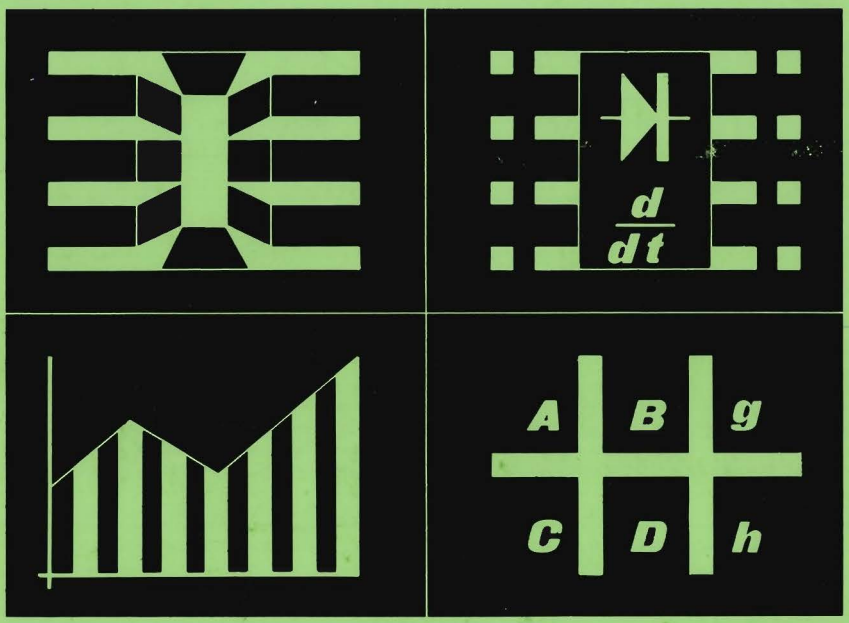
Piecewise-linear modelling and analysis

W.M.G. van Bokhoven

Piecewise–Linear Modelling and Analysis



PIECEWISE-LINEAR MODELLING AND ANALYSIS

PIECEWISE-LINEAR MODELLING AND ANALYSIS

PROEFSCHRIFT

ter verkrijging van de graad van doctor in de technische wetenschappen aan de Technische Hogeschool Eindhoven, op gezag van de rector magnificus, prof. ir. J. Erkelens, voor een commissie aangewezen door het college van dekanen in het openbaar te verdedigen op dinsdag 26 mei 1981 te 16.00 uur

door

Wilhelmus Maria Gezinus van Bokhoven
geboren te Drunen

DIT PROEFSCHRIFT IS GOEDGEKEURD
DOOR DE PROMOTOREN

Prof. Dr.-Ing. J. A. G. Jess
en
Prof. Dr. O. Wing

Aan Tonny

CONTENTS

1. Introduction and summary

In electrical engineering, the simulation of complex electronic networks has become one of the important and vital tasks in the design of large scale integrated circuits. Due to inevitable disturbing effects like for example stray capacitances, finite isolation or parasitic transistors it is no longer adequate or even possible to construct a physical equivalent circuit from discrete components in order to test the circuit-design for its correct operation by a sequence of measurements.

In a simulation of the circuit, these parasitic effects can in theory be incorporated easily and thus an early indication of the correctness of the design can be obtained before the circuit is actually integrated in silicon.

Then, although simulation seems to be the answer to a lot of problems, the difficulties are merely shifted from one place to another, i.e. the simulation task will become enormously complex. It has to cover logic simulation of digital circuits as well as timing simulation of these devices and in addition it must be able to predict the response of analog circuits for DC and transient excitations. Moreover, these traditionally different simulation tasks now have to be applied to the same device on account of the combination of both analog and digital circuits on a single chip. Such a complex task requires at least a hierarchical organization and subdivision in order to be completed successfully within a reasonable amount of time. Besides that, severe restrictions have to be imposed on the structure of the used database for the different levels of simulation, to enable an easy conversion of the calculated data between those levels. For example, the result of a simulation at register-transfer level must be translated to data required at the gate level for application in a timing simulation or the other way round.

Assuming all these problems to be solved, a simulation can still be unsuccessful because of instability arising by the interaction of the various simulation tasks at the different levels or types of simulation.

All these requirements indicate that an efficient simulation

can only be performed when organized in such a way that a common approach for all levels is obtained using a uniform database which avoids the conversion problems mentioned.

The main task at each simulation level is the solution of a set of nonlinear algebraic equations describing the network response after an appropriate modelling has been applied to the network components. Hence a common modelling approach for each simulation level is the key to a uniform simulation process and the associated database. Due to the diversity of the network components at the various levels with respect to the complexity of their internal structure or functional behaviour, a hierarchical and transparent organization of the simulation is necessary.

At the lowest level, a model must describe the electrical behaviour of bipolar or MOS transistors represented by a relatively simple set of nonlinear equations. At the gate level, dealing with logic signals represented by voltage or current levels, the operations on these signals are given by input-output relations in terms of tables. At a still higher level, macromodels are needed for the description of for example shift registers or full-adders in digital circuits as well as for operational amplifiers or oscillators in analog circuits. Thus, depending on the particular level or circuit type, the classical simulation of a circuit may require the solution of a set of nonlinear equations, a set of piecewise-linear equations, boolean function evaluation or even a combination of them.

Each of these topics has its own class of solution methods and related problems. The nonlinear equation set is generally solved by applying some version of the standard Newton-Raphson process [1] which is based on local linearizations valid in an infinitesimal small area of the solution space. The nonlinear equation set is then in fact transformed in a sequence of Linear equations in such a way that the sequence of solutions of these linear equations converges to a solution of the original problem. One of the main difficulties in this approach is that a sufficiently close estimate of the solution has to be supplied in advance in order to guarantee the convergence of these iterative solution algorithms.

This problem is almost as difficult as the determination of the solution itself. In particular for digital circuits where the DC state is generally non-unique, the resulting set of nonlinear equations will have multiple solutions and the standard Newton-Raphson method cannot easily cope with such a situation. In [2], Sandberg and Willson supplied a theory to test for the existence of a unique DC operating point for a certain class of networks with bipolar transistors. This test may reduce the simulation problems but is in fact of more theoretical importance as it shows whether a circuit has intrinsic possibilities to be used as a multi-stable element or can have a unique DC operating point exclusively, which is e.g. required for amplifiers and related analog circuit-blocks.

The evaluation of boolean functions in logic simulation is in principle not a difficult task. However the translation of these data in order to link up with transient simulation in some analog part of the circuit is not straightforward. An easy way out is provided by using threshold-gate equivalents which can be modelled by piecewise-linear functions.

From the foregoing observations it will be clear that a simulation program which includes different types of element-models requires a complex internal organization for dealing with the different solution methods and databases. Hence for an efficient simulation a choice has to be made for a unique type of representation for all element-models. As with increasing complexity, the DC response of electronic components shows a tendency to approach a piecewise-linear relationship, it seems most natural to use the piecewise-linear description as a basis for the modelling. Besides that it is far more easy to approximate a nonlinear function by a piecewise-linear function than the other way round.

This thesis is supposed to demonstrate that piecewise-linear modelling can be applied to solve many of the problems outlined above. Traditionally piecewise-linear modelling implies a number of problems.

As to the piecewise-linear equations, these were defined by piecewise-linear functions specified by a list of mappings and

polyhedral regions in some Euclidean space. In contrast with the Newton-Raphson method this yields a linearization valid in a collection of finite regions in the solution space. Within each region, the piecewise-linear equation set is then equivalent to a linear equation set which can easily be solved by standard methods from linear algebra. In this case the main problem is two-fold. First of all the required list of polyhedral regions suffers from the disadvantage that a lot of data is required to specify the boundaries of each region and the mapping of those regions. For an approximation of sufficient accuracy the number of required regions may become quite large. In fact the description is still a collection of local linear descriptions without any strong interrelationship except for some continuity demands at the region boundaries. It can be compared to the sequence of locally linearized descriptions which arise during the run of the Newton-Raphson algorithm, although the latter linearizations are only generated along a particular path in the solution space.

The second problem is situated in the method of localizing the particular region in which the solution is to be found. This process requires a search through the list of regions, guided by some algorithm and is time-consuming as well as inefficient. One cause of inefficiency is that the boundaries of a region are at least stored twice for each adjacent region and general information about adjacency of regions is not available in a direct way. Moreover, for any linear transformation applied on the set of piecewise-linear functions, the boundaries of the regions in the transformed space have to be determined one by one in order to be able to construct the list of the mapping for the transformed equation set. In particular, an electrical interconnection of two components would require the construction of a composite list for the combined piecewise-linear equation set from the separate lists of the equation sets for both individual components. This complex operation results in an exponential growth of the storage complexity and computation.

In order for a piecewise-linear approach to be successful, the forementioned disadvantages of the piecewise-linear equations have

to be overcome. As with nonlinear equations or mappings one should look for a comparable implicit global description of PL mappings, valid in all polyhedral regions of the definition space.

One of the main achievements of the research reported in this thesis is the presentation of such a new global implicit description of piecewise-linear mappings. This description is called the statemodel of the mapping. Basically the model consists of two functionally different parts. It combines linear algebraic methods with combinatorial methods in a single model.

The first part of the state-model actually describes the piecewise linear mapping in terms of affine transformations. The second (combinatorial) part is the so called state-equation which is responsible for the subdivision of the domain of the mapping in polyhedral regions. The well known linear complementarity problem is the essential tool which is used in the state-equation to obtain the required piecewise-linear mapping.

A principal advantage of the state-model over the standard list-oriented specification of PL mappings is the fact that the boundaries of the polyhedral regions are implicitly defined by the state-equation. Therefore it is no longer required to determine explicitly the transformed boundaries when some linear transformation is applied to the PL mapping. It appears that a number of equivalent formulations of the state-model can be given with the property that each equivalent formulation behaves like a kind of natural representation for one particular polyhedral region and still contains all necessary global information about the mapping. In each polyhedral region the state-model can be considered as a system being in a particular state. The natural descriptions for the different regions are obtained by transitions of the corresponding states. The mappings in adjacent regions in space are then described by adjacent states of the state-model. The neighbour relations between the various regions play an important role in the solution algorithms of PL equations applied on the state-model. These neighbour relations are mapped on a graph which is called the "structure" of the state-model. The structure contains information about mutually reachable states or regions.

In particular, a structural degeneracy of the state-model can lead to a non-connected graph, indicating that homotopy algorithms for finding a solution of the PL equations may not function properly. Moreover in this case the mapping is in general non-continuous.

Besides the application of this state-model to the modelling of electronic devices, it also appears to be a new valuable tool in the theory of PL homeomorphisms. Due to its formulation in terms of matrices and vectors instead of regions and corresponding mappings, new theorems about homeomorphism properties can be derived, which express the conditions for a homeomorphism in certain known matrix properties which can easily be verified. The results bear a strong connection with the theory developed by Sandberg and Willson [2].

To introduce the state-model, we start in chapter 2 from the description of piecewise-linear electrical networks containing ideal diodes. The various equivalent hybrid representations of a linear lumped memoryless multiport will form the basis of the state-model and the ideal diodes can be considered as to embody a real physical state. From this approach a physical background for the state-model is supplied which may help to understand the specific properties of the model. It appears that multiport networks which yield a positive output impedance at each port under all possible passive load conditions at the remaining ports can be expected to have a unique DC response when loaded by ideal diodes. In chapter 3 this expectation is substantiated by proving the equivalence of some mathematical conditions defining the so called Class P and P_0 matrices and certain physical properties of a multiport network regarding the positivity of a set of port immitancies. At the same time, the hybrid structure of a multiport network is introduced as a graph representing the various existing hybrid descriptions of the multiport. This graph will be used in the sequel as an adjacency structure for relating the different linear regions with respect to their topological properties such as neighbour relationship or connectivity. It is shown later on that a structural degeneracy of a multiport network indicated by singular principal submatrices of one of its

hybrid matrices implies that the represented mapping is not uniquely defined.

Based on the earlier observations of the properties of PL networks, the state-model is defined in chapter 4 as a mathematical model for PL mappings. The notions of state and equivalent representations are introduced and brought into relation with the subdivision of the domain of the mapping in polyhedral regions. In parallel with the standard state-space formulation of linear dynamic systems, the controllability and observability of the state-model is discussed. For later application in chapter 7 concerning homeomorphic mappings, the minimal state-model is introduced to restrict the class of PL mappings such that some homeomorphism theorems can be derived later on.

In chapter 5 the standard solution algorithms for the linear complementarity problem are discussed and brought into relation with the state-model formulation. In particular the application of Cottle's algorithm is extensively explained.

Chapter 6 deals with a new approach to the solution of the linear complementarity problem. By applying the so called modulus transformation, this problem is transformed into a specific type of nonlinear equations. The important advantage of this transformation is that the complementarity conditions disappear such that standard methods for the solution of nonlinear equations can be applied to solve the linear complementarity problem.

For a specific class of linear complementarity problems a contraction mapping algorithm is presented which runs in polynomial time-complexity, in contrast with the algorithms of chapter 5 which are known to be exponential in complexity. This new algorithm may save a lot of computation time in worst-case situations. Furthermore the modulus transformation is related to a global model of a PL mapping given by Chua [21], which latter model appears to be a special case of the state-model formulation.

In chapter 7 the state-model of the inverse mapping is derived by a simple linear transformation of the original model. The corresponding operation for a list-specified PL mapping would be much more complicated since the boundaries of the polyhedral regions in the range of mapping then have to be determined one by

one. The straightforward relation between the state-models of a mapping and its inverse mapping means an important advantage which is used to derive some new theorems about homeomorphic mappings. The previously defined structure of the state-model is extensively used in the proofs of the relevant theorems and appears to be a powerful tool. Furthermore explicit state-models are constructed for some relatively simple PL mappings, to be used later on in the formulation of more complex state-models.

Chapter 8 then deals with the construction of these complex state-models. To this purpose a state-model is considered as a description of a black box representing some PL system. Piecewise-linear systems of increasing complexity are then constructed by interconnection of PL subsystems just like the construction of a complex electronic circuit from standard electronic devices. The mathematical implementation of an interconnection amounts simply to the catenation of a linear relation between interconnected variables. Depending on the type of physical system that is to be represented by the state-models each interconnection may require a set of linear relations te be added to the system. For example a galvanic interconnection of electrical components requires some voltages to become equal and the sum of some currents te become zero. The proposed construction results in a hierarchically based modelling process in which each model can be used as a submodel or component in a more complex overall model.

The physical origin of the models is immaterial for the construction and analysis of the overall model, since we are always dealing with PL mappings. For example a PL segment originating from the description of a digital gate is indistinguishable from that of a bipolar transistor.

This property of the modelling is extremely important and allows for an efficient mixed-level simulation of electrical networks since a uniform database and solutuion algorithm can be applied.

The remaining part of chapter 8 is devoted to the implementation of dynamic PL systems, which appear to fit in a natural way to the

state-model description. By application of an integration formula, the differentiation operator can be replaced by an algebraic relation, yielding a state-model which may be considered as a PL equivalent of the transition matrix arising in the description of standard nonlinear dynamic systems.

It will be no surprise that time-cusp problems associated with the standard transient analysis of certain electronic circuits are completely absent in the state-model formulation, due to the fact that a sequence of two or more pivoting steps in one of the pivoting algorithms may yield a discontinuous response.

Finally in chapter 9 some DC models are constructed for a couple of common electronic devices to give an indication of the general line of the model construction. Two examples are included as an illustration of the capabilities of the presented theory.

2. Piecewise-linear electrical networks

As discussed in the introduction, our intention is to find some kind of closed form analytical description of a piecewise-linear mapping and a specific algorithm such that the solution of piecewise-linear equations can be determined by applying this algorithm to the model of a corresponding PL mapping. This analytical description or model must be able to cope with multi-dimensional PL mappings and has to cover PL dynamic systems as well. Furthermore, working with lists of linear systems that define the PL mapping in each PL region of the domain of the mapping should be avoided since these lists may grow exponentially long when a concatenation of mappings has to be constructed. Instead of this, one global description should cover the description of all PL regions.

It is obvious that these demands are not easily met when starting from a pure mathematical point of view. As in fact the PL equations of interest originate from linearizing nonlinear electrical networks, it is easy and more natural to attempt to construct the required global model from the description of a class of piecewise-linear networks. Hopefully such a model can be generalized to satisfy the properties required for less restricted PL mappings as well.

To this purpose, we focus our attention to the description of PL electrical networks. For the time being, we will restrict ourselves to PL resistive networks and cover dynamic networks later on.

2.1. The multiport network-model

In order to establish a PL behaviour, a network should contain PL elements besides all other kinds of linear static elements like resistors and independent or controlled linear sources. We will use the ideal diode, to be defined shortly, as the most primitive piecewise-linear element and consider all other PL elements as being constructed from these ideal diodes and the standard linear elements mentioned before. Let us define the ideal diode as a two-pole element with a voltage-current characteristic as given in fig.2.1. Note the orientation of the diode voltage which is opposite to the standard

reference direction. Then the voltage v and current i of the ideal diode satisfy

$$v \geq 0, \quad i \geq 0, \quad v \cdot i = 0 \tag{2.1}$$

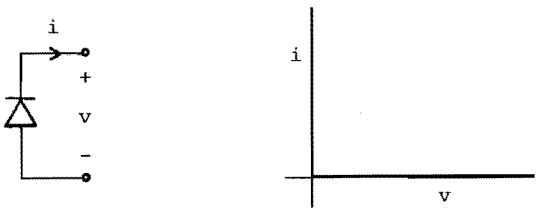


fig.2.1.

Due to the previous restriction concerning the allowed element types, we can depict the PL network as given in fig.2.2, in which all ideal diodes are pulled out of the network. The remaining multi-port network M is then a linear lumped memoryless network containing resistors and all types of fixed or controlled linear sources.

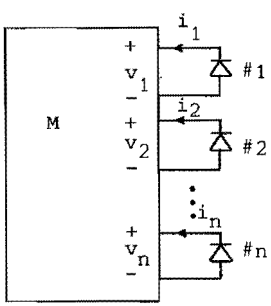


fig.2.2.

Let us assume that the network contains n ideal diodes. We denote the port voltages and port currents of the linear n-port network by

$$v = (v_1, v_2, \ldots, v_n)^t, \quad i = (i_1, i_2, \ldots, i_n)^t \tag{2.2}$$

It is known that the network M can be described by a so called constraint-matrix description [3], of the following form

$$C_1 v + C_2 i = b, \tag{2.3}$$

with $C_1 = (m \times n)$, $C_2 = (m \times n)$, $m \leq n$

If $m < n$, the network response for any allowed excitation with fixed sources at the ports is apparently non-unique. We will exclude these types of degenerate networks and assume that the linear multiport M is nondegenerate such that at least one excitation with voltage or current sources exists which yield a unique response. That is $m = n$ and at least one reordering of the columns of (2.3) exists which satisfies

$$Nw + Mz = b, \quad \det(N) \neq 0 \tag{2.4}$$

$$\text{and } w_k = v_k, \quad z_k = i_k \quad \text{or} \quad w_k = i_k, \quad z_k = v_k; \quad k=1,2,\ldots,m \tag{2.5}$$

The vectors w and z in (2.4) satisfying (2.5) are called complementary port vectors. From (2.4) we then obtain the familiar hybrid description

$$w = Hz + a, \tag{2.6}$$

with $H = -N^{-1}M$ and $a = N^{-1}b$

Due to (2.5), there are 2^n different partitions of v and i which may yield 2^n different but equivalent hybrid descriptions of the form (2.6). By virtue of the previous definition, a nondegenerate network has at least one hybrid representation.

As for the ideal diodes, we have by (2.1) and (2.5) for each ideal diode the relations $w_k \geq 0$, $z_k \geq 0$ and $w_k z_k = 0$, which is equivalent to

$$w \geq 0, \quad z \geq 0, \quad w^t z = 0 \tag{2.7}$$

the inequalities taken componentwise.

Hence the PL network of fig.2.2 is fully characterized by

$$w = Hz + a$$
$$w \geq 0, \quad z \geq 0, \quad w^t z = 0, \tag{2.8}$$

with w and z some particular complementary partition of the diode currents and voltages.

2.2. Some considerations about the network response

The network response is obtained by finding a solution pair w,z which satisfies (2.8). This general problem is known in the field of mathematical programming as the linear complementarity problem (LCP problem). Algorithms to solve this problem will be discussed in detail in chapter 5. However at this place it will be convenient to sketch the idea behind these algorithms in terms of operations on the network of fig.2.2.

It is obvious that an excitation of the network, which defines a vector a in (2.8), influences the particular states of the ideal diodes. That is, depending on the the vector a, some diodes will conduct ($v_k = 0$, $i_k \geq 0$) and the remaining will block ($v_k > 0$, $i_k = 0$). As a conducting diode forms a short-circuit and a blocking diode forms an open-circuit, the state of the diodes in turn influences the topology and thus the response of the network M. Then the network of fig.2.2 can be seen as the continuous counterpart of a finite state machine, with M corresponding to the combinatorial logic and the collection of diodes corresponding to the sequential part, memorizing the particular state (see fig.2.3).

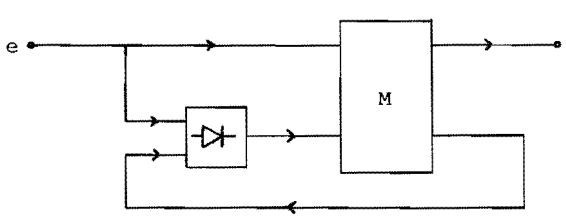


fig.2.3.

Calculation of the response to some excitation is now equivalent to the determination of the diode state, i.e. the specific conducting state of all diodes. A direct approach to do so would be to consider all 2^n possible diode states and solve for the response. If for some diode state a response of M will yield the same state as was initially assumed, then a solution is found. In fact, this suggests some iterative algorithm visualized in fig.2.4, which is obtained from fig.2.3 by opening the lower feedback loop.

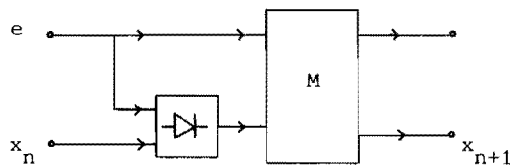


fig.2.4.

In fig.2.4 x_{n+1} is some function of x_n and the input e, say

$$x_{n+1} = f(x_n, e) \tag{2.9}$$

The solutions of (2.8) are then fixed-points of the mapping f(x,e) and, under certain conditions, equation (2.9) can be used as a contraction-mapping algorithm to find those solutions. The details of such an algorithm are discussed in chapter 6.

Another approach would be to move the input e gradually from a value e_0 for which a solution is known, to the value e^* for which the solution is required, meanwhile keeping track of all changes in the diode states. In this way the solution is continuously embedded in a (one) parameter family of solutions which can be found by continuation methods. A particular algorithm of this class is the well known Katzenelson algorithm [4].

All algorithms have in common that they run through a certain sequence of diode states ending in the required solution sooner or later.

To facilitate further discussions we make the following definitions.

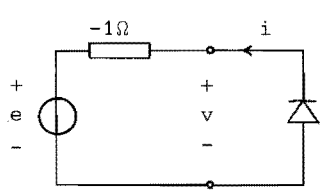
A port of a linear resistive multiport network is called a voltage (current) port if the excitation at that port is performed by a voltage (current) source. A port is an uncommitted port if the excitation is not specified. [See e.g. 5]

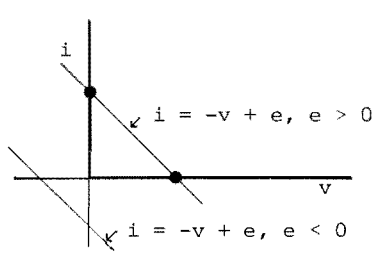
Let us from now on assign $z_k = v_k$ to a voltage port and $z_k = i_k$ to a current port, without leaving any uncommitted ports. If, for a particular port assignment, the matrix N in the relation corresponding to (2.4) is nonsingular, we say that the corresponding port assignment is admissible. Then for every excitation leading to an admissible port assignment, the network M has a unique response and there exists a hybrid representation of that network of the form (2.6).

Now, each diode state corresponds to a certain open- or short-circuit termination of the ports of M. An open-circuit can be seen as a current source of strength zero and a short-circuit as a voltage source of strength zero. In this way, each diode state is related to a particular port assignment. Then for a diode state yielding an admissible port assignment, we have from (2.6) that $w = a$ since $z = 0$.

According to (2.8), for this state to arise it is in addition required that $a \geq 0$. Hence we may find the solutions of (2.8) by investigating for all admissible port assignments whether they yield a nonnegative vector a in the corresponding hybrid representation (2.6).

Depending on the properties of M, multiple solutions or even no solution at all may exist as can be seen from the simple one-port example of fig.2.5. This circuit has two solution for $e > 0$ and no solution for $e < 0$.





It is already known that the LCP problem (2.8) has a unique solution for any vector a, if and only if $H \in P$ [6], [10], where the properties of the matrix class P have been given by Fiedler and Pták [8]. If the vector a is restricted to some subspace of $\mathbb{R}^n$, no such general statement is known by now. However, an upper bound for the number of different diode states has been given by Belevitch [9]. The properties of class P will be discussed in equivalent network terms in chapter 3.

2.3. Piecewise-linear two-pole elements

The construction of a PL equivalent of a nonlinear electronic network requires PL descriptions to be derived for various nonlinear devices such as for example bipolar transistors, diodes, tunnel diodes and MOS transistors. The nonlinear models of these devices currently employed in simulation packages have the property that the static nonlinearities can be incorporated by properly interconnected nonlinear two-pole elements. Then piecewise-linear modelling of these devices amounts to find PL models for one-dimensional nonlinear functions, describing the voltage-current characteristics of the particular two-pole elements. Let us assume that a certain two-pole element has been approximated by a finite number of linear segments in a sufficiently accurate way, yielding a continuous PL voltage-current relation with positive slopes for each segment (see fig.2.6). Then a network model in the sense of section 2.1 can be obtained as follows.

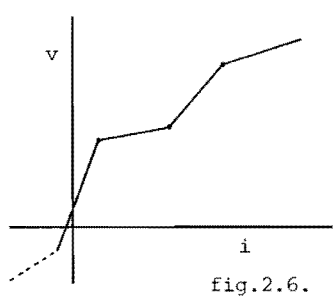


fig.2.6.

Let the modelling be conducted by going bottom-up along the voltage-axis and assume that a network model M_k has been derived which describes the required v-i curve for $v \le v_k$, having a fixed slope R_k for $v > v_{k-1}$. We then extend the network to M_{k+1} in such a way that the interval (v_{k-1}, v_k) is included with a slope R_{k+1} for $v > v_k$. Repetition of this step will then yield the required model after all segments have been processed.

When extending M, two situations may arise with either $R_{k+1} > R_k$ or $R_{k+1} < R_k$. In the first case we define M_{k+1} to be the network of fig.2.7. Since $R_{k+1} > R_k$, the resistor to be added is positive.

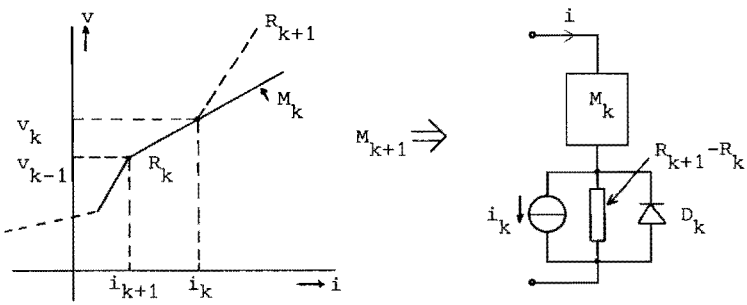


fig.2.7.

For $i \le i_k$, diode D_k conducts and the behaviour of M_{k+1} is identical to that of M_k. As soon as $i > i_k$ the diode blocks and the slope for $v > v_k$ becomes $R_k + (R_{k+1}-R_k) = R_{k+1}$, as required.

When $R_{k+1} < R_k$, the dual situation applies, leading to the network of fig.2.8.

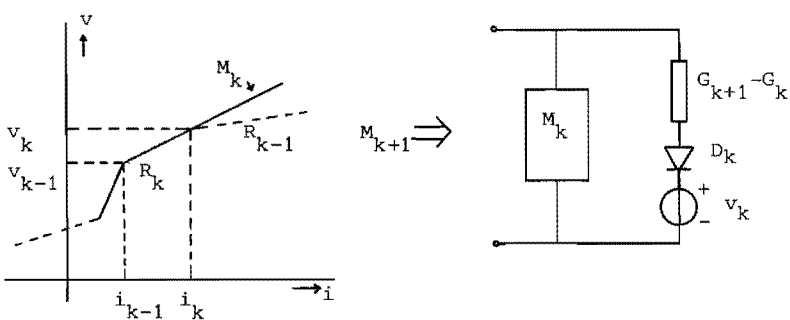


fig.2.8.

Since a model for the lowest (first) segment is a simple resistor in series with some fixed voltage source, the required process can be performed step by step. Note that the above synthesis also includes segments with zero or infinite slope which are just on the edge of realizability.

Hence monotonous PL voltage-current characteristics with non-negative slope can be modelled with positive resistors, fixed sources and ideal diodes only. This important property will be used later on when the solvability of PL networks will be discussed.

For voltage-current characteristics with negative slopes such as for example arising in tunnel diodes, the given synthesis leads to negative resistors. However, if the voltage-current characteristic under consideration is such that the voltage is a single-valued function of the current or the other way round, the synthesis can be performed with only one negative resistor.

As can be seen from fig.2.9, for the case $v = v(i)$ the circuit can be realized by a negative resistor $R^* = \min_i(R_i)$ in series with a network having a monotonously increasing v-i curve.

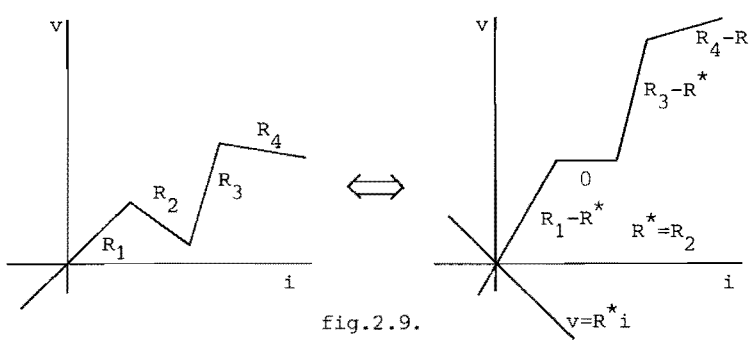


fig.2.9.

3. Equivalent network properties of P and P_0 matrices

As we showed in chapter 2, formulation of the network equations of a class of piecewise-linear networks results in an equation set known as the linear complementarity problem, given by

$$\begin{aligned} &w = Hz + a \qquad H = (n\times n) \\ &w \geq 0, \quad z \geq 0, \quad w^t z = 0 \qquad w,z \in \mathbb{R}^n \end{aligned} \tag{3.1}$$

In (3.1) the elements of the vectors w and z are the port voltages and port currents of a linear multiport network M and H represents a hybrid matrix describing the electrical behaviour of M. The independent port variables are collected in z and the complementary variables in w.

For the moment let us assume that the impedance matrix R of M does exist. With z = i and w = v, the network M is then described by

$$v = Ri + a \tag{3.2}$$

Equation (3.2) leads to a Thévenin equivalent of M as depicted in fig.3.1.

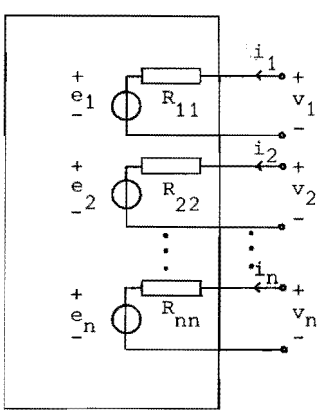


fig.3.1.

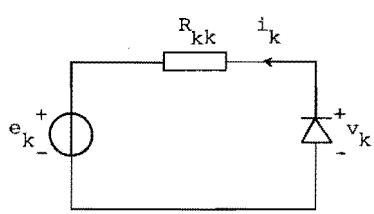


fig.3.2.

The internal port resistances are given by the diagonal elements of R and the controlled voltage sources e_k are defined by

$$e_k = \sum_{\substack{j=1 \\ j\neq k}}^{n} R_{kj} i_j + a_k \tag{3.3}$$

Next, the ports of M are loaded with ideal diodes, yielding for each port an equivalent circuit of the type given in fig.3.2. Now, for a single port the conducting state of the diode is unique for any fixed value of e_k if and only if $R_{kk} > 0$, because $R_{kk} \leq 0$ would lead to a situation already depicted in fig.2.5 and $R_{kk} > 0$, according to fig.3.3, yields a unique solution for any e_k.

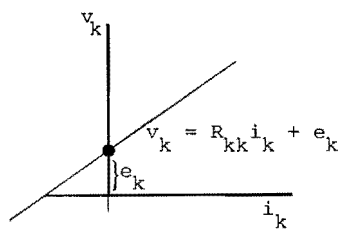


fig.3.3.

