the nonzero coefficients are algebraically independent. Therefore the function $f_T(I,J)$ can be evaluated efficiently by means of bipartite matchings, as has been explained in §1.1.2 (recall (1.7) in particular).

Combination: Each of $f_Q(I,J)$ and $f_T(I,J)$ enjoys a combinatorially nice property, being a variant of a valuated matroid. Therefore, the maximum can be computed by a straightforward application of an algorithmic scheme for the valuated matroid intersection problem, where the functions $f_Q(I,J)$ and $f_T(I,J)$ are evaluated polynomially many times (polynomial in n). If the stronger condition (MP-Q2) on $Q(s)$ is satisfied, the maximization can be reduced to an easier problem (the ordinary weighted matroid intersection problem).

For easy reference, let us give a temporary name, say "Algorithm D," to the above algorithm for computing the degree of determinant, while deferring a concrete description to Chap. 6 (cf. Remark 6.2.17).

With "Algorithm D" for the degree of determinant we can compute the index $\nu(A)$ of $A(s)$ based on the formula (1.3):

$$\nu(A) = \max_{i,j} \deg_s((i,j)\text{-cofactor of } A) - \deg_s \det A + 1.$$

A simplest way is to apply "Algorithm D" repeatedly to the whole matrix A and all the submatrices of order $n-1$ (which are n^2 in number) to obtain $\deg_s \det A$ and $\deg_s((i,j)\text{-cofactor of } A)$ for all (i,j). This naive method already gives a polynomial-time algorithm for $\nu(A)$, though an improvement for efficiency is possible.

Let us illustrate the above method for the matrix $A^{(2)}$ of (1.10), the second coefficient matrix of our electrical network. We regard it as a mixed polynomial matrix $A^{(2)}(s) = Q^{(2)}(s) + T^{(2)}(s)$ with

$Q^{(2)}(s) =$

ξ^1	ξ^2	ξ^3	ξ^4	ξ^5	η_1	η_2	η_3	η_4	η_5
1	−1	0	0	−1					
−1	0	1	1	1					
					−1	−1	0	−1	0
					0	1	1	0	−1
					0	0	−1	1	0
0	0	0	0	0	−1	0	0	0	0
0	0	0	0	0	0	−1	0	0	0
0	0	0	0	0	0	0	−1	0	0
0	0	0	0	0	0	0	0	−1	0
0	0	0	0	−1	0	0	0	0	0

,

$T^{(2)}(s) =$

ξ^1	ξ^2	ξ^3	ξ^4	ξ^5	η_1	η_2	η_3	η_4	η_5
0	0	0	0	0					
0	0	0	0	0					
					0	0	0	0	0
					0	0	0	0	0
					0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	R_1	0	0	0	0	0	0	0	0
0	0	R_2	0	0	0	0	0	0	0
0	0	0	sL	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	sC

under the assumption that R_1, R_2, L, and C are independent parameters. The matrix $Q^{(2)}(s)$, being free from s, satisfies the stronger condition (MP-Q2) trivially. On the right-hand side of (1.34) in Theorem 1.3.2 we can take

$$I = R \setminus \{\text{row } 8, \text{row } 9\}, \qquad J = C \setminus \{\text{column } 3, \text{column } 4\}$$

as a maximizer,[5] for which

$$\deg_s \det Q^{(2)}[I,J] = 0, \qquad \deg_s \det T^{(2)}[R \setminus I, C \setminus J] = 1.$$

Hence we obtain $\deg_s \det A^{(2)} = 1$. In a similar manner we obtain

[5] It can be checked easily that, if $\deg_s \det T^{(2)}[R \setminus I', C \setminus J'] = 2$, then $\deg_s \det Q^{(2)}[I', J'] = -\infty$ (i.e., $\det Q^{(2)}[I', J'] = 0$). In particular, this is the case with $I' = R \setminus \{\text{row } 7, \text{row } 8, \text{row } 9, \text{row } 10\}$ and $J' = C \setminus \{\text{column } 2, \text{column } 3, \text{column } 4, \text{column } 10\}$, which corresponds to the "spurious" quadratic terms (1.11)–(1.14).

$$\max_{i,j} \deg_s((i,j)\text{-cofactor of } A^{(2)}) = 2,$$

and therefore $\nu(A^{(2)}) = 2 - 1 + 1 = 2$ by (1.3). Thus, the present method gives the correct answer for the matrix $A^{(2)}$ for which the graph-theoretic method fails.

