THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE REARRANGEMENT OF NETWORKS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

THOMAS HENRY PUCKETT

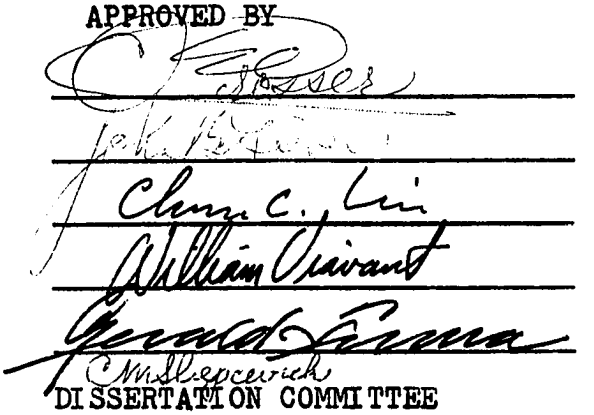
Norman, Oklahoma

1959

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/1441

THE REARRANGEMENT OF NETWORKS

APPROVED BY



DISSERTATION COMMITTEE

ACKNOWLEDGMENT

The author takes pleasure in acknowledging the advice and assistance of Dr. John B. Giever, formerly of the Department of Mathematics of the University of Oklahoma, now at the New Mexico State University.

TABLE OF CONTENTS

		Page
INTRODUCTION		1
Chapter		
I.	DIRECT-SUM DECOMPOSITION OF VECTOR SPACES	4
II.	THE ROTH FORMULATION OF THE NETWORK PROBLEM	39
III.	THE REARRANGEMENT OF NETWORKS	82
IV.	ANALYSIS OF A NETWORK DERIVED FROM A LINEAR HEAT FLOW PROBLEM	107
LIST OF REFERENCES		124

INTRODUCTION

The advent of large-scale digital computers has allowed larger and larger numerical problems to be handled economically, and has resulted in a general increase of interest in analytical techniques for handling large problems. While the works of Gabriel Kron, dating back to the 1930's, have often discussed the concept of solving a large physical system by first solving it in parts and then interconnecting the solutions to the parts, no useful specific technique for so doing has previously appeared. The presentation of such a technique is the purpose of this dissertation.

Network problems arise in their own right as problems of numerical interest. They also appear in another context. The numerical solution of a partial differential equation is usually based upon a set of finite difference approximations to the equation, and such a set often may be interpreted as the equations of an electric network. Two well-known cases are those of heat flow and wave propagation. The dissertation is therefore limited to the question of finding solutions of network problems only. Since the mathematical apparatus is to be that of finite-dimensional linear vector spaces,

only those systems which are describable by systems of linear equations will be considered.

J. Paul Roth, after an investigation of some of the concepts appearing in Kron's publications, has presented an outline of an abstract mathematical analysis of the technique of solving a network by considering it in parts, as well as some incomplete practical examples. Unfortunately his abstract analysis is inadequate and he gives no complete matrix formulas or proofs that the specified inversions can be performed. Based upon some of Roth's concepts a complete abstract analysis and matrix formulation will be given of the relationships between the solutions of two networks which differ from each other in a certain fashion, as well as a complete example.

The material to be presented is divided into three chapters. The first is concerned solely with some mathematical aspects of finite-dimensional vector spaces. The second chapter will demonstrate the utility of vector spaces in handling network problems, and will present a complete analysis of the usual techniques for expressing network problems and the methods of solution of network problems. The primary contribution of the dissertation will be found in the third chapter, which gives the analysis and matrix formulation of the relationships existing between a network and a torn version of the network, and the techniques whereby the solution of either network may be used to aid in the solution of the other. The fourth chapter gives an example of the use

of the techniques in the analysis of a network derived from a heat flow problem.

The general results of the analysis are that the specific numerical techniques are essentially those involved in inverting a matrix by either bordering or partitioning. If the two networks involved do not differ from each other by much, both networks may be solved for essentially the cost of solving one of the networks, or alternatively, the known solution of one may be used to find the solution of the other at small additional cost in computing time.

THE REARRANGEMENT OF NETWORKS

CHAPTER I

DIRECT-SUM DECOMPOSITION OF VECTOR SPACES

Preliminary Definitions and Discussion

This chapter presents a number of elementary results about finite-dimensional vector spaces which do not seem to be conveniently available elsewhere. In some cases the results given are simple abstract presentations of well-known facts about matrices, but for a portion of the results no prior abstract or matrix formulation is known.

It is assumed that the reader is familiar with the usual introductory material on vector spaces such as is given in Halmos (1). Certain conventions will be used. The word space will always mean a finite-dimensional vector space with either the real or the complex numbers as the scalar field. Since almost all of the discussion to follow will be independent of the particular scalar field in use, reference to the field will usually be omitted. The word zero or the symbol 0 will represent either the zero vector of the particular space under consideration or the zero element of the scalar field, the intended meaning being clear

from the context in which it appears in all cases. Upper-case roman letters will be used to indicate particular spaces, and corresponding lower-cases letters to indicate particular elements of the spaces. For instance, i_s stands for some element of the space I_s. (That is, if the discussion is concerned with abstract spaces. When matrix interpretations of abstract results are being considered, i_s will be taken to represent the coordinates of the corresponding vector with respect to some basis of the space. This convention will be given more specifically in the section on matrix interpretations.)

The dual space of a vector space is defined to be the space of all linear functionals acting from the original space to the scalar field, i.e., any element of the dual space is a linear function which assigns some element of the scalar field to each vector of the original space. It is indeed a finite-dimensional vector space, so the name dual space is appropriate, and it has the same dimension as the original space. When elements from a space and its dual both appear in an expression, the element from the original space will appear after the element from the dual space and be enclosed in parentheses to emphasize the functional idea. For instance if E represents the dual space to I_s, $e(i_s)$ stands for an element e of E acting on the element i_s of I_s to give some element of the scalar field. It is a standard discussion to show that the dual space of a dual space may

be regarded as the original space, i.e., the dual space of E may be taken to be I_s, and from this point of view it is usually immaterial which of the spaces is regarded as the original space and which the dual space. Given any non-zero element of a space, there is always some non-zero element of the dual space which when acting on the original element gives a non-zero scalar. In symbols, if $i_s \neq 0$, then there exists an e such that $e(i_s) \neq 0$.

Most of the discussion to follow will be devoted to the analysis of functions between spaces, and the usual word employed will be <u>map</u>: a linear, single-valued function from one vector space to another, and the usual symbol will be an upper-case roman letter. There will invariably be a diagram associated with a mapping, and on occasion mappings will be defined by just giving the appropriate diagram. For instance:

$$I \xrightarrow{F} J$$

identifies the map F as acting between spaces I and J in the direction indicated, assigning to every element of I some element of J, specifically written out when convenient as $j = Fi$.

When matrix interpretations are being considered, the map symbol will also be used to indicate a matrix of the map with respect to some specific pair of bases of the two spaces. Although the map itself, once defined, is unique, the corresponding matrix depends upon the bases in use and will generally change if the bases are changed. Therefore different symbols

will be necessary for the matrices of the map when more than one set of bases is being considered. The matrix conventions will be chosen so that the matrix equation corresponding to some abstract equation has exactly the same form as the abstract equation, so where convenient it will usually be permissible to regard a given abstract equation as also the corresponding matrix equation.

Given any pair of spaces with a map between them, there may be defined an adjoint map between the dual spaces which acts in the opposite direction to the direction of the original map, and which will be indicated by a prime on the original map symbol. The appropriate diagram is:

$$\begin{array}{ccc} & \overset{F}{\longrightarrow} & \\ I & & J \\ & \underset{F'}{\longleftarrow} & \\ E_s & & V \end{array}$$

where E_s and V are the dual spaces of I and J respectively, F is the original map, and F' is to be the adjoint map. The precise definition of F' is as follows: given any v, $e_s = F'v$ is to be that element which when acting on any i gives the same scalar as v acting on Fi, in symbols: $F'v(i) = v(Fi)$. As with dual spaces, the adjoint of an adjoint may be taken as the original map: $(F')' = F$. The diagram illustrates a convenient convention: any time two spaces are shown one immediately above the other, the lower will always represent the dual space of the upper, and correspondingly for maps: the lower map will be the adjoint of the upper.

Subspaces of a space will usually be indicated by the symbol of the original space with a subscript suffixed. For instance, those elements of J in the previous diagram which are the images of elements of I under the map F form a subspace of J, which will be called the image space of F and might be indicated by J_o. Likewise those elements of V which go into the zero element of E_s under F' form a subspace V_o, which will be called the null space of F'.

A primary use of subspaces is to form direct-sum decompositions of the original space. For instance, given any subspace J_o there exists at least one complementary subspace J_s which meets J_o only in the zero element of J, and such that any element j of J can be written in one and only one way as the sum of a pair of elements, one from J_o and one from J_s: $j = j_o + j_s$. Such a decomposition is indicated by $J = J_o \oplus J_s$. Correspondingly, V might be split up into the direct sum of V_o and some complementary space V_s: $V = V_s \oplus V_o$. Generally there will be many subspaces complementary to a given subspace.

Another example of a subspace is that of an annihilator: given a subspace of a space, its annihilator is that subspace of the dual space which maps everything in the given subspace into zero. There is a natural connection between annihilators and direct sums: suppose $J = J_o \oplus J_s$, and V_o is the annihilator of J_o and V_s the annihilator of J_s, then V may be decomposed into the direct sum of V_s and V_o: $V = V_s \oplus V_o$.

Such a decomposition will be called a dual decomposition of the two spaces.

Suppose the mapping under consideration to be

$$X \xrightarrow{F} Y \quad .$$

If F assigns only the zero element of X to the zero element of Y, it will be said to be a monomorphism. If every element of Y is the image of some element of X under F, F will be said to be an epimorphism. If F is both a monomorphism and an epimorphism, it will be said to be invertible.

A monomorphism takes distinct elements into distinct elements: if $x_1 \neq x_2$, then $Fx_1 \neq Fx_2$, which is an immediate consequence of the defining condition that if $Fx = 0$, then $x = 0$. The defining condition for an epimorphism takes the symbolic form that for any y, there exists some x such that $y = Fx$. If F is invertible, this implies there exists some unique map F^{-1} from Y to X such that FF^{-1} and $F^{-1}F$ are both identity maps. F^{-1} will be called the inverse of F. One special case is of considerable importance: if X and Y have the same dimension, F will be a monomorphism if and only if it is also an epimorphism, so in this case either property establishes invertibility. Any invertible map between two spaces is said to establish an isomorphism between the spaces, a one-to-one linear correspondence. Any two such spaces are said to be isomorphic, but the particular isomorphism will depend upon the map between them.