However, since e_k represents a controlled voltage source, its value cannot be chosen independently but is determined by the conducting state of the diodes at all other ports. Then a negative value of R_{kk} not necessarily yields multiple solutions since an alternate solution at port k may be prohibited by the conducting states of the diodes at the remaining ports. However for any R which is lower triangular, a necessary and sufficient condition for the existence of a unique solution for any vector a, is given by $R_{kk} > 0$ for all k, which can be proved as follows.
The first equation of (3.2) will give

$$\begin{aligned} &v_1 = R_{11} i_1 + a_1 \\ &v_1 \geq 0, \quad i_1 \geq 0 \qquad v_1 i_1 = 0, \end{aligned} \tag{3.4}$$

independent of the state of other diodes. The solution of (3.4) exists and is unique for any vector a if and only if $R_{11} > 0$.

Assume that a unique solution for the first k-1 equations exists for any vector a, independent of the state of the remaining diodes. The k-th equation now reads

$$\begin{aligned} v_k &= \sum_{j<k} R_{kj} i_j + R_{kk} i_k + a_k \\ v_k &\geq 0, \quad i_k \geq 0 \qquad v_k i_k = 0 \end{aligned} \tag{3.5}$$

Due to the assumption, the value of $\sum_{j<k} R_{kj} i_j + a_k$ is fixed and does not depend on v_m and i_m for $m \geq k$. Hence a unique solution of (3.5), and thus of the first k equations, exists for any vector a if and only if $R_{kk} > 0$. Then by induction the solution of (3.2) satisfying $v \geq 0$, $i \geq 0$ and $v^t i = 0$ exists and is unique for any vector a if and only if $\forall_k [R_{kk} > 0]$.

Of course, the above statement also holds true for any hybrid matrix which can be reordered into lower triangular form.

For non-triangular matrices the situation is more complicated. Instead of considering the value of e_k as a function of the voltages at all other ports, it is more suitable to consider the Thévenin equivalent of port k for any allowed open- or short-circuit termination at the other ports. Then for each hybrid representation H^j of M we have an equivalent circuit at port k given by one of both alternatives of fig.3.4, depending on the particular port-type assignment.

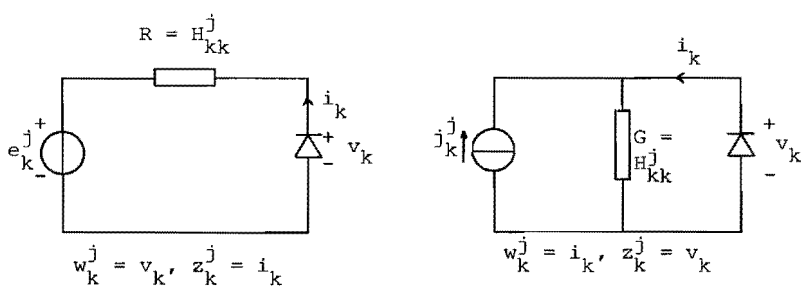


fig.3.4.

The sources in fig.3.4 are now fixed and on account of the previous arguments one might expect a unique solution to exist if the diagonal elements of all hybrid representations H^j are positive. In other words, for any port k, the output immittance should be positive for any open- or short-circuit termination at all other ports. We will call such a property the "zero load positive immittance property". In the next sections we will derive some theorems concerning the solution of (3.1) in terms of such properties.

3.1. Relations between hybrid representations

Consider again a partition of the constraint equations (3.1) induced by a particular admissible port-type assignment, with $z_k = i_k$, $w_k = v_k$ for a current port and $z_k = v_k$, $w_k = i_k$ for a voltage port, yielding

$$Nw + Mz = b, \quad \det(N) \neq 0$$

Hence

$$w = Hz + a \tag{3.6}$$

with H being a hybrid matrix of M. We are now concerned about the existence of other hybrid representations equivalent to (3.6). We are only interested in H and therefore take a=0. Furthermore we reorder the variables in (3.6) and then partition the vectors w and z in two parts,resulting in

$$\begin{pmatrix} w^1 \\ \hline w^2 \end{pmatrix} = \left(\begin{array}{c|c} H_{11} & H_{12} \\ \hline H_{21} & H_{22} \end{array}\right) \begin{pmatrix} z^1 \\ \hline z^2 \end{pmatrix} \tag{3.7}$$

Let H_{22} be nonsingular. Then the second equation set of (3.7) yields

$$z^2 = -H_{22}^{-1} H_{21} z^1 + H_{22}^{-1} w^2 \tag{3.8}$$

Substitution of (3.8) into the first equation set of (3.7) then

results in

$$w^1 = (H_{11} - H_{12}\, H_{22}^{-1}\, H_{21})z^1 + H_{11}\, H_{22}^{-1}\, w^2 \tag{3.9}$$

Then (3.8) and (3.9) can be combined into

$$\begin{pmatrix} w^1 \\ \hline z^2 \end{pmatrix} = \left(\begin{array}{c|c} H'_{11} & H'_{12} \\ \hline H'_{21} & H'_{22} \end{array}\right) \begin{pmatrix} z^1 \\ \hline w^2 \end{pmatrix} \quad \text{, with} \tag{3.10}$$

$$\begin{aligned} &H'_{11} = H_{11} - H_{12}\, H_{22}^{-1}\, H_{21}, \quad H'_{12} = H_{11}\, H_{22}^{-1} \\ &H'_{21} = -H_{22}^{-1}\, H_{22} \quad \text{and} \quad H'_{22} = H_{22}^{-1} \end{aligned} \tag{3.11}$$

Equation (3.11) will be referred to as a "principal transformation". Since $w' = \begin{pmatrix} w^1 \\ \hline z^2 \end{pmatrix}$, $z' = \begin{pmatrix} z^1 \\ \hline w^2 \end{pmatrix}$ is again a complementary partition satisfying (1.5), equation (3.10) defines a new hybrid description with hybrid matrix H'. Hence each nonsingular principal submatrix of H yields a new hybrid description. The converse statement is also true.

Lemma 1: Suppose both representations (3.7) and (3.10) to exist. Then H_{22} and H'_{22} are both nonsingular.

Proof: Subsitute $z^1 = 0$ in (3.7) and (3.10), which yields $w^2 = H_{22}\, z^2$ and $z^2 = H'_{22}\, w^2$, leading to $z^2 = H'_{22}\, H_{22}\, z^2$. Since (3.7) and (3.10) describe the same network, the last relation must hold for any z^2, i.e. we must have $H'_{22} \cdot H_{22} = I$. Therefore both H'_{22} and H_{22} are nonsingular and $H'_{22} = H_{22}^{-1}$.

As a consequence, given a hybrid matrix H of a linear multiport network M, the set of existing hybrid matrices is fully characterized by the set of nonsingular principal minors of H. After a principal transformation, any nonsingular principal minor yields an alternate hybrid matrix and a corresponding admissible distribution of open- and short-circuit terminations of the network ports. Due to lemma 1 a singular principal minor also means that the corresponding hybrid matrix does not exist.

In the sequel we will adopt the following notation to circumvent the reordering of the variables in (3.6).

With the set $I = \{m_1, m_2, \ldots, m_r\} \subset \{1,2,\ldots,n\}$ we define H_I as the principal submatrix of H with its elements defined by $(H_I)_{p,q} = H_{m_p,m_q}$. In other words, the elements of H_I are exclusively taken from the rows and columns $m_1, m_2, \ldots, m_r$ of H. Furthermore, $\det(H_I)$ denotes the determinant of H_I. If $I = \phi$, we define $\det(H_\phi) = 1$. Let H be some hybrid matrix of the linear multiport M. For all H_I with $\det(H_I) \neq 0$ we denote the corresponding hybrid representation of M obtained after a principal transformation with H_I by $H|I$. Then the following lemma holds.

Lemma 2: The diagonal elements h'_{ii} of $H|I$ are given by

$$h'_{ii} = \begin{cases} \dfrac{\det(H_{I\setminus\{i\}})}{\det(H_I)} & \text{for } i \in I \\[2ex] \dfrac{\det(H_{I\cup\{i\}})}{\det(H_I)} & \text{for } i \notin I \end{cases} \tag{3.12}$$

Proof: Consider the equation $H_I x = y$, with $y = 1_k^t = (0,0,\ldots 0,1,0,\ldots,0)^t$. Since H_I is nonsingular we have $x = H_I^{-1} y$ and hence by (3.11) $x_k = (H_I^{-1})_{kk} = h'_{i_k i_k}$. According to Cramer's rule, we also have

$$x_k = \frac{\det(H_{I\setminus\{i_k\}})}{\det(H_I)}, \text{ which proves (3.12) for } i \in I.$$

In particular, a principal transformation on a single diagonal element h_{ii} will yield

$$h'_{ii} = \frac{\det(H_\phi)}{\det(h_{ii})} = \frac{1}{h_{ii}} \tag{3.13}$$

Then the value h'_{jj} for $j \notin I$ can be obtained by a principal transformation on the indices $I \cup \{j\}$ leading to $\tilde{H}$, followed by a transformation on $\tilde{h}_{jj}$ yielding $h'_{jj} = \dfrac{1}{\tilde{h}_{jj}}$. Assume that $H_{I\cup\{j\}}$ is nonsingular. From the first part of the proof we then have

$$\tilde{h}_{jj} = \frac{\det(H_{I\cup\{j\}\setminus\{j\}})}{\det(H_{I\cup\{j\}})} = \frac{\det(H_I)}{\det(H_{I\cup\{j\}})}\text{, yielding}$$

$$h'_{jj} = \frac{1}{\tilde{h}_{jj}} = \frac{\det(H_{I\cup\{j\}})}{\det(H_I)}\ , \quad j \notin I \tag{3.14}$$

When $\det(H_{I\cup\{j\}}) = 0$ then h'_{jj} has to be zero, otherwise it would be possible to transform $H|I$ into a new $\tilde{H}$ with the portvariables with index j also interchanged. By lemma 1 the existence of $\tilde{H}$ would imply that $H_{I\cup\{j\}}$ is nonsingular, which is a contradiction. Thus (3.14) also holds for $\det(H_{I\cup\{j\}}) = 0$. End of proof.

A principal transformation using a single diagonal element will be called a one-step principal transformation. Using h_{ii} to this purpose, we find from lemma 2 that

$$h'_{ii} = \frac{1}{h_{ii}}\ , \quad h'_{jj} = \frac{\det(H_{\{i,j\}})}{h_{ii}}\ , \quad j \neq i \tag{3.15}$$

From (3.15) it can be seen that such a transformation is identical with a single Gauss-Jordan elimination step. One such step can be seen as changing the type of port from current port to voltage port or vice versa.

3.2. The hybrid structure of M

Given two existing hybrid matrices H and H', it will be useful to investigate whether or not H' can be derived from H by a sequence of one-step principal transformations. In fact the question is raised whether $H|I$, obtained by partial inversion of H on the matrix H_I, can be found by a sequence of Gaussian elimination steps on the diagonal pivots of H_I. To this purpose we define a partial ordering on the hybrid matrices of the network M. By lemma 1 this amounts to defining a partial ordering on all nonsingular principal submatrices of some hybrid matrix H. Let the deter-

minants of these submatrices,considered as functions of their coefficients,be the elements of a set V and let the relation $\prec$ be defined on V in the following way: $\det(H_I) \prec \det(H_J)$ if $I \subset J$. Then the relation $\prec$ satisfies

a) $\det(H_I) \prec \det(H_I)$

b) $\det(H_I) \prec \det(H_J)$ and $\det(H_J) \prec \det(H_K) \rightarrow \det(H_I) \prec \det(H_K)$

c) $\det(H_I) \prec \det(H_J)$ and $\det(H_J) \prec \det(H_I) \rightarrow \det(H_I) = \det(H_J)$

Hence $\prec$ indeed defines a partial ordering on the determinants of all principal minors of H, which can be visualized in a graph called the Hasse diagram. In this diagram, there is an edge between vertices corresponding to $\det(H_I)$ and $\det(H_J)$ iff

$\det(H_I) \prec x \prec \det(H_J) \rightarrow (x = \det(H_I)) \vee (x = \det(H_J))$

The principal minors corresponding to a pair of vertices connected by a single edge will be called adjacent minors. As an example fig.3.5 depicts the Hasse diagram for a 2×2 matrix H.

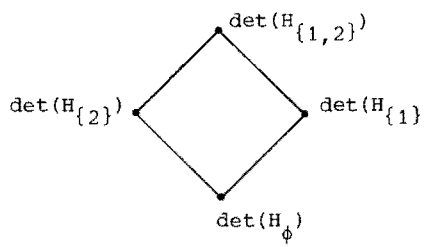


fig.3.5.

Now, by lemma 1, for each vertex with $\det(H_I) \neq 0$ there exists a corresponding hybrid representation $H|I$ of M. If in addition an adjacent minor H_J exists with $\det(H_J) \neq 0$, then the corresponding hybrid representation $H|J$ can be derived from $H|I$ by a one-step principal transformation on the diagonal element with index $k = (I \cup J) \setminus (I \cap J)$ of $H|I$. By lemma 2, this diagonal element then has a value $\theta(I,J)$ given by

$$\theta(I,J) = (H|I)_{kk} = \frac{\det(H_J)}{\det(H_I)} \tag{3.16}$$

with $k = (I \cup J) \setminus (I \cap J)$

On the other hand $H|I$ can also be obtained from $H|J$ by pivoting on

$$(H|J)_{kk} = \theta(J,I) = \frac{\det(H_I)}{\det(H_J)} = \frac{1}{\theta(I,J)}$$

We can use this property to assign an arbitrary direction and a corresponding real number θ to each edge in the Hasse diagram for which the determinants at the corresponding vertices are nonzero,as indicated in fig.3.6.

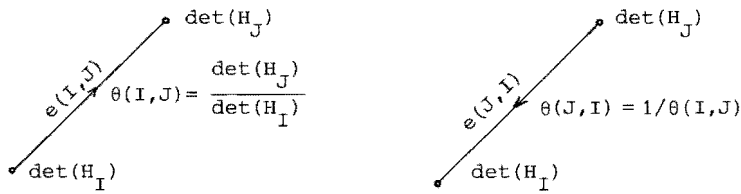


fig.3.6.

For those vertices with $\det(H_I) = 0$ we know that the hybrid matrix $H|I$ does not exist. From now on, we delete these vertices and the incident edges .

Whenever appropriate the above modified Hasse diagram will be called *the hybrid structure of H or the hybrid structure of the network M*. Note that the hybrid structure of M is not unique, however all hybrid structures are isomorphic. For a proof see appendix I ,where also the following more general relation is derived.

$$\det(H|I)_J = \frac{\det(H_{I-J})}{\det(H_I)} \tag{3.17}$$

As an example, the hybrid structure for $H = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ is given in fig.3.7.

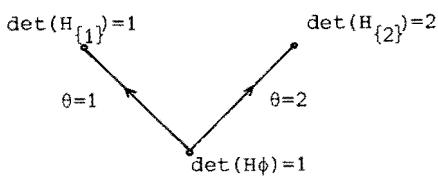


fig.3.7.

From the previous considerations we have found that the weights θ on the edges of the structure of H are given by all nonzero diagonal elements h^j_{ii} of all existing hybrid representations H^j of M. Due to its construction, the structure of H may be non-connected, that is in general its structure is a collection of disjunct but connected substructures.

Let us denote a directed edge between two adjacent minors H_I and H_J by e(I,J) (see fig.3.6). Furthermore, the number of incident edges at any vertex H_I in a subgraph of the hybrid structure is denoted by d(I). Within each connected substructure S we define a directed path P(I,J) between two vertices $\det(H_I)$ and $\det(H_J)$ as a subgraph of S satisfying

a) $P(I,I) = \phi$

b) $P(I,K) = P(I,L) \cup e(L,K)$. $\quad H_I, H_L, H_K \in S$

c) $P(I,K) \rightarrow d(I) = d(K) = 1$.

Then we have the following theorems.

Theorem 1: The nonzero diagonal elements of any hybrid representation H^j of M are positive if and only if, within each connected substructure of H, the determinants of all nonsingular principal submatrices of a particular hybrid representation H of M have the same sign.

Proof: Consider any two vertices $\det(H_I)$, $\det(H_J)$ in a connected substructure S. Then these vertices can be connected by a directed path P(I,J). Due to (3.16) the product of all weights $\theta(K,L)$ on the edges of the directed path P(I,J) equals $\dfrac{\det(H_J)}{\det(H_I)}$. Obviously if all $\theta(K,L) > 0$ then

$\mathrm{sign}(\det(H_I)) = \mathrm{sign}(\det(H_J))$.

On the other hand if $\forall_{K,L \in S}\ [\mathrm{sign}(\det H_K)) = \mathrm{sign}(\det(H_L))]$,

then each $\theta(K,L) = \dfrac{\det(H_L)}{\det(H_K)} > 0$.

The theorem then follows from the fact that, due to lemma 2, the set of all $\theta(K,L)$ is identical with the set of all nonzero diagonal elements of all existing hybrid representations of M.

Corollary: The product of all weights on the edges of a directed path between two vertices is the same for each directed path between those vertices.

Theorem 2: All 2^n hybrid descriptions H^J of M exist and have positive diagonal elements if and only if the determinants of all principal minors of some H^k are all positive.

Proof: Under the conditions of the theorem, the structure of M is connected and contains $\det(H^k_{\{i\}}) = h^k_{ii}$. Then the theorem follows from theorem 1.

Corollary: If the determinants of all principal minors of some H are positive, then the determinants of all principal minors of any other hybrid representation are also positive.

Due to the above theorms we have the important result that two hybrid representations $H' = H|I$ and $H'' = H|J$ can be obtained from each other by a sequence of one-step principal transformations if and only if there exists a path between the vertices corresponding to $\det(H_I)$ and $\det(H_J)$ in a hybrid structure of M.

3.3. The matrix classes P and P_0

As a generalization of the concept of positive definite and positive semi-definite matrices, Fiedler and Pták introduced two matrix classes called class-P and class-P_0, with the property that for a square matrix H

$$H \in P \quad \leftrightarrow \forall_{x\neq 0} \ \exists_k \ x_k \cdot (Hx)_k > 0 \tag{3.18}$$

$$H \in P_0 \quad \leftrightarrow \forall_{x\neq 0} \ \exists_k \ x_k \cdot (Hx)_k \geq 0 \wedge x_k \neq 0 \tag{3.19}$$

From these definitions it is obvious that $P \subset P_0$.
In [8] they derived a number of equivalent properties for a matrix to belong to class-P or class-P_0. The most important of them are listed below. Let D be a diagonal matrix. We denote $\forall_i d_{ii} > 0$ by $D > 0$ and analogously $\forall_i [d_{ii} \geq 0 \wedge D \neq 0]$ by $D \geq 0$.

a) $H \in P \leftrightarrow$ the determinant of every principal submatrix of H is positive.

b) $H \in P \leftrightarrow$ every real eigenvalue of H is positive.

c) $H \in P_0 \leftrightarrow$ the determinant of every principal submatrix of H is nonnegative. (3.20)

d) $H \in P_0 \leftrightarrow$ every real eigenvalue of H is nonnegative.

e) $H \in P_0 \leftrightarrow \det(H+D) \neq 0$ for any $D > 0$.

From the above properties it can be easily shown that

$$H \in P_0 \rightarrow \begin{cases} H + D \in P & \text{if } D > 0 \\ H + D \in P_0 & \text{if } D \geq 0 \end{cases} \tag{3.21}$$

Furthermore, if $H \in P(P_0)$ then any principal submatrix of H is in $P(P_0)$. The matrix classes P and P_0 play an important role in the solvability and uniqueness of solutions of network equations arising from electronic circuits with bipolar transistors modelled by their Ebers-Moll model, as well as in the linear complementarity problem. In particular we have the following theorem.

Theorem 3: The linear complementarity problem (3.1) has a unique solution for any vector a, if and only if $H \in P$.

For a proof see [6] and [10].

In this section we will derive some relations equivalent to (3.20) in terms of network properties. To this purpose we consider the matrix H to represent a (n×n) hybrid matrix of some sourceless multiport network M, described by

$$w = Hz, \tag{3.22}$$

with w and z some complementary partition of the port currents and voltages. Next we load the ports of M in the following way. For all ports except port k we connect an immittance d_i such that $w_i = -d_i z_i$ and connect a fixed source to port k such that $z_k = b_k$. See fig.3.8 for the case k = 1.

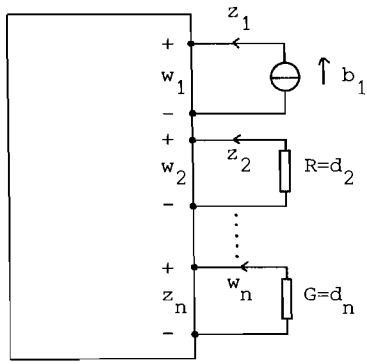


Fig.3.8.

Let the matrix D_k be defined by

$$D_k = \mathrm{diag}(d_1, d_2, \ldots, d_{k-1}, d_{k+1}, \ldots, d_n).$$

We say that D_k forms a positive load of M with respect to port k when $D_k > 0$. Next we define u_k as the immittance seen into port k under the previous load conditions, that is

$$u_k = \frac{w_k}{z_k} = \frac{w_k}{b_k} \tag{3.23}$$

With the above u_k we now define the following properties.
M has the "positive load nonnegative immittance property" ($M \in$ PLNNI) if and only if M is nondegenerate and for any existing hybrid matrix H and for any port k, every positive load with respect to port k yields a nonnegative immittance u_k. In other words

$$M \in \text{PLNNI} \leftrightarrow \forall_H \forall_k \forall_{D_k>0} [u_k \geq 0].$$

M has the "zero load positive immittance property" ($M \in$ ZLPI) if and only if M is nondegenerate and for some port k, every open- or short-circuit termination at all other ports is admissable and yields $u_k > 0$. With the above definitions we have the following theorems.

Theorem 4: Given M, then any existing hybrid matrix $H \in P_0$ if and only if $M \in$ PLNNI.

Proof: Necessity :

Assume $H \in P_0$ and arbitrarily take $k = 1$ and $D_k = D_1 > 0$. The network equations describing the loaded multiport network can now be written as

$$\begin{pmatrix} w_1 \\ \cdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = Hz + Dz \ , \tag{3.24}$$

$$\text{with } z_1 = b_1 \quad \text{and} \quad D = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & D_1 \end{array}\right)$$

Let us denote the matrix obtained from H after deleting the first row and first column by $H^{(1)}$. Then Cramer's rule applied to (3.24) gives

$$w_1 \det(H^{(1)} + D_1) = z_1 \det(H + D) \tag{3.25}$$

Since $H \in P_0 \rightarrow H^{(1)} \in P_0$, we have from (3.21) and $D_k > 0$, $D \geq 0$

$$H^{(1)} + D_1 \in P \quad \text{and } H + D \in P_0.$$

Hence $\det(H^{(1)} + D_1) > 0$ and $\det(H + D) \geq 0$. Then (3.25) yields

$$u_1 = \frac{w_1}{z_1} = \frac{w_1}{b_1} = \frac{\det(H + D)}{\det(H^{(1)} + D_1)} \geq 0.$$

Since k was arbitrarily set to one, we have

$$H \in P_0 \rightarrow \forall_k \forall_{D_{k>0}} [u_k \geq 0] \rightarrow M \in \text{PLNNI}$$

Sufficiency:

Let $H \notin P_0$. Then by (3.20e) there exists a diagonal matrix $\bar{D}$ such that

$$\det(H + \bar{D}) = 0, \ \bar{D} = \text{diag}(d_1, d_2, \ldots, d_n) > 0 \tag{3.26}$$

Next, for $i \neq 1$ load all ports i with immittances $d_i > 0$ and connect a source $z_1 = b_1$ to port 1. In the same way as for the necessary part we then find corresponding to (3.24)

$$\begin{pmatrix} w_1 \\ \cdot\cdot \\ 0 \end{pmatrix} = \left(\begin{array}{c:c} H_{11} & H_{12} \\ \hdashline H_{21} & H_{22} \end{array}\right) \begin{pmatrix} z_1 \\ \cdot\cdot \\ z_2 \end{pmatrix} + \left(\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & D_1 \end{array}\right) \begin{pmatrix} z_1 \\ \cdot\cdot \\ z_2 \end{pmatrix} \tag{3.27}$$

with $z_1 = b_1$ and $D_1 = \mathrm{diag}(d_2, d_3, \ldots, d_n) > 0$.

As in (3.25) Cramer's rule will yield

$$w_1 \det(H_{22} + D_1) = b_1 \det(H + \tilde{D}) \ , \tag{3.28}$$

with $\tilde{D} = \left(\begin{array}{c|c} 0 & 0 \\ \hline 0 & D_1 \end{array}\right)$.

Now, from (3.26) we find

$$\det(H + \bar{D}) = d_1 \det(H_{22} + D_1) + \det(H + \tilde{D}) = 0, \quad d_1 > 0 \tag{3.29}$$

Then from (3.29) either

a) both determinants in (3.29) are nonzero and have opposite signs;

b) both determinants are zero.

When case a) applies, we obtain from (3.28) that $u_1 = \frac{w_1}{b_1} < 0$.

In case b) we consider the value of h_{11}. If $h_{11} < 0$, we can find a sufficient high load D_1^* for the remaining ports such that the immittance seen port 1 is negative since $u_1 = h_{11} - H_{12}(H_{22} + D_1^*)^{-1} H_{21} < 0$ for D_1^* sufficiently large.

Thus we assume that $h_{11} \geq 0$ and have $h_{11} + d_1 > 0$. Next we load port 1 with the immittance d_1 and consider the (n-1)-port network formed by the other ports of M. By eliminating z_1 from the n-port equations it is easily seen that the new hybrid matrix describing the (n-1)-port becomes

$$w = \tilde{H} z$$

$$\tilde{H} = H_{22} - \frac{H_{21}H_{12}}{h_{11}+d_1}, \tag{3.30}$$

and $w = (w_2, w_3, \ldots, w_n)^t$, $z = (z_2, z_3, \ldots, z_n)^t$.

Now from (3.26) we have

$$\det(H + D) = (h_{11} + d_1)\cdot\det(H_{22} + D_1 - \frac{H_{21}H_{12}}{h_{11}+d_1}) = 0,$$

$$h_{11}+d_1 \neq 0.$$

Then with (3.30) we have

$$\det(\tilde{H} + D_1) = 0 \rightarrow \tilde{H} \notin P_0 \quad \text{since } D_1 > 0 \tag{3.31}$$

Then the above procedure can again be applied on the (n-1)-port network, ultimately yielding some $u_k < 0$ or ending in a one port network with hybrid matrix $h \notin P_0$, i.e. $h < 0$ in which case the immittance of this port is obviously negative. Hence we have

$H \notin P_0 \rightarrow \exists_k \exists_{D_k>0} [u_k < 0] \rightarrow M \notin$ PLNNI.

Theorem 5: Given M, then for any hybrid matrix H, $H \in P$ if and only if $M \in$ ZLPI.

Proof: Consider some port k of M. The values of u_k for all open- or short-circuit terminations at all other ports are now identical with the set of all diagonal elements h^j_{kk} of all hybrid matrices H^j of M with port k considered as voltage port and as current port. The theorem then follows immediately from theorem 2.

The conditions for theorem 4 can be relaxed in the following way. From (3.17) we observe that $\exists H \in P_0 \rightarrow$ all existing $H^j \in P_0$. Since the sufficiency proof for theorem 4 holds for every existing H^j the PLNNI property can be relaxed to

$M \in$ PLNNI $\leftrightarrow \exists_H \forall_k \forall_{D_k>0} [u_k \geq 0]$.

In [5] some additional theorems have been derived along a different line. From theorem 4 we conclude that a multiport network M with a hybrid matrix $H \in P_0$ can never become "active" at a certain

port when loaded by positive resistors in the sense that no "small signal gain" can be obtained at the input of the port where the excitation is connected to. Thus the network shows a certain kind of passivity which we will call P_0-passivity. Furtermore , a purely resistive network constructed from PL two-pole elements with a monotone increasing voltage-current characteristic always has a unique solution for any excitation. This is so because according to section 2.3, these elements can be constructed with fixed sources, positive resistors and ideal diodes only. In the model of fig.2.2, the network M is then a multiport network containing only positive resistors and fixed sources and thus its hybrid matrix H is positive definite (PD). Since PD $\subset$ P, we have H $\in$ P and by theorem 3 the network solution is unique.

4. The state-model of a piecewise-linear mapping

In chapter 2 and 3 we discussed some properties of PL networks formed by linear resistive n-port networks loaded by ideal diodes. It appeared that a hybrid description of the resistive n-port network M played an essential role in the formulation of the network equations and resulted in the LCP-problem (2.8) describing the conducting state of the ideal diodes. Once this state has been determined, the voltages and currents at the ports of M are known and then it is also possible to determine any other voltage or current in the inside of the network M.

Obviously these internal voltages and currents are also piecewise-linear functions of the excitation of the network M and the intention of this chapter is to use these variables to model piecewise-linear mappings. Without loss of generality we may consider the internal excitations and responses to be applied or measured on M at some additional network ports. Hence for a PL mapping $f : x \in \mathbb{R}^k \to y \in \mathbb{R}^m$, let us represent the variables x and y as voltages or currents at some ports of M. In particular let x represent a set of current sources applied to a set of ports #1, let y represent the open-circuit voltages at a set of ports #2 and assume that the ideal diodes are connected to a set of ports #3, as indicated in fig.4.1.

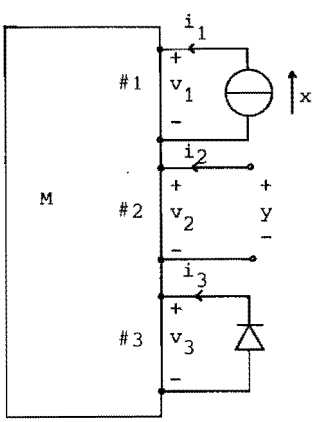


fig.4.1.

Furthermore, let us assume that the impedance matrix Z of M exists and is given by

$$\begin{pmatrix} v_1 \\ \hline v_2 \\ \hline v_3 \end{pmatrix} = \left(\begin{array}{c|c|c} z_{11} & z_{12} & z_{13} \\ \hline z_{21} & z_{22} & z_{23} \\ \hline z_{31} & z_{32} & z_{33} \end{array}\right) \begin{pmatrix} i_1 \\ \hline i_2 \\ \hline i_3 \end{pmatrix} + \begin{pmatrix} e_1 \\ \hline e_2 \\ \hline e_3 \end{pmatrix} \tag{4.1}$$

With the load conditions of fig.4.1 we then have

$$i_1 = x, \quad i_2 = 0, \quad v_2 = y \tag{4.2}$$

As we are not interested in v_1, the remaining equations of (4.1) and (4.2) yield

$$\begin{aligned} y &= z_{21}x + z_{23}i_3 + e_2 \\ v &= z_{31}x + z_{33}i_3 + e_3 \end{aligned} \tag{4.3}$$

Of course, some voltages of y may be measured at some ports of #1, in which case z_{21} in (4.3) would become some submatrix of $\left(\begin{array}{c|c} z_{11} & z_{12} \\ \hline z_{21} & z_{22} \end{array}\right)$

Let us rename the matrices and variables in (4.3) such that this equation system is rewritten as

$$\begin{aligned} y &= Ax + Bi + g \\ v &= Cx + Di + h \end{aligned} \tag{4.4}$$

Due to the ideal diodes connected to the set of ports #3, the network equations of interest finally become

$$y = Ax + Bi + g \tag{4.5a}$$

$$v = Cx + Di + h \tag{4.5b}$$

$$i \geq 0, \quad v \geq 0, \quad i^t v = 0, \tag{4.5c}$$

with $A = (m \times k)$, $B = (m \times n)$, $C = (n \times k)$ and $D = (n \times n)$.

In analogy with the state-space formulation of linear dynamic systems given by the equation set

$$y = Ax + Bu$$

$$\dot{u} = Cx + Du,$$

we call the system (4.5) *the state-model of the mapping* $f : x \in \mathbb{R}^k \to y \in \mathbb{R}^m$, and denote the system by $S(A,B,C,D,g,h)$ or shortly S.

Given x, the image vector y is obtained as follows. First for a given x, the pair v,i in the state-model is determined by solving the LCP problem

$$v = Cx + Di + h, \quad v \geq 0, \quad i \geq 0, \quad v^t i = 0 \tag{4.6}$$

This pair v,i then determines the conducting state of the connected diodes, i.e. the state of the system (4.5). Once this state is known, the vector y is obtained from (4.5a). Thus the LCP problem (4.6) plays the same role in the state-model as the state-equation $\dot{u} = Cx + Du$ in the description of linear dynamic systems. Therefore we will call the equation set (4.5b) and (4.5c) *the state-equations of the state-model* S. As an example of such a formulation consider the network of fig.4.2, with the nonlinear resistive elements R_1 and R_2 as given in fig.4.3.

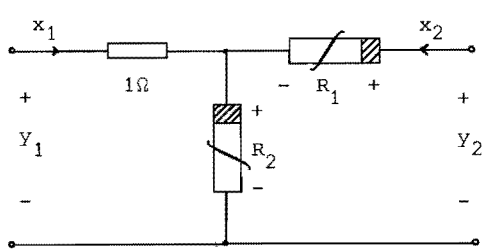


fig.4.2.