Finally, we mention a duality theorem which forms a basis for "Algorithm D" for computing the maximum in (1.34), and which provides a systematic method for index reduction. This theorem is a concrete manifestation of a general duality theorem for the valuated matroid intersection problem.

Consider a transformation of $A(s)$ to

$$\tilde{A}(s) = \operatorname{diag}(s; -p_R) \cdot A(s) \cdot \operatorname{diag}(s; p_C)$$

with two integer vectors $p_R \in \mathbf{Z}^R$ and $p_C \in \mathbf{Z}^C$, where $\operatorname{diag}(s;p)$ for a vector $p = (p_i)$ means a diagonal matrix $\operatorname{diag}[s^{p_1}, s^{p_2}, \cdots]$. This is quite a natural transformation that corresponds to rewriting the system of equations $A\boldsymbol{x} = \boldsymbol{b}$ into $\tilde{A}\tilde{\boldsymbol{x}} = \tilde{\boldsymbol{b}}$ in terms of

$$\tilde{\boldsymbol{x}} = \operatorname{diag}(s; -p_C) \cdot \boldsymbol{x}, \qquad \tilde{\boldsymbol{b}} = \operatorname{diag}(s; -p_R) \cdot \boldsymbol{b}.$$

The transformation yields another mixed polynomial[6] matrix $\tilde{A}(s) = \tilde{Q}(s) + \tilde{T}(s)$ with

$$\tilde{Q}(s) = \operatorname{diag}(s; -p_R) \cdot Q(s) \cdot \operatorname{diag}(s; p_C),$$
$$\tilde{T}(s) = \operatorname{diag}(s; -p_R) \cdot T(s) \cdot \operatorname{diag}(s; p_C).$$

For any choice of $p_R \in \mathbf{Z}^R$ and $p_C \in \mathbf{Z}^C$ we obviously have

$$\max_{\substack{|I|=|J| \\ I\subseteq R, J\subseteq C}} \{\deg_s \det \tilde{Q}[I,J] + \deg_s \det \tilde{T}[R\setminus I, C\setminus J]\}$$
$$\leq \max_{\substack{|I|=|J| \\ I\subseteq R, J\subseteq C}} \deg_s \det \tilde{Q}[I,J] + \max_{\substack{|I|=|J| \\ I\subseteq R, J\subseteq C}} \deg_s \det \tilde{T}[R\setminus I, C\setminus J]. \quad (1.35)$$

A natural question here is whether the inequality in (1.35) turns into an equality for an appropriate choice of p_R and p_C.

The duality theorem asserts the existence of a pair of vectors $p_R \in \mathbf{Z}^R$ and $p_C \in \mathbf{Z}^C$ ("dual variables") which makes the inequality of (1.35) into an equality. Combining this with Theorem 1.3.2 we obtain the following theorem (cf. Theorem 6.2.15).

Theorem 1.3.3. *For a nonsingular mixed polynomial matrix $A(s) = Q(s) + T(s)$, there exist $p_R \in \mathbf{Z}^R$ and $p_C \in \mathbf{Z}^C$ such that*

$$\tilde{A}(s) = \operatorname{diag}(s; -p_R) \cdot A(s) \cdot \operatorname{diag}(s; p_C) = \tilde{Q}(s) + \tilde{T}(s)$$

[6] Strictly speaking, "polynomial" is to be replaced by "Laurent polynomial" (cf. §2.1.1), since negative powers of s can appear in $\tilde{A}(s)$.

satisfies

$$\deg_s \det \tilde{A} = \max_{\substack{|I|=|J| \\ I\subseteq R,\, J\subseteq C}} \deg_s \det \tilde{Q}[I,J] + \max_{\substack{|I|=|J| \\ I\subseteq R,\, J\subseteq C}} \deg_s \det \tilde{T}[R\backslash I, C\backslash J]. \quad (1.36)$$

A further condition (i) $p_R \geq \mathbf{0}$, $p_C \geq \mathbf{0}$, *or* (ii) $p_R \leq \mathbf{0}$, $p_C \leq \mathbf{0}$, *may be imposed on* p_R *and* p_C. □

The "Algorithm D" for computing $\deg_s \det A$ consists of finding such a pair of vectors (p_R, p_C) to transform $A(s)$ to $\tilde{A}(s)$ as well as an optimal pair of subsets (I, J) in the right-hand side of (1.34).