As an example of a map which is neither a monomorphism nor an epimorphism, consider a map F which takes every element of X (assumed to have dimension of at least one) into the zero element of Y (also assumed to have dimension of at least one). There will be at least two elements of X with the same image in Y, and at least one element of Y which is not the image of any X.

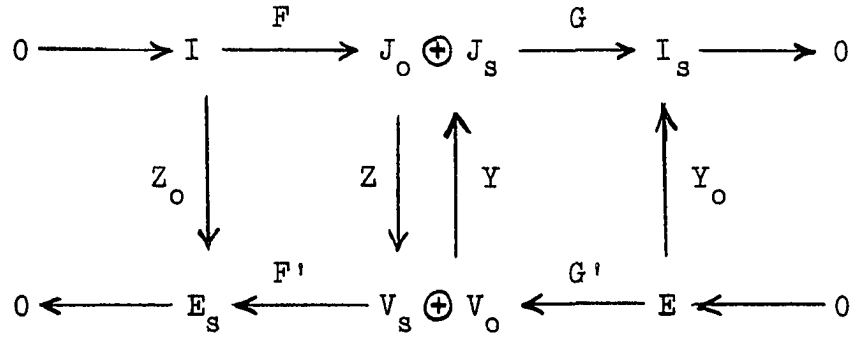
As an example of an epimorphism, a <u>projection</u> of J onto a subspace J_o may be constructed as follows. Let J be decomposed into a direct sum of J_o and some complementary space J_s, $J = J_o \oplus J_s$. Any element of J may then be expressed in one and only one way as $j = j_o + j_s$. The projection map of J onto J_o along J_s is then defined by assigning to the element j the element j_o of the subspace. Since it was necessary to choose an arbitrary complementary subspace J_s to make the projection, there is nothing particularly unique in general about the image of an element of J. However, any projection always assigns the same image to a j which is already in the subspace J_o, namely itself, and of course the elements of the complementary subspace make up the null space of the mapping. The essence of a projection, in contrast to just any epimorphism from J to J_o, is its property of assigning to elements of J_o their natural images, namely themselves.

The preceding discussion may be turned around to give an example of a monomorphism: the <u>injection</u> of a subspace back into the original space, in which each element of the

subspace is assigned itself as an image, considered as an element of the larger space. Again, while many monomorphisms could be constructed between a subspace and the original space, the essence of an injection is its natural character, so to speak.

Three Theorems on Decompositions of Spaces

The following diagram, essentially due to Roth (2), will be analysed in detail throughout the remainder of this chapter, and the results used in the next chapter in the analysis of the network problem. It will be called the Roth Diagram. The notation has been chosen with an application to electric networks in mind. In the theorems only those parts of the diagram specifically referred to in each theorem will be assumed to necessarily exist for the purposes of the theorem. The zeros on the ends of the diagram are for a notational convenience to be discussed later.



One aspect of the theorems of some importance is that, although it is assumed that the spaces have been decomposed into direct sums, it is not assumed that the decompositions are dual to each other. Specifically, it is not taken that

J_s and V_s are annihilators of each other, but just any spaces complementary to J_o and V_o respectively, which are taken as mutual annihilators.

The theorems will utilize two concepts not previously mentioned: commutativity of a diagram and an exact sequence of maps. It is seen that there are two maps from I to E_s: Z_o and F'ZF. If the two maps are identical: Z_o = F'ZF, the diagram is said to be commutative. On the other hand, if the image space of F is the same as the null space of G, the upper part of the diagram is said to form an exact sequence.

Theorem 1: If F is a monomorphism, then:

(A) F' is an epimorphism, and conversely;

and if also J_o is the image space of F, and V_o its annihilator, then:

(B) F establishes an isomorphism between I and J_o,

(C) F' establishes an isomorphism between E_s and V_s,

(D) V_o is the null space of F',

(E) there exist epimorphisms F_r from J to I such that F_rF is an identity map, and each such F_r will be called a reverse of F,

(F) there exist monomorphisms $(F')_r$ from E_s to V such that $F'(F')_r$ is an identity map, and each such $(F')_r$ will be called a reverse of F',

(G) the adjoint of a reverse is a reverse of the adjoint, in symbols: $(F_r)' = (F')_r$, and where no confusion will arise the symbol F_r' will be used to

represent reverses of F';

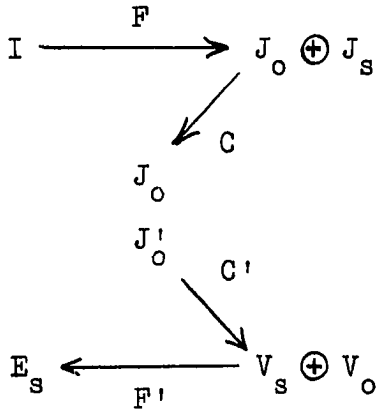
and if also the maps Z and Z_o exist such that $Z_o = F'ZF$ then:

(H) a necessary and sufficient condition that Z_o be invertible is that for any non-zero j_{oa} there exist some j_{ob} such that $Zj_{oa}(j_{ob}) \neq 0$.

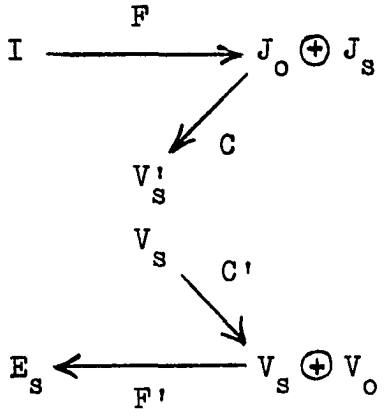
Proofs: (A) For any non-zero i, $Fi \neq 0$, so there exists some v such that $v(Fi) = v(j) \neq 0$. Suppose F' not an epimorphism, i.e., the image space of F' does not span E_s. The annihilator of this image space will then be a non-zero subspace of I, say I_o. For any v, $0 = F'v(I_o) = v(Fi_o)$, so $Fi_o = 0$ since only the zero element annihilates everything in V. But if i_o is non-zero, the result is that F takes a non-zero element into zero, which contradicts the assumption that F be a monomorphism. To demonstrate the converse, assume F' an epimorphism. Then for any non-zero i, there exists e_s such that $e_s(i) \neq 0$, and by the assumption of F' some v such that $e_s = F'v$, so $F'v(i) \neq 0$. But this may be written as $v(Fi) \neq 0$, so $Fi \neq 0$ and F must be a monomorphism. (B) F is both a monomorphism and an epimorphism as far as J_o is concerned and therefore establishes the isomorphism. (C) Since I and J_o are isomorphic under F, they must have the same dimension, and likewise E_s and V_s must have this same dimension. It is therefore only necessary to show that for any e_s there is some v_s such that $e_s = F'v_s$, since this implies that F' considered as restricted to acting on V_s is an epimorphism, which with the fact that E_s and V_s have the same dimension will establish

the desired isomorphism. Since F' is an epimorphism, for any e_s there is some v such that $e_s = F'v$, and since V is a direct sum of V_s and V_o this may be written as $e_s = F'vs + F'v_o$. Now considering e_s acting on any i: $e_s(i) = F'v_s(i) + F'v_o(i)$, or $e_s(i) = F'v_s(i) + v_o(Fi) = F'v_s(i) + v_o(j_o) = F'v_s(i)$, which implies that the image space of F' acting on V_s spans E_s since no assumptions were made about e_s and i in the equation. (D) The immediately preceding calculation showed that V_o is contained in the null space of F', so it only remains to show that every element of the null space is also contained in V_o. If v is in the null space of F', then $0 = F'v$, so $0 = F'v_s + F'v_o = F'v_s$, and since F' restricted to V_s is a monomorphism $v_s = 0$, so v must lie entirely in V_o. (E) Let C be any projection of J onto J_o. The map CF establishes an isomorphism between I and J_o and is therefore invertible since I and J_o have the same dimension. A direct calculation will show that $(CF)^{-1}C$ is indeed a reverse for F. (F) For this case take C' to be the injection of any V_s into V. The map F'C' will be invertible, and $C'(F'C')^{-1}$ will be a reverse for F'. (G) The calculations in this case are slightly more complicated than those above, even though the same equations apply. The difficulty is illustrated by the fact that the adjoint of a projection of J onto J_o has no direct interpretation as an injection of V_s into V, unless restrictions are placed upon the choice of V_s. The two cases will be considered separately. Let C be any projection of J onto J_o,

let J_o' be the dual space to J_o and C' the adjoint of C. The diagram is then:



where the source of complication may be identified as the fact that V_s is not necessarily the image space of C'. Nevertheless the map F'C' establishes the necessary isomorphism between J_o' and E_s and is therefore invertible. Suppose j_o' is non-zero. There is then some j_o such that $j_o'(j_o) \neq 0$. Now since I and J_o are isomorphic under the map CF there is some i such that $j_o = CFi$, and the previous expression may be written as $0 \neq j_o'(j_o) = j_o'(CFi) = F'C'j_o'(i)$, i.e., if j_o' is non-zero its image in E_s is also non-zero. F'C' is therefore a monomorphism and invertible since the dimensions of J_o' and E_s are the same. For the other case let C' be defined as the injection of V_s into V, V_s' as the dual space of V_s and C as the adjoint of C'. The diagram is then:



and it is desired to show that CF is invertible, which will follow immediately from a demonstration that C establishes an isomorphism between J_o and V_s'. Suppose j_o non-zero. Then $v_s(Cj_o) = C'v_s(j_o) = v_s(j_o)$ since C' is the injection of V_s into V. But this expression is surely not identically zero for all possible v_s since V_s is a space complementary to the annihilator of J_o. Cj_o is therefore non-zero and C acts as a monomorphism from J_o to V_s', and the desired result follows since J_o and V_s' have the same dimension. (H) To show necessity, assume Z_o invertible and therefore a monomorphism. If i_a is non-zero, then $j_{oa} = Fi_a$ will also be non-zero since F is a monomorphism. Since this element must eventually map into a non-zero element of E_s, and also everything in V_o goes into zero under F', the image of j_{oa} under Z must have a non-zero component in V_s, say v_{sa}. Now since v_{sa} lies in a subspace complementary to the annihilator of J_o and is non-zero there must be some j_b such that $v_{sa}(j_b) \neq 0$. Even further, there must be some j_{ob} such that $v_{sa}(j_{ob}) \neq 0$, since a possible choice for J_s is the annihilator of V_s. Making the

calculation corresponding to the condition of interest: $Zj_{oa}(j_{ob}) = v_{sa}(j_{ob}) + v_{oa}(j_{ob})$, where v_{oa} is the component of Zj_{oa} in V_o, and since V_o is the annihilator of J_o the second term is zero, giving $Zj_{oa}(j_{ob}) = v_{sa}(j_{ob}) \neq 0$, which establishes the necessity. The sufficiency follows from the fact that the condition quarantees that Z maps images of non-zero elements of I into elements with non-zero components in V_s, and since V_s is isomorphic to E_s the final image in E_s is also non-zero. Z_o is therefore guaranteed to be a monomorphism, and invertible since I and E_s have the same dimension.