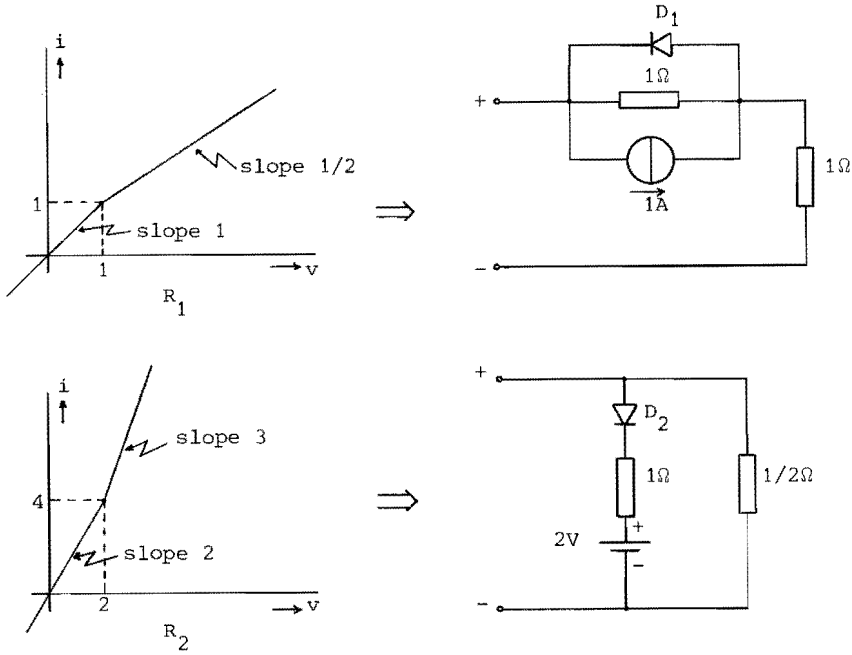


fig.4.3.

The network M is for this case given in fig.4.4 and yields the impedance matrix

$$v = \begin{pmatrix} 3/2 & 1/2 & 0 & -1/2 \\ 1/2 & 5/2 & 1 & -1/2 \\ 0 & 1 & 1 & 0 \\ -1/2 & -1/2 & 0 & 3/2 \end{pmatrix} i + \begin{pmatrix} 0 \\ -1 \\ -1 \\ 2 \end{pmatrix} \qquad (4.7)$$

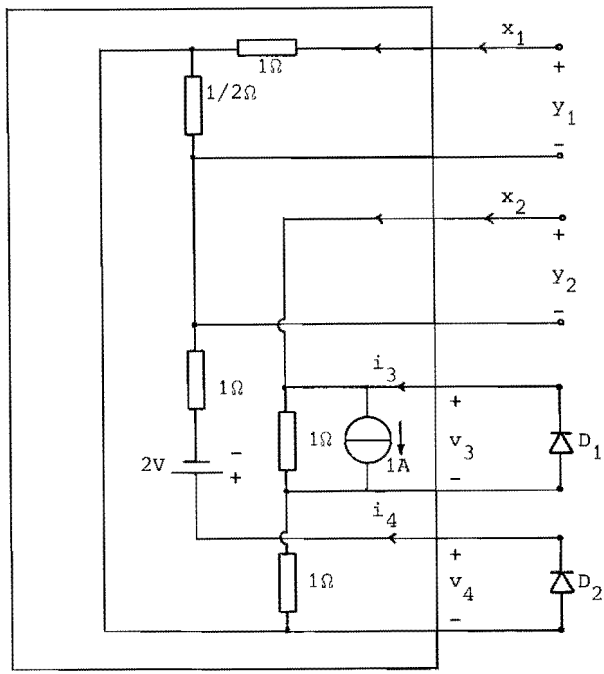


fig.4.4.

From (4.7) the state-model then becomes after renaming the vectors v_3, i_3 and v_4 , i_4 in v_1, i_1 and v_2, i_2

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 3/2 & 1/2 \\ 1/2 & 5/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 & -1/2 \\ 1 & -1/2 \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1/2 & -1/2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 3/2 \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} + \begin{pmatrix} -1 \\ 2 \end{pmatrix} \qquad (4.8)$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0.$$

For example, with $x_1 = 1$ and $x_2 = 1$, a solution is given by

$$v_1 = 0, \quad i_1 = 0, \quad v_2 = 1, \quad i_2 = 0; \quad y_1 = 2, \quad y_2 = 2.$$

Observe that (4.8) itself describes a PL two-port network by its impedance matrix. It can again be used as a PL element for constructing more complex PL networks.

Up to now, we did not show that the mapping f as modelled by (4.5) is indeed PL. This will be postponed until section 4.2. At this place it is however convenient to discuss some properties of (4.5) in network terms.

As in chapter 3, we observe that the $(n \times n)$ matrix D considered as a hybrid matrix may yield maximally 2^n different descriptions equivalent to (4.6). Of course for each description the term $Cx + h$ representing the combined effect of all sources at the ports #3 is in general different too. Let us assume that $Cx + h = a \geq 0$. Then a solution of (4.6) is obviously given by $v = a$, $i = 0$, yielding $y = Ax + g$ from (4.5a). Hence in a region $R_0 \subset \mathbb{R}^k$ with $R_0 = \{x \mid Cx + h \geq 0\}$ we have

$$x \in R_0 \rightarrow y = Ax + g.$$

Let us denote the other hybrid representations equivalent to (4.6) by

$$w^j = C^j x + h^j + D^j z^j.$$

Then for these hybrid descriptions we have a corresponding equation for y given by

$$y = A^j x + B^j z^j + g^j.$$

Then

$$x \in R^j = \{x \mid C^j x + h^j \geq 0\} \rightarrow y = A^j x + g^j,$$

since $w^j = C^j x + h^j$, $z^j = 0$ is a solution of the state-equations. Thus, we have maximally 2^n regions in $\mathbb{R}^k$ where (4.5) defines maximally 2^n affine mappings $f^j : x \rightarrow y$. Then, in principle the system (4.5) using n diodes is equivalent to a list of 2^n mappings in 2^n regions.

Should two systems S and $\tilde{S}$ each with n_1 respectively n_2 diodes be interconnected, i.e. some concatenation of two PL mappings be defined, then the overall system contains $n_1 + n_2$ diodes. This system may describe $2^{n_1+n_2} = 2^{n_1} \cdot 2^{n_2}$ affine mappings in as many

regions. Hence, the complexity of the state-model description grows linearly while a list of regions and corresponding mappings would grow exponentially. This is a very important advantage of the state-model description over all other descriptions that work with a list of all mappings.

As to the various regions, it is possible that these regions intersect, that is

$$x \in R_i \cap R_j \rightarrow y = A^i x + g^i \vee y = A^j x + g^j,$$

and the mapping is one to many. The particular value of y which will be found, will then in general depend on the history of x, i.e. on the particular way in which we arrived at x. Such a system has an inherent memory and systems of this type will be used later to model static electronic memory devices like flip-flops or Schmitt-triggers. It may also happen that two different diode states yield the same vector y, which means that some diode states may become unobservable. On the other hand, if $\bigcup_i R_i \neq \mathbb{R}^k$, there are some regions in $\mathbb{R}^k$ for which the mapping $x \rightarrow y$ is not defined. In network terms this means that no hybrid matrix exists for that particular excitation. Furthermore, if $\exists_i\ R_i = \phi$, this means that the corresponding diode state is uncontrollable by the input vector x, for example due to the fact that a particular part of the network becomes disconnected for that x or two or more diodes are in some way connected in series such that they all have to conduct at the same time.

4.1. The structure of the state-model

From now on, we consider the equation set (4.5) as a pure mathematical model describing a piecewise-linear mapping $f : x \rightarrow y$, without any specific electrical equivalent in mind for the related variables. However all other equivalent descriptions that originated from our previous network approach, will still be called *equivalent hybrid representations of the state-model* (4.5). These equivalent representations are based on the different partitions

of the vector pairs v,i in complementary pairs w,z. Due to (4.5c) it is obvious that there exist in total 2^n different classes of complementary pairs v,i if we only consider the distribution of the zero components over i and v and pay no attention to the specific numerical value of their nonzero components. We now say that the system (4.5) can be in one of 2^n different states, depending on the particular complementary partition upon which a certain hybrid description or representation is based.

To facilitate the notation we we will map these partitions on the set of 2^n boolean n-vectors s_p, $p = 0,1,\ldots,2^n-1$ of the boolean n-cube, i.e.

$$s_p = (s_{p_1}, s_{p_2}, \ldots, s_{p_n}),$$

with $s_{p_i} \in \{0,1\}$ and $p = 0,1,2,\ldots,2^n-1$.

The vectors s_p are called *state-vectors* and are numbered such that the index p equals the decimal equivalent of s_p considered as a binary number. For example for n = 3 we have $s_6 = (1,1,0)$.

The particular mapping is defined as follows

$$\begin{aligned} i_k > 0, \quad v_k = 0 &\rightarrow s_{p_k} = 1 \\ i_k = 0, \quad v_k \geq 0 &\rightarrow s_{p_k} = 0 \end{aligned} \tag{4.9}$$

For example for n = 4 and a complementary solution of (4.5b) and (4.5c) given by i = (2,0,1.5,0), v = (0,0.7,0,1.2) we have from (4.9) $s = (1,0,1,0) = s_{10}$. For each s_p there may exist a corresponding so called *natural equivalent hybrid representation* for (4.5), which describes the same PL mapping and is obtained from (4.5) by a principal transformation using D_I, with $I = \{m \mid s_{p_m} = 1\}$. This means that all v_m and i_m for which $s_{p_m} = 1$ have to be interchanged. In state s_p the natural equivalent hybrid representation of (4.5) is then defined by

$$y = A^p x + B^p z^p + g^p$$
$$w = C^p x + D^p z^p + h^p \tag{4.10}$$
$$w^p \geq 0, \quad z^p \geq 0 \quad \text{and} \quad w^{p^t} z^p = 0,$$

$$\text{with} \left.\begin{matrix} w_k^p = i_k \\ z_k^p = v_k \end{matrix}\right| \text{for } s_{p_k} = 1 \quad \text{and} \quad \left.\begin{matrix} w_k^p = v_k \\ z_k^p = i_k \end{matrix}\right| \text{for } s_{p_k} = 0 \tag{4.11}$$

The superscript p in (4.10) refers to state s_p and the representation (4.10) is denoted by S_p. From (4.11) we observe that (4.5) is the representation in state S_0. In the above way, each solution of (4.5b) and (4.5c) is coupled to a particular hybrid representation (4.10) by means of its corresponding state-vector.
We now investigate for which set of vectors x a particular class of vector-pairs v,i is produced as a solution of the state equations. To this purpose let us consider the class v,i which maps on state s_p. Due to (4.9) and (4.11) this class satisfies $z^p = 0$, $w^p \geq 0$ and hence from representation S_p this implies

$$C^p x + h^p \geq 0 \qquad C = (n \times k) \tag{4.12}$$

Hence, when x satisfies (4.12), the particular class v,i which maps on state s_p contains a solution of the state equations. The n inequalities in (4.12) represent n halfspaces bounded by k-dimensional hyperplanes in $\mathbb{R}^k$. The rows of C^p represent the normal vectors on these hyperplanes. Then (4.12) is satisfied in the intersection of these halfspaces which defines a convex polyhedron in $\mathbb{R}^k$. Let us denote these polyhedral regions by R_p, with

$$R_p = \{x \mid C^p x + h^p \geq 0\} \tag{4.13}$$

In a polyhedron R_p we have $w^p \geq 0$ and $z^p = 0$. Using the first equation of (4.10), this yields

$$x \in R_p \rightarrow y = A^p x + g^p \tag{4.14}$$

Therefore in each polyhedron R_p, the domain x is mapped on the range y by an affine mapping, yielding the following theorem.

Theorem 6: The state-model represents a piecewise-linear mapping $f : x \in \mathbb{R}^k \rightarrow y \in \mathbb{R}^m$ over a union of polytopes $\bigcup_i R_i \subset \mathbb{R}^k$.

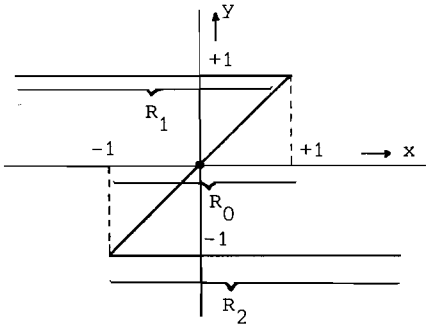
Of course, all S_p are equivalent, that is each representation (4.10) can be used to describe the PL mapping in all regions, which means that the state-model is a global model.

As to the regions (4.13), the same considerations as in the previous section apply. Hence there may be empty regions, unobservable or uncontrollable states or states s_p for which the representation S_p does not exist. In the latter case it means that a certain principal transformation on D is not possible. In terms of the definitions of chapter 3 we then have

$$S_p \text{ exists} \leftrightarrow \det(D_{I_p}) \neq 0 \tag{4.15}$$

with $I_p = \{m \mid s_{p_m} = 1\}$

We can use (4.15) to relate each existing S_p to nonsingular principal submatrices of D and visualize this relation by mapping these S_p or their corresponding regions R_p on the vertices of the hybrid structure of D as defined in chapter 3. The structure obtained in this way with the vertices $\det(D_I)$ now labelled by S_p or R_p according to (4.15), will be called *the structure of the state-model* S. A path between two vertices S_p and S_q then implies that representation S_q can be obtained from S_p by a sequence of one-step principal transformations using only diagonal elements of intermediate matrices D. As a simple example, the state-model of a one-dimensional mapping $f : x \in \mathbb{R} \rightarrow y \in [-1,+1]$ is given in fig.4.5.



$$\begin{array}{c|c|c} A & B & g \\ \hline C & D & h \end{array} \Rightarrow S_0 : \begin{array}{c|cc|c} 1 & -1 & 1 & 0 \\ \hline 1 & -1 & 1 & 1 \\ -1 & 1 & -1 & 1 \end{array}$$

$$S_1 : \begin{array}{c|cc|c} 0 & 0 & -1 & 1 \\ \hline 0 & 0 & -1 & 2 \\ -1 & 1 & -1 & 1 \end{array}$$

$$S_2 : \begin{array}{c|cc|c} 0 & 1 & 0 & -1 \\ \hline 1 & -1 & 1 & 1 \\ 0 & -1 & 0 & 2 \end{array}$$

fig.4.5.

The structure of the state-model of fig.4.5 is given in fig.4.6 below.

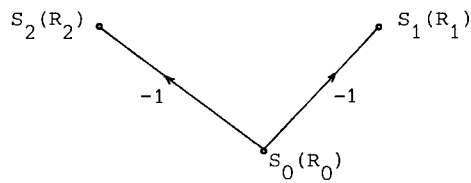


fig.4.6.

The various "polytopes" are obtained from the given S_i, yielding

$$R_0 = \{x \mid \quad x + 1 \geq 0, \quad -x + 1 \geq 0\}$$

$$R_1 = \{x \mid 0.x + 2 \geq 0, \quad -x + 1 \geq 0\}$$

$$R_2 = \{x \mid \quad x + 1 \geq 0, \; 0.x + 2 \geq 0\}$$

Observe that S_3 does not exist since $\det(D_{\{1,2\}}) = 0$. Furthermore, R_0 overlaps with R_1 and R_2. As we will see later, this is due to the negative weights on the edges of the structure of S.

4.2. Adjacent regions in a minimal state-model

In this section, we will discuss some properties of the state-model S for a piecewise-linear mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^m$, given by

$$\begin{aligned} y &= Ax + Bi + g \\ v &= Cx + Di + h \qquad\qquad A = (m\times m), \quad B = (m\times n) \\ v &\geq 0, \quad i \geq 0, \quad v^t i = 0 \qquad C = (n\times m), \quad D = (n\times n) \end{aligned} \tag{4.16}$$

To facilitate further notations let us define H^p and I_p in representation S_p by

$$H^p = \left(\begin{array}{c|c} A^p & B^p \\ \hline C^p & D^p \end{array}\right) \tag{4.17}$$

$$I_p = \{k \mid s_{p_k} = 1\} \tag{4.18}$$

By the definition of S_p, any other existing H^q can be derived from H^p by a principal transformation on D^p_M, with

$$M = I_p - I_q \triangleq (I_p \cup I_q)\backslash(I_p \cap I_q) \tag{4.19}$$

From (3.10) and (3.11) this will yield

$$A^q = A^p - B^p_M \, D^{p\,-1}_M \, C^p_M \, , \tag{4.20}$$

with B^p_M consisting of all columns of B^p with their indices in M considered as an ordered set and C^p_M consisting of all rows for C^p under the same conditions. For a particular example see fig.4.7.

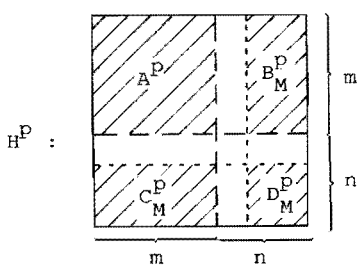


fig.4.7.

Next consider region R_p given by

$$R_p = \{x \mid C^p x + h^p \geq 0\} \tag{4.21}$$

Let the hyperplane $K^p_k = \{x \mid C^p_{k.} x + h^p_k = 0\}$, with $C^p_{k.}$ representing the k-th row of C^p, be a boundary of R_p. If for some $l \neq k$ $K^p_l = K^p_k$,

the boundary hyperplane K_k^p would be redundantly determined, yielding an in some sense unnecessarily complicated state-model. Therefore we define a *minimal state-model* as a state-model for which $\forall_p \forall_{l \neq k} [K_l^p \neq K_k^p]$ and in addition for each boundary hyperplane the corresponding diagonal element of D is nonzero. In the remaining of this section we will only consider such minimal models.
Then we have

$$\exists_{x \in R_p \cap K_k^p} [C_{k.}^p \bar{x} + h_k^p = 0 \wedge C_{l.}^p \bar{x} + h_l^p > 0, \quad l \neq k] \tag{4.22}$$

Let us now apply a one-step principal transformation using D_{kk}^p, which yields a new representation S_1 with $H = H^q$ and $k = I_p - I_q$.
In this representation we have by (3.10) and (3.11)

$$A^q = A^p - \frac{B_{.k}^p \, C_{k.}^p}{D_{kk}^p} \tag{4.23a}$$

$$C_{r.}^q = \begin{cases} C_{r.}^p - \dfrac{D_{rk}^p}{D_{kk}^p} C_{k.}^p, & r \neq k \\ -\dfrac{1}{D_{kk}^p} C_{k.}^p, & r = k \end{cases} \tag{4.23b}$$

$$h_r^q = \begin{cases} h_r^p - \dfrac{D_{rk}^p}{D_{kk}^p} h_k^p, & r \neq k \\ -\dfrac{h_k^p}{D_{kk}^p}, & r = k \end{cases} \tag{4.23c}$$

and $$g^q = g^p - \frac{h_k^p}{D_{kk}^p} B_{.k}^p \tag{4.23d}$$

From (4.23b) and (4.23c) this yields

$$C^q_{l\cdot}x + h^q_l = C^p_{l\cdot}x + h^p_l - \frac{D^p_{lk}}{D^p_{kk}}(C^p_{k\cdot}x + h^p_k) \quad , \; l \neq k \tag{4.24}$$

Next consider again the points $\bar{x}$ from (4.22). For these $\bar{x}$ we have by (4.24)

$$\begin{aligned} C^q_{l\cdot}\bar{x} + h^q_l &= C^p_{l\cdot}\bar{x} + h^p_l > 0 \,, \quad l \neq k \\ C^q_{k\cdot}\bar{x} + h^q_k &= C^p_{k\cdot}\bar{x} + h^p_k = 0 \end{aligned} \tag{4.25}$$

Hence $K^p_k = \{x \mid C^p_{k\cdot}x + h^p_k = 0\}$ is also a boundary hyperplane of R_q. Then R_p and R_q are adjacent regions which have at least a subset of the hyperplane $K^p_k = K^q_k$ in common (see fig.4.8).

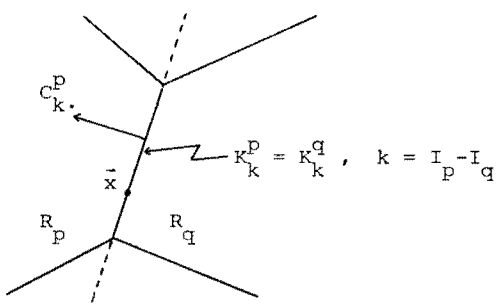


fig.4.8.

As to the affine mappings in both regions we have from (4.23a) and (4.23d)

$$A^q x + g^q = A^p x + g^p - \frac{1}{D^p_{kk}} B^p_{\cdot k}(C^p_{k\cdot}x + h^p_k) \tag{4.26}$$

Then for any $\bar{x}$ on the boundary between R_p and R_q, we have from (4.22) and (4.26)

$$A^q\bar{x} + g^q = A^p\bar{x} + g^p \tag{4.27}$$

Hence a PL mapping f modelled by a minimal state-model is continuous since it is continuous within each region as well as on the boundaries between the regions. Furthermore the representation in adjacent regions can be found by a one-step principal transformation using the diagonal element of D with the same index as the common boundary hyperplane which separates them.

In network terms we could think of each boundary hyperplane as being coupled to a single ideal diode which has to reverse its state when the exitation traverses from one region to an adjacent region. In a nonminimal model at least one other ideal diode could change its state at the same time, which in general implies that the mapping f is not continuous.

The minimal state-models will play an important role in describing homeomorphisms as we will see in chapter 7.

5. Complementary pivoting methods

In the previous sections the state-model was presented as a global model for a PL mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^k$. The calculation of the image vector y for a given vector x by this model is reduced to the determination of a solution of the linear complementarity problem given by the state equations

$$\begin{aligned} &v = Di + Cx + h \\ &v \geq 0, \quad i \geq 0, \quad v^t i = 0 \qquad D = (n\times n), \quad C = (n\times m) \end{aligned} \tag{5.1}$$

The existance of efficient solution algorithms for the LCP problem is then of vital importance for a practical application of the state-model. Fortunately various such solution algorithms exist and there is a vast amount of literature on this topic (see e.g. [11] and [12]).

The algorithms to solve the LCP problem can in general be divided into three different classes, depending on the particular equivalent formulation of the LCP problem.

The first class of algorithms are so called homotopy algorithms. In these algorithms, the problem (5.1) is embedded in a one-parameter family of problems $H(v,i,\lambda)$ defined on $\mathbb{R}^n \times [0,1]$, in such a way that the solution $H(v,i,0)$ is known and the solution of $H(v,i,1)$ is a solution of (5.1). The algorithms in this class generate a family of solutions $v(\lambda), i(\lambda)$ along some path in $\mathbb{R}^n \times [0,1]$.

In the most familiar homotopy methods, the solution path is generated by a sequence of one-step principal transformations (Gauss-Jordan pivoting steps), in each step extending the already obtained path by some segment, meanwhile keeping the solution complementary or almost complementary. Such methods will be called complementary pivoting methods and will be the subject of this chapter. In particular the Katzenelson algorithm, the Cottle algorithm and the Lemke algorithm will be discussed.

A second class of algorithms is formed by variational algorithms which minimize some constrained cost-function in $\mathbb{R}^n$ which is formulated in such a way that its solution is equivalent to a solution of (5.1). As an example, Cryer [13] reformulated (5.1)

for symmetric D into

$$\text{minimize} \quad f(i) = \frac{1}{2}\, i^t D i + a^t i \ ,$$

$$\text{subject to} \quad i \geq 0$$

In this way a lot of constraint minimization-algorithms can be used to solve the LCP problem. However in most situations these algorithms only work properly when f(i) is convex, which limits the applicability of these methods. Therefore we will not discuss these methods any further.

The third class of methods are the iterative methods. In this case, the LCP problem is transformed into an equivalent set of nonlinear equations in $\mathbb{R}^n$. By using adequate fixed-point algorithms, the LCP problem can be solved iteratively. A special representative of this class is the modulus algorithm which will be treated in chapter 6. Other algorithms of this type can be found in the simplicial fixed-point methods [14].

5.1. The Katzenelson algorithm

Katzenelson [4] originally devised his algorithm for solving a set of PL equations f(x) = y defined by

$$x \in R_i \rightarrow f(x) = f^i = A^i x + b^i, \quad A^i = (n \times n), \quad i = 0,1,\ldots,k, \tag{5.2}$$

with R_i representing convex polyhedral regions in $\mathbb{R}^n$.
The particular set of PL equations evolved from a description of a PL resistive network containing uncoupled resistors with strictly monotonously increasing v-i characteristics. For these equations it could easily be shown that

$$\forall_{i,j} \ \operatorname{sign}(\det A^i) = \operatorname{sign}(\det A^j) \tag{5.3}$$

From our discussions in section 3.3 we already know that in such a case always a unique solution exists.

In [15] and [16] Fujisawa and Kuh respectively Kuh and Chien extended this algorithm for less restricted conditions than (5.3).

In this section we will describe a version of the algorithm which can be used to solve the LCP problem (5.1) for matrices C and D such that the pivots that are taken in the algorithm are always positive. A sufficient condition to this purpose is given by $D \in P$.

Let us assume that we are interested in the solution of (5.1) for $x = x^*$. We start from a point x_0 for which the solution of (5.1) is known. Let this point be such that $Cx_o + h > 0$, yielding $v = Cx_0 + h$, $i = 0$. Next we construct a path $x(\lambda) = x_0 + \lambda(x^*-x_0)$, $0 \le \lambda \le 1$ in $\mathbb{R}^n$ and try to find the solution of (5.1) for any λ within the given interval. From (5.1) and the given $x(\lambda)$ we then have

$$\begin{aligned} &v(\lambda) = Di(\lambda) + Cx_0 + h + \lambda C(x^*-x_0) \ , \\ &Cx_0 + h > 0, \quad i \ge 0, \quad v \ge 0, \quad i^t v = 0 \end{aligned} \tag{5.4}$$

If $Cx_0 + h + C(x^*-x_0) \ge 0$, we take $\lambda = 1$ and immediately find a solution

$$v(1) = Cx^* + h, \quad i(1) = 0$$

Hence we assume that some components of $C(x^*-x_0)$ are negative such that

$$\exists\ !_s\ [\lambda_s = \min_k \lambda_k \,|\, (C_{k.}x_0 + h_k + \lambda_k C_{k.}(x^*-x_0) = 0) \wedge 0 < \lambda_s < 1] \tag{5.5}$$

The situation that the index s in (5.5) is nonunique will be discussed later. Then from (5.5) we have

$$\begin{aligned} &C_{k.}x(\lambda_s) + h_k > 0, \quad k \ne s \\ &C_{s.}x(\lambda_s) + h_s = 0, \quad C_{s.}(x^*- x_0) < 0 \end{aligned} \tag{5.6}$$

Hence $x = x(\lambda_s)$ is a point on the boundary hyperplane K_s^0 of R_0 given by

$$K_s^0 = \{x \mid C_{s\cdot}x + h_s = 0\}$$

Increasing λ above the value λ_s means that a new region R_p ($p = 2^{n-s}$) is entered for which the natural hybrid representation can be found from (5.4) by pivoting on the diagonal element D_{ss}. The representation in question then yields

$$w^p = D^p z^p + C^p x(\lambda) + h^p, \tag{5.7}$$

with $p = 2^{n-s}$, $w^p \geq 0$, $z^p \geq 0$ and $w^{p^t} z^p = 0$

From the relation (4.23) we easily find

$$C_{k\cdot}^p x(\lambda) + h_k^p = C_{k\cdot}x(\lambda_s) + h_k + (\lambda-\lambda_s)(C_{k\cdot} - \frac{D_{ks}}{D_{ss}} C_{s\cdot})(x^*-x_0) \tag{5.8a}$$

for $k \neq s$

and

$$C_{s\cdot}^p x(\lambda) + h_s^p = -(\lambda-\lambda_s) \frac{C_{s\cdot}(x^*- x_0)}{D_{ss}} \tag{5.8b}$$

As stated before, we assume $\forall_{p,k}[D_{kk}^p > 0]$. Then by (5.6) and (5.8) we have

$$C^p x(\lambda) + h^p > 0 \quad \text{for } \lambda > \lambda_s \quad \text{and } (\lambda-\lambda_s) << 1.$$

Hence we may continue to increase λ until either some boundary hyperplane K_t^p of R_p is reached for $\lambda = \lambda_t$, indicated by

$$\lambda_t = \min_k (\lambda_k \mid C_{k\cdot}^p x_0 + h_k^p + \lambda_k C_{k\cdot}^p (x^*-x_0) = 0 \wedge \lambda_k > \lambda_s) ,$$

or λ_t does not exist and λ can be increased indefinitely.

In the last case as well as for $\lambda_t > 1$, we take $\lambda = 1$ and from (5.7) have found a solution $w^p = C^p x(1) + h^p$, $z^p = 0$. If $\lambda_t < 1$ we repeat the previous process by pivoting on D_{tt}^p. After each pivoting, we enter a new region in which the value of λ can again be increased. This guarantees that the algorithm never re-enters the same region (see the corresponding proof in [16]). Since the

number of regions is finite, sooner or later a region R_q is entered for which $x(1) = x^*$ is an internal point, yielding $w^q = C^q x^* + h^q$, $z^q = 0$ as a solution of (5.1).

Should the index s in (5.5) be nonunique, then at least two boundary hyperplanes are crossed at the same time, in which case the path $x(\lambda)$ hits some corner of a region. In that case the initial point x_0 can either be perturbated such that this situation no longer happens or other actions equivalent to those described in [16] can be taken, which will not be repeated here.

5.2. The algorithm of Cottle

In [17] R.W. Cottle described a complementary pivoting algorithm for the solution of (5.1) when $D \in P$. In our terminology of hybrid matrices and ideal diodes, this algorithm can be seen as a method to determine the solution of a multiport network loaded with ideal diodes by gradually connecting these diodes one by one to the multiport network. The algorithm to solve (5.1) a given $x = x^*$ goes as follows.
Without loss of generality we assume that

$$\begin{aligned} &v = Di + a, \quad a = Cx^* + h \\ \text{with } &a_j \geq 0 \quad \text{for } j \leq k \\ &a_j < 0 \quad \text{for } k < j \leq n \end{aligned} \tag{5.9}$$

From the above representation with k nonnegative components in the vector a, we try to find a new representation by a sequence of one-step principal transformations (simplex pivoting steps) such that k + 1 components of a are nonnegative. When we are able to do so, the solution is obtained by repeated application of the same step until no negative components are left in the vector a.

It would be convenient if an exchange between v_{k+1} and i_{k+1} could do the job. Taking $D_{k+1,k+1}$ as a pivot in any case guarantees that that the new a_{k+1} becomes positive since, after the pivoting has

taken place, its new value becomes $-a_{k+1}/D_{k+1,k+1}$ and $D_{k+1,k+1} > 0$ because $D \in P$. However, performing that pivoting step may produce negative values among the first k components of the new vector a, such that nothing is gained by such a pivoting step. In order to overcome this problem let us view upon the one-step principal transformation by the pivot $D_{k+1,k+1}$ as a continuous proces instead of an immediate exchange between v_{k+1} and i_{k+1}. To this purpose we allow the variable i_{k+1} to grow continuously from zero to positive values, keeping all other components of i equal to zero. Furthermore, for the time being we focus our attention to the first k + 1 equations of (5.9) only. With $i_j = 0$ for $j \le k$ we may write these equations as

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \\ \cdots \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \\ \cdots \\ a_{k+1} \end{pmatrix} + i_{k+1} \begin{pmatrix} D_{1,k+1} \\ D_{2,k+1} \\ \vdots \\ D_{k,k+1} \\ \cdots\cdots \\ D_{k+1,k+1} \end{pmatrix} \tag{5.10}$$

In (5.10) every component v_j is given as a linear function of i_{k+1}. These functions $v_j(i_{k+1})$ can be plotted as depicted in fig.5.1 for some values of j. For $i_{k+1} = -a_{k+1}/D_{k+1,k+1} = d_{k+1}$ we have $v_{k+1}(d_{k+1}) = 0$. The corresponding values $v_j(d_{k+1})$ for $j \le k$ appear to become the components $\tilde{a}_j$ of the new vector $\tilde{a}$ which is obtained from (5.9) after exchanging v_{k+1} and i_{k+1}. In the case that $v_j(d_{k+1}) \ge 0$ for $j \le k$, this pivoting step immediately leads to the desired k + 1 nonnegative components of $\tilde{a}$. It may however happen that one or more components v_j become negative before v_{k+1} reaches zero.