In addition to the algorithmic significance, the transformation to $\tilde{A}(s)$ has another meaning of index reduction transformation. Recall that the index is equal to one for a system of purely algebraic equations and to zero for a system of ordinary differential equations in the normal form.

Corollary 1.3.4. *The index of* $\tilde{A}(s)$ *is at most one:* $\nu(\tilde{A}) \leq 1$.

Proof. For any $I \subseteq R$ and $J \subseteq C$, consider the expression of $\deg_s \det \tilde{A}[I,J]$ as in Theorem 1.3.2. This shows that $\deg_s \det \tilde{A}[I,J]$ is bounded by the right-hand side of (1.36), which is equal to $\deg_s \det \tilde{A}$. Then (1.3) establishes the claim. ■

Theorem 1.3.3 is illustrated below for the matrix $A^{(2)}$ of (1.10) of our electrical network. By "Algorithm D" we can find

$$p_R = (0,0,0,0,0;0,0,0,0,1), \qquad p_C = (1,0,0,0,1;0,0,0,0,0)$$

as the "dual variables," which yields

$$\tilde{A}^{(2)}(s) = \begin{array}{c|ccccc|ccccc} Q^1 & \xi^2 & \xi^3 & \xi^4 & Q^5 & \eta_1 & \eta_2 & \eta_3 & \eta_4 & \eta_5 \\ \hline s & -1 & 0 & 0 & -s & & & & & \\ -s & 0 & 1 & 1 & s & & & & & \\ \hline & & & & & -1 & -1 & 0 & -1 & 0 \\ & & & & & 0 & 1 & 1 & 0 & -1 \\ & & & & & 0 & 0 & -1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & R_1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & R_2 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & sL & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & C \end{array} . \quad (1.37)$$

As asserted in (1.36), we have

$$\deg_s \det \tilde{A}^{(2)} = 2 = 1 + 1 = \max_{I,J} \deg_s \det \tilde{Q}^{(2)}[I,J] + \max_{I,J} \deg_s \det \tilde{T}^{(2)}[I,J]$$

in contrast to

$$\deg_s \det A^{(2)} = 1 < 0 + 2 = \max_{I,J} \deg_s \det Q^{(2)}[I,J] + \max_{I,J} \deg_s \det T^{(2)}[I,J].$$

In accordance with the change of variables $\tilde{\boldsymbol{x}} = \mathrm{diag}\,(s; -p_C) \cdot \boldsymbol{x}$, the 1st and 5th columns of $\tilde{A}^{(2)}$ are indexed by

$$Q^1 = s^{-(p_C)_1}\xi^1 = \int \xi^1 \mathrm{d}t \quad \text{(charge supplied by the source)},$$
$$Q^5 = s^{-(p_C)_5}\xi^5 = \int \xi^5 \mathrm{d}t \quad \text{(charge stored in the capacitor)},$$

respectively. The last row of $\tilde{A}^{(2)}$ represents the constitutive equation $Q^5 = C\eta_5$ ("charge is proportional to voltage") rather than $\xi^5 = C\dot{\eta}_5$ ("current is proportional to the time derivative of voltage"), which corresponds to the last row of $A^{(2)}$. The "dual variables" p_R and p_C admit such a physical interpretation.

Remark 1.3.5. The structural index computed by a "structural algorithm," whether graph-theoretic or matroid-theoretic, is equal to the true index value only in the generic case ("almost surely") and there is no guarantee of equality for a particular (numerically specified) matrix. "Combinatorial relaxation algorithms," to be described in Chap. 7, are refinements of the "structural algorithms" such that they first compute the generic values by combinatorial algorithms and then invoke a minimum amount of numerical arithmetic operations to always guarantee the true index value. □

Remark 1.3.6. Theorem 1.3.2 implies as an immediate corollary that A is nonsingular if and only if there exist $I \subseteq R$ and $J \subseteq C$ such that both $Q[I,J]$ and $T[R \setminus I, C \setminus J]$ are nonsingular. Application of this result to submatrices yields a rank identity:

$$\mathrm{rank}\, A = \max_{\substack{|I|=|J| \\ I \subseteq R,\, J \subseteq C}} \{\mathrm{rank}\, Q[I,J] + \mathrm{rank}\, T[R \setminus I, C \setminus J]\},$$

which is true for a mixed matrix $A = Q + T$ in general. This identity acts as a first bridge between mixed matrices and matroids, as will be seen in Chap. 4 (cf. Theorem 4.2.8, in particular). □

1.3.3 Block-triangular Decomposition

Kirchhoff's laws can be written in many different ways, on which the success and failure of the graph-theoretic method depends. This is what we have seen with the coefficient matrices $A^{(1)}$ and $A^{(2)}$ of our electrical network of Fig. 1.1; $A^{(1)}$ is good for the graph-theoretic method, while $A^{(2)}$ is not.