Theorem 2: If G is an epimorphism, then:

(A) G' is a monomorphism, and conversely;

and if also J_o is the null space of G, and V_o its annihilator, then:

(B) G establishes an isomorphism between J_s and I_s,

(C) V_o is the image space of G',

(D) G' establishes an isomorphism between E and V_o,

(E) there exist monomorphic and epimorphic reverses for G and G' respectively, such that GG_r and $(G')_rG'$ are each identity maps, and $(G_r)' = (G')_r$ as in Theorem 1;

and if also the maps Y and Y_o exist such that $Y_o = GYG'$, then:

(F) a necessary and sufficient condition that Y_o be invertible is that for any non-zero v_{oa} there exist some v_{ob} such that $v_{ob}(Yv_{oa}) \neq 0$.

This theorem is primarily a restatement of Theorem 1 to

apply to the other end of the diagram, and serves as a convenient statement of results for later applications. The only part requiring independent proof is part (C), which is essentially the converse of the corresponding part (D) of Theorem 1. The proof is easy. Every image of an element of E lies in V_o since $G'e(j_o) = e(Gj_o) = 0$, and since E and V_o have the same dimension and G' is a monomorphism the image space of G' must span V_o.

Theorem 3 is essentially Theorems 1 and 2 combined. It serves to introduce the ohmic condition for the existence of the various inverses. A map Z from J to V is said to be ohmic if for j non-zero, $Zj(j) \neq 0$.

Theorem 3: If J_o is both the image space of F and the null space of G (exact sequence), and V_o is the annihilator of J_o, and F is a monomorphism and G an epimorphism, and the maps Z, Z_o, Y and Y_o all exist satisfying the commutative conditions of Theorems 1 and 2, then:

(A) all the results of Theorems 1 and 2 hold, and the lower sequence is exact,

(B) there exist reverses of F and G such that the reverse sequence is also exact, and for these reverses the map $FF_r + G_rG$ is an identity map, with corresponding statements for the dual spaces and adjoint maps;

and if also Z and Y are each other's inverses, then:

(C) Z_o is invertible if and only if Y_o is invertible;

and if also for any non-zero j, $Zj(j) \neq 0$ (ohmic), then:

(D) Y is ohmic,

(E) Z_o is invertible, and both it and its inverse are ohmic,

(F) Y_o is invertible, and both it and its inverse are ohmic.

Proofs: (A) is satisfied, since the conditions of this theorem meet all the conditions of the previous theorems, and V_o is both the image space of G' and the null space of F'. (B) Let C be the projection of J onto J_o along J_s, and likewise let D be the injection of J_s into J. Reverses for F and G are then given by $F_r = (CF)^{-1}C$ and $G_r = D(GD)^{-1}$. The image space of G_r is therefore J_s, which is also the null space of F_r, so the reverse sequence is exact. To show that $FF_r + G_rG$ is an identity map, let any j be split into j_o and j_s: $j = j_o + j_s$. Now FF_r acting on j gives $FF_rj = FF_rj_o + FF_rj_s$ or $FF_rj = FF_rj_o = F(CF)^{-1}Cj_o = F(CF)^{-1}j_o$. $(CF)^{-1}$ identifies the element of I which maps onto j_o under F, so $FF_rj = j_o$. Likewise $G_rGj = j_s$, so $FF_rj + G_rGj = j$. The treatment of the adjoints follows exactly the same pattern. (C) To show that Y_o is invertible by Theorem 2 it must be demonstrated that, given any non-zero v_{oa}, there exists some v_{ob} such that $v_{ob}(Yv_{oa}) \neq 0$, or since Y and Z are inverses, $v_{ob}(Z^{-1}v_{oa}) \neq 0$. This may be written out in the form that the j which maps onto any non-zero v_o under Z must have a non-zero component in J_s, i.e., does not lie entirely in the annihilator of V_o.

But this is immediate, since the condition on Z that Z_o be invertible guarantees that the image under Z of any non-zero j_o must have a non-zero component in V_s, so the image cannot lie wholly in V_o (recall J_o is the image space of F, and also the annihilator of V_o). The converse statement follows by symmetry. (D)-(E)-(F) The inverse of an ohmic map is also ohmic: suppose Z ohmic and v non-zero, with $j = Yv$, then $v(Yv) = Zj(j) \neq 0$. Note also that any ohmic map between a space and its dual space must be invertible, since it is by the ohmic condition also a monomorphism. To obtain the stated results the only significant condition remaining is to show that ohmicness is hereditary, for instance, if Z is ohmic then so is Z_o. Making the calculation, for i non-zero, $Z_o i(i) = F'ZFi(i) = ZFi(Fi) \neq 0$, since Fi is non-zero by the fact that F is a monomorphism.

The ohmic condition seems to have been first utilized by Roth, following a suggestion of Steenrod (3). The name is derived from the analogy between the condition and Ohm's Law as used in electric network analysis, of which an example is given in the next chapter. It essentially represents the restriction on Z that the necessary and sufficient condition of Theorem 1 be satisfied no matter how J is carved up into a direct sum, which is to say no matter what the image space J_o of F might be. It is certainly a sufficient condition for invertibility of Z_o, but is necessary only under the additional stipulation that Z_o be invertible for any monomorphism F.

Another point of interest is that it is only the pair of subspaces J_o and V_o, each the annihilator of the other, that has entered into the preceding calculations in a natural fashion. In the network problem, for instance, it will be seen that J_o and V_o are, so to speak, natural subspaces, while J_s and V_s are just complementary subspaces, with no other necessary restriction.

The use of zeros on the ends of the diagram is a convenient notation that, with the specification that the entire sequence be exact, automatically implies, for instance, that F is a monomorphism and G an epimorphism. Consider the left end. Exactness and the zero coming into I from the left imply that F maps only zero into zero, and is therefore a monomorphism. On the other end, the zero implies that all of I_s goes into zero and is therefore all in the image space of G, which is therefore an epimorphism.

Some Comments on the Relationships Between the Center Spaces and the End Spaces

The previous theorems have demonstrated that, subject to the various mappings, the elements of the various spaces are certainly not independent. A typical question might be: given some j, is there any particular corresponding i and i_s? An answer to this is: it is possible to find reverses F_r and G_r, and use them to determine a pair i and i_s, such that the sum of the images of i and i_s in J give the original element j. Suppose F_r and G_r are constructed utilizing injections

and projections exactly as described earlier, and therefore form an exact sequence (it is assumed that the full diagram is being considered under the conditions of Theorem 3). The end space elements are then taken as $i_s = Gj = Gj_s$ and $i = F_r j = F_r j_o$, where $j = j_o + j_s$. Calculating the j_a corresponding to these end space elements as follows: $j_a = Fi + G_r i_s = FF_r j + G_r Gj = j$ by Theorem 3. It is therefore possible to go out to the end spaces and back and recover the original element, although the I component is somewhat arbitrary, depending upon the particular F_r used, which in turn depends upon the fairly arbitrary choice of the J_s space.

The converse question: given i and i_s, is there some corresponding j, has an exactly similar answer. Using the same reverses as before, if $j = Fi + G_r i_s$, then $Gj = i_s$ and $F_r j = i$. In this case j is not unique, again due to the arbitrariness of J_s and therefore G_r.

A question answered by Roth (2), which is essentially the solution to the network problem of the next chapter, is: given i_s and e_s, what are the corresponding j and v? More precisely, given the entire diagram and conditions of Theorem 3, the elements e_s and i_s, what are j and v such that $v = Zj$, $Gj = i_s$ and $F'v = e_s$? It is immediately verifiable that the following is indeed a solution, assuming the inverses to exist: $j = FZ_o^{-1} e_s + YG'Y_o^{-1} i_s$, and $v = Zj$. The matter of the existence of the inverses has already been treated. If the

solution exists, it is unique. Suppose j_a and v_a to represent the differences between two solutions to the same problem, so $v_a = Zj_a$, $Gj_a = 0$ and $F'v_a = 0$. Since $Gj_a = 0$, j_a is in the null space of G so $j_a = Fi$ for some i. Then $0 = F'v_a$ or $0 = F'Zj_a = F'ZFi = Z_o i$, so i must be zero since Z_o is a monomorphism. Therefore $j_a = Fi = 0$, and the assumed different solutions to the same problem must be identical.

The question considered by Roth can be split into parts, corresponding to the earlier treatment by Theorems 1 and 2 of each end of the diagram separately. Given the map Z (Y not necessarily existing), the left end of the diagram as in Theorem 1, e_s and the condition that $i_s = 0$, what are j and v such that $v = Zj$, $Gj = 0$ and $F'v = e_s$? The answer: $j = FZ_o^{-1}e_s$ and $v = Zj$. Uniqueness follows as above.

Alternatively, given Y (Z not necessarily existing), the right end of the diagram as in Theorem 2, i_s and the condition that $e_s = 0$, what are j and v such that $j = Yv$, $F'v = 0$, and $Gj = i_s$? The answer: $v = G'Y_o^{-1}i_s$ and $j = Yv$, and again uniqueness follows as above. Note that in each case the solution is essentially just the Roth solution with the appropriate e_s or i_s set equal to zero.

In the case treated by Roth there are no particular unique elements of I and E associated with a given problem. For instance j will in general not be an image of some i, since it must map into a non-zero i_s under G. This does give a hint that in the other two cases this might not be

so, which is the situation. For instance if $i_s = 0$, the solution is $j = FZ_o^{-1}e_s$, which is the image of an element of I, namely $i = Z_o^{-1}e_s$, which can therefore be naturally associated with the problem. If $e_s = 0$, then $e = Y_o^{-1}i_s$ can be naturally associated with the problem.

Inversion of a Map Between Two Spaces by Extension of the Spaces

Suppose there is given the spaces I and E_s, and the map Z_o between them from I to E_s, and it is desired to find Z_o^{-1}, which is to say, given an element of E_s, what is the corresponding element of I such that $e_s = Z_o i$? Rather than finding Z_o^{-1} directly, the following discussion gives an indirect procedure in which the given spaces and map are embedded in a "larger" problem, whose solution is used to find the desired inverse. The procedure will be called abstract bordering, since it is essentially an abstract treatment of the operation to be given later of inverting a matrix by bordering.