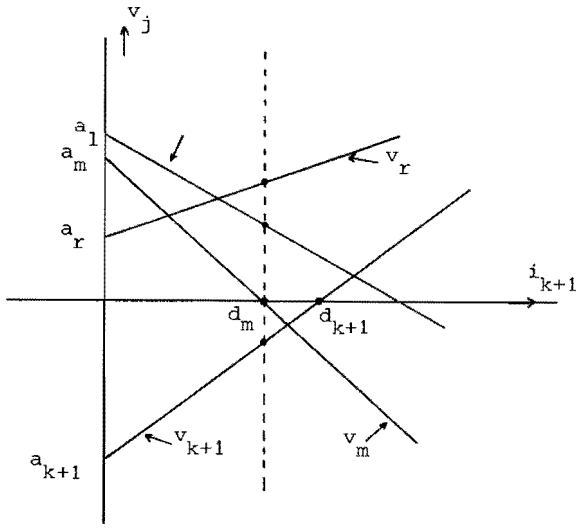


fig.5.1.

With $d_j = -a_j/D_{j,k+1}$

and $d_m = \min_j (d_j \mid d_j \geq 0)$, (5.11)

this happens when $d_m < d_{k+1}$. In this case we keep $i_{k+1} = d_m$ and take $D_{m,m}$ as a pivot, exchanging i_m and v_m. The new values $v_j = v_j(d_m)$, $j \neq m$ can be found from the intersections with the dotted axis in fig.5.1. It will be clear that for $i_{k+1} = d_m$ we have

$$v_m = 0, \quad i_m = 0 \quad \text{and } v_j > 0 \quad \text{for } k \geq j \neq m$$

Furthermore, due to $D_{k+1,k+1} > 0$, v_{k+1} has become less negative at this point. The performed principal transformation results in the new equation system

$$w = \tilde{D}z + \tilde{a}, \quad w \geq 0, \quad z \geq 0, \quad w^t z = 0 \tag{5.12}$$

with

$$\left.\begin{array}{l} w_j = v_j \\ z_j = i_j \end{array}\right\} \text{for } j \neq m \quad \text{and} \quad \begin{array}{l} w_m = i_m \\ z_m = v_m \end{array}$$

For $z_j = 0$, $j \le k$ the equations equivalent to (5.10) are now given by

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_k \\ \cdots \\ v_{k+1} \end{pmatrix} = \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_1 \\ \vdots \\ \tilde{a}_k \\ \cdots \\ \tilde{a}_{k+1} \end{pmatrix} + i_{k+1} \begin{pmatrix} \tilde{D}_{1,k+1} \\ \tilde{D}_{2,k+1} \\ \vdots \\ \tilde{D}_{k,k+1} \\ \cdots\cdots \\ \tilde{D}_{k+1,k+1} \end{pmatrix} \tag{5.13}$$

The various coefficients in (5.13) are easily obtained from (4.23) yielding

$$\tilde{a}_m = -a_m/D_{m,m} \qquad \tilde{D}_{m,k+1} = -D_{m,k+1}/D_{m,m}$$

$$\left.\begin{aligned} \tilde{a}_j &= a_j - a_m D_{m,j}/D_{m,m} \\ \tilde{D}_{j,l} &= D_{j,l} - D_{j,m} D_{m,l}/D_{m,m} \end{aligned}\right\} \; j,l \neq m \tag{5.14}$$

From (5.14) and $d_m = -a_m/D_{m,m}$ we find

$$v_j(d_m) = a_j + d_m D_{j,k+1} = \tilde{a}_j + d_m \tilde{D}_{j,k+1} = w_j(d_m), \quad j = 1,2,\ldots,k+1$$

Hence the values of v_j and i_j as a function of i_{k+1} join continuously at $i_{k+1} = d_m$ when a change in the basis between v_m and i_m is performed at this point. Therefore the picture of fig.5.1 may be extended to that given in fig.5.2.

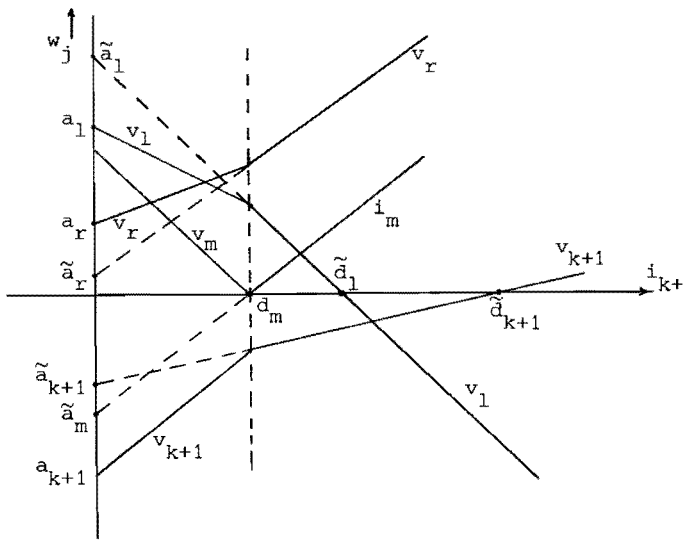


fig.5.2.

From fig.5.2 we may see that the pivoting step with $D_{m,m}$ may result into negative components of the vector $\tilde{a}$ as determined by the dotted extensions of the new linear relations. However, this is no problem since the relevant values of v_j or i_j, $j \neq k$ are all positive immediately to the right of $i_{k+1} = d_m$, until a new intersection of one of these lines with the i_{k+1}-axis occurs. As before we should look for this possible intersection among the lines with negative slope, i.e. among the equations with negative entries in the k + 1-th column of $\tilde{D}$. We thus look for a $\tilde{d}_l$ satisfying

$$\tilde{d}_l = \min_j \, (\tilde{d}_j = -\tilde{a}_j/\tilde{D}_{j,k+1} \mid \tilde{d}_j > 0) \tag{5.15}$$

If such a $\tilde{d}_l$ does not exist or if $\tilde{d}_l > \tilde{d}_{k+1}$, we take $\tilde{D}_{k+1,k+1}$ as a pivot and obtain a vector $\bar{a}$ with at least k + 1 positive elements. However if $\tilde{d}_l < \tilde{d}_{k+1}$ we are again forced to perform a one-step principal transformation with $\tilde{D}_{l,l}$ as pivot and repeat the above procedure on the new representation. This process ultimately leads to a basis in which v_{k+1} can be exchanged for i_{k+1}, because the

piecewise-linear curve $v_{k+1}(i_{k+1})$ arising in fig.5.2, from all pivoting steps (base transformations), always has a positive slope since in every representation (base) $\tilde{D}_{k+1,k+1} > 0$ when $D \in P$. Then this monotonically increasing function guarantees that no set of base variables occurs twice since the maximum value of v_{k+1} is uniquely determined by the particular base. Hence in the worst case the algorithm runs through all possible bases before a new representation is produced with at least k + 1 positive components in the corresponding vector a.

Should the vector a at the beginning of the algorithm satisfy $a < 0$ then any pivoting step produces at least one positive entry in the vector a. Hence the algorithm can always start with $k = 1$.

During the run of the algorithm it may happen that the index l in (5.15) is not unique, i.e. two or more variables w_j become zero at the same time, which is in fact equivalent to the corner problem in the Katzenelson algorithm. The question is then which pivots should be taken in order to get all slopes of these variables positive for $i_{k+1} > \tilde{d}_l$. To this purpose let us assume without loss of generality that the m variables $w_1, w_2, \ldots, w_m$ all become zero if $i_{k+1} = \tilde{d}_l$ and consider the subsystem

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = \begin{pmatrix} \tilde{a}_1 \\ \tilde{a}_2 \\ \vdots \\ \tilde{a}_m \end{pmatrix} + i_{k+1} \begin{pmatrix} \tilde{D}_{1,k+1} \\ \tilde{D}_{2,k+1} \\ \vdots \\ \tilde{D}_{m,k+1} \end{pmatrix} + \begin{pmatrix} \tilde{D}_{11} & \tilde{D}_{12} & \cdots & \tilde{D}_{1m} \\ \tilde{D}_{21} & \tilde{D}_{22} & \cdots & \tilde{D}_{2m} \\ \vdots & \vdots & & \vdots \\ \tilde{D}_{m1} & \tilde{D}_{m2} & \cdots & \tilde{D}_{mm} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix}, \tag{5.16}$$

With $\tilde{a}_j + \tilde{d}_l \tilde{D}_{j,k+1} = 0 \qquad j = 1,2,\ldots,m \qquad m < k$

For $i_{k+1} > \tilde{d}_l$ we then have

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{pmatrix} = (i_{k+1} - \tilde{d}_1) \begin{pmatrix} \tilde{D}_{1,k+1} \\ \tilde{D}_{2,k+1} \\ \vdots \\ \tilde{D}_{m,k+1} \end{pmatrix} + \begin{pmatrix} \tilde{D}_{11} & \tilde{D}_{12} & \cdots & \tilde{D}_{1m} \\ \tilde{D}_{21} & \tilde{D}_{22} & \cdots & \tilde{D}_{2m} \\ \vdots & \vdots & & \vdots \\ \tilde{D}_{m1} & \tilde{D}_{m2} & \cdots & \tilde{D}_{mm} \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{pmatrix} \quad (5.17)$$

In order for all slopes of $w_j(i_{k+1})$ to become positive we have to pivot the system (5.17) such a way that the vector $(\tilde{D}_{1,k+1},\ldots\tilde{D}_{m,k+1})^t$ becomes positive. However this problem is exactly of the same type as the problem we are currently solving but of smaller dimension. Hence the same Cottle algorithm can be applied to make all slopes nonnegative. Then for $i_{k+1} = \tilde{d}_1$ all other variables w_j with $j > m$, $j \neq k + 1$ have remained unchanged after the pivoting performed on (5.17) since $w_j(\tilde{d}_1) = 0$ for $j \leq m$.

In a practical application of the above algorithm it is of advantage to check after each pivoting step whether by accident the current number q of nonnegative components of the vector a already exceeds k. In such a case the algorithm becomes more efficient when restarted on the last obtained representation with q nonnegative components. Furthermore in selecting the negative component of the vector a which has to be made positive it is advisable to choose that one for which the corresponding d_{k+1} is as small as possible. This reduces the chance that other pivots have to be taken before v_{k+1} and i_{k+1} can be exchanged and benefits the numerical stability of the algorithm.

5.3. The Lemke algorithm

The most familiar algorithm to solve the LCP problem is the well known Lemke algorithm [7]. A review of the matrix classes on which the algorithm can be applied succesfully is given by Karamardian [18] and includes class P matrices. For a given x and

$$w = Dz + a, \quad a = Cx + h$$
$$w \geq 0, \quad z \geq 0, \quad w^t z = 0, \quad D = (n \times n), \quad C = (n \times m), \tag{5.18}$$

the algorithm runs as follows.
First of all, the vector a in (5.18) is extended by a multiple z_0 of some positive vector e, yielding

$$w = Dz + a + z_0 e, \quad e > 0 \tag{5.19}$$

Next by defining

$$\underline{w} = \begin{pmatrix} w_0 \\ \cdots \\ w \end{pmatrix}, \quad \underline{z} = \begin{pmatrix} z_0 \\ \cdots \\ z \end{pmatrix}, \quad M = \begin{pmatrix} 0 & \vdots & 0 \;..\; 0 \\ & \vdots & \\ \cdots & \cdots & \cdots \\ e & \vdots & D \end{pmatrix} \text{ and } \underline{a} = \begin{pmatrix} 0 \\ \cdots \\ a \end{pmatrix}$$

We consider the problem

$$\underline{w} = \underline{a} + M\underline{z} \tag{5.20a}$$

$$\underline{w} \geq 0 \qquad \underline{z} \geq 0 \tag{5.20b}$$

$$\underline{w}^t \underline{z} = 0 \tag{5.20c}$$

Obviously, if after some pivoting steps a solution of (5.20) is obtained for which $\underline{z}_j = z_0 = 0$, then the components w and z of that solution also satisfy (5.18).

Before continuing our description we make the following definitions. In any pivoted equivalent relation (5.20a), the variables in $\underline{w}$ will be called basic variables and those in $\underline{z}$ are nonbasic variables. Furthermore a pair $\underline{w}, \underline{z}$ which satisfies (5.20a) and (5.20b) will be called *feasible* and a pair $\underline{w}, \underline{z}$ for which $\underline{w}^t \underline{z} = 0$ and $\underline{z}$ contains exactly one pair of components w_j, z_j will be called *almost complementary*.

Let us now consider a value z_0 such that

$$z_0^* = \min_j \left(-\frac{a_j}{e_j} \mid a_j < 0\right) \tag{5.21}$$

and let this minimum be obtained for $j = k$. Then from (5.19) and (5.20) we obtain for $z = 0$

$$\begin{aligned} w_j &= a_j + z_0^* e_j > 0 \qquad j \neq k \\ w_k &= 0 \end{aligned} \tag{5.22}$$

Next the system (5.20) is pivoted on the element $M_{k,1}$ yielding

$$\underline{w}' = \underline{a}' + M'\underline{z}' , \tag{5.23}$$

with the basic variables $\underline{w}'$ and the nonbasic variables $\underline{z}'$ given by

$$\begin{aligned} \underline{w}' &= (w_0, w_1, \ldots, w_{k-1}, z_0, w_{k+1}, \ldots, w_n)^t \\ \underline{z}' &= (w_k, z_1, \ldots, z_{k-1}, z_k, z_{k+1}, \ldots, z_n)^t \end{aligned} \tag{5.24}$$

From (5.22) to (5.24) and $z_0 = z_0^*$ we then have that an almost complementary feasible solution of (5.23) is given by

$$w' = a', \quad z' = 0, \quad a' > 0$$

since

$$a_j' = a_j + z_0^* e_j > 0 \qquad j \neq k \quad \text{and } a_k' = z_0^* > 0 \tag{5.25}$$

From (5.24) we observe that the variable w_k has become nonbasic. We then start to make its complement z_k positive until some component $\underline{w}_r' = \underline{a}_r' + M_{rk}' z_k = 0$. Then the system (5.23) is pivoted on M_{rk}' yielding

$$\underline{w}'' = \underline{a}'' + \underline{M}''\underline{z}''$$

$$\begin{aligned} \text{with } \underline{w}'' &= (w_0, w_1, \ldots, w_{r-1}, z_k, w_{r+1}, \ldots, z_0, \ldots)^t \\ \underline{z}'' &= (w_k, z_1, \ldots, z_{r-1}, w_r, z_{r+1}, \ldots, z_r, \ldots)^t \end{aligned} \tag{5.26}$$

As before it is easy to show that $\underline{a}'' > 0$ and $\underline{w}'' = \underline{a}''$, $\underline{z}'' = 0$ is then an almost complementary solution of (5.20) since $\underline{z}''$ contains w_r as well as z_r. Again the complement z_r of the variable w_r which

became nonbasic is increased until some other component of $\underset{\sim}{w}''$ becomes zero, which is then exchanged with z_r. This process is repeated until either z_0 becomes again nonbasic or no component of $\underset{\sim}{w}$ can be made zero since the column of M corresponding the the variable to be increased, only contains nonnegative elements.

In the first case we have found a complementary feasible solution of (5.20) for which the elements w_i, z_i satisfy (5.18) and thus a solution is found.

In the second case the algorithm ends in a so called almost complementary ray. Lemke has shown that the latter case never arises when $D \in P$ and also showed that no cycling can occur in this situation.

A disadvantage of the algorithm is that the pivots to be taken are not principal pivots which makes the implementation of the algorithm more difficult, especially when sparse matrix techniques are to be used.

6. The modulus algorithm

The complementary pivoting algorithms as described in chapter 5 all work reasonable efficient in practice when applied to those classes of matrices for which they are known to produce a solution. As to the complexity of these methods it is known that the Katzenelson- Cottle- and the Lemke-algorithm are nonpolynomial in time-complexity [19], [20]. Yet as for the simplex method in linear programming, the average complexity is polynomial.

In this section we will describe an algorithm which is based on contraction methods and has the property that for a specific class of matrices with a specified bounded condition number the time-complexity is polynomial in the dimension of the problem. At the same time, the formulation of the LCP problem in this particular algorithm can be used to present an alternate state-model description for which the global model as presented by Chua [21] is included as a special case.

6.1. The modulus transformation

Consider the LCP problem

$$\begin{aligned} &v = Mi + a \qquad M = (m\times m) \\ &v \geq 0, \quad i \geq 0, \quad v^t i = 0 \end{aligned} \tag{6.1}$$

We will transform (6.1) into another equivalent problem by the so called modulus transformation. We will define this modulus transformation on the general nonlinear complementary problem:

Given $v,i \in \mathbb{R}^m$, with $v = v(i)$. Determine the points v,i satisfying

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0. \tag{6.2}$$

Then we apply this transformation on (6.1) as a special case. For this purpose, with $z \in \mathbb{R}^m$ we define

$$|z| = (|z_1|,|z_2|,\ldots,|z_m|)^t \tag{6.3}$$

Furthermore, for a scalar function $\varphi : \mathbb{R} \to \mathbb{R}$ and a vector $x \in \mathbb{R}^m$ we define $\varphi[x]$ by

$$\varphi[x] = (\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_m))^t$$

With the above definitions we have the following theorem.

Theorem 7: For any strictly monotone $\theta(t) : \mathbb{R}_+ \to \mathbb{R}_+$, satisfying $\theta(0) = 0$, the nonlinear complementarity problem is completely equivalent to the solution of a set of m nonlinear equations in $\mathbb{R}^m$ given by
$|\theta[v] - \theta[i]| = \theta[v] + \theta[i]$

Proof: Let us denote the property "i,v is a solution of (6.2)" by CP. We will first show that

$$CP \leftrightarrow \exists_{z \in \mathbb{R}^m} [i = h[|z|+z] \wedge v = h[|z|-z]] \tag{6.4}$$

with a strictly monotone increasing $h(t) : \mathbb{R}_+ \to \mathbb{R}_+$ and $h(0) = 0$. For this purpose we assume that we have a vector i satisfying (6.2). For this vector i we define the vectors z and w according to

$$z = \frac{1}{2}(\theta[i]-\theta[v]), \quad w = \frac{1}{2}(\theta[i]+\theta[v]) \tag{6.5}$$

and θ as given in the theorem. Due to the definition of $\theta(t)$, its inverse on $\mathbb{R}_+$ exists and is denoted by $h(t)$. Then from (6.5) we obtain for a pair i,v satisfying (6.2)

$$\begin{aligned} z_k &= \frac{1}{2}\theta(i_k), \quad w_k = \frac{1}{2}\theta(i_k) \quad \text{if} \quad v_k = 0, \quad i_k \geq 0 \\ z_k &= -\frac{1}{2}\theta(v_k), \quad w_k = \frac{1}{2}\theta(v_k) \quad \text{if} \quad i_k = 0, \quad v_k \geq 0 \end{aligned} \tag{6.6}$$

Hence from (6.6) we obtain

$$CP \to w = |z|. \tag{6.7}$$

Subsitution of (6.7) in (6.5) and eliminating $\theta[i]$ and

$$\theta[i] = |z| + z$$
$$\theta[v] = |z| - z \tag{6.8}$$

With the previous h(t), (6.7) and (6.8) result into

$$CP \rightarrow \exists_z [i = h[|z|+z] \wedge v = h[|z|-z]],$$

with the particular z given by (6.5).
On the other hand, if the r.h.s of (6.4) holds, it is immediately obvious that $i \geq 0$, $v \geq 0$ and $i^t v = 0$, due to the definition of h(t). This concludes the proof of (6.4).

Next, from the r.h.s of (6.4) we eliminate $|z|$ and z yielding

$$\exists_{z \in \mathbb{R}^m} [i = h[|z|+z] \wedge v = h[|z|-z] \leftrightarrow$$
$$|z| = \frac{1}{2}(\theta[i] + \theta[v]),$$
$$z = \frac{1}{2}(\theta[i] - \theta[v]) \tag{6.9}$$

Eliminating z from the r.h.s of (6.9) then yields with (6.4)

$$CP \leftrightarrow |\theta[i] - \theta[v]| = \theta[i] + \theta[v],$$

which concludes the proof of theorem 7.

The result of theorem 7 is closely related to a theorem of Mangasarian [22] who proved that for the same conditions on $\theta(t) : \mathbb{R} \rightarrow \mathbb{R}$

$$CP \leftrightarrow \theta[|v-i|] = \theta[v] + \theta[i]$$

From the proof of theorem 7 we also derive the following theorem.

<u>Theorem 8</u>: If $h : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a strictly increasing function with $h(0) = 0$, then

$$i := h[|z| + z]$$
$$v := h[|z| - z]$$

is a solution of the nonlinear complementarity problem iff z is a solution of the system of m nonlinear equations

For the special case that $h(t) := t$, the equations in theorem 8 are a different notation for the equivalent formulation of the nonlinear complementarity problem given by Megiddo in terms of the so called extension [23, p.113].

Definition: The transformation i,v defined by $i := h[|z| + z]$, $v := h[|z| - z]$ with a strictly increasing $h : \mathbb{R}_+ \to \mathbb{R}_+$ and $h(0) = 0$, is called *the modulus transformation.*

By theorem 8, application of the modulus transformation to a function $v = v(i)$ automatically guarantees that $i \geq 0$, $v \geq 0$ and $i^t v = 0$.

Theorem 9: The linear complementarity problem (6.1) with $\det(I+M) \neq 0$ is equivalent to the determination of a vector $z \in \mathbb{R}^m$ which satisfies

$$z = (I+M)^{-1}(I-M)|z| - (I+M)^{-1}a.$$

The solutions z^j of this equation determine the solutions v^j, i^j of the LCP problem by $v^j = |z^j| - z^j$, $i^j = |z^j| + z^j$.

Proof: With $h(t) := t$ we apply the modulus transformation on $v = Mi + a$, yielding

$$|z| - z = M(|z|+z) + a, \quad \text{with } v = |z| - z, \quad i = |z| + z.$$

Rearranging terms then gives

$$(I+M)z = (I-M)|z| - a \tag{6.10}$$

By a suitable scaling of i in (6.1) we can always arrange that $I + M$ is nonsingular. Hence from (6.10) we find

$$z = (I+M)^{-1}(I-M)|z| - (I+M)^{-1}a\ , \tag{6.11}$$

with $v = |z| - z$, $i = |z| + z$.

The theorem then follows from theorem 8.

In the next section we will present an algorithm to solve the nonlinear equation (6.11) by some contraction-mapping algorithm . Due to theorem 9, we do not have to bother about the complementarity

conditions for v and i and in fact we now have the large field of algorithms to solve nonlinear equations at our disposal for solving the LCP problems as well.

6.2. The modulus algorithm as a contraction mapping

In this section we will show that (6.11) can be solved by contraction mapping for certain conditions on the matrix M.

Let us rewrite (6.11) in the form

$$z = b + D|z|, \tag{6.12}$$

with $b = -(I+M)^{-1}a$ and $D = (I+M)^{-1}(I-M)$.

In the sequel we will use an absolute norm over $\mathbb{R}^m$ satisfying

$$\begin{aligned}
&\text{a)}\ \ \|x\| \geq 0,\ \|x\| = 0 \ \text{ iff } x = 0\\
&\text{b)}\ \ \|\alpha x\| = |\alpha|\cdot\|x\|\\
&\text{c)}\ \ \|a+b\| \leq \|a\| + \|b\|\\
&\text{d)}\ \ \|\,|a|\,\| = \|a\|\\
&\text{e)}\ \ 0 \leq a \leq b \ \rightarrow\ \|a\| \leq \|b\|
\end{aligned} \tag{6.13}$$

As an example the standard $\|\cdot\|_\infty$ and $\|\cdot\|_2$ norm are absolute norms. Furthermore, the norm $\|A\|$ of a matrix A is defined to be the induced norm

$$\|A\| = \sup_{x\neq 0} \frac{\|Ax\|}{\|x\|} .$$

<u>Lemma 3</u>: $\|\,|a| - |b|\,\| \leq |a-b|$, $a,b \in \mathbb{R}^m$, the inequality holding componentwise.

<u>Proof</u>: We partition a and b in four parts such that any possible sign combination of the components of a and b is represented. With $a_i, b_i \geq 0$ this is accomplished by

$$a = (a_1, a_2, -a_3, -a_4)^t \text{ and } b = (b_1, -b_2, b_3, -b_4)^t$$

Then we have

$$p = \|a| - |b\| = (|a_1-b_1|,\ |a_2-b_2|,\ |a_3-b_3|,\ |a_4-b_4|)^t \tag{6.14}$$

and

$$q = |a-b| = (|a_1-b_1|,\ |a_2+b_2|,\ |a_3+b_3|,\ |a_4-b_4|)^t \tag{6.15}$$

From (6.14) and (6.15) we see that every component of p is less or equal to the corresponding component of q and thus $0 \le p \le q$, which proves the lemma.

Theorem 10: If $\|D\| < 1$, the iterative process $z^{n+1} = b + D|z^n|$ converges and the limit z^* of the sequence $\{z^n\}$ for $n \to \infty$ is the unique solution of $z = b + D|z|$.

Proof: With $f(z) = b + D|z|$, equation (6.12) can be written as

$$z = f(z) \tag{6.16}$$

Now consider the function $f(z)$ for which we derive

$$\|f(z^1)-f(z^2)\| = \|D(|z^1|-|z^2|)\| \le \|D\|\cdot\||z^1|-|z^2|\| \tag{6.17}$$

By lemma 3 and the norm properties d) and e) we have

$$\||z^1|-|z^2|\| = \|\,|\,|z^1|-|z^2|\,|\,\| \le \||z^1-z^2\| = \|z^1-z^2\|$$

Thus (6.17) yields

$$\|f(z^1)-f(z^2)\| \le \|D\|\cdot\|z^1-z^2\|\ ,$$

which means that f is Lipschitz bounded for all $z^1, z^2 \in \mathbb{R}^m$ since $\|D\|$ does not depend on z^1 or z^2.

Then if $\|D\| < 1$, the iteration $z^{n+1} = f(z^n)$ converges by the Banach contraction mapping theorem [24] and $z^* = \lim_{n\to\infty} z^n$ is the unique solution of $z = f(z)$, which proves theorem 10.

Definition: The iteration $z^{n+1} := b + D|z^n|$ will be called *the modulus algorithm.*

Theorem 11: If $\|(I+M)^{-1}(I-M)\| < 1$, the modulus algorithm $z^{n+1} = (I+M)^{-1}(I-M)|z^n| - (I+M)^{-1}a$ yields the unique solution of the LCP problem

$$v = Mi + a, \quad v \geq 0, \quad i \geq 0, \quad v^t i = 0 \quad \text{via}$$

$$z^* = \lim_{n\to\infty} z^n \text{ and } v = |z^*| - z^*, \; i = |z^*| + z^*.$$

Proof: The proof is obvious from theorem 9 and theorem 10.

6.3. A polynomial algorithm

In this section we will bound the number of operations of the modulus algorithm to obtain an exact solution of the LCP problem in the case of symmetric positive definite matrices M.

In the sequel $\|\cdot\|$ will denote the standard $\|\cdot\|_2$ norm unless stated otherwise.

Definition: The radius $\theta(A)$ of a matrix A is defined by

$$\theta(A) := \|(I+A)^{-1}(I-A)\|_2$$

Next we present the main theorem of this section.

Theorem 12: Given a symmetric positive definite matrix $M = (m\times m)$ with a bounded condition number $K = \|M\|\cdot\|M^{-1}\|$. Then the LCP problem $v = Mi + a$, $v \geq 0$, $i \geq 0$, $v^t i = 0$ is solvable in polynomial time by the modulus algorithm and the number N of multiplications or additions is asymptotically bounded by

$$N \sim \frac{1}{7} K^{1/2} m^{7/2}.$$

Proof: Since M is symmetric and positive definite, M has positive real eigenvalues, i.e. $\lambda_i(M) > 0$. The eigenvalues μ_i of $D = (I+M)^{-1}(I-M)$ are given by

$\mu_i(D) = \frac{1-\lambda_i}{1+\lambda_i}$ and hence the μ_i are real and satisfy $|\mu_i| < 1$ since $\lambda_i > 0$. Let $\rho(A)$ denote the maximum modulus of the eigenvalues of A. Then since D is also symmetric, we have $\theta = \theta(M) = \|D\|_2 = \sqrt{\rho(D^tD)} = \sqrt{\rho(D)^2} = \rho(D) = \max_i|\mu_i| < 1$ and by theorem 10 the iteration $z^{n+1} = b + D|z^n|$ converges to a unique point z^*. The contraction mapping theorem then yields

$$\|z^*-z^n\| \le \frac{\theta^n}{1-\theta}\|z^1-z^0\|, \tag{6.18}$$

with z^* satisfying

$$z^* = b + D|z^*| \tag{6.19}$$

From theorem 10 we know that the solution z^* is unique. From (6.19) we find

$$\|z^*\| = \|b + D|z^*|\| \ge \|b\| - \|D|z^*|\| \ge \|b\| - \|D\|\cdot\|z^*\|$$

Thus

$$\|z^*\| \ge \frac{\|b\|}{1+\|D\|} = \frac{\|b\|}{1+\theta} = \gamma\|b\| \tag{6.20}$$

If we start the iteration with $z^0 = 0$ we obtain $z^1 = b$ and thus by (6.18)

$$\|z^*-z^n\| \le \frac{\theta^n}{1-\theta}\|b\| = \delta(n)\|b\| \tag{6.21}$$

Then $\|z^*\| - \|z^n\| \le \|z^*-z^n\| \le \delta(n)\|b\|$,

or using (6.20)

$$\|z^n\| \ge \|z^*\| - \delta\|b\| \ge (\gamma-\delta)\|b\| = \varepsilon(n)\|b\| \tag{6.22}$$

From $\|z^n\|_2 \ge \varepsilon(n)\|b\|_2$, we have

$$\exists_i \left[|z_i^n| \geq \frac{\varepsilon(n)\,||b||}{\sqrt{m}} \right] \tag{6.23}$$

On the other hand, from (6.21) we find

$$|z_i^* - z_i^n| \leq \delta(n)\,||b|| \tag{6.24}$$

As soon as $\frac{\varepsilon(n)}{\sqrt{m}} > \delta(n)$, we see from (6.23) and (6.24) that the component z_i^* of z^* must have the same sign as the component z_i^n of z^n. This happens as soon as

$$\frac{1}{1+\theta} > (1+\sqrt{m})\,\frac{\theta^n}{1-\theta} \quad , \text{ or}$$

$$n > N_0 = \left\lceil \frac{\ln \frac{1-\theta_0}{1+\theta_0} - \ln(1+\sqrt{m})}{\ln \theta_0} \right\rceil , \quad \theta_0 = \theta(M) < 1 \tag{6.25}$$

Hence after N_0 iterations, the sign of at least one component of z^* is known. Then from (6.11) we also know whether v_k or i_k in the complementary solution has to be zero.

In the case that $i_k = 0$, we delete the k-th equation from (6.1) as well as the k-th column of M and are left with an identical problem of dimension m - 1 with a matrix $\bar{M} = (m-1)\times(m-1)$. Should $v_k = 0$ we exchange v_k and i_k in (6.1) and again reduce the problem to dimension m - 1 with a matrix $\tilde{M}$.

This phase concludes cycle number zero of the algorithm. Now we apply the same process in each following cycle. In appendix II it is shown that $\theta(M_j) < 1$ for all matrices M_j that arise this way in every subsequent cycle of the algorithm. Hence the process results in the signs of all components of z^* after running through m such cycles. In the case that $v_k + i_k = 0$ for some k, we get a reduced system sooner or later for which the b column is zero,

leading to $z = 0$ and hence $v_k = i_k = 0$. From (6.25) the number of iterations in cycle k is limited by

$$N_k \le \left\lceil \frac{\ln \frac{1-\theta_k}{1+\theta_k} - \ln(1+\sqrt{m-k})}{\ln \theta_k} \right\rceil \quad , \quad k=0,1,2,\ldots,m-1$$

With $K_1 = \dfrac{\ln \frac{1-\bar{\theta}}{1+\bar{\theta}}}{\ln \bar{\theta}} > 0$, $K_2 = \dfrac{-1}{\ln \bar{\theta}}$ and $\bar{\theta} = \max_i \theta_i$,

this yields

$$N_k \le K_1 + K_2\sqrt{m-k} \tag{6.26}$$

Of course, the reduced problem in the k+1-th cycle is most efficiently obtained by eliminating the component of z^*, whose sign was determined in the k-th cycle, from the equations in the k-th cycle. This amounts in total for at most $O(m^3)$ multiplications (long operations). The number of long operations in the required contractions equals

$$N_C = \sum_{i=0}^{m-1} N_i (m-i)^2 \le \sum_{i=0}^{m-1} (K_1 + K_2\sqrt{m-i})(m-i)^2$$

By using the Euler-Maclaurin summation formula we find

$$N_C \le \frac{2}{7} K_2 m^{7/2} + \text{lower order terms} \tag{6.27}$$

In appendix II it is derived that

$K_2 \le \frac{1}{2} (\|M\| \cdot \|M^{-1}\|)^{1/2} = \frac{1}{2} K^{1/2}$ can be realized by a suitable scaling of M, yielding

$$N \sim \frac{1}{7} K^{1/2} m^{7/2}$$

This concludes the proof of theorem 12.

From our previous considerations about networks containing uncontrolled resistive elements with strictly monotone increasing characteristics, it will be clear that the above algorithm will

yields a polynomial-time solution process since the hybrid matrix of a purely resistive network (e.g. the impedance matrix) is symmetric and positive definite.