The matroid-theoretic method, i.e., the method of mixed polynomial matrices, works for both $A^{(1)}$ and $A^{(2)}$. It is stable against arbitrary choices in the representation of KCL and KVL, giving a correct answer for any legitimate

representation of Kirchhoff's laws. Though the "embarrassing phenomenon" has been resolved by the matroid-theoretic method, no answer has yet been given to the question: Why is $A^{(1)}$ better than $A^{(2)}$?

In this section we shall answer this question by considering *block-triangular decompositions* of the coefficient matrices. In so doing we intend to introduce another mathematical aspect of mixed (polynomial) matrices.

Recall from §1.1.3 that the discrepancy $\nu_{\rm str}(A^{(2)}) \neq \nu(A^{(2)})$ is caused by the "spurious" quadratic terms (1.11)–(1.14) in $\det A^{(2)}_{\rm str}$, whereas those terms do not appear in $\det A^{(1)}_{\rm str}$. The reason for this difference is apparent from the following block-triangular forms of $A^{(1)}$ and $A^{(2)}$ obtained through reorderings of rows and columns:

$$\bar{A}^{(1)} = \begin{array}{c|ccccccccccc}
 & \xi^1 & \xi^2 & \xi^3 & \xi^4 & \eta_2 & \eta_3 & \eta_4 & \xi^5 & \eta_5 & \eta_1 \\
\text{KCL} & 1 & -1 & 0 & 0 & & & & -1 & & \\
\text{KCL} & -1 & 0 & 1 & 1 & & & & 1 & & \\
\text{KVL} & & & & & 1 & 1 & 0 & & -1 & \\
\text{KVL} & & & & & 0 & -1 & 1 & & & \\
\text{branch 2} & 0 & R_1 & 0 & 0 & -1 & 0 & 0 & & & \\
\text{branch 3} & 0 & 0 & R_2 & 0 & 0 & -1 & 0 & & & \\
\text{branch 4} & 0 & 0 & 0 & sL & 0 & 0 & -1 & & & \\
\text{branch 5} & & & & & & & & -1 & sC & \\
\text{KVL} & & & & & & & & & -1 & -1 \\
\text{branch 1} & & & & & & & & & & -1
\end{array},$$

$$\bar{A}^{(2)} = \begin{array}{c|ccccccccccc}
 & \xi^1 & \xi^2 & \xi^3 & \xi^4 & \eta_2 & \eta_3 & \eta_4 & \xi^5 & \eta_5 & \eta_1 \\
\text{KCL} & 1 & -1 & 0 & 0 & & & & -1 & & \\
\text{KCL} & -1 & 0 & 1 & 1 & & & & 1 & & \\
\text{KVL} & & & & & 1 & 1 & 0 & & -1 & \\
\text{KVL} & & & & & 0 & -1 & 1 & & 0 & \\
\text{branch 2} & 0 & R_1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \\
\text{branch 3} & 0 & 0 & R_2 & 0 & 0 & -1 & 0 & 0 & 0 & \\
\text{branch 4} & 0 & 0 & 0 & sL & 0 & 0 & -1 & 0 & 0 & \\
\text{branch 5} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & sC & \\
\text{KVL} & & & & & -1 & 0 & -1 & & 0 & -1 \\
\text{branch 1} & & & & & & & & & & -1
\end{array}.$$

In $\bar{A}^{(1)}$ the entry "sC" cannot contribute to $\det \bar{A}^{(1)}$ for a combinatorial reason that it lies in an off-diagonal block, whereas "sC" disappears in $\det \bar{A}^{(2)}$ as a result of cancellation. This is why "sC" does not appear in $\det A^{(1)}_{\rm str}$ but does survive in $\det A^{(2)}_{\rm str}$.