The diagram considered before is applicable. Suppose the given Z_o, I and E_s to be embedded, or extended, in a commutative fashion to form the whole diagram such that $Z_o = F'ZF$, etc. Given Z_o, Z, I, E_s, J and V, note that it is not necessarily trivial to determine the injection map F, so to speak, such that $Z_o = F'ZF$. However, once this is done, it should be straightforward to set up the other side of the diagram by just letting G be a projection of J onto J_s along J_o, and defining $Y_o = GYG'$ (recall J_o is the image space of

F, and J_s any complementary space). It is now shown that Z_o^{-1} can be expressed in terms of the various maps of the diagram, a reverse of F, and the inverses of Z and Y_o.

The basic question has already been stated: given e_s, what is i such that $e_s = Z_o i$? The corresponding j will be an element of J_o, namely Fi, and therefore the corresponding i_s will be zero by the exactness of the sequence. The general procedure will be as follows: since F' is an epimorphism, given any e_s there is some v such that $e_s = F'v$, and as a matter of fact any reverse of F' may be used to find such a v. However, since the correct answer is unique, the v found by $v = F'_r e_s$ will almost certainly not be the correct one, so a correction must be made. The correction will be determined by finding the j corresponding to the incorrect v, and then using G to determine the corresponding i_s which in general will be non-zero since the element of j will not necessarily be also an element of J_o. The actual correction will then be made by adding to $F'_r e_s$ a term from E chosen to make the net I_s component zero, which is found by inverting Y_o. The element of J corresponding to this corrected element of V must then be also an element of J_o, and so any F_r will determine the unique i which maps onto it under F. This i will answer the original question, since its corresponding element in V will be $F'_r e_s + G'e$, which will give e_s when mapped to E_s by F', the second term going to zero by the exactness of the lower sequence. Therefore an i has been

found whose image under $Z_o = F'ZF$ is e_s, and it must be the correct i since there is only one. The detailed calculations follow.

Let F_r' and F_r be any reverses of F' and F respectively. Then a trial element of V will be $v_t = F_r'e_s$, and the corresponding element of J will be $j_t = YF_r'e_s$, and the resulting $i_{st} = GY'F_r'e_s$. The E component which will cancel this is $e_c = - Y_o^{-1}i_{st} = - Y_o^{-1}GYF_r'e_s$, and adding the image of this to v_t the final v is $v = F_r'e_s - G'Y_o^{-1}GYF_r'e_s$. The final element of J is then $j = Yv$, and since this must also be an element of J_o, any F_r may be used to determine the answer to the original question, the element of I: $i = F_r(Y - YG'Y_o^{-1}GY)F_r'e_s$, or $Z_o^{-1} = F_r(Y - YG'Y_o^{-1}GY)F_r'$.

This procedure for finding Z_o^{-1} requires, from one possible point of view, the inversions of Z and Y_o, and the finding of a reverse of F. It will be seen when actual matrix manipulations are considered that the finding of the reverse of F is often completely trivial, so it will be given no further consideration here. One possible way of giving an estimate of the utility of this procedure might then be to compare in some way the work necessary to invert Z_o directly with the work necessary to invert Z and Y_o. Now in general, the cost in computing time to invert a map increases as the dimension of the spaces involved increases, and in fact a common estimate is that cost increases as the cube of the dimension. In the procedure given above the dimension of J

is greater than the dimension of I, specifically, the dimension of J equals the sum of the dimensions of I and I_s. The procedure therefore replaces the inversion of Z_o with the inversion of Z, a map between spaces of greater dimension, and the auxiliary inversion of Y_o. From this point of view it would certainly be most uneconomical to use the procedure.

There are two cases where the procedure has possible merit, however. For instance, the inverse of Z might already be known due to previous computational effort, so the only problems would be the determination of F such that $Z_o = F'ZF$ and the inversion of Y_o. If the dimension of J is less than twice the dimension of I, then the inversion of Y_o should be less expensive than the direct inversion of Z_o. Another possibility is that for some reason the inversion of Z might be trivial, in which case again the inversion of Y_o would most likely be the primary expense. While this possibility may seem unlikely, it will be seen in the next chapter that this is exactly what commonly happens in problems derived from an electric network.

Finally, since the diagram is skew-symmetric, essentially the same procedure may be applied to determine the inverse of Y_o in terms of the inverses of Y and Z_o. The result is $Y_o^{-1} = G_r'(Z - ZFZ_o^{-1}F'Z)G_r$.

Inversion of a Map Between Two Spaces by Splitting of the Spaces into Smaller Spaces

The matter to be considered now is that of finding the inverse of a map by splitting the spaces involved and inverting some "smaller" maps. Specifically, using the previous diagram again, the question is to find the inverse of Z in terms of the inverses of Z_o and Y_o. More precisely, Y may be calculated by finding Z_o^{-1}, using Z_o^{-1} and Z to find Y_o^{-1} directly, then inverting this to find Y_o, and finally working with Z_o^{-1} and Y_o to find Y. The procedure will be called abstract partitioning since it is essentially an abstract treatment of the technique to be given later of inverting a matrix by partitioning.

The basic question is: given some v, what is j such that v = Zj? It will be necessary to work with reverses of F and G again, and it will be necessary to add the specification that the reverses be chosen as described earlier so that the reverse sequences are also exact. This will allow elements of E_s and E to be determined from the element of V such that when they are mapped back into V the original v is obtained. In effect, V may then be regarded as the direct sum of E_s and E, so to speak. Let F_r, F_r', G_r and G_r' represent such reverses.

Since the calculations are slightly involved, let v be split into its components in the two subspaces:

$v = v_s + v_o$. If j_1 and j_2 are then found such that $v_s = Zj_2$ and $v_o = Zj_1$, then $j = j_1 + j_2$ will satisfy $v = Zj$. More specifically, with the reverse maps chosen as specified above, it is only necessary that j_1 and j_2 satisfy the conditions (A) $G_r'Zj_1 = G_r'v = e$ and simultaneously $F'Zj_1 = 0$, and (B) $F'Zj_2 = F'v = e_s$ and simultaneously $G'Zj_2 = 0$. The sufficiency of these two conditions follows from the choice of reverses, since using them $e_s = F'v = F'v_s$, $e = G_r'v = G_r'v_o$ and $v = F_r'e_s + G'e$. Likewise recall that if $i = F_r j$ and $i_s = Gj$, then $j = Fi + G_r i_s$.

(A) For this case it is necessary to determine a j such that it gives a specified e component and a zero e_s component when mapped down. This may be used to determine a condition on the i and i_s components of j as follows: $e_s = 0 = F'Zj_1$ or $0 = F'ZFi_1 + F'ZG_r i_{s1}$, which may be solved for i_1 as a function of i_{s1}, giving $i_1 = -Z_o^{-1}F'ZG_r i_{s1}$. Now assume Y_o available as a result of inverting its inverse, found from the knowledge of Z and Z_o^{-1}. Note $Y_o = GYG'$, even though Y is not yet known, so the i_{s1} can be immediately calculated as $i_{s1} = Y_o e$, and substituting in the equations above the result for j_1 is $j_1 = (G_r Y_o G_r' - FZ_o^{-1}F'ZG_r Y_o G_r')v_o$.

(B) In this case the element of J is to be found such that it gives a zero E component and a specified E_s when mapped down. The technique used in (A) is not directly applicable, since inverses would be called for whose existence is doubtful. The procedure will be used instead of

just taking the trial values of zero for i_s and $Z_o^{-1}e_s$ for i, which will map down into e_s correctly, and using the non-zero result at E to determine the necessary correction. If $i_t = Z_o^{-1}e_s$ and $i_{st} = 0$, the error term is $e_a = G_r'ZFZ_o^{-1}e_s$. It is now necessary to find the correction terms i_c and i_{sc} which map down into zero at E_s and $-e_a$ at E. But this is precisely the situation considered in (A) above, so those results may be used to obtain $i_c = -Z_o^{-1}F'ZG_r i_{sc}$ and $i_{sc} = -Y_o G_r'ZFZ_o^{-1}e_s$. The final J component is then $j_2 = (FZ_o^{-1} + FZ_o^{-1}F'ZG_r Y_o G_r'ZFZ_o^{-1} - G_r Y_o G_r'ZFZ_o^{-1})F'v_s$. Adding the two components together and defining $Y_a = G_r Y_o G_r'$ and $Y_b = FZ_o^{-1}F'$, the final expression is then:

$$Y = Y_a - Y_b Z Y_a + Y_b + Y_b Z Y_a Z Y_b - Y_a Z Y_b \quad .$$

Correspondingly, if $Z_a = F_r' Z_o F_r$ and $Z_b = G'Y_o^{-1}G$, then the expression for Z becomes:

$$Z = Z_a - Z_b Y Z_a + Z_b + Z_b Y Z_a Y Z_b - Z_a Y Z_b \quad .$$

These somewhat formidable looking results actually hold more promise than the previous ones. In terms of inversions, the original inversion of Z is replaced by a pair of inversions between spaces the sum of whose dimensions is the dimension of the original spaces. Assuming that the cost of an inversion was proportional to the cube of the dimension of the spaces involved, and taking the dimensions of I and I_s to be, for example, each half the dimension of J, the cost of this procedure would be one-fourth the cost of a direct inversion. The cost of the other manipulations involved

would of course be difficult to determine accurately at best, and in any case it is better treated after the above results have been put in matrix form, the subject of the next section.

Matrix Interpretations

Matrices represent the only efficient method of dealing with the actual numerical calculation of maps between spaces of large dimension. Since the actual matrix of a map depends upon the choice of bases for the two spaces involved, it is necessary to make a few comments about bases for vector spaces before actually introducing matrices.

A primary property of any basis of a space is that any element of the space can be expressed as some linear combination of the basis elements, using coefficients from the scalar field. There are generally an indefinite number of possible bases for a given space. Any single non-zero element of the space can be used as a member of some basis for the space. Any set of linearly independent elements of the space can be used as part of a basis. Any basis of a subspace of the space can be used as part of a basis for the space. The basis elements of a pair of complementary subspaces may be joined to form a basis for the entire space.

Given any basis of a space, there is a unique dual basis for the dual space. Say $\{j_1, j_2, \ldots, j_n\}$ is some basis for J. Then there is a unique dual basis for V, $\{v_1, v_2, \ldots, v_n\}$, such that $v_i(j_k) = 0$ unless $i = k$, in which case $v_i(j_i) = +1$. Dual bases might be thought of as

a dual direct-sum decomposition of the two spaces carried to an extreme.