6.4. Implementation aspects and variants of the modulus algorithm

In a lot of practical situations, the matrix M in (6.1) will be sparse and the iteration $z^{n+1} = (I+M)^{-1}(I-M)|z^n| - (I+M)^{-1}a$ is not efficient since $(I+M)^{-1}(I-M)$ may become non-sparse. To be able to use sparse matrix techniques it is better to use the iteration in the equivalent formulation of

$$(I+M)(z^{n+1}+|z^n|) = 2|z^n| + a \tag{6.28}$$

If M is sparse, then I + M is also sparse and an L\U factorization of (I+M) enables us to solve (6.28) by sparse matrix techniques with a minimum amount of work. Furthermore it will be convenient during the iteration to monitor the residue vector e^n given by

$$e^n = Mi^n + a - v^n \tag{6.29}$$

with $i^n = |z^n| + z^n$, $v^n = |z^n| - z^n$. From (6.29) we then find

$$e^n = M(|z^n|+z^n) + a - |z^n| + z^n = (M+I)z^n - (I-M)|z^n| + a,$$

or $\quad (I+M)^{-1}e^n = z^n - (I+M)^{-1}(I-M)|z^n| + (I+M)^{-1}a = z^n - z^{n+1}$

Hence

$$e^n = (I+M)(z^n-z^{n+1}) \tag{6.30}$$

From (6.30) we obtain

$$\|e^n\| \le \|I+M\|\,\|z^n-z + z - z^{n+1}\| \le \|I+M\|\left(\frac{\theta^n}{1-\theta} + \frac{\theta^{n+1}}{1-\theta}\right)\|b\|, \tag{6.31}$$

with $b = (I+M)^{-1}a$.

Then, with $\theta < 1$, (6.31) results into a relative error

$$\frac{\|e^n\|}{\|a\|} \le 2\|I+M\| \cdot \|(I+M)^{-1}\| \frac{\theta^n}{1-\theta} \tag{6.32}$$

For a prescribed accuracy, the required number of iterations can now be found from (6.32).

As to the contraction mapping iteration $z^{n+1} = D|z^n| + b$ to solve $z = D|z| + b$, it will be clear that refinements to this brute force attack should be possible which can increase the convergence speed of this algorithm, depending on the properties of D. From the derivations in appendix II, we may obtain that the eigenvalues of each principal submatrix of D are less than one in modulus if M is symmetric and positive definite. Hence all diagonal elements of D then satisfy $|D_{kk}| < 1$. For that case it may be interesting to consider the iteration

$$z^{n+1} - L|z^{n+1}| = U|z^n| + b \ , \tag{6.33}$$

with $L + U = D$ and $U_{ij} = 0$ for $i \le j$.
Then the left hand side of (6.33) can be uniquely solved for z^{n+1} by going from top to bottom through this "triangular" system since each equation is of the form

$$z_k^{n+1} - L_{kk}|z_k^{n+1}| = \sum_{j=k+1}^{m} U_{kj}|z_j^n| + \sum_{j=1}^{k-1} L_{kj}|z_j^{n+1}| \quad k=1,2,\ldots,m$$

with $|L_{kk}| < 1$.

Practical experiments indeed showed a speed-up of this Gauss-Seidel-like iteration. A different approach to the solution of (6.1) is found by applying the modulus transformation with $h(t) = t^2$.
We then have

$$\begin{aligned} i &= h[|z|+z] = (|z|+z)^2 = 2z^2 + 2z|z| \ , \\ v &= h[|z|-z] = (|z|-z)^2 = 2z^2 - 2z|z| \ , \end{aligned} \tag{6.34}$$

with the multiplications in z^2 and $z|z|$ understood to be taken componentwise. Substitution of (6.34) in (6.1) then yields

$$z|z| = (I+M)^{-1}(I-M)z^2 - \frac{1}{2}(I+M)^{-1}a \tag{6.35}$$

From (6.35) we immediately observe that this nonlinear set of equations, written in the form $F(z) = 0$ is continuously differentiable for all $z \in \mathbb{R}^m$ since the scalar functions $t|t|$ and t^2 both have this property. Hence we can apply Newton-Raphson iteration on (6.35) to find the solution z^* and benefit from the quadratic convergence. A disadvantage is of course that some initial guess, sufficiently close to the final solution, must be given .

6.5. Relations with the global model of Chua

The modulus transformation as defined in section 6.1 can of course also be used in the state-model to describe some PL mapping $f : x \to y$. To this purpose we apply the transformation $i = |z| + z$, $v = |z| - z$ to the equations

$$\begin{aligned} y &= Ax + Bi + g \\ v &= Cx + Di + h\ , \quad v \geq 0, \quad i \geq 0, \quad v^t i = 0 \end{aligned} \tag{6.36}$$

yielding

$$y = \bar{A}x + \bar{B}|z| + \bar{g} \tag{6.37a}$$

$$z = \bar{C}x + \bar{D}|z| + \bar{h} \tag{6.37b}$$

with

$$\begin{aligned} \bar{A} &= A - B(I+D)^{-1}C, & \bar{B} &= 2B(I+D)^{-1} \\ \bar{C} &= -(I+D)^{-1}C\ , & \bar{D} &= (I+D)^{-1}(I-D) \\ \bar{h} &= -(I+D)^{-1}h\ , & \bar{g} &= g - B(I+D)^{-1}h \end{aligned} \tag{6.38}$$

Due to our previous derivations, the equation system (6.37) is completely equivalent to the state-model (6.36). As an additional advantage no extra complementarity conditions are necessary for the state-equations (6.37b). A specific state in (6.36) is now equivalent to a specific sign distribution over the components z_i of z in (6.37b). As with the state-model (6.36), there may exist 2^n different equivalent descriptions of the form (6.37) which also may be classified by a state vector s_p corresponding to

$$z_k > 0 \rightarrow s_{p_k} = 1$$

$$z_k \leq 0 \rightarrow s_{p_k} = 0$$

Now consider the special situation that the matrix D in (6.36) is equal to the unit matrix I. The state equations in (6.36) then become

$$v = Cx + h + i, \quad C = (n \times m)$$

from which we obtain for each component

$$v_k = (Cx+h)_k + i_k = C_{k.}x + h_k + i_k, \quad k=1,2,\ldots,n \tag{6.39}$$

Equation (6.39) then yields

$$\begin{aligned} &C_{k.}x + h_k \geq 0 \rightarrow v_k = C_{k.}x + h_k, \quad i_k = 0 \\ &C_{k.}x + h_k < 0 \rightarrow v_k = 0, \qquad\quad , \quad i_k = -(C_{k.}x+h_k) \end{aligned} \tag{6.40}$$

From (6.40) we see that the domain $\mathbb{R}^m$ is divided into a number of halfspaces by hyperplanes $C_{k.}x + h_k = 0$, which extend through whole $\mathbb{R}^m$ (see fig.6.1).

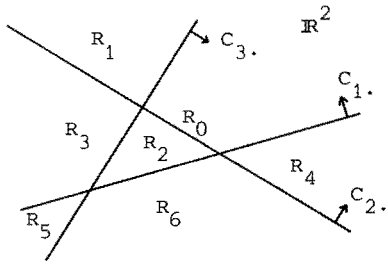


fig.6.1.

In each region, we only have to determine on which side of each hyperplane the region is in order to find a solution of the state equations. In our previous terminology we have with

$$I_p = \{m \mid s_{p_m} = 1\}$$

$$R_p = \{x \mid C_{i.}x + h_i \geq 0,\ i \notin I_p \ \wedge\ C_{i.}x + h_i < 0,\ i \in I_p\} \quad (6.41)$$

which is precisely the definition of the regions in Chua's global model [21] . Then for this specific case with $D = I$ we obtain an alternate state model from (6.37) and (6.38) given by

$$y = \bar{A}x + \bar{B}|\bar{C}x + \bar{h}| + \bar{g}\ , \quad (6.42)$$

since $\bar{D} = 0$. The expression for a single component $y_j(x)$ from eq. (6.42) is exactly identical to eq. (3) in [21].

It is easy to see that the global description (6.42) became possible because the state equations (6.39) or (6.37b) could explicitly be solved in this special case. However the model (6.37) is more general than Chua's global model, due to the implicit equation (6.37b) which defines z. As we already showed in section 6.4, the state equation can also be solved explicitly if $\bar{D}$ is upper or lower triangular, yielding explicit expressions for the corresponding PL mappings. Moreover when dealing with a mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^k$ and $k > 1$, the global model of Chua cannot be used in general although the regions R_i satisfy (6.41), as can be seen from the example in fig.6.2, for which the state-model is given by

$$y = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} i$$

$$v = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 2 & 0 \end{pmatrix} x + \begin{pmatrix} 1 & 1/2 & 1 \\ 1/2 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} i \qquad (6.43)$$

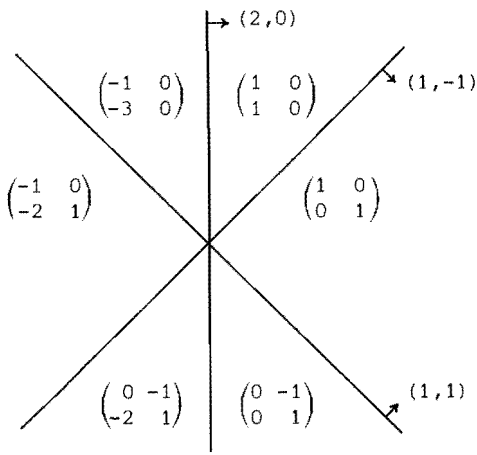
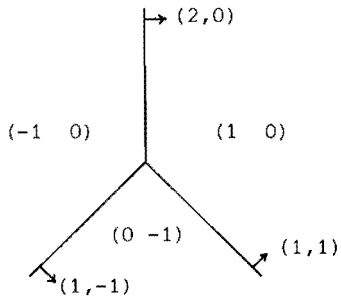
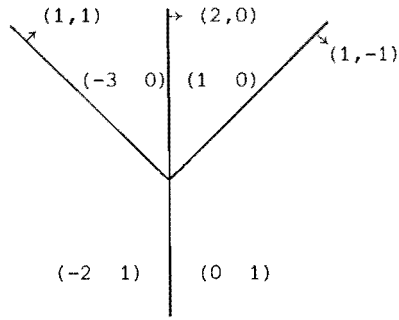


fig.6.2.

The corresponding affine mapping $y = A^p x$ in each region R_p is indicated in the figure. The maps of the component functions $y_1(x)$ and $y_2(x)$ can be easily obtained from fig.6.2 and are depicted in fig.6.3. From this figure, it is immediately obvious that (6.41) does not hold for the regions of the components $y_1(x)$ and $y_2(x)$, hence modelling by Chua's global model is not possible although the regions of the vector function $y(x)$ do satisfy (6.41). In this case the state equation in (6.43) cannot explicitly be solved to obtain eq. (6.42).



$y_1(x)$



$y_2(x)$

fig.6.3.

As we will show in the next chapter, the state-model formulation has additional advantages which makes it more versatile than Chua's model.

7. The inverse mapping and some homeomorphism conditions

Up to now, the state-model has been used to define a PL mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^k$ and some algorithms have been described to obtain the image vector y of a given vector x. These state-model descriptions can be used to model static nonlinear electronic devices such as bipolar transistors, digital gates etc. With these models available, it will be possible to describe a static non-linear electrical network constructed with these devices by an overall state-model. For general nonlinear networks, the calculation of the DC network response due to some applied excitation amounts to solving a set of nonlinear equations. In the same way, for piecewise-linear networks with a given state-model description, the DC network response is obtained by solving a set of piecewise-linear equations. With the state-model defining $y = f(x)$, it is then required to determine the vector x for a prescribed vector y, which is however equivalent to the determination of the inverse mapping $x = f^{-1}(y)$. If we are able to determine the state-model of this inverse mapping, the same algorithms as before can be used to find the image x for a given y. In the next sections we will show how to construct such a state-model and discuss some homeomorphism conditions for a PL function $y = f(x)$.

7.1. The state model of the inverse mapping

When discussing an inverse mapping in an Euclidean space, it is obvious that the domain and the range of the particular mapping have to be of the same dimension, otherwise the inverse mapping cannot be defined properly. Hence let us consider a PL mapping f from $\mathbb{R}^m$ into $\mathbb{R}^m$, defined by the state-model S(A,B,C,D,g,h).

$$
\begin{aligned}
&y = Ax + Bi + g \\
&v = Cx + Di + h \\
&v \geq 0, \quad i \geq 0, \quad v^t i = 0 \qquad (7.1) \\
&A = (m\times m), \quad B = (m\times n), \quad C = (n\times m), \quad D = (n\times n) \\
&x,y \in \mathbb{R}^m, \quad v,i \in \mathbb{R}^n
\end{aligned}
$$

The domain X of f consists of a number of polytopes R_i in which f is a linear mapping, i.e.

$$x \in R_i \rightarrow y = A^i x + g^i \ , \quad A^i = (m \times m) \ , \tag{7.2}$$

with A^i and g^i defined by the equivalent representation S_i.
For a proper inverse mapping to exist, it is required that at least one of the matrices A^i in (7.2) is nonsingular, otherwise the whole domain $X = \bigcup_i R_i$ will be mapped into a subspace of $\mathbb{R}^m$ and the inverse mapping is nowhere defined. Without loss of generality we assume that in representation S_0: $\det(A) = \det(A^0) \neq 0$.
From the first equation of (7.1) we then eliminate x, yielding

$$x = A^{-1}y - A^{-1}Bi - A^{-1}g \tag{7.3}$$

Substitution of (7.3) in the remaining equations then results into

$$\begin{aligned} x &= \tilde{A}y + \tilde{B}i + \tilde{g} \\ v &= \tilde{C}y + \tilde{D}i + \tilde{h} \\ v &\geq 0, \quad i \geq 0, \quad v^t i = 0 \ , \end{aligned} \tag{7.4}$$

with

$$\begin{aligned} &\tilde{A} = A^{-1}, \quad \tilde{B} = -A^{-1}B, \quad \tilde{C} = CA^{-1}, \quad \tilde{D} = D - CA^{-1}B \\ &\tilde{g} = -A^{-1}g, \quad \tilde{h} = h - CA^{-1}g \end{aligned} \tag{7.5}$$

Since (7.4) is in fact nothing else than the equation set (7.1), written in a different way, it is equivalent to (7.1) and describes the same mapping. However (7.4) is also a state-model describing a PL mapping $h : y \in \mathbb{R}^m \rightarrow x \in \mathbb{R}^m$, hence (7.4) is the state-model of the required inverse mapping f^{-1} and is denoted by
$S^{-1}(A,B,C,D,g,h)$ or $S(\tilde{A},\tilde{B},\tilde{C},\tilde{D},\tilde{g},\tilde{h})$.

Observe that the existance of only one nonsingular matrix A^i was sufficient to obtain S^{-1}. All other A^i may have been singular, which means that $f^{-1} : y \rightarrow x$ may be a one to many mapping. If some

$\tilde{A}^i$ is singular, the same holds for $f : x \to y$, yielding the somewhat surprising result that a mapping $f : x \to y$ which is one to many, has an inverse mapping $f^{-1} : y \to x$ which can still be described by a single state-model.

As an example consider the mapping $f : \mathbb{R} \to \mathbb{R}$ defined by

$$\begin{aligned} y &= x - 2i \\ v &= -x + i + 1 \ , \quad v \geq 0, \quad i \geq 0, \quad v \cdot i = 0 \end{aligned} \tag{7.6}$$

From (7.5) the state-model S^{-1} becomes

$$\begin{aligned} x &= y + 2i \\ v &= -y - i + 1 \ , \quad v \geq 0, \quad i \geq 0, \quad v \cdot i = 0 \end{aligned} \tag{7.7}$$

By considering all representations S_i and S_i^{-1}, the particular mappings f and f^{-1} can be depicted as indicated in fig.7.1.

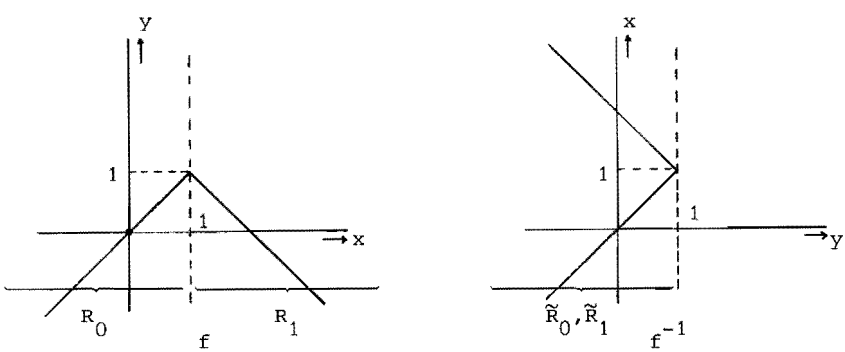


fig.7.1.

From this figure we see that $f^{-1} : y \to x$ is one to many because $\tilde{R}_0$ and $\tilde{R}_1$ overlap. Furthermore from (7.7) is is easily found that the state-equation $v = -y + 1 - i$ has no solution for $y > 1$, indicating that f^{-1} is not defined for $y > 1$, which is obvious from fig.7.1.

As for the state-model S, the state-model S^{-1} also has a number of equivalent representations S_i^{-1}, yielding a natural description for the mapping f^{-1} in polyhedral regions $\tilde{R}_i$. It is easy to see that a pair $\bar{v}, \bar{i}$ which satisfies the state equation of S for a given x_0 also satisfies the state-equations of S^{-1}, for

$y = Ax_0 + B\bar{i} + g$. Hence the regions R_i and $\tilde{R}_i$ both correspond to the same state in S and S^{-1} respectively. However when some representation S_k does not exist, this does not mean that the corresponding S_k^{-1} does not exist, as can be seen from the following example with the various regions depicted in fig.7.2.

$$\begin{pmatrix} y \\ \hline v \end{pmatrix} = \left(\begin{array}{cc|cc} 1 & 0 & 3 & 2 \\ 0 & 2 & -2 & -1 \\ \hline -1 & 0 & 1 & 1 \\ 0 & -1 & 1 & 1 \end{array}\right) \begin{pmatrix} x \\ \hline i \end{pmatrix} \rightarrow \begin{pmatrix} x \\ \hline v \end{pmatrix} = \left(\begin{array}{cc|cc} 1 & 0 & -3 & -2 \\ 0 & 1/2 & 1 & 1/2 \\ \hline -1 & 0 & 4 & 3 \\ 0 & -1/2 & 0 & 1/2 \end{array}\right) \begin{pmatrix} y \\ \hline i \end{pmatrix} \quad (7.8)$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0$$

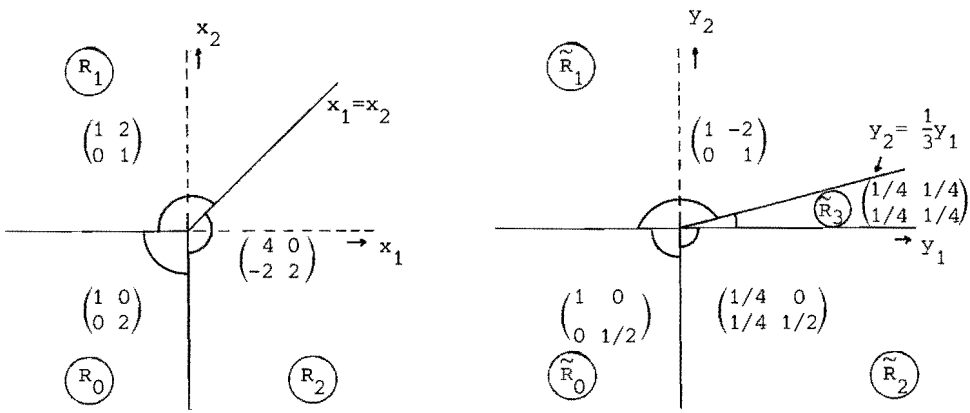


fig.7.2.

From (7.8) we find $\det(D_{\{1,2\}}) = 0$, $\det(\tilde{D}_{\{1,2\}}) \neq 0$. Hence S_3 does not exist while S_3^{-1} does. Using the relations (3.17) it can easily be shown that for every k for which S_k^{-1} exists and S_k does not exist, the determinant of $\tilde{A}^k$ will be zero. Hence such regions $\tilde{R}_k$ will be mapped on some subspace (hyperplane or corner) of the X-space. In the particular example $\tilde{R}_3$ is mapped into the hyperplane $x_1 = x_2 = 1/4(y_1+y_2)$. Thus although S_3 does not exist, there is a region R_3 corresponding to $\tilde{R}_3$ given by $x_1 - x_2 = 0$. Observe that the mapping $f : x \rightarrow y$ is not continuous on the boundary between R_1 and R_2, due to the fact that two pivoting steps are necessary to enter region R_2 from region R_1.

From the above considerations it will be clear that the state-model description of a PL mapping has important advantages over descriptions based on a list of mappings. In the latter case, the description of the inverse mapping requires the boundaries of all regions $\tilde{R}_i$ corresponding to R_i to be determined before such an explicit list of regions for f^{-1} can be produced. Moreover, due to the linear-algebraic nature of the state-model descriptions, it is a simple matter to model all kinds of linear transformations of a PL mapping by performing these transformations on its state-model description. In particular the model for a concatenation $t \circ f$ of two PL mappings $f : x \to z$ and $t : z \to y$ is easily derived. To this purpose considere f and t to be defined by

$$
\begin{aligned}
z &= Ax + Bi_1 + g & y &= Ez + Fi_2 + k \\
v_1 &= Cx + Di_1 + h & v_2 &= Gz + Hi_2 + l \\
v_1 &\geq 0, \quad i_1 \geq 0, \quad v_1^t i_1 = 0 & v_2 &\geq 0, \quad i_2 \geq 0, \quad v_2^t i_2 = 0
\end{aligned}
\tag{7. 9}
$$

Then a simple substitution of the first model into the second one will yield the state-model $\bar{S}$ for $t \circ f : x \to y$, given by

$$
\begin{aligned}
y &= \bar{A}x + \bar{B}i + \bar{g} \\
v &= \bar{C}x + \bar{D}i + \bar{h} \\
v &= \begin{pmatrix} v_1 \\ \hline v_2 \end{pmatrix} \geq 0 \ , \quad i = \begin{pmatrix} i_1 \\ \hline i_2 \end{pmatrix} \geq 0 \ , \quad v^t i = 0
\end{aligned}
\tag{7.10}
$$

with

$$
\begin{aligned}
&\bar{A} = EA, \quad \bar{B} = (EB \mid F) \\
&\bar{C} = \begin{pmatrix} C \\ \hline GA \end{pmatrix}, \quad \bar{D} = \left(\begin{array}{c|c} D & 0 \\ \hline GB & H \end{array}\right) \\
&\bar{g} = k + Eg, \quad \bar{h} = \begin{pmatrix} h \\ \hline 1+Gg \end{pmatrix}
\end{aligned}
\tag{7.11}
$$

In the next section we will use the state-model S^{-1} for presenting some homeomorphism conditions for a mapping $f : \mathbb{R}^m \to \mathbb{R}^m$.

7.2. Homeomorphic mappings

Consider a PL mapping $f : \mathbb{R}^m \to \mathbb{R}^m$ which can be modelled by a state-model S such that S^{-1} exists. The mapping f is said to be a homeomorphism if f is continuous, one to one and onto.

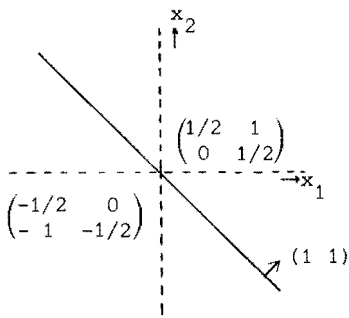
Homeomorphic mappings $f : x \to y$ play an important role in the analysis of nonlinear or piecewise-linear electrical networks. If, for an excitation y and a response x, the particular network equations have the form $y = f(x)$, then this response exists and is unique for any excitation y if f is a homeomorphism. Thus, circuits with multi-stable DC operating points like flip-flops or circuits with hysteresis effects cannot be described by a homeomorphic mapping. With the availability of the state-models S and S^{-1} it is then of interest to derive conditions on the matrices of these models, which will guarantee that the modelled PL mapping is a homeomorphism.

Without any constraints on the state-model in connection with the class of PL mappings to be modelled, it is a difficult task to find general necessary and sufficient conditons for a homeomorphism. However a general sufficient condition is easily obtained and is given by the following theorem.

Theorem 13: Let the PL mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^m$ be modelled by a state-model S for which S^{-1} exists. Then f is a homeomorphism if the matrices D and $\tilde{D}$ from S and S^{-1} satisfy $D, \tilde{D} \in P$.

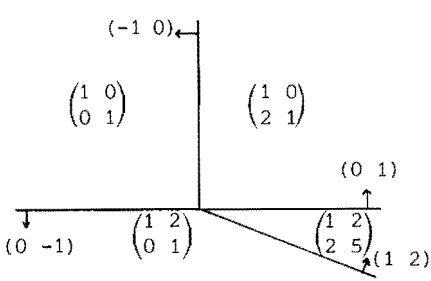
Proof: From theorem 3 in chapter 3.3 and (7.4), we know that the solution of the state-equations for each $x, y \in \mathbb{R}^m$ exists and is unique when $D, \tilde{D} \in P$. Hence for each $x \in \mathbb{R}^m$ the model S produces only one $y = y(x)$ and for each $y \in \mathbb{R}^m$ the model S^{-1} only yields one $x = x(y)$. Then the mapping f is one to one and onto. The continuity of f is implied by the unique solutions of the state equations.

The sufficient conditions of theorem 13 differ from those given by Fujisawa and Kuh [15] and the conditions given by Chien [25], as can be seen from the following two examples, borrowed from Chien. Example I (fig.7.3) satisfies the sufficient conditions of Chien and does not satisfy those of Fujisawa and Kuh, with the reverse situation for example II from fig.7.4. However both examples satisfy the condition of theorem 13 as can be found from the given state-models.



$$\begin{pmatrix} y_1 \\ y_2 \\ \hline v_1 \end{pmatrix} = \left(\begin{array}{cc|c} 1/2 & 1 & 1 \\ 0 & 1/2 & 1 \\ \hline 1 & 1 & 1 \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ \hline i_1 \end{pmatrix}$$

fig.7.3.



$$\begin{pmatrix} y_1 \\ y_2 \\ \hline v_1 \\ v_2 \end{pmatrix} = \left(\begin{array}{cc|cc} 1 & 2 & 0 & -2 \\ 0 & 1 & 2 & 0 \\ \hline -1 & -2 & 1 & 2 \\ 0 & -1 & 0 & 1 \end{array}\right) \begin{pmatrix} x_1 \\ x_2 \\ \hline i_1 \\ i_2 \end{pmatrix}$$

fig.7.4.

One might expect that the conditions of theorem 13 are necessary as well because $D \notin P$ implies that for at least one value of the vector a in $v = a + Di$, two different solutions of the state-

equations exist [6], which could possibly yield two different vectors y for a given vector x. However this need not be the case since the particular vector a may not belong to the subspace spanned by $Cx + h$ for $x \in \mathbb{R}^m$. Furthermore, if it does belong to that subspace, the second solution of the state-equations not necessarily yields a different value for y because the difference of both solutions may ly in the null-space of the matrix B.

To illustrate the problems we are faced with, we will discuss some examples.

As a first example we again consider the mapping given by (7.8) and depicted in fig.7.2, for which we find $D \in P_0$ and $\tilde{D} \in P$. The mapping is not a homeomorphism although the conditions are close to the sufficient conditions of theorem 13.

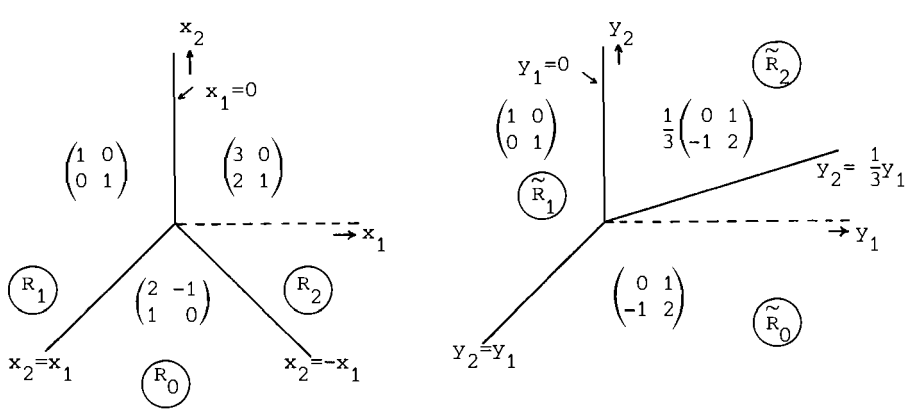
The fact that the conditions of theorem 13 are not necessary can be seen from the next example given in (7.12) and depicted in fig.7.5, for which both matrices D and $\tilde{D}$ satisfy $D, \tilde{D} \in P_0$ and yet the mapping is a homeomorphism.

$$y = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} x + \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} i \qquad x = \begin{pmatrix} 0 & 1 \\ -1 & 2 \end{pmatrix} y + \begin{pmatrix} -2 & -1 \\ -2 & -1 \end{pmatrix} i$$

$$\Rightarrow$$

$$v = \begin{pmatrix} -1/2 & -1/2 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 & 1/2 \\ 2 & 1 \end{pmatrix} i \qquad v = \begin{pmatrix} 1/2 & -3/2 \\ 1 & -1 \end{pmatrix} y + \begin{pmatrix} 3 & 3/2 \\ 2 & 1 \end{pmatrix} i$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0 \tag{7.12}$$



In contrast with example I, the homeomorphism is due to the fact that in this case the matrix $\tilde{D}$ has the same singular principal submatrices as the matrix D.

Finally, the third example is given by

$$y = x + (2 \quad -7)i \qquad\qquad x = y + (-2 \quad 7)i$$

$$v = \begin{pmatrix}1\\1\end{pmatrix} x + \begin{pmatrix}1 & -2\\ 3/4 & -1\end{pmatrix} i \quad \Rightarrow \quad v = \begin{pmatrix}1\\1\end{pmatrix} y + \begin{pmatrix}-1 & 5\\ -5/4 & 6\end{pmatrix} i$$

The corresponding mapping is depicted in fig.7.6.

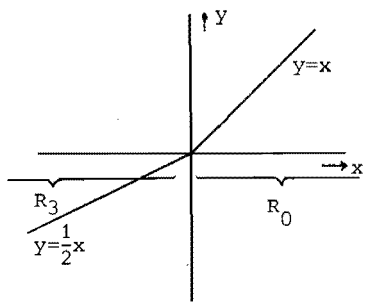


fig.7.6.

Although $D \notin P, P_0$ and $\tilde{D} \notin P, P_0$ the mapping is a homeomorphism. However, the model is redundant since a one-dimensional state-equation would already be sufficient to divide the domain in two half-spaces.

The given three examples all have the common property that there is at least one boundary hyperplane between two adjacent regions for which two pivoting steps have to be performed in order to obtain the natural equivalent representations of both adjacent regions from each other. In other words, although two regions are adjacent in space, their corresponding vertices in the structure of D are not adjacent, which indicates some redundancy in the state-model.

In order to eliminate this situation, we again use the minimal state-model of section 4.2 for a class of piecewise-linear mappings, which is now more accurately defined as follows.

Definition: A state-model is called a *minimal state-model* if and only if the natural equivalent representations of two adjacent regions in space yield different mappings and can be obtained from each other by a single pivoting step (one-step principal transformation).

Now consider a state-model S with a state-equation $v = Cx + h + Di$, $C = (n \times m)$, for which the k-th component of $Cx + h$ is denoted by $C_{k.}x + h_k$. The region R_j corresponding to representation S_j is then defined by

$$R_j = \{x \mid C^j_{k.}x + h^j_k \geq 0, \quad k=1,2,\ldots,n\} \tag{7.13}$$

Let some hyperplanes K^j_k, defined by

$$K^j_k = \{x \mid C^j_{k.}x + h^j_k = 0\} \tag{7.14}$$

be depicted in fig.7.7.

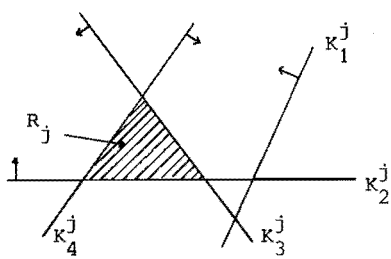


fig.7.7.

From this figure it is obvious that the hyperplane K^j_1 is not an actual boundary of the region R_j. In terms of linear algebra this means that the inequality $C^j_{1.}x + h^j_1 \geq 0$ depends on the remaining inequalities of (7.13) in such a way that it is satisfied if the remaining inequalities are satisfied. This implies that the row $C^j_{1.}$ is a positive linear combination of the remaining rows of C^j and hence rank $(C^j) < n$. We will call such hyperplanes *virtual boundary hyperplanes*.

Definition: The hyperplane K_k^j is called a *proper boundary hyperplane* of R_j, if and only if $R_j \cap K_k^j$ is non-empty.