The above observation leads naturally to a further question: Which description of Kirchhoff's laws gives a finest block-triangular decomposition? This may be rephrased to the following problem: Given a coefficient matrix A of the form

$$A = \begin{array}{|c|c|}\hline \text{K C L} & O \\ \hline O & \text{K V L} \\ \hline \multicolumn{2}{|c|}{\text{constitutive eqns}} \\ \hline \end{array},$$

such as $A^{(1)}$ and $A^{(2)}$, find a finest block-triangular decomposition using a transformation of the form

$$\bar{A} = P_{\mathrm{r}} \cdot \begin{array}{|c|c|c|}\hline S_{\mathrm{KCL}} & O & O \\ \hline O & S_{\mathrm{KVL}} & O \\ \hline O & O & I \\ \hline \end{array} \cdot \begin{array}{|c|c|}\hline \text{K C L} & O \\ \hline O & \text{K V L} \\ \hline \multicolumn{2}{|c|}{\text{constitutive eqns}} \\ \hline \end{array} \cdot P_{\mathrm{c}}, \tag{1.38}$$

where S_{KCL} and S_{KVL} are nonsingular matrices, I is an identity matrix, and P_{r} and P_{c} are permutation matrices.

It will be shown in Chap. 4 that there exists a canonical (finest) block-triangular decomposition under the transformation (1.38) and it can be computed by an efficient algorithm. The canonical form of $A = A^{(2)}$ is given by

$$\bar{A}^{(3)} = \begin{array}{c|cccccccccc} & \xi^1 & \xi^2 & \xi^3 & \xi^4 & \eta_2 & \eta_3 & \eta_4 & \xi^5 & \eta_5 & \eta_1 \\ \hline \text{KCL} & 1 & -1 & & & & & & -1 & & \\ \text{KCL} & & -1 & 1 & 1 & & & & & & \\ \text{KVL} & & & & & 1 & 1 & 0 & & -1 & \\ \text{KVL} & & & & & 0 & -1 & 1 & & & \\ \text{branch 2} & & R_1 & 0 & 0 & -1 & 0 & 0 & & & \\ \text{branch 3} & & 0 & R_2 & 0 & 0 & -1 & 0 & & & \\ \text{branch 4} & & 0 & 0 & sL & 0 & 0 & -1 & & & \\ \text{branch 5} & & & & & & & & -1 & sC & \\ \text{KVL} & & & & & & & & & -1 & -1 \\ \text{branch 1} & & & & & & & & & & -1 \end{array} \tag{1.39}$$

that is obtained from $A^{(2)}$ through (1.38) with

$$S_{\mathrm{KCL}} = \begin{array}{|cc|}\hline 1 & 0 \\ 1 & 1 \\ \hline \end{array}, \qquad S_{\mathrm{KVL}} = \begin{array}{|ccc|}\hline -1 & 0 & 0 \\ -1 & -1 & -1 \\ -1 & 0 & -1 \\ \hline \end{array}.$$

The canonical form does not depend on how Kirchhoff's laws are initially written (in particular, $A^{(1)}$ and $A^{(2)}$ have a common canonical form (1.39)), and therefore, it can be regarded as representing a certain combinatorial structure inherent in our electrical network.

Returning to the determinant, note first that the transformation (1.38) changes the determinant only by a constant factor (i.e., $\det \bar{A} = \pm \det S_{\mathrm{KCL}} \cdot \det S_{\mathrm{KVL}} \cdot \det A$). The determinant of $\bar{A}$ is obviously the product of the determinants of the diagonal blocks. Hence the off-diagonal entries, such as "sC" in (1.39), do not appear in the determinant. Furthermore, a theorem in Chap. 4 (Theorem 4.5.4) states that any system parameter contained in a diagonal block appears in $\det \bar{A}$. This explains, in our example, why R_1, R_2 and sL appear in

$$\det A^{(1)} = \det A^{(2)} = -\det \bar{A}^{(3)} = R_1 R_2 + sL \cdot R_1 + sL \cdot R_2.$$

For the degree of $\det \bar{A}^{(3)}$, we may apply the method of §1.3.2 to each diagonal block and sum up the degrees thus obtained. This will be more efficient than to apply the same method to the whole matrix.

Block-triangular decompositions of the above kind are one of the major topics studied in this book; among which are the Dulmage–Mendelsohn decomposition in Chap. 2 and the combinatorial canonical form (CCF) in Chap. 4.

Notes. The concept of a mixed matrix was introduced in Murota–Iri [238] with the observation on two kinds of numbers explained in §1.2.1, while the dimensional analysis in §1.2.3 is due to Murota [200]. This chapter is an improved version of a presentation (Murota [223]) at ICIAM 95.