When dual bases are used, the calculation $v(j)$ takes a particularly simple form. Suppose $v = \sum_i a_i v_i$ and $j = \sum_k b_k j_k$. Then $v(j) = \sum_i a_i b_i$. This should not be confused with the concept of a scalar product. Scalar products do have a connection with dual spaces, but since scalar products are not needed in the discussions to follow, they will not be considered here.

To determine the matrix of a mapping it is necessary to select and order bases for both of the spaces involved. Different orderings of basis elements give rise to different matrices for the same mapping. Since for the most part only one basis is ever used at a time for each space, the map symbol itself will be used to indicate the matrix of the map.

Suppose F to be the map from I to J, and $\{j_1, j_2, \ldots, j_m\}$ is the selected basis for J and $\{i_1, i_2, \ldots, i_n\}$ is the selected basis for I. The image of any i_q may be expressed in the form $Fi_q = \sum_p a_{pq} j_p$. The matrix of F with respect to this pair of bases is then the matrix with a_{pq} as the p,q entry of the matrix. Note that the q'th column of the matrix gives the coordinates of the image of i_q with respect to the chosen basis for J. If the coordinates of an element of I with respect to the chosen basis are now ordered in the same order as the basis elements and arranged as a column

matrix, the coordinates of the image of this element with respect to the chosen basis for J will be obtained by multiplying the column matrix from the left by the matrix of the mapping as determined above.

Since usually only one set of bases for the various spaces will be considered at a time, it will be possible to simplify the notation and allow a given equation to represent either the abstract map equation or a corresponding matrix equation, if the convention is made that when the equation is considered as a matrix equation a symbol for an element of a space will be interpreted as the column vector of the coordinates of the vector with respect to the basis in use.

One standard matrix property may be stated immediately. Suppose F is a map between I and J, and that dual bases are in use for both I and E_s, and J and V. Then the matrix of F' will be the transpose of the matrix of F. Exhaustive discussions of various matrix properties may be found in many standard texts (4, 5). The following discussion will be limited to the subjects treated on an abstract level above, inversion by bordering and partitioning.

While it would be adequate simply to state that the previous equations are now to be considered as matrix equations, considerable simplification in form can be achieved by judicious selection of the bases to be used. First let it be specified that dual bases are to be used in each space and dual space. Now consider the map F from I to J, which

establishes an isomorphism between I and J_o. Since F is a monomorphism, the images of the basis elements of I will form a basis for J_o, so let it be specified that these elements be part of the basis for J. Under these specifications the matrix of F will then have the form

$$\begin{bmatrix} U \\ 0 \end{bmatrix}$$

where U is an identity matrix of appropriate order, and 0 a zero matrix likewise. This has essentially established an identity mapping from I to J_o so corresponding elements of the two spaces will have exactly the same coordinates, i.e., as matrices $i = j_o$. The two might as well be identified in this case and the point of view taken that J is the direct sum of I and some J_s. Further, since the matrix of F' will be the transpose of the above, E_s might as well be identified with V_s, since, as matrices, $e_s = v_s$. Using this same technique on the other end of the diagram, identifications can be made between J_s and I_s, and V_o and E.

In the general case suppose that bases for J and V have been chosen by joining bases for J_o and J_s, and V_s and V_o. If L then represents the matrix of Z with respect to these bases, then the mapping may be explicitly written out in matrix form:

$$\begin{bmatrix} v_s \\ v_o \end{bmatrix} = \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} \begin{bmatrix} j_o \\ j_s \end{bmatrix}$$

where L has been partitioned to agree with the natural partitioning of the coordinate column matrices. Likewise if P represents the inverse matrix of L, the Y mapping with respect to these bases, then the converse equation becomes:

$$\begin{bmatrix} j_o \\ j_s \end{bmatrix} = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{bmatrix} v_s \\ v_o \end{bmatrix}$$

with the corresponding partitions of P.

If the "matched base" specifications are now utilized, the coordinates of I and J_o will be the same, i.e., $i = j_o$, $j_s = i_s$, etc., so the above equations may be put in the form:

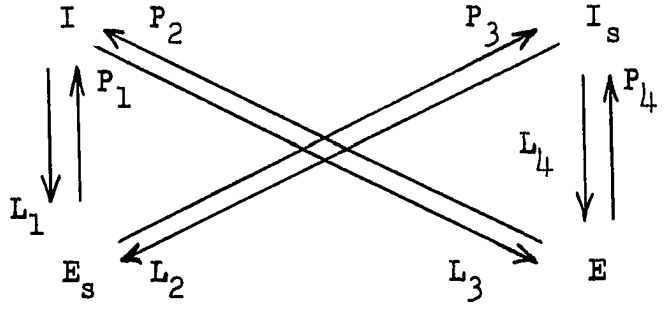
$$\begin{bmatrix} i \\ i_s \end{bmatrix} = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{bmatrix} e_s \\ e \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} e_s \\ e \end{bmatrix} = \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} \begin{bmatrix} i \\ i_s \end{bmatrix},$$

which will allow a considerable simplification of the final matrix formulas.

In this formulation the Z_o map is represented by the matrix L_1 and Y_o by P_4. It will be instructive to apply the procedure used to find the inverses of Z_o and Y_o directly to the matrix formulation given here. Say it is desired to find the inverse of P_4, that is, given i_s, what is the e such that $i_s = P_4 e$. In terms of the entire matrix P, the question must be phrased: given i_s, what is the e such that

$$\begin{bmatrix} i \\ i_s \end{bmatrix} = P \begin{bmatrix} e_s \\ e \end{bmatrix}, \text{ under the condition that } e_s = 0?$$

Note that nothing is said about i, which as a matter of fact will be adjusted to produce $e_s = 0$. Knowing the L matrix, the condition becomes $L_1 i + L_2 i_s = 0$, so take $i = -L_1^{-1} L_2 i_s$. The required e may then be calculated as $e = L_3 i + L_4 i_s$ or $e = (L_4 - L_3 L_1^{-1} L_2) i_s$, so $P_4^{-1} = L_4 - L_3 L_1^{-1} L_2$. The corresponding calculation for L_1^{-1} follows exactly the same pattern, and gives $L_1^{-1} = P_1 - P_2 P_4^{-1} P_3$. The expressions are of course essentially the same as those given earlier from the abstract treatment, as examination will disclose. The following sketch is useful in such comparisons:



Other useful matrix expressions are easily obtained. For instance, since subspaces of a direct-sum decomposition meet only in the zero element of the space, the following expressions result immediately: $P_3 L_1 + P_4 L_3 = 0$ and $L_1 P_2 + L_2 P_4 = 0$.

The matrix partitioning results could be derived from the previous expressions for Z and Y in terms of the inverses of Z_o and Y_o, but it would be somewhat tedious. The essential point is to notice for instance, that Z_a effectively acts only on components of J_o, components of J_s going into zero under F_r, and the result of applying Z_a is always an

element of V_s. A much quicker procedure is just to rearrange the matrix formulas given above to obtain the following expressions for the partitions of P, for instance:
$P_4 = (L_4 - L_3L_1^{-1}L_2)^{-1}$, $P_3 = -P_4L_3L_1^{-1}$, $P_2 = -L_1^{-1}L_2P_4$ and $P_1 = L_1^{-1} + P_2P_4^{-1}P_3$.

The corresponding expression for the partitions of L are: $L_1 = (P_1 - P_2P_4^{-1}P_3)^{-1}$, $L_3 = -P_4^{-1}P_3L_1$, $L_2 = -L_1P_2P_4^{-1}$ and $L_4 = P_4^{-1} + L_3L_1^{-1}L_2$.

A very common numerical technique for matrix inversion due to Woodbury may be expressed by the matrix equation $(A + TBS)^{-1} = A^{-1} - A^{-1}T(B^{-1} + SA^{-1}T)^{-1}SA^{-1}$. If A^{-1} is already known, the right hand side is often significantly cheaper to compute than the left. To show that this is essentially equivalent to matrix bordering, suppose it is desired to find the inverse of P_4 above, so it has been bordered by P_1, P_2 and P_3 as indicated with P_1 non-singular. The inverse in then given by $P_4^{-1} = L_4 - L_3L_1^{-1}L_2$ as discussed above.

Suppose matrices A, B, S and T are now defined as $B = P_1$, $S = B^{-1}P_2$, $T = P_3B^{-1}$ and $A = P_4 - TBS$, and these expressions turned around to give $P_1 = B$, $P_2 = BS$, $P_3 = TB$ and $P_4 = A + TBS$. Making the partitioned inverse calculations, it is immediately found that the result for P_4^{-1} is Woodbury's formula.

Really substantial savings in computing cost with bordering can only be guaranteed in those cases in which the given matrix can be recognized as a submatrix of a not too

much larger matrix whose inverse is already known. Conversely with partitioning, it must be possible to identify in the given matrix a relatively large submatrix whose inverse is already known. For arbitrary matrix inversion problems there is no reason to expect either condition to be met easily. However, it will be seen in Chapter III that in some electrical network problems the conditions can be met quite easily.

CHAPTER II

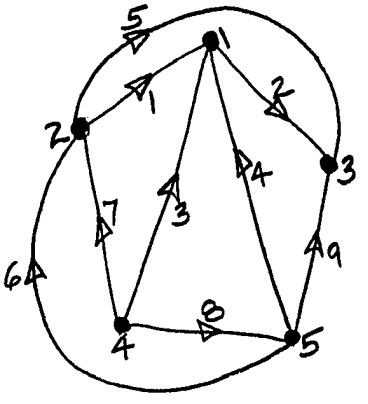
THE ROTH FORMULATION OF THE NETWORK PROBLEM

The most precise and useful formulation of the network problem seems to be one outlined by Roth (2) which will be developed and justified here. The necessary electrical background may be found in Guillemin (6). As an aside, the electrical reader should realize that the dual spaces and maps used in this treatment have no necessary connection with duality as presented by Guillemin. While this formulation is quite satisfactory for the purposes of considering electrical duality, no assumptions are made that the networks being considered are necessarily planar and therefore have dual networks in the electrical sense. On the other hand, corresponding discussions of voltage and current relationships will be phrased as symmetrically as possible.

Fundamentally, network analysis is based upon the application of Ohm's and Kirchoff's laws to a given physical situation. While Ohm's Law may be applied in a straightforward fashion, it will be seen that the proper application of Kirchoff's Laws requires a certain amount of manipulation,

and in fact is equivalent to a demonstration that the Roth Diagram is an adequate representation of the network problem. The point of view which will be taken is the usual one utilized to write network equations, that the net source voltage around a tieset must equal the net branch voltage around the same tieset, and that the net source current across a cutset must equal the net branch current across the same cutset. The network will be considered to exist as an entity independent of the voltage and current sources which might be connected to the network, and as a matter of fact the sources will be represented by separate spaces from the branch voltage and current spaces of the network.