The proper boundary hyperplanes of a region R_j can be determined by an algorithm given by Tschernikov [26], which determines whether a system of linear inequalities has a solution and in addition detects the redundant inequalities. Next consider a minimal state-model S for which S^{-1} exists. Then for each proper boundary hyperplane K_k^j, the corresponding diagonal element D_{kk}^j is non-zero otherwise the corresponding representation in the adjacent region cannot be obtained by a single pivoting step. For the same reason, $K_k^j \neq K_l^j$ for $k \neq l$. As each proper boundary hyperplane K_k^j is coupled to a diagonal element D_{kk}^j which in turn is related to an edge in the structure of S, we are able to modify this structure by deleting all edges that do not correspond to proper boundary hyperplanes. Furthermore, we also delete all vertices and incident edges corresponding to empty regions caused by uncontrollable states. The remaining structure will be called *the proper structure of S*.

Theorem 14: Let the mapping $f : x \to y$ be modelled by a minimal state-model S. Then f is a homeomorphism from $\mathbb{R}^m$ onto $\mathbb{R}^m$ only if the proper structure of S is connected and each edge in the proper structure has a positive weight.

Proof: Let f be a homeomorphism. Then y is defined for any $x \in \mathbb{R}^m$ and any pair of points x_1, x_2 can be connected by a line in $\mathbb{R}^m$ which traverses certain regions by crossing proper boundary hyperplanes. Then any pair of regions can be reached from each other by a sequence of one-step principal transformations and hence the proper structure is connected.

Now, for a homeomorphism f, consider without loss of generality representation S_0 and assume R_0 is non-empty. Let us take a point $x_0 \in R_0$, arbitrarily close to a proper boundary hyperplane K_k. We then have the state-equations $v = Cx_0 + h + Di$, with

$$C_{k\cdot}x_0 + h_k = \varepsilon > 0,\ \varepsilon \text{ arbitrarily small}$$
$$C_{l\cdot}x_0 + h_l > 0 \quad , l \neq k \tag{7.15}$$

Next, this system is pivoted on D_{kk} yielding a new representation S_q for the adjacent region R_q, with a state equation

$$v = C^q x_0 + h^q + D^q i \tag{7.16}$$

and

$$C^q_{l\cdot} = C_{l\cdot} - C_{k\cdot}\frac{D_{lk}}{D_{kk}}, \quad h^q_l = h_l - h_k\frac{D_{lk}}{D_{kk}}, \ l\neq k$$
$$C^q_{k\cdot} = -\frac{C_{k\cdot}}{D_{kk}}, \quad h^q_k = -\frac{h_k}{D_{kk}} \tag{7.17}$$

From (7.15) and (7.17) we obtain

$$C^q_{k\cdot}x_0 + h^q_k = \frac{-\varepsilon}{D_{kk}}$$
$$C^q_{l\cdot}x_0 + h^q_l = (C_{l\cdot}x_0 + h_l) - \frac{\varepsilon D_{lk}}{D_{kk}}, \ l\neq k \tag{7.18}$$

Now assume $D_{kk} < 0$. Then from (7.18) we can always have $C^q x_0 + h^q > 0$ by taking ε sufficiently small, which means that $x_0 \in R_q$. Hence the adjacent regions R_0 and R_q overlap each other yielding two different images for a single point x_0. Then f cannot be a homeomorphism and thus D_{kk} has to be positive. The same arguments hold for any representation S_j with R_j non-empty, which proves the theorem.

Next we will derive a relation between the determinants of two matrices A^j and A^k from the affine mappings in R_j and R_k. To this purpose, consider a state-model S, for which S^{-1} exists, and S and H^p defined by the relations (4.16) and (4.17) in section 4.2. With (4.18) and (4.20) we obtain with p = 0

$$A^q = A - B_M D_M^{-1} C_M \,,$$
$$M = \{k \mid s_{qk} = 1\} \tag{7.19}$$

From (7.19) we have

$$\det(A^q) = \det(A - B_M D_M^{-1} C_M) = \det(A)\cdot\det(1 - A^{-1} B_M D_M^{-1} C_M) \tag{7.20}$$

Using the relation $\det(1 + PQ) = \det(1 + QP)$ this yields

$$\det(A^q) = \det(A)\cdot\det(1 - D_M^{-1} C_M A^{-1} B_M) =$$
$$= \det(A)\cdot\det(D_M - C_M A^{-1} B_M)/\det(D_M) \tag{7.21}$$

However from (7.5) we obtain $\det(D_M - D_M A^{-1} B_M) = \det(\tilde{D}_M)$, which results with (7.21) into

$$\frac{\det(A^q)}{\det(A)} = \frac{\det(\tilde{D}_M)}{\det(D_M)} \tag{7.22}$$

With $J = \{k \mid s_{rk} = 1\}$, eq(7.22) then yields

$$\frac{\det(A^q)}{\det(A^r)} = \frac{\det(\tilde{D}_M)}{\det(\tilde{D}_J)} \cdot \frac{\det(D_J)}{\det(D_M)} \tag{7.23}$$

Now assume that R_r and R_q are adjacent with a common boundary hyperplane $K_k^q = K_k^r$, i.e. $k = M - J$. Then (7.23) yields with (3.12)

$$\frac{\det(A^q)}{\det(A^r)} = \frac{\det(\tilde{D}_{J-\{k\}})}{\det(\tilde{D}_J)} \cdot \frac{\det(D_J)}{\det(D_{J-\{k\}})} = \frac{\tilde{D}_{kk}^{\,r}}{D_{kk}^{\,r}} \tag{7.24}$$

We are now able to prove the following theorem.

<u>Theorem 15</u>: Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be modelled by a minimal state-model S for which S^{-1} exists. Then f is a homeomorphism only if the proper structures of S and S^{-1} are isomorphic and the weights on all edges are positive.

Proof: Let f be a homeomorphism, which requires that $\det(A^j) \neq 0$ for each non-empty region R_j. Then each non-empty region R_j is mapped by $y = A^j x + g^j$ onto a region $\tilde{R}_j$ in Y-space, corresponding to an existing vertex in the proper structure of S^{-1}. By theorem 14 and (7.24), an edge in S^{-1} which corresponds to a proper boundary hyperplane between two regions $\tilde{R}_r$ and $\tilde{R}_q$ then has a nonzero weight $\tilde{D}^r_{kk}$. Furthermore, in Y-space there cannot exist a region $\tilde{R}_p$ adjacent to $\tilde{R}_q$ which has no corresponding region R_p in S, because this could only be the case when $\det(\tilde{A}^p) = 0$.

By (7.24), using $S = (S^{-1})^{-1}$, this would yield a zero weight on the corresponding edge in S which contradicts the minimality of S. Hence the proper structures of S and S^{-1} are isomorphic. In addition for a homeomorphism, adjacent regions in Y-space may not overlap, which results in the requirement that all weights on the edges of the proper structure of S^{-1} have to be positive as well.

Corollary: Let f be modelled by a minimal statemodel for which S^{-1} exists. Then f is a homeomorphism only if the determinants of the Jacobians A^j in the regions R_j have the same sign.

This corollary easily follows from theorem 15 and relation (7.24) and has already been proved by Fujiwawa and Kuh [15]. However the conditions of theorem 15 are stronger since two corresponding edges in S and S^{-1} with negative weight would also yield the same sign for the determinants of the adjacent regions. By theorem 15 this is not allowed for a homeomorphism.

Next we consider a PL mapping $y = f(x)$, mapping $\mathbb{R}^m$ into $\mathbb{R}^m$. For a given $\bar{y}$ there are at most finitely many solutions of the equation $\bar{y} = f(x)$. Let these solutions be given by $\bar{x}^1, \bar{x}^2, \ldots, \bar{x}^k$. With $\bar{x}^j \in R_{n_j}$, $j=1,2,\ldots,k$, the Jacobian of the mapping in each region R_{n_j} is given by A^{n_j}. Following Ohtsuki et al [27], we define the degree of the above mapping f at $\bar{y}$, as the integer

$$\deg(f,\bar{y}) = \sum_{j=1}^{k} \operatorname{sign}(\det(A^{n_j})) \tag{7.25}$$

Now, the following invariance property holds

Let $f : \mathbb{R}^m \to Q \subset \mathbb{R}^m$ and $\bar{y}^1, \bar{y}^2 \in Q$.
Then $\deg(f,\bar{y}^1) = \deg(f,\bar{y}^2)$, if $\bar{y}^1$ and $\bar{y}^2$ can be connected by a path in Q. (7.26)

We next prove the following theorem.

Theorem 16: Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be modelled by a minimal state-model S for which S^{-1} exists. Then f is a homeomorphism from $\mathbb{R}^m$ onto $\mathbb{R}^m$ if and only if

a) the proper structures of S and S^{-1} are isomorphic, connected and have only positive weights on the edges,

b) there exists a pair $\bar{x},\bar{y}$ with $\bar{y} = f(\bar{x})$ for which $|\deg(f,\bar{y})| = |\deg(f^{-1},\bar{x})| = 1$.

Proof: Necessity: the necessity is obvious from theorem 15 and the definition (7.25).
Sufficiency: When condition a) holds, the system S produces a vector y for any vector $x \in \mathbb{R}^m$ because Katzenelson's algorithm always works since only positive pivots will show up during the run of the algorithm which always allows for an increase of the parameter λ. As the same holds for S^{-1} and any $y \in \mathbb{R}^m$, the mapping is onto. Conditon b), together with (7.26) then guarantees that the mappings is also one to one because on account of a) the determinants of the Jacobian have the same sign. Thus the mapping is a homeomorphism.

To demonstrate the application of the foregoing theorems we consider an example given by Kuh [15] which is depicted in fig.7.8 and can be described by:

$$y = \begin{pmatrix} 0 & -1/\sqrt{3} \\ \sqrt{3} & 0 \end{pmatrix} x + \begin{pmatrix} -4/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & -1 & 1 \end{pmatrix} i$$

$$v = \begin{pmatrix} 0 & 1 \\ -\sqrt{3} & 1 \\ \sqrt{3} & 1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} i \qquad (7.27)$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0$$

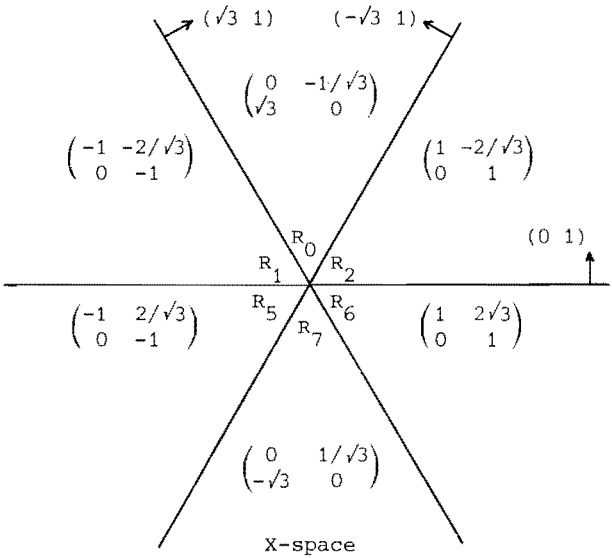


fig.7.8.

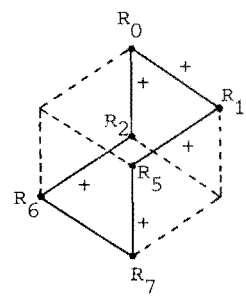


fig.7.9.

The proper structure corresponding to (7.27) is given in fig.7.9. The inverse system can be found from (7.27) and yields

$$x = \begin{pmatrix} 0 & 1/\sqrt{3} \\ -\sqrt{3} & 0 \end{pmatrix} y + \begin{pmatrix} 0 & 1/\sqrt{3} & -1/\sqrt{3} \\ -4 & 1 & 1 \end{pmatrix} i$$

$$v = \begin{pmatrix} -\sqrt{3} & 0 \\ -\sqrt{3} & -1 \\ -\sqrt{3} & 1 \end{pmatrix} y + \begin{pmatrix} -3 & 1 & 1 \\ -4 & 1 & 2 \\ -4 & 2 & 1 \end{pmatrix} i \qquad (7.28)$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0$$

The corresponding regions in Y-space and the proper structure of S^{-1} are given in fig.7.10 and 7.11 respectively.

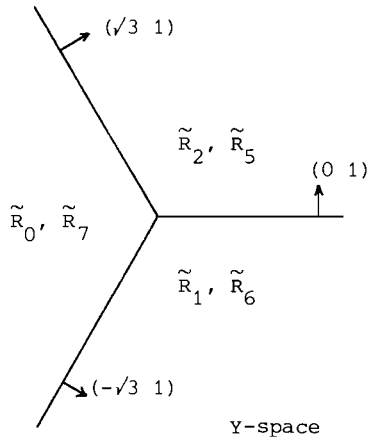


fig.7.10

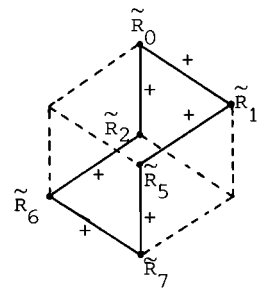


fig.7.11

By this mapping the whole X-space is mapped twice on Y-space as can be seen from fig.7.10. Hence $\forall_{\bar{y}} \in \mathbb{R}^2$ $\deg(f,\bar{y}) = 2$. Furthermore both proper structures of S and S^{-1} satisfy condition a) of theorem 16. Due to the violation of condition b) the mapping is not a homeomorphism.

The conditions in theorem 16 can be relaxed for C = (n×m) when rank(C) = n. For such a case we have the following theorem.

Theorem 17: Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be modelled by a minimal state model S for which S^{-1} exists and let rank(C) = n, for C = (n×m). Then f is a homeomorphism from $\mathbb{R}^m$ onto $\mathbb{R}^m$ if and only if $D,\tilde{D} \in P$.

Proof: Necessity: Assume that $D \notin P$. Since rank(C) = n, we can construct any vector $a \in \mathbb{R}^n$ by some particular x from $\mathbb{R}^m$. Next we pivot the system S in any possible state. Then each state s_p corresponds to a region $R_p \subset \mathbb{R}^m$ which is non-empty since we can always find a vector $\bar{x}$ such that $C^p\bar{x} + h^p = \bar{a} > 0$. As a result each row of $C^p x + h^p = 0$ has to be a proper boundary

hyperplane of R_p. Furthermore all representations S^p have to exist, otherwise we would be able to find a region R_q and a particular proper boundary hyperplane K_j^q which corresponds to a zero diagonal element of D^q, contradicting the minimality of S. Hence the proper structure of S is a full Hasse diagram with 2^n vertices in which each edge represents a proper boundary hyperplane. When $D \notin P$, at least one edge will have a negative weight and then by theorem 15 the mapping is not a homeomorphism. The same arguments hold for S^{-1}.

Sufficiency: Let $D, \tilde{D} \in P$. Then as before the proper structures of S and S^{-1} are both full Hasse-diagrams representing 2^n states with a positive weight on all edges. Thus condition a) of theorem 16 is satisfied. When $D \in P$, the solution of the state equations of S is unique for any x and thus each vector x produces only one vector y, i.e. $\deg(f^{-1},x) = \pm 1$. In the same way $\tilde{D} \in P$ will yield $\deg(f,y) = \pm 1$. Then by theorem 16 the mapping is a homeomorphism.

As an example, the mappings depicted before in fig.7.3 and 7.4 are homeomorphism on account of theorem 17.

7.3. The construction of the state-model for some simple mappings

Up to now a general construction of the state-model from a list-specified PL mapping is not known. For a relatively small number of segments or regions the state-model can readily be found by trial and error. Only for those mappings for which the regions are defined by parallel hyperplanes a general method does exist and will be discussed below. A specific example of this class of mappings is given by one-dimensional PL functions, which class will be treated first. The models of more complex mappings have to be constructed by concatenation and interrelation of these one-dimensional mappings, which is treated in chapter 8.

Let a one-dimensional continuous PL function $y = f(x)$ be defined as depicted in fig.7.12. The linear segments are defined in the intervals $[-\infty,a_1]$, $[a_1,a_2]$,..., $[a_{n-1},a_n]$, $[a_n,\infty]$ and have a slope e_k in interval $[a_{k-1},a_k]$.

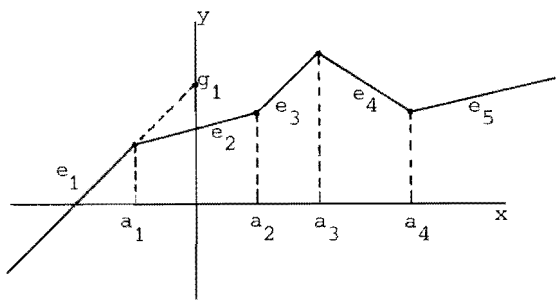


fig.7.12.

In order to find the state-model for $f(x)$ we could use the method of chapter 2.3 to construct a PL two-pole element having a v-i characteristic corresponding to $v = f(i)$, which can then be modelled by some hybrid matrix description of the obtained resistive network loaded with ideal diodes. However a faster method is obtained by considering the following state-model.

$$y = e_1 x + (b_1\ b_2\ \dots\ b_n)i + g_1$$

$$v = \begin{pmatrix} -1 \\ -1 \\ \cdot \\ \cdot \\ -1 \end{pmatrix} x + \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & & & \cdot \\ \cdot & & 1 & & \cdot \\ \cdot & & & 1 & 0 \\ 0 & \cdot & \cdot & 0 & 1 \end{pmatrix} i + \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_n \end{pmatrix} \qquad (7.29)$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0$$

with $a_1 < a_2 < \dots < a_n$

From (7.29) we easily see that the natural hybrid representation of (7.29) for $x \in [a_{k-1},a_k]$ is found by pivoting on all pivots D_{jj}

with $j < k$ since each pivot D_{jj} only affects the relation between v_j and i_j. Hence when x runs from the left to right along the x-axis, a pivot D_{kk} must be used when x crosses the boundary a_k. From (7.29) the slope e_{k+1} of $y = f(x)$ in the new interval $[a_k, a_{k+1}]$ then becomes $e_{k+1} = e_k + b_k$. Then, in order for (7.29) to be the state-model of f(x) in fig.7.12, we have to take $b_k = e_{k+1} - e_k$, $k=1,2,\ldots,n$. The value of g_1 is obtained from the representation in state s_0 when $x \le a_1$, yielding $y = e_1 x + g_1$.

Observe that the pivoting of D proceeds from top-left to bottom-right along the diagonal when x goes from left to right along the x-axis. Hence the states corresponding to existing regions are given by the state-vectors

$$(0,0,\ldots,0),\ (1,0,\ldots,0),\ (1,1,0,\ldots,0),\ \ldots,\ (1,1,1,\ldots,1)$$

All other states will yield empty regions.
Obviously the mapping $y = f(x)$ will be a homeomorphism if all slopes in fig.7.12 have the same sign, with no slope being zero. Using $e_{k+1} = e_k + b_k$, this means that (7.29) represents a homeomorphism if and only if

$$\operatorname{sign}(e_1 + \sum_{k=1}^{m} b_k) = \operatorname{sign}(e_1),\ m=1,2,\ldots,n \tag{7.30}$$

Next for a PL mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^m$ such that each region in y-space is bounded by parallel hyperplanes, the state-model takes the following form

$$y = Ax + Bi + g$$

$$v = \begin{pmatrix} -n^t \\ -n^t \\ \cdot \\ \cdot \\ -n^t \end{pmatrix} x + 1i + \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_k \end{pmatrix} \tag{7.31}$$

$$v \ge 0,\quad i \ge 0,\quad v^t i = 0$$

and $a_1 < a_2 < \ldots < a_k$, $A = (m \times m)$, $B = (m \times k)$

In (7.31) the parallel hyperplanes are given by $n^t x = a_j$, $j=1,2,\ldots,k$, with n representing the common normal vector on these hyperplanes.

Let us consider a vector $\bar{x} \in R_0$ satisfying $n^t\bar{x} < a_1$. Then R_0 is a halfspace bounded by the hyperplane $n^t x = a_1$. From $\bar{x}$ we are going to trace a line $x = \bar{x} + \lambda n$, orthogonal to all parallel hyperplanes and traversing all strip-shaped regions in x-space. Each time a boundary hyperplane is crossed, the corresponding diagonal element of the unit matrix (matrix D) is used as a pivot to obtain the new natural representation in the adjacent region, resulting in an update of the affine mapping. Let us denote the regions and hyperplanes encountered along the previously given line as given in fig.7.13.

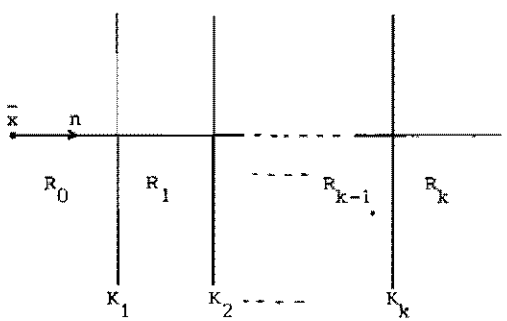


fig.7.13.

With the mapping in R_k defined by $y = A_k x + g_k$ we then find from (7.31) and the discussed process

$$A_{l+1} = A_l + B_{l+1} n^t \tag{7.32}$$

with B_l representing the l-th column of B.
Thus the matrix A and the columns of B are easily found from the various mappings in the different regions. For mappings of the above type, the following theorem holds.

Theorem 18: A continuous PL mapping $f : x \in \mathbb{R}^m \to y \in \mathbb{R}^m$ such that the regions in x-space are bounded by parallel hyperplanes is a homeomorphism if and only if the determinants of the Jacobian in all regions have the same sign.

Proof: From the foregoing arguments we know that the mapping can be modelled by the state-model of eq.(7.31). Obviously $y = f(x)$ is a homeomorphism if and only if $\tilde{y} = Mf(Nx)$ is a homeomorphism for some nonsingular pair of matrices M,N because M and N only perform linear transformations on the whole x- and y-spaces and do not affect the one to one and onto properties of the mapping. In particular we take $M = NA^{-1}$ and define $\tilde{x}$ by $\tilde{x} = Nx$, with

$$N = \begin{pmatrix} -n_1 & -n_2 & \cdot & \cdot & \cdot & -n_m \\ 0 & 1 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & 1 & \cdot & \cdot & 0 \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{pmatrix} , \tag{7.33}$$

such that $n^t N^{-1} = (-1\ 0\ 0\ \ldots\ 0)$.
With (7.31) this results in a state-model for $\tilde{y} = Mf(\tilde{x})$ given by

$$\tilde{y} = 1_m \tilde{x} + \tilde{B}i + g$$

$$v = \begin{pmatrix} -1 & 0 & 0 & \cdot & \cdot & 0 \\ -1 & 0 & 0 & \cdot & \cdot & 0 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ -1 & 0 & 0 & \cdot & \cdot & 0 \end{pmatrix} \tilde{x} + 1_k i + \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_k \end{pmatrix} \tag{7.34}$$

with $v \geq 0$, $i \geq 0$, $v^t i = 0$

In stead of the original problem we now prove that (7.34) is a homeomorphism if and only if the determinants

of the Jacobian in all regions have the same sign.
To this purpose we partition the vectors $\tilde{x}$ and $\tilde{y}$ as follows

$$\tilde{y} = \begin{pmatrix} \tilde{y}_1 \\ \hline \tilde{y}_r \end{pmatrix} \begin{matrix} \} \text{ 1 component} \\ \\ \} \text{ m-1 components} \end{matrix} \qquad \tilde{x} = \begin{pmatrix} \tilde{x}_1 \\ \hline \tilde{x}_r \end{pmatrix}$$

This partition then yields with (7.34)

$$\tilde{y}_1 = \tilde{x}_1 + \tilde{B}_1 i + g_1 \tag{7.35a}$$

$$v = \begin{pmatrix} -1 \\ -1 \\ \cdot \\ \cdot \\ -1 \end{pmatrix} \tilde{x}_1 + 1_k i + \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_k \end{pmatrix} \tag{7.35b}$$

$$\tilde{y}_r = \tilde{x}_r + \tilde{B}_r i + g_r \tag{7.35c}$$

$v \geq 0, \quad i \geq 0, \quad v^t i = 0,$

with $\tilde{B} = \begin{pmatrix} \tilde{B}_1 \\ \hline \tilde{B}_r \end{pmatrix}$ and $g = \begin{pmatrix} g_1 \\ \hline g_r \end{pmatrix}$

A comparison of (7.35a) and (7.35b) with (7.29) shows that $\tilde{y}_1$ is a one-dimensional PL function of $\tilde{x}_1$, which will be denoted by $\tilde{y}_1 = h(\tilde{x}_1)$. We next claim that (7.35a) to (7.35c) is a homeomorphism if and only if $\tilde{y}_1 = h(\tilde{x}_1)$ is a homeomorphism. Obviously we only have to prove the sufficiency of the above claim. Hence let us assume that $\tilde{y}_1 = h(\tilde{x}_1)$ is a homeomorphism. Then for each $\tilde{x}_1 \in \mathbb{R}$ we have a unique $\tilde{y}_1 \in \mathbb{R}$ and because $1_k \in P$ we also have a unique solution of the state equation (7.35b). For any $\tilde{x}_r \in \mathbb{R}^{m-1}$ we then produce by (7.35c) a unique $\tilde{y}_r \in \mathbb{R}^{m-1}$. Thus any $\tilde{x} \in \mathbb{R}^m$ will yield a unique $\tilde{y} \in \mathbb{R}^m$.

On the other hand, for each $\tilde{y}_1 \in \mathbb{R}$ we find a unique $\tilde{x} \in \mathbb{R}$ and again a unique solution of the state equations, which produces by (7.35c) a unique $\tilde{x}_r \in \mathbb{R}^{m-1}$ for any $\tilde{y}_r \in \mathbb{R}^{m-1}$. Thus any $\tilde{y} \in \mathbb{R}^m$ produces a unique $\tilde{x} \in \mathbb{R}^m$. Then the mapping is one to one and onto, i.e. it is homeomorphism.

From (7.35) with $\tilde{A}^0 = 1_m$ and (7.30) we know that $\tilde{y}_1 = h(\tilde{x}_1)$

is a homeomorphism if and only if $\operatorname{sign}(\tilde{A}^j_{11}) = \operatorname{sign}(\tilde{A}^0_{11})$ for all regions R_j with representation S_j. However from (7.34) we also observe that pivoting on $D = 1_k$ only affects the first column of $\tilde{A}_1$ which remains of the form

$$\tilde{A}^j = \left(\begin{array}{c|ccccc} \tilde{A}^j_{11} & 0 & 0 & . & . & 0 \\ \hline \tilde{A}^j_{21} & 1 & 0 & . & . & 0 \\ . & 0 & 1 & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ \tilde{A}^j_{m1} & 0 & . & . & . & 1 \end{array}\right)$$

Hence $\det(\tilde{A}^j) = \tilde{A}^j_{11}$. Then $\tilde{y}_1 = h(\tilde{x}_1)$ is a homeomorphism if and only if $\operatorname{sign}\det(\tilde{A}^j) = 1$ for all regions R_j. By our previous considerations this implies that $y = f(x)$ is a homeomorphism if and only if the determinants of the Jacobians of all regions have the same sign. End of proof.

The above theorem was also given by Chien [25]. However the proof given there relies on another theorem for which the proof seems to be doubtful; at least the proof is not as obvious as was suggested.

With the above theorem we conclude our discussions on homeomorphism properties of mappings modelled by the state-model.

8. Interconnection of PL systems and dynamic PL systems

Standard nonlinear electrical networks are in general constructed by interconnection of basic network elements or building blocks such as for example transistors, operational amplifiers, gates etc. In order to construct a piecewise-linear description of such a network in the form of an overall state-model, a method has to be supplied to assemble the required state-model from the individual state-models for the employed building blocks. Since the state-model formulation has a standard structure independent of the kind of mapping which has been modelled, such an approach then allows for a hierarchical modelling and description of electrical networks, as each state-model on its own can be considered as a macro-model. Furthermore, a combination of analog and digital circuits can easily be modelled since the state-model itself has no inherently different nature for both types of circuits and a single algorithm is used to solve the PL equations which arise in the analysis of such circuits. Thus the state-model description allows for a uniform description for different types of building blocks of diverse complexity, resulting in a uniform database and solution-algorithm as well.

In the next section we will present a method to obtain such an overall state-model for a collection of interconnected sub-systems. For reasons of simplicity each in- or output variable is considered as a "signal" passed from one block to the other, where as in practical network simulations the interconnection of two nodes will require two voltages to be set equal and the sum of two currents to be set to zero.

8.1. The state-model construction

Consider a collection of components $C^1, C^2, \ldots, C^k$ each of which has been modelled by an appropriate state-model $S^1, S^2, \ldots, S^k$ such that the corresponding PL mapping gives a sufficiently accurate description of the nonlinear static input-output relation of each component. Let the in- and output variables (signals) of component C^m

be denoted by the vectors x^m and y^m respectively. Interconnection of these components then amounts to equate the input and output variables which correspond to the particular interconnection.

For example the interconnections of the "circuit" of fig.8.1 are determined by $x_2^1 = y_2^2$ and $x_1^2 = y_2^1$. In addition we may define the overall in- and output variables by $I_1 = x_1^1$, $I_2 = x_2^2$, $O_1 = y_1^1$ and $O_2 = y_1^2$

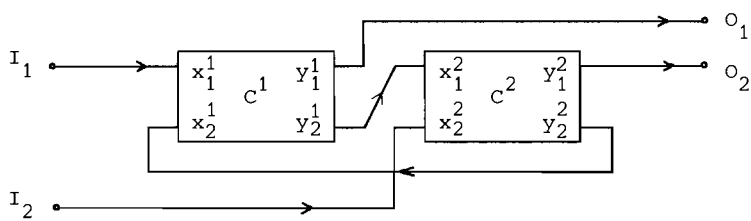


fig.8.1.

Together with the PL description of C^1 and C^2 these relations fully define the behaviour of the overall circuit with inputs I_1, I_2 and output O_1, O_2. Our task is now to obtain the state-model description of such a compound circuit.

To this purpose we shall not limit ourselves to the above type of interconnections but allow for more general relations of the form

$$\sum_{i,j} a_{ij}x_j^i + \sum_{i,j} b_{ij}y_j^i = 0 \tag{8.1}$$

Using (8.1) we also have the possibility to interconnect two inputs or two outputs. In standard electronic networks such a connection in general makes no sense. However in our modelling by the state-model the difference between input variables and output variables is more or less artificial since a pivoting step on the element A_{kl} of some state-model S exchanges y_k for x_l and hence it then appears as if y_k is an input- and x_l is an output variable.

Therefore we will consider the construction of compound models or macro-models as symbolized in fig.8.2, with the overall in- and output vectors denoted by x^0 and y^0 and the interconnection realized within the block labelled #INT and described by linear relations of the form (8.1).

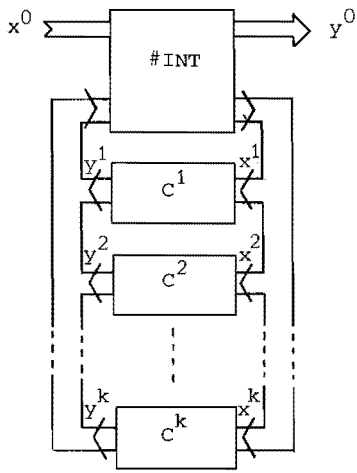


fig.8.2.

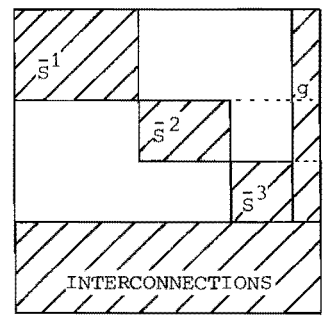


fig.8.3.

First of all we rewrite the state-model equations of the individual blocks C^m in the form

$$\bar{S}^m \cdot \begin{pmatrix} y^m \\ v^m \\ x^m \\ i^m \end{pmatrix} + g^m = 0$$

Then the coefficients of the matrices $\bar{S}^m$ and the vectors g^m are entered in a matrix $\bar{S}$ as indicated in fig.8.3 for k = 3. At the same time a pointer system is set up to memorize the exact locations of all variables in $\bar{S}$. In the next step all interconnections are processed by appending $\bar{S}$ with a number of rows representing equations of the type (8.1) for each interconnection. In fig.8.3 these rows are depicted in the lower border-block.

The overall model is now easily obtained from $\bar{S}$ by eliminating all variables of the vectors x^m and y^m which do not belong to x^0 or y^0, by using pivots from the lower border block. The remaining equations are then reordered and eventually pivoted to obtain the standard form

$$\bar{S}^0 \cdot \begin{pmatrix} y^0 \\ v^0 \\ x^0 \\ i^0 \end{pmatrix} + g^0 = 0$$

As a simple example consider the circuit of fig.8.4 which contains a saturating amplifier in C^1 and a PL feedback network in C^2, with the diode in C^2 considered to be an ideal diode.

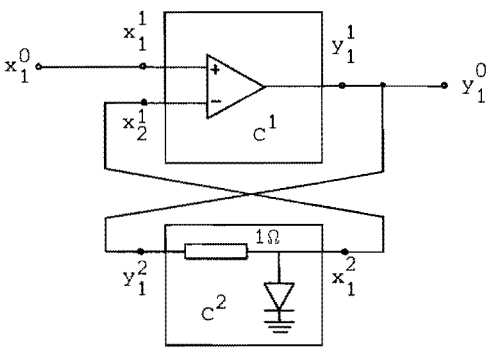


fig.8.4.

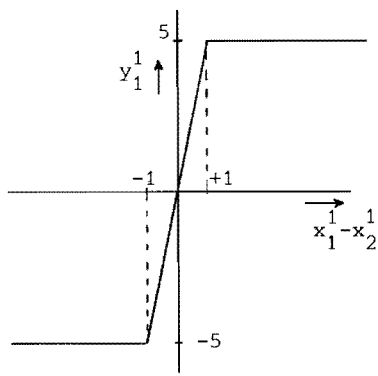


fig.8.5.

The variables in x and y represent voltages with respect to ground. Let the amplifier in C^1 have a characteristic as given in fig.8.5, for which the state-model is easily obtained as

$$y_1^1 = (5 \quad -5) \begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix} + (-5 \quad 5) \begin{pmatrix} i_1^1 \\ i_2^1 \end{pmatrix}$$

$$\begin{pmatrix} v_1^1 \\ v_2^1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1^1 \\ x_2^1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} i_1^1 \\ i_2^1 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{8.2}$$

Furthermore the state-model of C^2 is found to become

$$y_1^2 = x_1^2 + i_1^2$$
$$v_1^2 = -x_1^2 \tag{8.3}$$

With the given interconnections the matrix $\bar{S}$ now becomes

y^1	v^1		x^1		i^1		y^2	v^2	x^2	i^2	g
-1	0	0	5	-5	-5	5	0	0	0	0	0
0	-1	0	-1	1	1	0	0	0	0	0	1
0	0	-1	1	-1	0	1	0	0	0	0	1
0	0	0	0	0	0	0	-1	0	1	1	0
0	0	0	0	0	0	0	0	-1	-1	0	0
0	0	0	0	-1	0	0	0	0	1	0	0
-1	0	0	0	0	0	0	1	0	0	0	0

The first five rows of $\bar{S}$ contain the descriptions of C^1 and C^2 and the last two rows contain the interconnections. Then using row 1 and row 4, we elimintate y_1^1 and y_1^2 from the last interconnection equations yielding the following interconnection block.

0	0	0	0	-1	0	0	0	0	1	0	0
0	0	0	-5	5	5	-5	0	0	1	1	0
				x_2^1					x_1^2		

(8.4)

Next we use the two encircled columns in (8.4) to eliminate x_2^1 and x_1^2 from the upper five rows of $\bar{S}$. After deleting the fourth row since we are not interested in y_1^2 and reordering the remaining columns we obtain the overall state-model

$$y_1^0 = 5/6\ x_1^0 + (-5/6 \quad 5/6 \quad 5/6)\, i^0$$

$$v^0 = \begin{pmatrix} -1/6 \\ 1/6 \\ -5/6 \end{pmatrix} x_1^0 + \begin{pmatrix} 1/6 & 5/6 & -1/6 \\ 5/6 & 1/6 & 1/6 \\ 5/6 & -5/6 & 1/6 \end{pmatrix} i^0 + \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{8.5}$$

with $v^0 = (v_1^1, v_2^1, v_1^2)^t$, $i^0 = (i_1^1, i_2^1, i_1^2)^t$

and $v^0 \geq 0$, $i^0 \geq 0$, $v^{0t} i^0 = 0$.

The mapping presented by (8.5) is depicted in fig.8.6.

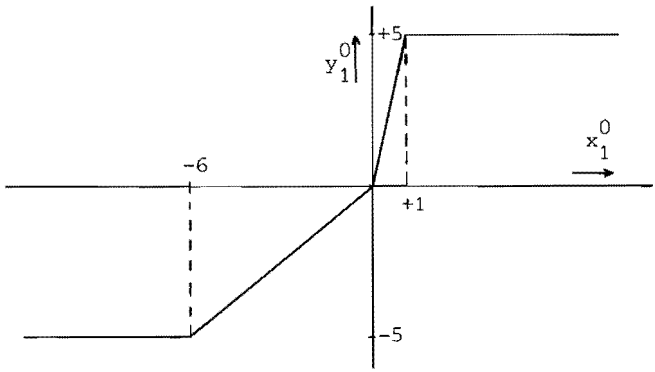


fig.8.6.

As can be seen for $x_1^0 > 0$ the input-output relation is the same as for the amplifier since in this case the diode in C^2 is conducting which prevents feedback from output to input. For $x_1^0 < 0$ the diode does not conduct and hence there is a full feedback around the amplifier yielding a gain of 5/6 until the input voltage becomes less than -6 Volt in which case the amplifier goes into saturation.

It will be clear that the number of overall state equations equals the sum of all state equations of the individual state-models of the components $C^1, C^2, \ldots, C^k$. Thus the storage complexity of the description grows linearly with the storage complexity of the components. Furthermore it will be obvious that linear components can be added as well since their algebraic input-output relations can simply be considered as a state-model description without state equations.

In the above way we are able to construct an overall state-model or macro-model for a system of interconnected PL subsystems which allows for a hierarchically structured modelling process for which the nature and the complexity of the components is immaterial.

8.2. Dynamic PL systems

Up to now the state-model was used to describe static PL mappings and could serve as a macro-model for nonlinear static electronic components. Since most electronic networks include energy storage elements like capacitors and inductors, the state-model has to be extended in such a way that these elements can be included in order to become a feasible network analysis tool. As these energy storage elements have voltage-current relationships which are expressed by means of time-derivatives of some physical quantities like charge or flux, the extended state-model must contain some variables with associated time derivatives. Rather than trying to model separate capacitors or inductors we take a more general approach by including standard linear dynamic networks that are described by a common state-space formulation of the type

$$\begin{aligned} y &= Ex + Fu \\ \dot{u} &= Gx + Hu \end{aligned} \tag{8.6}$$

Systems with the above description can be considered as linear algebraic systems with an input vector $\begin{pmatrix} x \\ \hline u \end{pmatrix}$ and an output vector $\begin{pmatrix} y \\ \hline \dot{u} \end{pmatrix}$. From this point of view, these linear systems can easily be interconnected with a static PL system using the methods presented in the previous chapter. After performing this interconnection, the final state-model description then takes the following form

$$\begin{aligned} \begin{pmatrix} y \\ \hline \dot{u} \end{pmatrix} &= A \begin{pmatrix} x \\ \hline u \end{pmatrix} + Bi + g \\ v &= C \begin{pmatrix} x \\ \hline u \end{pmatrix} + Di + f \\ v &\geq 0, \quad i \geq 0, \quad v^t i = 0 \end{aligned} \tag{8.7}$$

Relation (8.7) is symbolized as depicted in fig.8.7.

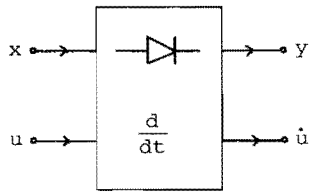
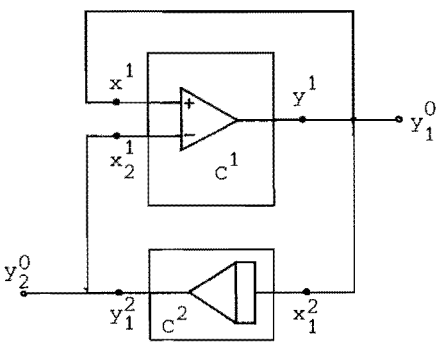


fig.8.7.

These dynamic PL state-models can now be interconnected just like the standard static PL systems.

As an example we will derive the dynamic state-model for the multivibrator circuit of fig.8.8, which consists of a saturating amplifier with a non-inverting integrator in the feedback loop.



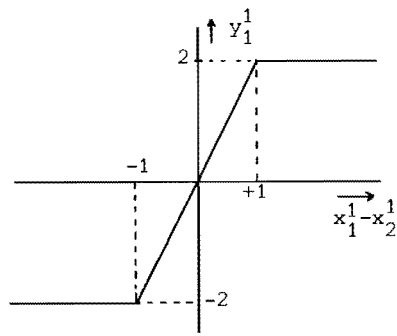


fig.8.8.

fig.8.9.

The characteristic of the saturating amplifier is depicted in fig.8.9 and yields in the shorthand notation $\begin{array}{c|c|c} A & B & g \\ \hline C & D & f \end{array}$ the following coefficients for the state-model S(A,B,C,D,g,f).

	x_1^1	x_2^1	i		g/f
y_1^1	2	-2	-2	2	0
v	-1	1	1	0	1
	1	-1	0	1	1

(8.8)

The integrator in C^2 is defined by $y_1^2 = u$, $\dot{u} = x_1^2$, resulting in the following array of coefficients.

	x_1^2	u	g/f
y_1^2	0	1	0
$\dot{u}$	1	0	0

(8.9)

Interconnection of (8.8) and (8.9) using $x_2^1 = y_1^2$, $x_1^1 = y_1^1$ and $y_1^1 = x_1^2$ then yields after elimination of x_1^1, x_2^1 and x_1^2 the following coefficient array for the overall dynamic state-model.

	u	i		g/f
y_1^0	2	2	-2	0
y_2^0	1	0	0	0
$\dot{u}$	2	2	-2	0
v	-1	-1	2	1
	1	2	-1	1

(8.10)

The above model describes an autonomous system since no input vector x is left in the model (8.10).

For a final transient analysis of PL dynamic systems described by (8.7), a relation between $\dot{u}$ and u has to be added to the model equations. Like in the case of transient analysis of standard nonlinear networks, an integration formula is applied to eliminate the derivatives. This results in a pure algebraic PL system describing the behaviour of the dynamic system at discrete time instants.

Because the differential equations originating from descriptions of electronic networks are in general stiff equations, the applied integration formula should at least be stiffly stable. Furthermore it is of advantage to use linear multistep formulas since methods are available to implement these formulas with automatic order and stepsize control for a good compromise between accuracy, stability and efficiency [28].

Because we are dealing with piecewise-linear systems for which the first derivatives will be non-continuous, it is obvious that integration methods of order higher than two do not improve, the integration process since these formulas rely on the property that locally a sufficient accurate Taylor-series expansion of the derivatives must exist [29]. Thus maximum accuracy is obtained with second order linear multistep methods like for example the trapezoidal rule. For simplicity, without loss of generality we will use the first order backward Euler integration formula with fixed stepsize h. To this purpose let us consider the system (8.7) at time instant $t_{n+1} = t_n + h$, which is denoted by adding the subscript n+1 to each variable. Partitioning A, B, C and g as

$$\left(\begin{array}{c|c} A_1 & A_2 \\ \hline A_3 & A_4 \end{array}\right), \quad \left(\begin{array}{c} B_1 \\ \hline B_2 \end{array}\right), \quad (C_1 \mid C_2), \text{ and } \left(\begin{array}{c} g_1 \\ \hline g_2 \end{array}\right),$$

we have from (8.7)

$$y_{n+1} = A_1 x_{n+1} + A_2 u_{n+1} + B_1 i + g_1 \tag{8.11a}$$

$$\dot{u}_{n+1} = A_3 x_{n+1} + A_4 u_{n+1} + B_2 i + g_2 \tag{8.11b}$$

$$v = C_1 x_{n+1} + C_2 u_{n+1} + Di + f \tag{8.11c}$$

$$v \geq 0, \quad i \geq 0, \quad v^t i = 0$$

The backward Euler formula is given by

$$u_{n+1} = u_n + h\dot{u}_{n+1}, \tag{8.12}$$

which introduces a discretization error of $O(h^2)$.
Using (8.11b) and (8.12) we find

$$u_n = -hA_3 x_{n+1} + (1 - hA_4) u_{n+1} - hB_2 i - hg_2,$$

which equation is used to replace (8.11b) in the above state-model, leading to

$$
\begin{aligned}
y_{n+1} &= A_1 x_{n+1} + A_2 u_{n+1} + B_1 i + g_1 \\
u_n &= -hA_3 x_{n+1} + (1 - hA_4) u_{n+1} - hB_2 i - hg_1 \\
v &= C_1 x_{n+1} + C_2 u_{n+1} + Di + f \\
v \geq 0, \quad & i \geq 0, \quad v^t i = 0
\end{aligned}
\tag{8.13}
$$

Equation (8.13) now describes a PL mapping $\psi : (x_{n+1}, u_{n+1}) \rightarrow (y_{n+1}, u_n)$. Pivoting (8.13) on the full matrix $(1 - hA_4)$, exchanges the variables u_{n+1} and u_n yielding the final mapping $\bar{\psi} : (x_{n+1}, u_n) \rightarrow (y_{n+1}, u_{n+1})$, denoted by

$$
\begin{aligned}
\begin{pmatrix} y_{n+1} \\ \hdashline u_{n+1} \end{pmatrix} &= \bar{A} \begin{pmatrix} x_{n+1} \\ \hdashline u_n \end{pmatrix} + \bar{B} i + \bar{g} \\
v &= \bar{C} \begin{pmatrix} x_{n+1} \\ \hdashline u_n \end{pmatrix} + \bar{D} i + \bar{f} \\
v \geq 0, \quad & i \geq 0, \quad v^t i = 0
\end{aligned}
\tag{8.14}
$$

Equation (8.14) can be seen as the overall state-model of the compound system given in fig.8.10, with the lower box representing the backward Euler formula.

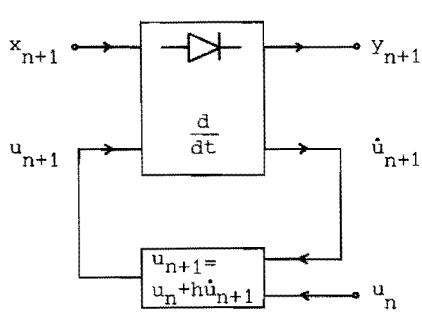


fig.8.10.

From (8.14), for prescribed x_{n+1} and u_n, the response y_{n+1} and u_{n+1} can be found by using one of the previously discussed solution methods for the LCP problem. In fact eq. (8.14) represents the PL analogon of the transition matrix description of nonlinear dynamic networks, which allows us to proceed by one time-step. Systems modelled by (8.14) can again be interconnected by the methods of section 8.1, allowing for the construction of models for larger systems. Observe however that the time-step h has been fixed by the transformation from (8.7) to (8.14). For maximum flexibility it is advisable to perform the latter transformation only on the final model.

Let us now continue our example of the multivibrator given by (8.10). With h = 0.1 we obtain in analogy with (8.13)

	u_{n+1}	i		g/f
y_1^0	2	2	-2	0
y_2^0	1	0	0	0
u_n	0.8	-0.2	0.2	0
v	-1	-1	2	1
	1	2	-1	1

Pivoting on the indicated pivot then yields the final model

	u_n	i		g/f
y_1^0	2.5	2.5	-2.5	0
y_2^0	1.25	0.25	-0.25	0
u_{n+1}	1.25	0.25	-0.25	0
v	-1.25	-1.25	2.25	1
	1.25	2.25	-1.25	1

(8.15)

The response calculated by (8.15) using $u_0 = 0.9$ is depicted in fig.8.11.

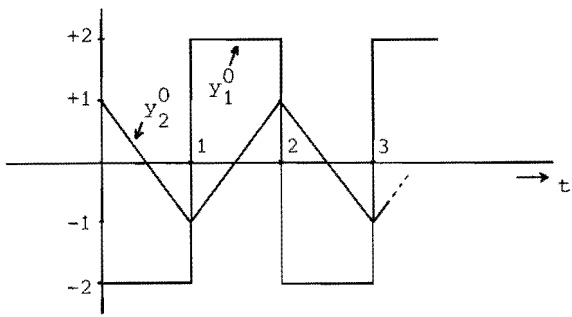


fig.8.11.

Nonlinear energy storage elements like e.g. nonlinear capacitors or inductors can easily be included in the state-model by constructing a PL approximation of their charge-voltage or flux-current characteristic.

With the above tools, common electronic devices can be modelled by the state-model. In the next chapter some examples will be given to aid the implementation of these models in an analysis package and to demonstrate the general line in the modelling proces.

9. Piecewise-linear models for some electronic devices

In this section we present PL models for a voltage controlled switch, a MOS transistor and some digital gates. In addition we will give the results of the PL analysis of some small circuits as an indication of the applicability of the PL approach.

A voltage controlled switch can be realized by the network given in fig.9.1, with the diodes considered to be ideal.

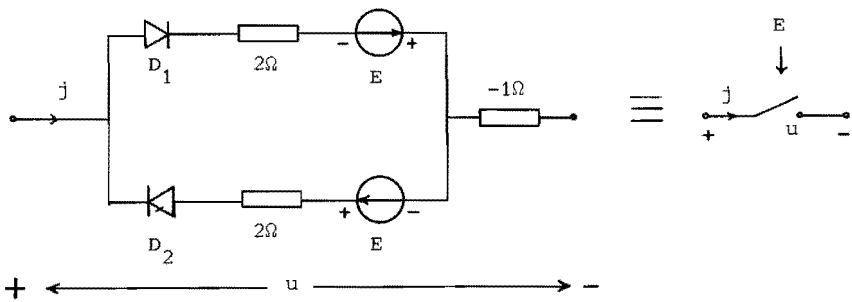


fig.9.1.

For $E > 0$, both diodes will conduct as long as $|j| \leq E$ and the circuit is then equivalent to a shortcircuit. For $E < 0$, both diodes do not conduct as long as $|u| \leq -E$ and the circuit is an open circuit. From fig.9.1 the state-model is readily found to become

	u	E	i_1	i_2	
j	0	0	1	-1	0
v_1	-1	-1	1	1	0
v_2	1	-1	1	1	0

(9.1)

The mapping corresponding to (9.1) is given in fig.9.2 and indeed represents a voltage controlled switch as long as the electrical parameters u and j remain within the above given bounds.

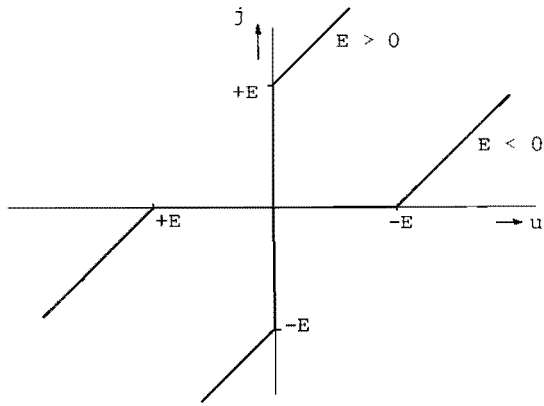


fig.9.2.

The PL model for the voltage controlled switch can for example be used in modelling switched capacitor filters.

A model for a MOS transistor can be obtained as follows. We start from a symmetric model for a standard MOS transistor [30], which is depicted in fig.9.3.

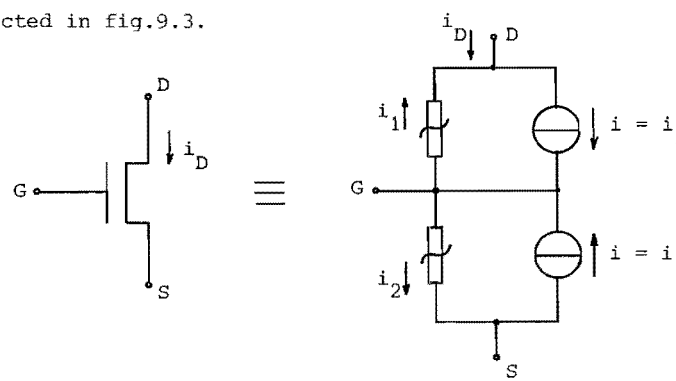


fig.9.3.

The nonlinear resistive elements in the above model are defined by

$$i_1 = f(V_{GD}, V_T) \quad \text{and} \quad i_2 = f(V_{GS}, V_T),$$

with V_T representing the threshold voltage. In the simplest approximation f is given by

$$f(v_1,v_2) = K[v_1-v_2]_+^2 \,,$$
$$\text{with } [x]_+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{9.2}$$

For some $v_2 = V_T$ this characteristic is sketched in fig.9.4 and in analogy with the Ebers-Moll model for bipolar transistors it can be seen as the voltage-current characteristic of a "MOS diode".

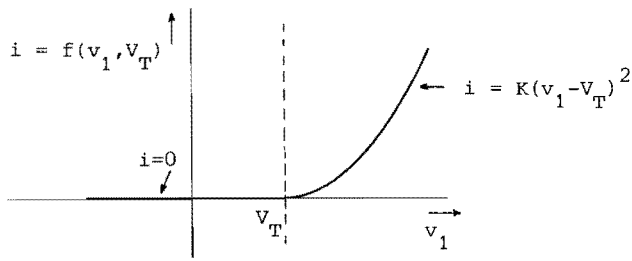


fig.9.4.

Then the overall characteristic of the MOS transistor can apparently be constructed from a PL approximation of the one-diomensional mapping of fig.9.4. To this purpose we consider a PL Tchebychev-like approximation $\bar{y}(x)$ of the mapping $y = [x]_+^2$ as indicated in fig.9.5 and fig.9.6.

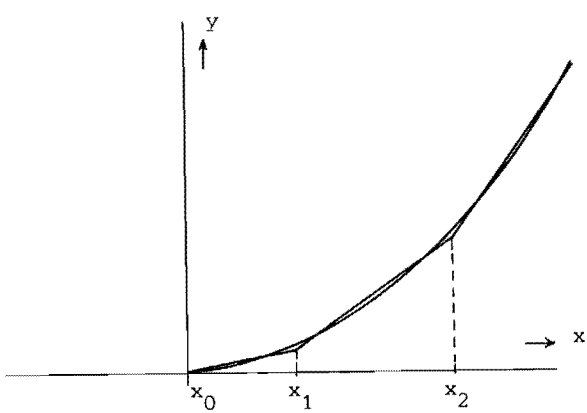


fig.9.5.

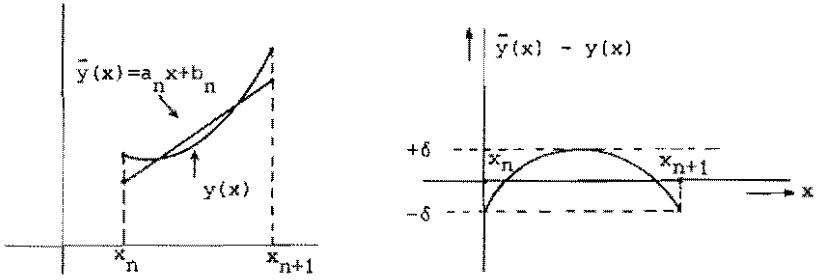


fig.9.6.

For a given x_n and δ we easily derive that the approximation in the next interval is given by

$$\bar{y}(x) = a_n x + b_n, \quad x_n \le x \le x_{n+1},$$

with

$$a_n = 2x_n + \varepsilon$$

$$x_{n+1} = x_n + \varepsilon$$

$$b_{n+1} = \delta - a_n^2/4 \quad \text{and } \varepsilon = \sqrt{8\delta}$$

For the first interval we find $a_0 = 2\sqrt{\delta}$, $x_1 = 1/2\ (a_0+\varepsilon)$.
For $\delta = 0.16$ this yields in two digits accuracy

$$\varepsilon = 1{,}13, \quad a_0 = 0.8, \quad x_1 = 0.97, \quad a_{n+1}-a_n = 2.26, \quad x_{n+1}-x_n = 1.13$$

and results in a state-model given by

	x	i_1	i_2	i_3	i_4	i_5	i_6	
y	0.8	0.8	2.26	2.26	2.26	2.26	2.26	0
v_1	1	1						0
v_2	-1							0.97
v_3	-1							2.10
v_4	-1							3.23
v_5	-1							4.36
v_6	-1						1	5.49

(9.3)

Eq.(9.3) is a PL approximation of $y = [x]_+^2$ in 7 segments with an absolute error $\delta \le 0.16$ for $x \le 6.62$.

Next the full MOS transistor model can be constructed by interconnection of the blocks given in fig.9.7.

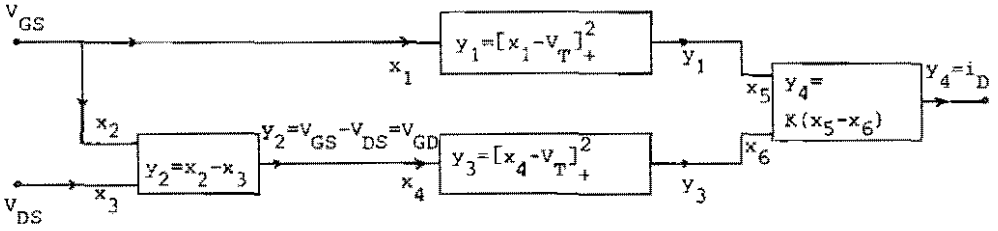


fig.9.7.

For $K = 1A/V^2$ and $V_T = 2V$ the interconnection method of chapter 8 yields with (9.3) the following state-model (9.4)

	V_{GS}	V_{DS}	i_1	i_2	i_3	i_4	i_5	i_6	i_7	i_8	i_9	i_{10}	i_{11}	i_{12}	
I_D	0	0.8	0.8	2.26	2.26	2.26	2.26	2.26	-0.8	-2.26	-2.26	-2.26	-2.26	-2.26	0
v_1	1	0	1												-2
v_2	-1	0		1											2.97
v_3	-1	0			\										4.10
v_4	-1	0				\									5.23
v_5	-1	0					\								6.36
v_6	-1	0						\							7.49
v_7	1	-1							\						-2
v_8	-1	1								\					2.97
v_9	-1	1									\				4.10
v_{10}	-1	1										\			5.23
v_{11}	-1	1											\		6.36
v_{12}	-1	1												1	7.49

An indication of the accuracy of the model (9.4) is given by the table of fig.9.8 which compares the current i_{DPL} obtained from (9.4) with the value i_{DN} obtained from the MOS model using (9.2)

V_{GS}	V_{DS}	i_{DPL}	i_{DN}
2	*	0	0
3	0.5	0.48	0.75
3	1	0.88	1
3	2	0.88	1
4	0.5	1.53	1.75
4	1	3.06	3
4	1.5	3.54	3.75
4	2	3.94	4
4	3	3.94	4
5	1	5.11	5
5	2	8.17	8
5	3	9.05	9
5	4	9.05	9
6	1	7.02	7
6	2	12.2	12
6	3	15.2	15
6	4	16.1	16

fig.9.8.

In practice $f(v_1,v_2)$ differs from (9.2) but can always be modelled as long as it remains a one-dimensional function of v_1-v_2. Furthermore, the dependence of the threshold voltage v_2 of the source to substrate voltage can also be modelled by a PL model and included in the above model (9.4).

Next we will derive a model for a digital nand-gate. The model is derived from the circuit given in fig.9.9, using ideal diodes and a 3-segment ideal invertor.

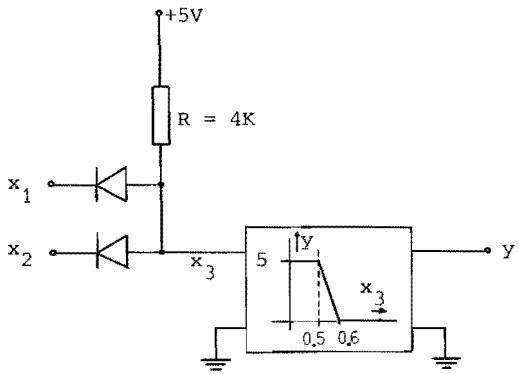


fig.9.9.

From this circuit we observe that x_3 is defined by

$$x_3 = \min(x_1, x_2)$$

which can be modelled by the mapping depicted in fig.9.10.

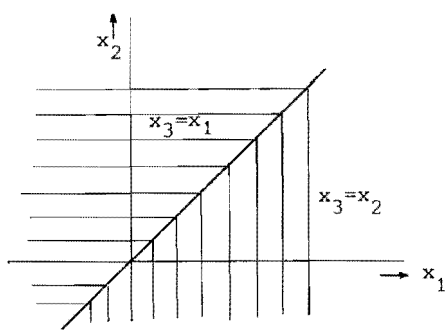


fig.9.10.

The state-model of this mapping becomes

	x_1	x_2	i_1
x_3	0	1	-1
v_1	1	-1	1

(9.5)

Furthermore, the invertor is modelled by

	x_3	i_2	i_3	
y	-50	-50	50	30
v_2	1	1	0	-0.5
v_3	- 1	0	1	0.6

(9.6)

The overall model for the circuit of fig.9.9 using (9.5) and (9.6) then becomes

	x_1	x_2	i_1	i_2	i_3	
y	0	-50	50	-50	50	30
v_1	1	- 1	1	0	0	0
v_2	0	1	- 1	1	0	-0.5
v_3	0	- 1	1	0	1	0.6

(9.7)

The corresponding mapping is depicted in fig.9.11. Note the non-convex region in which y = 5, which is due to an unobservable state coupled to the "boundary" $x_1 = x_2$.

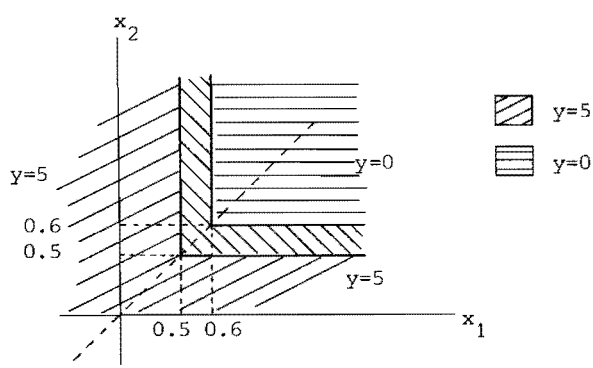


fig.9.11.

Next let us consider a model for the so-called threshold gates which can be used to realize various types of digital circuit blocks like e.g. flip-flops or ring-counters and lend themselves particularly well for a PL modelling. These gates are symbolized as given in fig.9.12 and their output is defined by

$$y = \begin{cases} 1 & \text{if} \quad a_1x_1+a_2x_2+\ldots+a_nx_n \geq t_1 \\ 0 & \text{if} \quad a_1x_1+a_2x_2+\ldots+a_nx_n \leq t_2 \end{cases} \tag{9.8}$$

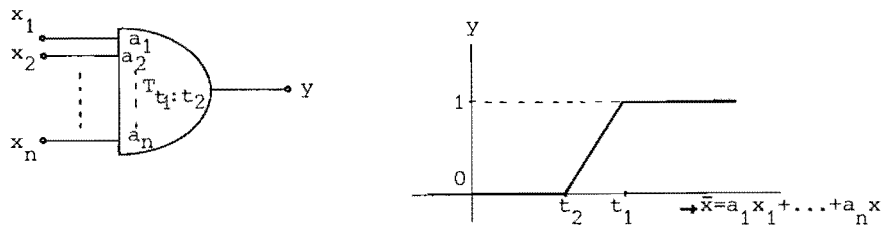


fig.9.12.

From (9.8) the general state-model is easily derived to become

	x_1	x_2	$\cdots$	x_n	i_1	i_2	
y	$\frac{a_1}{t_1-t_2}$	$\frac{a_2}{t_1-t_2}$	$\cdots$	$\frac{a_n}{t_1-t_2}$	$\frac{1}{t_1-t_2}$	$\frac{-1}{t_1-t_2}$	$\frac{-t_2}{t_1-t_2}$
v_1	a_1	a_2	$\cdots$	a_n	1	0	$-t_2$
v_2	$-a_1$	$-a_2$	$\cdots$	$-a_n$	0	1	t_1

(9.9)

Using (9.9) a set-reset flip-flop could for example be modelled from the circuit given in fig.9.13.

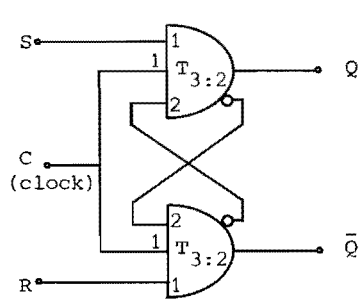


fig.9.13.

We will conclude by presenting the results obtained by calculating the response of two small circuits which have been modelled by the state-model approach. The first circuit is depicted in fig.9.14 and forms a Wien-bridge oscillator, stabilized by an amplitude control-loop using a MOS transistor.