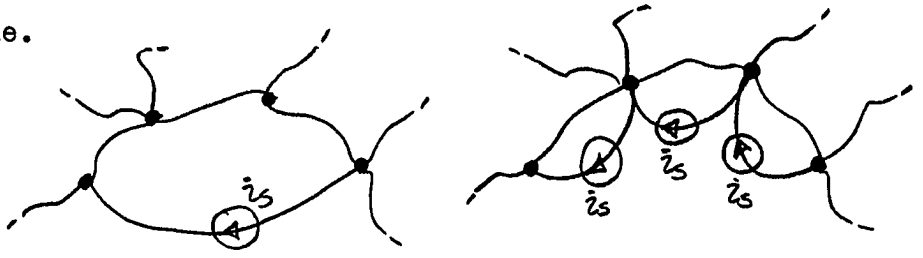
In general a network is represented by a linear graph of oriented branches and unoriented nodes, such as the following:



The arrow on each branch may conveniently be used to indicate the reference directions chosen for the branch voltage and current. A branch voltage or current refers only to the voltage across or current through the ordinary impedance of the branch, and does not include the effects of voltage or

current sources. For instance, if there is a battery in series with branch 9 above, the voltage across branch 9 will then not be equal to the voltage between nodes 3 and 5, but differ from it by the battery voltage. There may be current sources bridged across various nodes, and batteries placed in various branches, but these will be treated on a separate basis from the branch variables.

Since it is desired that the topological relations among the branches remain fixed no matter what voltage and current sources are connected to the network, it will be stipulated that voltage sources are to be inserted only in series with the network branches, and current sources are to be bridged only across the network nodes. Often an even more explicit convention for current sources will be convenient: a given current source must be bridged across some single branch of the network, rather than between any arbitrary pair of the network nodes. This convention allows greater symmetry in the voltage and current discussions, and represents no actual additional restriction on the allowable current sources. Any arbitrary node-pair source current may be rearranged into the desired form by inspection as indicated by the following example.

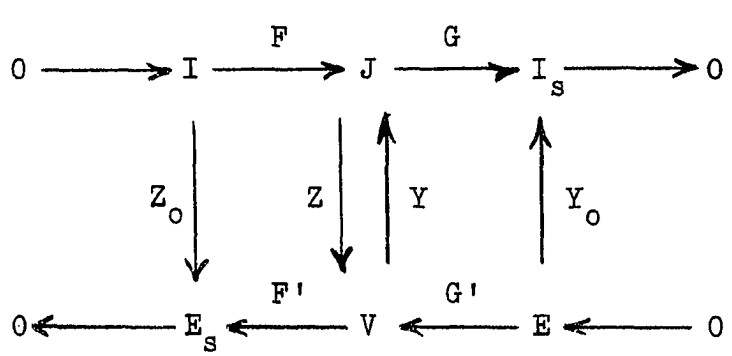


The source spaces will be the node-pair source currents and the loop source voltages. Finally, for convenience, it will be assumed that the network has only one part, since separated electrical networks may be joined without any significant change in their operation.

The topological structure of a network may be conveniently summarized by use of an incidence matrix: the i,j entry is to be zero if branch j is not connected to node i, +1 if branch j is connected to node i and the arrow on the branch points toward the node, and likewise -1 if the arrow points away from the node. Using the conventions that a blank entry represents a zero, a plus sign a +1 and a minus sign a -1, the incidence matrix for the example above is then:

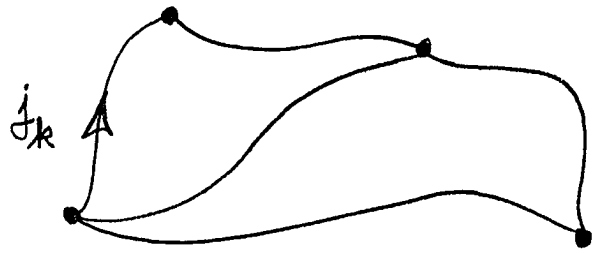
	1	2	3	4	5	6	7	8	9
1	+	-	+	+					
2	-				-	+	+		
3		+			+				+
4			-				-	-	
5				-		-		+	-

The Roth Diagram is repeated here for convenience:



and the following discussion will show that this is a complete representation of the network problem. The upper sequence will be constructed as a direct representation of the current relationships, and it will then be shown that the dual spaces and adjoints maps of the lower sequence properly represent the voltage relationships. The sequences will be exact and the diagram commutative. To establish the representation it will be necessary to work on the matrix level in part, but once established the representation may be considered on the purely abstract level.

The J space will be defined first, in terms of a particular basis. Let j_k represent a current of one ampere through branch k in the reference direction, and zero current in the remainder of the branches. (If the question is raised as to the physical realizability of such a current, it is only necessary to point out that the current balance will be achieved by some current source connected across the nodes of the network, a matter to be discussed shortly.) The space J is then defined as the space created by taking all linear combinations of the j_k symbols with coefficients from the scalar field, so its dimension will equal the number of branches of the network. Since the space has been defined in terms of a particular basis, this basis will be called the natural basis of the space. Any element of J will be referred to as a network current. For a simple network some j_k might be conveniently indicated in the following schematic fashion:



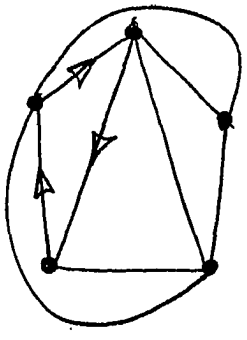
In this case the arrow on the branch is not used to indicate a reference direction necessarily, but to indicate that the particular branch has a current of one ampere in the direction indicated. The other branches, of course, have zero current, and could be replaced by open-circuits.

The space I_s is to represent the source currents so it will be constructed as a direct image of the elements of J, which will automatically make G an epimorphism. Specifically, let i_{snk} represent the necessary source current which must be connected to the network when the network current is a natural basis element, i.e., j_k of J, plus one ampere in branch k and zero in the other branches. Then if $j = \sum_k a_k j_k$, the map is defined as $Gj = \sum_k a_k i_{snk}$, i.e., the usual extension by linearity. This set of images forms a vector space. The dimension of I_s will usually be less than that of J since the set of i_{snk}'s, the images of the natural basis for J, will not be indépendent in general.

By way of examples, consider the calculation of the source current necessary for the basis element j_1 of the network given earlier. If the current through branch 1 is to be one ampere in the reference direction, with no current in the

other branches, then some combination of current sources must be connected across the network nodes which results in taking one ampere from node 1 and feeding one ampere to node 2. This particular net source current will be i_{sn1}. Note that there are several ways in which "physical" current sources could be bridged across the network nodes to produce this particular net source current.

Suppose the network current to be one-half ampere in the reference direction in branch 1 and one ampere in the reference direction in branch 2, $j = (1/2)j_1 + (1)j_2$. The i_{sn2} for this network is one ampere out of node 3 and one ampere into node 1, and so $Gj = (1/2)i_{sn1} + (1)i_{sn2}$, which represents one-half ampere into node 2, one-half ampere into node 1 and one ampere out of node 3. As a final example, consider a network current of one ampere in the reference direction in branches 1 and 7 and one ampere against the reference direction in branch 3: $j = (1)j_1 + (-1)j_3 + (1)j_7$, and $Gj = 0$, so this current can exist with no current source at all connected to the network (although quite possibly it might be necessary to have a voltage source or two).



Note that these calculations are all completely independent of the sources or causes, so to speak, of the currents, or the branch impedances or voltages. They are relationships which must exist solely for the satisfaction of Kirchoff's current law.

It will be convenient to give the informal designation of tieset to a set of branches, such as branches 1, 3 and 7 of the example, such that the implied corresponding current can exist with no source current connected to the network. A simple example is that of a branch shorted upon itself. The currents of this nature themselves require a more formal definition. Specifically, let J_o be the null space of the map G as defined above. Then a loop current will mean any element of J_o. Exactness of the upper sequence may now be immediately obtained by defining I to be the null space J_o, and F as the injection of I into J.

The incidence matrix may now be used to find some informal geometric interpretations of these spaces. The calculation of the source current necessary for an element of the natural base of J is essentially the operation of determining the node-pair which bounds the given branch which is precisely the information contained in the columns of the incidence matrix. A tieset will then be a set of branches with zero boundary, i.e., zero incidence on every node. Making an identification between the node-pairs and the elements of I_s (for instance, i_{snl} before would be identified with node-pair

n_1-n_2, where n_1 represents one ampere out of node 1, etc.) does allow one useful result to be obtained by an elementary inductive proof: the dimension of I_s is equal to the number of nodes of the network less one. From this it also is immediate that the dimension of I (the tieset space, so to speak) plus the dimension of I_s (the node-pair space, so to speak) must equal the number of branches of the network, which is essentially what is probably the earliest theorem of the field of algebraic topology. Because of this identification of the current sources with the node-pairs, the elements of I_s will be called node-pair current sources.

The identification of I_s with the node-pairs of the network is quite convenient in that the selection of a basis for I_s can be accomplished by specifying an independent set of node-pairs. For example, the selection of node-pairs n_1-n_5, n_2-n_5, n_3-n_5 and n_4-n_5 is equivalent to a selection of a basis for I_s made up of i_{sn4}, i_{sn6}, i_{sn9} and $-i_{sn8}$. The same set of node-pairs could equally well be said to imply other bases for I_s of course, for instance, i_{sn4} might be replaced by $i_{sn1} + i_{sn7} - i_{sn8}$. Note however that $-i_{sn4} + i_{sn1} + i_{sn7} - i_{sn8} = 0$, since the corresponding branches form a tieset. On the other hand, the use of tiesets to aid in the selection of a basis for I is obvious.

For illustrations, let a basis be selected for I_s corresponding to node-pairs n_1-n_4, n_2-n_4, n_3-n_4 and n_5-n_4; and a basis for I corresponding to the following tiesets: 1-3-7,

3-4-8, 2-4-9, 6-7-8, and 1-2-5, i.e., the first basis element is to be $j_1-j_3+j_7$, etc. If the natural basis is used for J, then the matrices of G and F are

$$G = \begin{array}{c|ccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline 1 & + & - & + & + & & & & & \\ 2 & - & & & & - & + & + & & \\ 3 & & + & & & + & & & & + \\ 4 & & & & - & & - & & + & - \end{array}$$

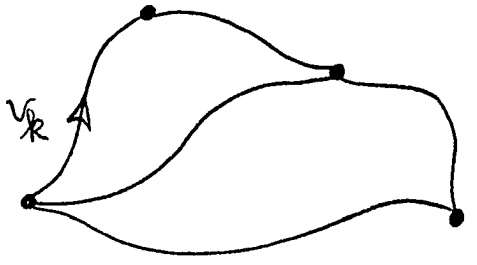
and

$$F = \begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & + & & & & + \\ 2 & & & + & & + \\ 3 & - & + & & & \\ 4 & & - & + & & \\ 5 & & & & & - \\ 6 & & & & + & \\ 7 & + & & & - & \\ 8 & & - & & + & \\ 9 & & & - & & \end{array}$$

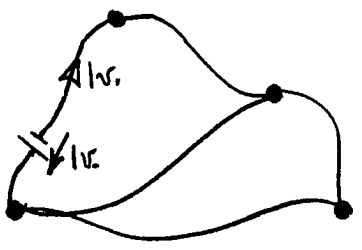
Note that the matrix of G is just the incidence matrix with the fourth row deleted, and the exactness of the sequence is illustrated by the fact that the matrix product GF = 0.

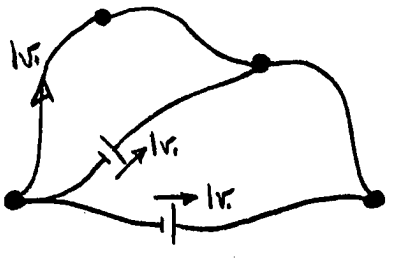
The lower sequence of the diagram is now defined as the dual spaces and adjoint maps of the upper sequence, and it will be necessary to demonstrate that under the identifications to be made the spaces and maps properly represent the network voltages in accordance with Kirchoff's voltage law. The dual basis to the natural basis for J will be called the natural basis of V. Let v_k represent the k'th element of

this basis. Since the corresponding j_k represents a current of one ampere in branch k, it is tentatively assumed that v_k represents a voltage of one volt in the reference direction across branch k, and zero volts across the other branches. For a simple network such a voltage might be indicated by the following sketch:



In this case the diagram implies that there is to be one volt across the marked branch in the direction indicated and zero volts across the other branches, which could be replaced by short-circuits if convenient. If the question is raised as to how such a voltage can exist physically, the answer may be given that naturally some voltage source will be necessary, which might be physically supplied in several ways by insertion of batteries into the network. The following sketches indicate two possibilities of many:



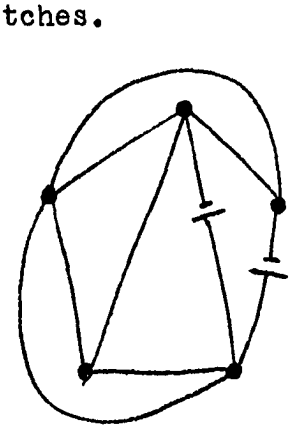


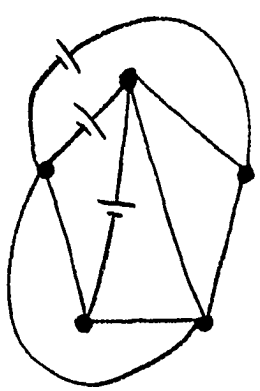
V is now defined as all linear combinations of the natural basis symbols v_k with coefficients from the scalar field, and

any element of V will be called a network voltage.

The elements of the space E_s are to represent the effective source voltages present around the various loops of the network. Suppose a basis has been chosen for I, with i_k a typical element, and the corresponding tieset identified. The tentative identification is made that e_{sk}, the corresponding element of the dual basis for E_s, represents a net plus one volt of loop source voltage around the loop defined by the same tieset, due to physical voltage sources being present in the various branches which make up the tieset, and zero volts around the other tiesets of the basis. Elements of E_s will be called loop voltage sources.

For instance, in the example considered earlier the dual basis element to loop current i_3 would be called e_{s3}, and represents one volt of source potential around tieset 3, and zero volts around the other tiesets corresponding to the other basis elements chosen in I. Two of many possible physical source arrangements are indicated by the following sketches.





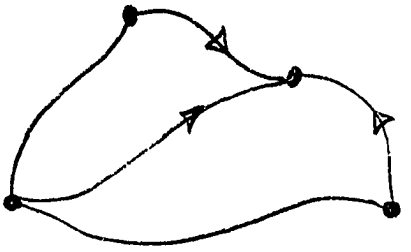
Additional discussion of the implications of these choices for bases is deferred to a later section of this chapter.

To show that the above tentative identifications are correct, it is necessary to show that given some network voltage v, the map F' assigns to v the correct loop voltage source e_s in accordance with Kirchoff's voltage law. The procedure will be to show that the correct e_s is obtained if the network voltage is an element v_k of the natural basis for V, and then the usual extension by linearity will guarantee that the correct result is obtained for any element of V.

Suppose dual bases are in use in I, J, V and E_s, and the natural bases in J and V. The matrix of F' will then be the transpose of the matrix of F, and the k'th column of F' will be the same as the corresponding row of F and indicate which of the various tiesets corresponding to the basis of I the k'th branch is a member. Physically, the condition $e_s = F'v_k$ may be arranged by replacing all branches but the k'th by short circuits and putting a one volt battery in series with branch k, as described earlier. The net loop voltage sources which must then be present will be one volt in every loop which, so to speak, passes through the k'th branch, or making allowance for relative orientations and multiple inclusions of a branch in a loop, plus or minus one volt times the number of times the loop passes through the branch. Now the image of v_k under F' is determined by examining the k'th column of F', and this is precisely the information contained therein, recalling that the given set of tiesets was used to

define the dual bases of both I and E_s. The tentative identifications are therefore justified.

Attention is now directed towards E, the null space of F', and by the discussions of Chapter I, essentially those network voltages which can exist with zero voltage sources. Suppose dual bases in use in J, V, I_s and E, and the natural bases in J and V, and for the sake of discussion the basis in I_s is determined by the specification that it correspond to node-pairs n_1-n_m, n_2-n_m, ..., n_{m-1}-n_m, where m is the number of nodes of the network. The matrix of G will then be the incidence matrix with the bottom row deleted, and the matrix of G' will be the transpose. The rows of the matrix of G may be examined to determine the images of the dual base elements of E corresponding to this choice of node pairs, with the result that these elements of E are essentially those network voltages corresponding to the branches incident upon nodes 1 through m-1 with due regard for orientation. The following example shows the image of a typical element of E for a simple network:



The corresponding set of branches will be called a <u>cutset</u> of the network, so no voltage sources will be needed for any network voltage implied by a cutset. A simple example is

that of an open-circuited branch. Since the basis elements of E can essentially be specified as above by the specification of a set of node-pairs, the elements of E will be called node-pair voltages.

The representation of a network by the diagram is now completed with the stipulation that the map Z shall represent Ohm's Law, i.e., given any j, $v = Zj$ shall be the branch voltage corresponding to the given branch current, or conversely for Y. Recall that both do not need to exist for the diagram to have solutions of interest, and it has been previously stipulated that $Z_o = F'ZF$ and $Y_o = GYG'$, if they exist.

The Roth network problem may now be stated: given e_s, i_s and the various maps, what are j and v such that $e_s = F'v$, $i_s = Gj$ and $v = Zj$? His solution to this problem has been given in Chapter I and shown to be unique when it exists: $j = FZ_o^{-1}e_s + YG'Y_o^{-1}i_s$, and $v = Zj$.

As discussed before, the existence of both Z and Y is not necessary for the following reduced network problems: given only Z and the condition that $i_s = 0$, $j = FZ_o^{-1}e_s$ and $v = Zj$ is a solution; and given only Y and the condition that $e_s = 0$, $v = G'Y_o^{-1}i_s$ and $j = Yv$ is a solution, both solutions being unique if they exist. Y will not exist if any branch is actually a short-circuit, nor Z if any branch is an open-circuit.

A trivial converse network problem might be stated: given v and j such that $v = Zj$, what are e_s and i_s such that

$e_s = F'v$ and $i_s = Gj$? The question answers itself.

Before developing the matrix form of the above results, the ohmicness property (which is a sufficient condition that the inverses called for above exist) will be illustrated, since it will also point up the necessity for careful distinction between vectors and the coordinates of the vectors with respect to a basis.

Suppose every branch of the network to contain a finite, non-zero, passive resistor, R_k representing the resistance of the k'th branch. Let v_k and j_k represent elements of the dual natural bases as before. The image of any j_k under Z may then be written as $Zj_k = R_k v_k$, which is likely to result in some confusion on the part of the electrical reader without detailed knowledge of vector space manipulations, since Ohm's Law for a resistor is usually written as $v = Rj$. This last form is correct if v and j represent the coordinates of vectors of V and J with respect to a pair of bases. The equation involving j_k and v_k above is an abstract equation between the vectors themselves, and as such is correct. For instance, j_k represents a current of one ampere in branch k, and v_k a potential of one volt across branch k. If the resistance is three ohms, for example, and the current is one ampere, the voltage will be three volts, or $3v_k$.

On the other hand, suppose j^k and v^k represented the coordinates of some j and v with respect to these bases. If the vector is j_k, this corresponds to $j^k = 1$ and the other

coordinates zero, and the corresponding $3v_k$ is represented by $v^k = 3$ and the other coordinates zero, and it is seen that in this case $v^k = R_k j^k$.

Suppose now the scalar field in use to be the real numbers, which would be the usual case for a resistive network, so the j^k's and v^k's mentioned before would be real. Any element of J can then be expressed as $j = \sum_k j^k j_k$, and $v = Zj = \sum_k j^k R_k v_k$. If the calculation of the ohmicness condition is now made, using the fact that the j_k's and v_k's are elements of a pair of dual bases, the result is $Zj(j) = \sum_k (j^k)^2 R_k$, which will certainly be non-zero for any non-zero j. Therefore when each branch of the network is made up of a finite, non-zero, passive resistor, the necessary inversions can always be made and solutions to the network problem will always exist, no matter how the branches may be connected together. Note also that in this case ohmicness is equivalent to the net power dissipated in the branches being non-zero, although not necessarily positive. For instance, the necessary non-zero result for any non-zero j would still be obtained if all of the resistances were negative, i.e., active elements. Even further, the gyrator may be mentioned as an excellent example of a network which is not ohmic, but still can be solved in many cases of interest.