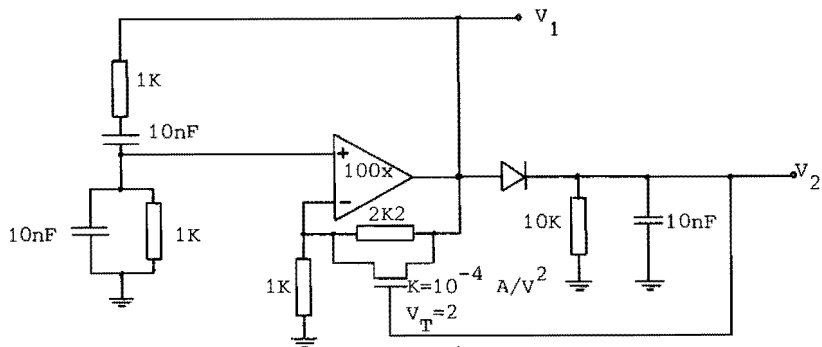
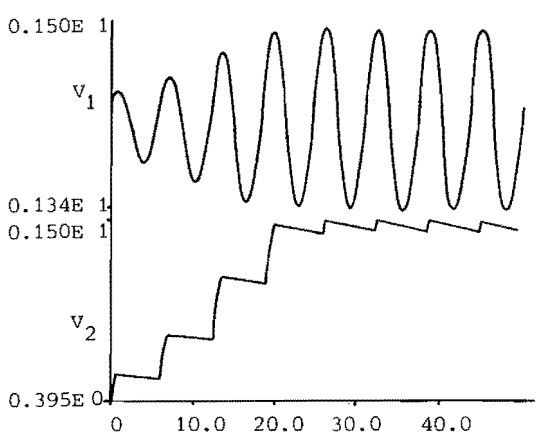


fig.9.14.

Transient analysis of the above circuit produced the response given in fig.9.15. As can be seen the amplitude becomes stationary in about 4 cycles.



The second example is given by a PL van der Pol oscillator as depicted in fig.9.16.

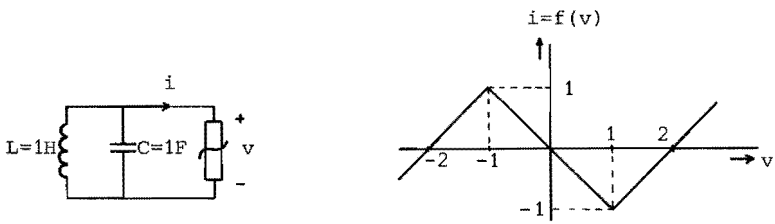


fig.9.16.

The nonlinear resistive element shows a negative resistance in the range $|v| \leq 1$. The response is given in fig.9.17 and shows a highly nonlinear current flow through the PL resistor, producing a large third harmonic current component.

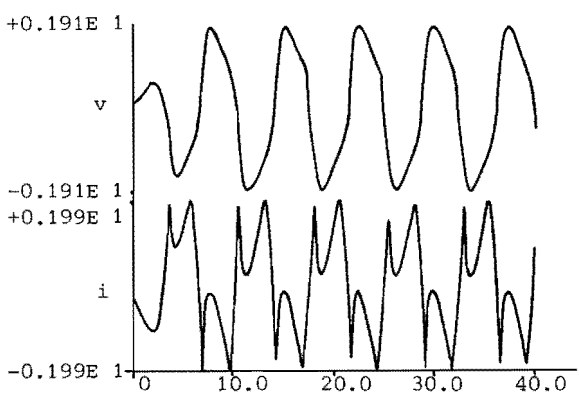


fig.9.17.

More complex circuits like dynamic 4-bit shift registers and phase-locked loops have also been analysed and gave a response in agreement with the theory. Due to the non-sparse implementation, the complexity of the circuits which could be modelled was limited such that the overall model had to fit into a 42×42 matrix. A sparse

implementation will eliminate this constraint and an adequate package based on the state-model approach can deliver the designer of electronic circuits a powerful tool for a mixed level analysis of complex combined analog-digital networks.

Appendix I

Consider a hybrid structure for the matrix $\tilde{H} = H|I$, derived from H. We are going to map this structure on the structure of H by mapping the vertices of $\tilde{H}$ onto those of H according to

$$\psi\colon \det((H|I)_J) \to \det(H_{I-J}), \tag{A1}$$

with $I-J = (I \cup J) \setminus (I \cap J)$.

Then we have:

For any edge in the structure of $\tilde{H}$ there is a corresponding edge in the structure of H with the same weight.

Proof: Consider the vertices $\det(\tilde{H}_J)$ and $\det(\tilde{H}_{J-\{k\}})$ in the structure of $\tilde{H}$. The mapping ψ from (A1) maps these vertices on $\det(H_{I-J})$ and $\det(H_{I-J-\{k\}})$ of the structure of H (See fig.A1).

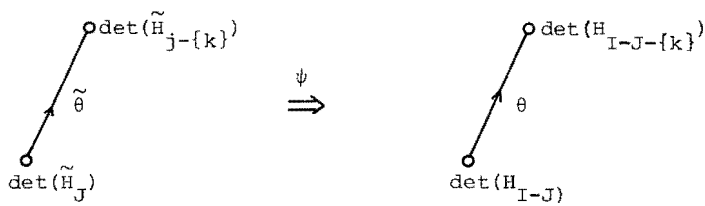


fig.A1.

Hence adjacent vertices remain adjacent under this mapping. Next consider the weights $\tilde{\theta}$ and θ in fig.A1. Using (3.12) we have

$$\tilde{\theta} = \frac{\det(\tilde{H}_{J-\{k\}})}{\det(\tilde{H}_J)} = (\tilde{H}|J)_{kk} \tag{A2}$$

However $(\tilde{H}|J)_{kk} = ((H|I)|J)_{kk} = (H|I-J)_{kk}$, due to the definition of $(H|I)$. Furthermore, we have

$$\theta = \frac{\det(H_{I-J-\{k\}})}{\det(H_{I-J})} = (H|I-J)_{kk}$$

and thus $\tilde{\theta} = \theta$. End of proof.

Hence the structures of $\tilde{H}$ and H are isomorphic and the weights on the edges are invariant under the mapping ψ.
Next we consider a path $\tilde{P}$ in the structure of $\tilde{H}$ from $\tilde{H}_\phi$ to $\tilde{H}_J$.
Then holds

$$\frac{\det(\tilde{H}_J)}{\det(\tilde{H}_\phi)} = \frac{\det((H|I)_J)}{\det((H|I)_\phi)} = \Pi \quad \text{(all weights on the edges of the path } \tilde{P}) \tag{A3}$$

From the corresponding path P in the structure of H and the invariancy of the weights we then obtain using (A1) and (A3)

$$\frac{\det(\tilde{H}_J)}{\det(\tilde{H}\ )} = \Pi \quad \text{(all weights on the edges of the path P)} = \\ = \frac{\det(H_{I-J})}{\det(H_I)} \tag{A4}$$

Because by definition $\det(\tilde{H}_\phi) = 1$, equ.(A4) yields

$$\det((H|I)_J) = \det(\tilde{H}_J) = \frac{\det(H_{I-J})}{\det(H_I)}$$

Appendix II

Let us assume that the reduced problem in the jth cycle is obtained from M by exchanging the variables $i_{l_1}, i_{l_2}, \ldots, i_{l_k}$ with the corresponding $v_{l_1}, v_{l_2}, \ldots, v_{l_k}$ $(k \leq j)$. By a reordering we may arrange that i_{l_1} to i_{l_k} form the first k components of i. Next partition the reordered equations as

$$\begin{pmatrix} v_1 \\ \cdot\cdot \\ v_2 \end{pmatrix} = \left(\begin{array}{c:c} M_1 & M_2 \\ \hdashline M_3 & M_4 \end{array}\right) \begin{pmatrix} i_1 \\ \cdot\cdot \\ i_2 \end{pmatrix} + \begin{pmatrix} a_1 \\ \cdot\cdot \\ a_2 \end{pmatrix}, \quad M_1 = (k \times k), \quad M_2 = M_3^t \tag{B1}$$

From (B1) we derive the equivalent formulation

$$\begin{pmatrix} i_1 \\ \cdot\cdot \\ v_2 \end{pmatrix} = \left(\begin{array}{c:c} M_1^{-1} & -M_1^{-1}M_2 \\ \hdashline M_3M_1^{-1} & M_4 - M_3M_1^{-1}M_2 \end{array}\right) \begin{pmatrix} v_1 \\ \cdot\cdot \\ i_2 \end{pmatrix} + \begin{pmatrix} -M_1^{-1}a_1 \\ \cdots\cdots \\ a_2 - M_3M_1^{-1}a_1 \end{pmatrix} \tag{B2}$$

In the sequel we use the notation $A \subset B$ to denote that A is a principal submatrix of B.
The reduced matrix M_j in the jth cycle is now a principal submatrix of $\bar{M} = M_4 - M_3M_1^{-1}M_2$ since at this state $v_1 = 0$ and (j-k) components of i_2 are zero, hence

$$M_j \subset \bar{M} = M_4 - M_3M_1^{-1}M_2 = \bar{M}^t \tag{B3}$$

Next, from (B1) consider the matrix M^{-1}, which is easily found to become

$$M^{-1} = \begin{pmatrix} (M_1 - M_2M_4^{-1}M_3)^{-1} & . & -M_1^{-1}M_2(M_4 - M_3M_1^{-1}M_2)^{-1} \\ \cdots & \cdots & \cdots \\ -M_4^{-1}M_3(M_1 - M_2M_4^{-1}M_3)^{-1} & . & (M_4 - M_3M_1^{-1}M_2)^{-1} \end{pmatrix} =$$

$$= \begin{pmatrix} \tilde{M}_1 & \tilde{M}_2 \\ \tilde{M}_3 & \tilde{M}_4 \end{pmatrix} \tag{B4}$$

Note that M^{-1}, M_1^{-1} and M_4^{-1} exist since M is positive definte.

Then from (B3) and (B4) we have

$$M_j \subset \bar{M} = \tilde{M}_4^{-1} \ , \quad \text{with } \tilde{M}_4 \subset M^{-1} \tag{B5}$$

Now, due to (B4), $\tilde{M}_4$, $\tilde{M}_4^{-1}$ and M_j are symmetric and positive definite since M^{-1} is symmetric and positive definite. Therefore all eigenvalues of M_j, $\tilde{M}_4^{-1}$ and $\tilde{M}_4$ are positive real. For a symmetric positive definite matrix A let us denote the samlles eigenvalue by $\sigma(A)$ and the largest eigenvalue by $\rho(A)$. Furthermore the interval $[\sigma(A),\rho(A)]$ is denoted by $R(A)$. Then since $\tilde{M}_4 \subset M^{-1}$, by the Sturmian separation theorem [31], we have

$$R(\tilde{M}_4) \subset R(M^{-1}), \text{ i.e. } \sigma(\tilde{M}_4) \geq \sigma(M^{-1}), \ \rho(\tilde{M}_4) \leq \rho(M^{-1}) \tag{B6}$$

Then for $\bar{M} = \tilde{M}_4^{-1}$ we have, due to (B6)

$$R(\bar{M}) = \left[\frac{1}{\rho(\tilde{M}_4)}, \frac{1}{\sigma(\tilde{M}_4)}\right] \subset \left[\frac{1}{\rho(M^{-1})}, \frac{1}{\sigma(M^{-1})}\right] \equiv \left[\sigma(M), \rho(M)\right] \tag{B7}$$

or $R(\bar{M}) \subset R(M)$

By the same Sutrm theorem, we have from (Bt) and (B7)

$$R(M_j) \subset R(\bar{M}) \subset R(M) \tag{B8}$$

Hence the eigenvalues $\lambda_i(M_j)$ of the matrix M_j from the jth cycle satisfy

$$\sigma(M) \leq \lambda_i(M_j) \leq \rho(M) \tag{B9}$$

Then the corresponding eigenvalues of μ_i of $D_j = (1+M_j)^{-1}(1-M_j)$ satisfy

$$\mu_i(D_j) = \frac{1-\lambda_i(M_j)}{1+\lambda_i(M_j)} \tag{B10}$$

hence

$$|\mu_i(D_j)| \le \max \left(\left|\frac{\sigma(M)-1}{\sigma(M)+1}\right| , \left|\frac{\rho(M)-1}{\rho(M)+1}\right| \right) = \theta_M < 1 \qquad \text{(B11)}$$

Then since D_j is symmetric, we have

$$\|D_j\|_2 = \max_i |\mu(D_j)| \le \theta_M < 1$$

or $\theta(M_j) < 1$

When the matrix M is scaled in advance by a factor α, we have from (B11)

$$\sigma(M_j) \le \max \left(\left|\frac{\alpha\sigma(M)-1}{\alpha\sigma(M)+1}\right| , \left|\frac{\alpha\rho(M)-1}{\alpha\rho(M)+1}\right| \right)$$

For $\alpha = (\sigma(M)\rho(M))^{-1/2}$ this yields

$$\theta(M_j) \le \theta_M = \frac{\rho(M)^{1/2}-\sigma(M)^{1/2}}{\rho(M)^{1/2}+\sigma(M)^{1/2}} = \frac{1-d}{1+d} ,$$

with $d = \sqrt{\frac{\sigma(M)}{\rho(M)}} = K^{-1/2}$ and K being the condition number of M.

Then $1/\ln(1/\theta_M) = \frac{1}{\ln\frac{1+d}{1-d}} \le \frac{1}{2d} = \frac{1}{2}K^{1/2}$.

The values of $\sigma(M)$ and $\rho(M)$ can also be bound by

$$\rho(M) \le \|M\|_\infty \quad \text{and} \quad \sigma(M) \ge \frac{1}{\|M^{-1}\|_\infty} \qquad \text{(B12)}$$

From (B9), (B10) and (B12) it is then easily found that

$$\|D_j\|_2 \le \bar{\theta} = \max \left(\left|\frac{\|M^{-1}\|_\infty -1}{\|M^{-1}\|_\infty +1}\right| , \left|\frac{\|M\|_\infty -1}{\|M\|_\infty +1}\right| \right) < 1$$

which yields an easy computable bound for all $\|D_j\|_2$.

References

[1] J.M. Orthega and W.C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, New York, Academic Press, 1970.

[2] I.W Sandberg and A.N. Willson, "Some theorems on properties of DC equations of nonlinear networks", Bell Syst. Tech. J., Vol. 48, January 1969, pp. 1-34.

[3] P.M. Lin, "Formulation of hybrid matrices for linear multiports containing controlled sources", IEEE Trans. on Circ. and Syst., Vol. CAS-21, March 1974, pp. 169-175.

[4] J. Katzenelson, "An algorithm for solving nonlinear resistive networks", Bell Syst. Tech. J., Vol. 44, October 1965, pp. 1605-1620.

[5] W.M.G. van Bokhoven and J.A.G. Jess, "Some new aspects of P and P_0 matrices and their application to networks with ideal diodes", Proc. IEEE ISCAS, 1978, pp. 806-810.

[6] G. Verkroost, "On the existence and uniqueness of solutions for networks containing linear memoryless elements and ideal diodes", Proc. of IEE Europ. Conf. on Circ. Th., IEE 116, 1974.

[7] C.E. Lemke, "On complementary pivot theory", in: Mathematics of the decision sciences, Part I, Eds. G.B Dantzig and A.F. Veinott, Jr., New York, Academic Press, 1970.

[8] M. Fiedler and V. Pták, "Some generalisations of positive definiteness and monotonicity", Num. Math., Vol. 9, December 1966, pp. 163-172.

[9] V. Belevitch, "The number of states of rectifier networks", IRE Trans. on Circ. Th., vol. CT-9, 1962, pp. 93-94.

[10] H. Samuelson et al., "A partition theorem for Euclidean n-space", Proc. Amer. Math. Soc., Vol. 9, 1958, pp. 805-807.

[11] *Complementarity and fixed point problems*. Mathematical Programming Study 7, Eds. M.L. Balinski and R.W. Cottle, North-Holland Publ. Comp., Amsterdam, 1978.

[12] L.T. Watson, "Solving the nonlinear complementarity problem by a homotopy method", SIAM J. of Contr. and Opt., Vol. 17, No. 1, 1979, pp. 36-46.

[13] C.W. Cryer, "The solution of a quadratic programming problem using systematic overrelaxation", SIAM J. of Contr., Vol. 9, No. 3, 1971, p. 385.

[14] M.L. Fisher and F.J. Gould, "A simplicial algorithm for the nonlinear complementarity problem", Math. Progr., Vol. 6, 1974, pp. 281-300.

[15] T. Fujisawa and E.S. Kuh, "Piecewise-linear theory of nonlinear networks", SIAM J. Appl. Math., Vol. 22, No. 2, March 1972.

[16] M.J. Chien and E.S. Kuh, "Solving piecewise-linear equations for resistive networks", Int. Jrn. Circ. Th. and Appl., Vol. 4, 1976, pp. 3-24.

[17] R.W. Cottle and G.B. Dantzig, "Complementary pivot theory of mathematical programming", Lin. Algebra and its Appl., Vol. 1, 1968, pp. 103-125.

[18] S. Karamardian, "The complementarity problem", Math. Progr., Vol. 2, 1972, pp. 107-129.

[19] Y. Fathi, "Computational complexity of LCPs associated with positive definite symmetric matrices", Math. Progr. Vol. 17, 1979, pp. 335-344.

[20] K.G. Murty , "Computational complexity of complementary pivot methods", in: Mathematical Programming Study 7, Eds. M.L. Balinski and R.W. Cottle, North-Holland Publ. Comp., Amsterdam, 1978.

[21] S.M. Kang and L.O. Chua, "A global representation of multi-dimensional piecewise-linear functions with linear partitions", IEEE Trans. on Circ. and Syst., Vol. CAS-25, November 1978, pp. 938-940.

[22] O.L. Mangasarian, "Equivalence of the complementarity problem to a system of nonlinear equations", SIAM J. Appl. Math., Vol. 31, No. 1, July 1976, pp. 89-91.

[23] N. Megiddo and M. Kojima, "On the existence and uniqueness of solutions in nonlinear complementarity theory" , Math. Progr., Vol. 12, 1977, pp. 110-130.

[24] L.B. Rall, *Computational solution of nonlinear operator equations*, J. Wiley & Sons Inc., New York, 1969, p. 65.

[25] M.J. Chien, "Piecewise-linear theory and computation of solutions of homeomorphic resistive networks", IEEE Trans. on Circ. and Syst., Vol. CAS-24, No. 3, March 1977, pp. 118-127.

[26] S.N. Tschernikow, *Lineare Ungleichungen*, Hochschulbücher für Mathematik, VEB Verlag der Wissenschaften, 1971.

[27] T. Ohtsuki, T, Fujisawa and S. Kumagai, "Existence theorems and a solution algorithm for piecewise-linear resistor networks", SIAM J. Math. Anal., Vol. 8, No. 1, February 1977, pp. 69-99.

[28] W.M.G. van Bokhoven, "Linear implicit differentiation formulas of variable step and order", IEEE Trans. on Circ. and Syst., Vol. CAS-22, No. 2, February 1975, pp. 109-115.

[29] P. Henrici, *Discrete variable methods in ordinary differential equations*, J. Wiley & Sons Inc., New York, 1962.

[30] J.E. Meyer, "MOS models and circuit simulation", RCA Rev., Vol. 32, March 1971, pp. 42-63.

[31] R. Bellman, *Introduction to matrix analysis*, McGraw-Hill Book Comp. Inc., New York, 1960, pp. 115-116.

Samenvatting

In de elektrotechniek moet de simulatie van complexe elektronische schakelingen als een vitaal onderdeel van het huidige ontwerpproces van LSI schakelingen beschouwd worden. Vanwege onvermijdelijke effecten,zoals strooi capaciteiten of parasitaire transistoren,die in een geïntegreerde schakeling optreden is het niet langer mogelijk of aanbevelenswaardig om de schakeling zodanig uit discrete componenten op te bouwen dat het ontwerp door middel van metingen volledig op zijn correctheid getoetst kan worden.

Deze parasitaire effecten kunnen echter in theorie vrij eenvoudig in een simulatie worden meegenomen waardoor een vroegtijdige indicatie betreffende de correctheid van het ontwerp verkregen kan worden voordat de schakeling daadwerkelijk in silicium geïntegreerd wordt.

Alhoewel simulatie dus het antwoord lijkt op een groot aantal problemen,worden de moeilijkheden in feite slechts doorgeschoven naar deze simulatie,die dan zeer complex zal worden. Naast logische simulatie van digitale circuits zal de simulatie in staat moeten zijn ten behoeve van een correcte timing het tijdsverloop van de diverse logische signalen te berekenen. Tevens zal de responsie van analoge circuits ten aanzien van inschakel verschijnselen en gelijkspanningsgedrag moeten kunnen worden bepaald.

Omdat in de huidige toepassingen veelal analoge circuits in combinatie met digitale circuits op één chip geïntegreerd worden, moeten traditioneel verschillende simulatie taken nu op een enkel circuit uitgevoerd worden.Dit probleem stelt stringente eisen aan de onderdelen en de interne organisatie van een simulatie programma. Een hierarchische structuur van zo'n programma is een eerste vereiste. Verder dienen de diverse taken binnen een simulatie programma dusdanig op elkaar afgestemd te worden dat een zo uniform mogelijke datastructuur toegepast kan worden teneinde conversie problemen bij de vertaling van gegevens benodigd voor de verschillend gerichte taken te voorkomen.

Een van de belangrijkste onderdelen van een simulatie programma is te vinden in het oplossen van stelsels niet-lineaire vergelijkingen welke,door toepassing van geschikte modellen voor de aanwezige componenten,ontstaan bij het formuleren van de netwerkvergelijkingen.

Voor schakelingen met bipolaire of MOS transistoren kan het gelijkspanningsgedrag van deze componenten door middel van een relatief klein stelsel niet-lineaire vergelijkingen vastgelegd worden. Het stelsel van niet-lineaire vergelijkingen dat het gedrag van de totale schakeling beschrijft wordt dan meestal opgelost door toepassing van een zekere variant van de Newton-Raphson methode.

Het gedrag van digitale schakelingen wordt daarentegen vaak vastgelegd door middel van tabellen of logische functies en vereist in het algemeen een heel andere aanpak,zowel ten aanzien van de modellering als van de analyse methodes. Dit bemoeilijkt de onderlinge uitwisselbaarheid van de berekende responsies. Door echter gebruik te maken van modellen gebaseerd op de zogenaamde drempel-logica ontstaat een beschrijving met stuksgewijs-lineaire vergelijkingen welke beter bij de voorafgaande beschrijving aansluit.

In complexe analoge schakelingen ten slotte wordt het gedrag van bijvoorbeeld operationele versterkers,hysterese schakelingen e.d. vaak weergegeven door een zogenaamd macro-model dat te beschrijven is met behulp van stuksgewijs-lineaire karakteristieken.

Ieder van bovenstaande modelleringstechnieken leidt tot een specifieke analyse methode met de daaraan verbonden eigen problematiek zoals convergentie van de oplossings algoritme en complexiteit van de datastructuur.

Het zal duidelijk zijn dat de eerder gewenste uniforme aanpak van de simulatie alleen bereikt kan worden indien de modellering op ieder analyse-niveau van dezelfde basis structuur uitgaat.

In dit proefschrift is gekozen voor een aanpak op basis van stuksgewijs-lineaire afbeeldingen. Enerzijds omdat de meer complexe circuits zich gemakkelijker lenen voor een stuksgewijs-

lineaire beschrijving dan voor een beschrijving door middel van algemene niet-lineaire functies en anderzijds omdat het zeker eenvoudiger is om niet-lineaire functies te benaderen door stuksgewijs-lineaire functies dan omgekeerd. Vooral de relatief eenvoudige wijze van macro-modellering van complexe bouwstenen opent aantrekkelijke mogelijkheden voor de simulatie van grote electronische schakelingen. Te meer daar deze macro-modellen gecombineerd kunnen worden met modellen voor afzonderlijke transistoren in termen van stuksgewijs-lineaire afbeeldingen,zodat een analyse van deze schakelingen op gemengd niveau kan worden uitgevoerd op basis van een gemeenschappelijke datastructuur.

De traditionele beschrijving van stuksgewijs-lineaire afbeeldingen is echter behept met een aantal nadelen. Allereerst worden deze afbeeldingen meestal vastgelegd door middel van een lijst waarin zowel de diverse gebieden (polyeders) in een Euclidische ruimte als de bijbehorende lineaire afbeeldingen aangegeven zijn. Dit leidt tot een redundante formulering van de afbeelding omdat begrenzende hypervlakken minstens twee maal in de lijst voorkomen. Daarnaast kunnen burenrelaties tussen de verschillende gebieden niet op een impliciete wijze worden vastgelegd,waardoor de gangbare methodes voor het oplossen van stuksgewijs-lineaire vergelijkingen tot een inefficiënt zoekproces in de betreffende datastructuur leiden. Tevens is een lineaire transformatie op de variabelen van een stuksgewijs-lineaire afbeelding moeilijk te verwerken in de desbetreffende lijst en leidt een sequentie van stuksgewijs-lineaire afbeeldingen zelfs tot een exponentiële groei van deze lijst.

De basis van dit proefschrift wordt gevormd door een nieuwe beschrijvingsmethode voor stuksgewijs-lineaire afbeeldingen die de meeste van de voornoemde bezwaren wegneemt. Daartoe wordt een toestands-model voor een stuksgewijs-lineaire afbeelding geïntroduceerd. Dit toestands-model levert een globale impliciete beschrijving van de afbeelding die voor alle gebieden geldig is.Het desbetreffende model is opgebouwd te denken

uit twee verschillende onderdelen. Het ene deel beschrijft de lineaire afbeelding voor elk gebied terwijl het andere deel (de zogenaamde toestands-vergelijking) verantwoordelijk is voor de opdeling van het definitie gebied van de afbeelding in afzonderlijke polyeders. Het lineaire complementariteits-probleem vormt de essentiële grondslag van het model,zodat bekende algoritmen voor de oplossing van dit probleem nu ook gebruikt kunnen worden om stuksgewijs-lineaire vergelijkingen op te lossen.

Naast de behandelde toepassing in de elektrotechniek kan het toestands-model ook met vrucht gebruikt worden in de theorie van homeomorfismen van stuksgewijs-lineaire afbeeldingen.

In de diverse hoofdstukken van dit proefschrift komen de volgende zaken aan de orde.

Hoofdstuk 2 behandelt de beschrijving van stuksgewijs-lineaire netwerken met ideale diodes. Deze netwerkbeschrijving wordt later gebruikt om het toestands-model te definieren.

In hoofdstuk 3 worden,aan de hand van de mogelijke hybriede beschrijvingen van een lineair geheugenloos multipoort netwerk, enkele theorema's geformuleerd die de equivalentie aantonen van de eigenschappen van klasse P en P_0 matrices en zekere passiviteits eigenschappen van voornoemde netwerken. Verder wordt een graaf gedefinieerd die de "structuur" van het netwerk genoemd wordt,waarmee burenrelaties tussen diverse representaties vastgelegd worden.

Hoofdstuk 4 behandelt het toestands-model van een stuksgewijs-lineaire afbeelding. De begrippen "toestand" en "equivalente representatie" worden uiteen gezet en gerelateerd met de in hoofdstuk 3 behandelde hybriede representaties. Tevens wordt een minimaal toestands-model gedefinieerd waarvan later in hoofdstuk 7 gebruik gemaakt wordt bij de afleiding van homeomorfisme eigenschappen van stuksgewijs-lineaire afbeeldingen.

In hoofstuk 5 worden een drietal bekende algoritmen voor de oplossing van het lineaire complementariteits-probleem beschreven. Deze algoritmen kunnen worden toegepast om stuksgewijs-lineaire vergelijkingen op te lossen.

Hoofstuk 6 beschrijft een nieuwe algoritme voor de oplos-

sing van het lineaire complementariteits-probleem. Door de zogenaamde "modulus transformatie" wordt dit probleem eerst omgezet in een stelsel niet-lineaire vergelijkingen welke vervolgens door contractie afbeelding kunnen worden opgelost. Voor een bepaalde klasse van problemen leidt dit,in tegenstelling met de algoritmen uit hoofdstuk 5, tot een polynome tijdcomplexiteit van de algoritme. Tevens wordt deze methode in verband gebracht met het "globale" model van Chua.

Hoofdstuk 7 behandelt het toestands-model van de inverse afbeelding. Er worden een aantal theorema's over homeomorfismen van stuksgewijs-lineaire afbeeldingen afgeleid,die tot een uitbreiding van de bestaande theorie leiden. Daarnaast worden enkele eenvoudige expliciete toestands-modellen bepaald.

In hoofdstuk 8 komt de constructie van het toestands-model voor stuksgewijs-lineaire systemen aan de orde. De modellering geschiedt door de modellen voor de diverse componenten te combineren aan de hand van een specificatie van de onderlinge connecties. Tevens wordt het toestands-model uitgebreid tot een model voor stuksgewijs-lineaire dynamische systemen door gebruik te maken van een lineaire meerstaps integratie formule.

Tenslotte worden in hoofdstuk 9 enige voorbeelden van toestands-modellen gegeven voor enkele gangbare elektronische componenten. Aan de hand van een Wien-brug oscillator worden de perspectieven van een stuksgewijs-lineaire modellering gedemonstreerd.

Curriculum Vitae

De schrijver van dit proefschrift werd op 9 juli 1942 geboren te Drunen. In 1959 behaalde hij het HBS-b diploma aan het Dr. Moller College te Waalwijk. Van 1962 tot 1966 was hij als kort-verband officier verbonden aan de Koninklijke Luchtmacht in het dienstvak Bewapening en Elektronica. Van november 1966 tot oktober 1970 was hij als werkstudent en als Technisch Ambtenaar werkzaam op de Technische Hogeschool te Eindhoven. In september 1970 behaalde hij aldaar met lof het diploma Elektrotechnisch Ingenieur. Van oktober 1970 tot februari 1972 was hij als medewerker verbonden aan het Natuurkundig Laboratorium van de N.V. Philips, waar hij onderzoek verrichtte aan niet-lineaire oscillatoren. Sinds februari 1972 is hij als Wetenschappelijk Medewerker verbonden aan de vakgroep Automatisch Systeem Ontwerpen in de afdeling der Elektrotechniek van de Technische Hogeschool te Eindhoven.

STELLINGEN BIJ HET PROEFSCHRIFT VAN

W.M.G. van BOKHOVEN.

26-5-1981

I

De uitspraak van D.Meyer-Ebrecht "Jedoch wurde bisher keine allgemeine Theorie der amplituden-geregelten harmonischen Oszillator aufgestellt" is onwaar.

Meyer-Ebrecht,D.
Schnelle Amplitudenregelung Harmonischer Oszillatoren.
Diss. T.U. Braunschweig,1974,p.2

II

Bij de behandeling van de gesynchroniseerde Van der Pol oscillator heeft Nayfeh de invloed van de amplituderegellus niet correct verwerkt. De door hem aangegeven voorwaarde $a > 1$ voor stabiele synchronisatie bij snelle regeling is daardoor onjuist en dient vervangen te worden door $a > \sqrt{2}$.

Nayfeh,A.H.
Forced Oscillations of the Van der Pol Oscillator with delayed Amplitude Limiting.
IEEE Trans. on Circ.Th.,CT-15,1968,pp.192-200

III

Volgens de gevoeligheidsanalyse van 2e orde aktieve RC-filters met één positieve versterker zoals gepubliceerd door Weyten,zou de belangrijke door Fleischer behandelde klasse van aktieve filters niet kunnen functioneren. Dit is onjuist en in tegenspraak met de praktijk.

Weyten,L.
Q- and ω_0-Sensitivities in Positive Gain Second-order RC Active Filters.
Proc.IEEE,vol.60,1972,pp.1462-1463

IV

Een minimum-fase spanningsoverdrachtsfunctie waarvan de Bode amplitude-plot slechts uit segmenten bestaat met hellingen van 0, +6 of -6 dB/oct., kan gerealiseerd worden met behulp van klasse-M immittantie functies.

Bokhoven,W.M.G. van
A Canonical Synthesis for Immittance Functions that can be decomposed into the Product of an RL and an RC Impedance.
IEEE Trans. on Circ. and Syst.,CAS-23,1976,pp.36-39.

V

De wijdverbreide opvatting dat de spanningsoverdrachtsfunctie van 2e orde aktieve RC-filters met één versterker bij negatieve terugkoppeling ongevoeliger zou zijn voor variaties in versterking dan bij positieve terugkoppeling, is onjuist.

Bokhoven, W.M.G. van

Sensitivity Bounds and Explicit Polynomial Decomposition for Biquadratic Single Amplifier Active Filters.

Proc. 1976 IEEE Int. Symp. on Circ. and Syst., pp.118-121

VI

De directe toepassing van een aantal Newton-Cotes integratieformules bij het oplossen van gewone differentiaalvergelijkingen leidt tot een instabiele oplossingsmethode. De op deze formules gebaseerde impliciete Runge-Kutta procedures leiden daarentegen tot A-stabiele integratieformules.

Bokhoven, W.M.G. van

Efficient higher order implicit one-step methods for integration of stiff differential equations.

BIT 20, 1980, pp.34-43.

VII

Door toepassing van nieuwe op systeemtheorie gebaseerde CAD hulpmiddelen zal de rol van de soldeerbout in het ontwerpproces van elektronische schakelingen voor een groot deel door een lichtpen worden overgenomen.

VIII

De wijziging van het verenigingsrecht en de recente plannen van de regering om bestuursleden van verenigingen hoofdelijk aansprakelijk te stellen voor financiële tekorten in de exploitatie, brengt het voortbestaan van vele amateur-sportverenigingen in gevaar doordat op de bereidheid van de leden om tot dan toe zonder vergoeding diensten te verrichten een te grote aanslag gepleegd wordt.