In general the calculation of the power dissipated in the branches of the network takes slightly different forms depending upon the scalar field and whether transient or

steady-state conditions are being considered, for instance. Assuming the scalar field to be the reals, and J and V to represent actual instantaneous currents and voltages, $v(j)$ will be the correct power calculation. This can be put into another form by use of a dual direct-sum decomposition of J and V as follows: $v(j) = v_s(j_o+j_s) + v_o(j_o+j_s) = v_s(j_o) + v_o(j_s)$ or $v(j) = v_s(Fi) + G'e(j_s) = F'v_s(i) + e(Gj_s) = e_s(i) + e(i_s)$, which may be interpreted as the power dissipated in the branches must equal the power supplied by the sources.

If the dual basis stipulation is not made, but only that J_o and V_o be mutual annihilators, then the result is $v(j) = e_s(i) + e(i_s) + v_s(j_s)$. If there are only voltage sources present $j_s = 0$, and the result becomes $v(j) = e_s(i)$. If there are only current sources present $v_s = 0$, and the result becomes $v(j) = e(i_s)$.

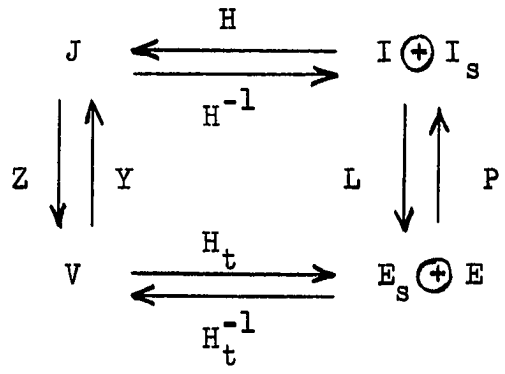
Matrix Interpretation

As might be expected from the discussions of Chapter I, the proper choice of bases for the various spaces will simplify the treatment. The usual method of stating the Ohm's Law expressions for the branches of a network is in terms of the natural bases for V and J, so Z will be used to indicate the matrix of these expressions, the J to V mapping. Likewise Y will be used to indicate the V to J mapping with respect to the natural bases.

The useful bases for the center spaces turn out to be those which allow the subspaces of the center spaces to be

identified with the end spaces in the fashion of Chapter I. Specifically, a split base for J is defined as one made up in part by the images under F of whatever basis is used for I, with corresponding ordering, and likewise a split base for V is defined as one made up in part of the images under G' of whatever basis is used for E, with corresponding ordering. If the further stipulation is made that the bases be dual bases in all cases then the coordinates of a vector in I will be the same as the coordinates of its image in J_o, so the two spaces may be identified, and likewise for J_s and I_s, E_s and V_s, and V_o and E. The dual basis restriction will be somewhat relaxed in a later discussion.

Let L represent the matrix of Z with respect to such a dual set of bases, and likewise P the matrix of Y. Let H be the matrix whose columns give the expressions for the elements of the split base of J in terms of the natural basis. Since dual bases have been specified, the transpose of H, H_t, will be the change of basis matrix from the natural basis for V to the split basis, i.e., the rows of H will give the expressions for the elements of the natural basis for V in terms of the elements of the dual split basis. Then $L = H_t Z H$ and $P = H^{-1} Y (H_t)^{-1}$. These relations may be conveniently summarized by the sketch:



The matrix expressions of Chapter I were

$$\begin{bmatrix} e_s \\ e \end{bmatrix} = \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} \begin{bmatrix} i \\ i_s \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} i \\ i_s \end{bmatrix} = \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{bmatrix} e_s \\ e \end{bmatrix}.$$

The network problem then takes the form, given e_s and i_s, find e and i such that these expressions hold.

Working with the L matrix first, the top part may be solved for i, giving $i = L_1^{-1}(e_s - L_2 i_s)$, and this put into the bottom part to get $e = L_3 L_1^{-1}(e_s - L_2 i_s) + L_4 i_s$, which solves the problem. Notice that if $i_s = 0$ the top part is just $e_s = L_1 i$, which identifies L_1 as the usual loop matrix of the network, with respect to this set of bases. Note that L_2 acts on i_s to produce an equivalent set of voltage sources.

The corresponding operations with the P matrix give $e = P_4^{-1}(i_s - P_3 e_s)$ and $i = P_1 e_s + P_2 P_4^{-1}(i_s - P_3 e_s)$, and P_4 is the usual node matrix of the network with respect to this set of bases, since if $e_s = 0$, $i_s = P_4 e$. Note that P_3 acts on e_s to produce an equivalent set of current sources.

A number of results concerning these matrices may be found in Chapter I, including the explicit expressions for

the inverse of the loop matrix in terms of the inverse of the node matrix, and vice-versa.

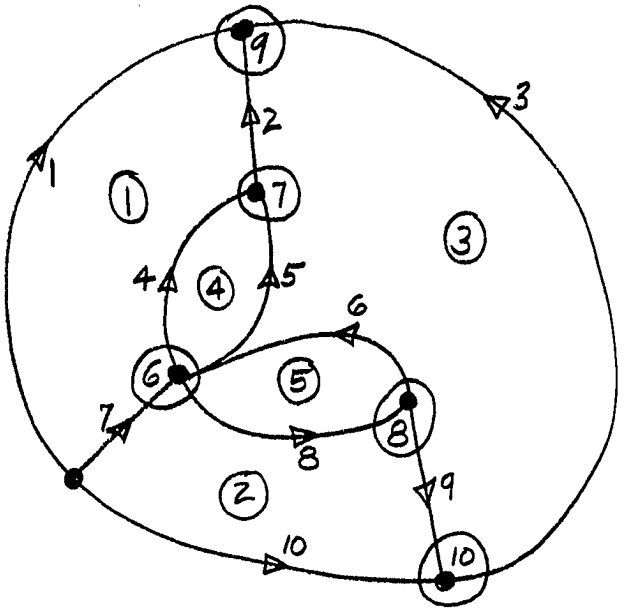
As far as practical circuit solutions are concerned, the Roth formulation adds nothing in the sense it does not provide a means of obtaining the solution to a network problem by the inversion of a matrix smaller than the usual loop or node matrix. It does explicitly give the relationships between the voltage and current spaces of the network, and in particular gives the expressions for the inverses of the loop and node matrices each in terms of the other, which is occasionally of considerable value. However, the real contribution of the Roth formulation is to the theoretical understanding of the network problem, and to the analysis of more complicated situations such as are considered in the next chapter.

In general a network will be said to be solved if either the loop or node matrix has been inverted, since then all the remaining manipulations can be performed without the need of additional inversions. However, this statement must be qualified in the case that the P and L matrices may not be equally easy to obtain due to an unusually complicated set of mutual couplings between the branches. Such cases are not common. If the network contains energy storage elements, an additional step is necessary--the inversion of a Laplace transform. In terms of large-scale computers this inversion may or may not require a significant amount of computing effort.

If only sinusoidal steady-state answers are needed, the calculations are relatively simple. Transient calculations will require that the characteristic polynomial be factored to obtain the natural frequencies of the network.

In order to consolidate the discussion to this point, two examples will now be given in which completely dual bases will be used. The following section will then treat a number of other possible manipulations of bases.

For the network shown, let the dual bases for the various spaces be as indicated by the encircled numbers.



For the bases implied by the circles, the change of basis matrix H may take the form:

$H = J$

	I					I_s				
	1	2	3	4	5	6	7	8	9	10
1	+								+	
2	-		+							
3			-							
4	-			+			+			
5			+	-						
6			+		-					
7	-	+				+	+	+		
8		+			-			+		
9		+	-							
10		-								+

which may be inverted and transposed to give

$H_t^{-1} = V$

	E_s					E				
	1	2	3	4	5	6	7	8	9	10
1									+	
2	-						-		+	
3	-	-	-	-	-				+	-
4						-	+			
5				-		-	+			
6					-	+		-		
7						+				
8						-		+		
9		+						-		+
10										+

As an aid in checking these matrices the diagram on page 58 is useful. Some comments on the implications of this procedure may be found in the next section.

If each branch of the network is taken to contain a one ohm resistor, the Z and Y matrices are identity matrices. The

L and P matrices are then found to be:

$L =$

		I				I_s				
		1	2	3	4	5	6	7	8	9	10
E_s	1	4	-	-	-		-	$\bar{2}$	-	+	
	2	-	4	-		-	+	+	2		-
	3	-	-	5	-	-					
	4	-		-	2			+			
	5		-	-		2			-		
E	6	-	+				+	+	+		
	7	$\bar{2}$	+		+		+	2	+		
	8	-	2			-	+	+	2		
	9	+								+	
	10		-								+

$P =$

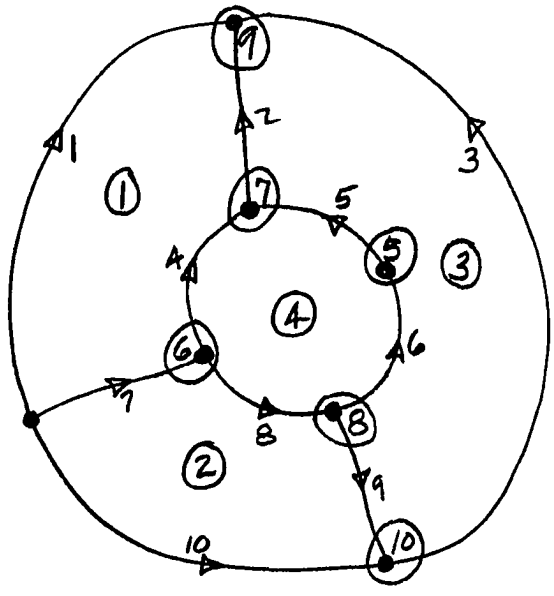
		E_s					E				
		1	2	3	4	5	6	7	8	9	10
I	1	2	+	+	+	+		+		$\bar{2}$	+
	2	+	2	+	+	+			-	-	2
	3	+	+	+	+	+				-	+
	4	+	+	+	2	+	+	-		-	+
	5	+	+	+	+	2	-		+	-	+
I_s	6				+	-	5	$\bar{2}$	$\bar{2}$		
	7	+			-		$\bar{2}$	3		-	
	8		-			+	$\bar{2}$		3		-
	9	$\bar{2}$	-	-	-	-		-		3	-
	10	+	2	+	+	+			-	-	3

The inverses of the loop and node matrices are:

$$L_1^{-1} = \frac{1}{99}\begin{vmatrix} 47 & 25 & 27 & 37 & 26 \\ 25 & 47 & 27 & 26 & 37 \\ 27 & 27 & 45 & 36 & 36 \\ 37 & 26 & 36 & 86 & 31 \\ 26 & 37 & 36 & 31 & 86 \end{vmatrix}$$

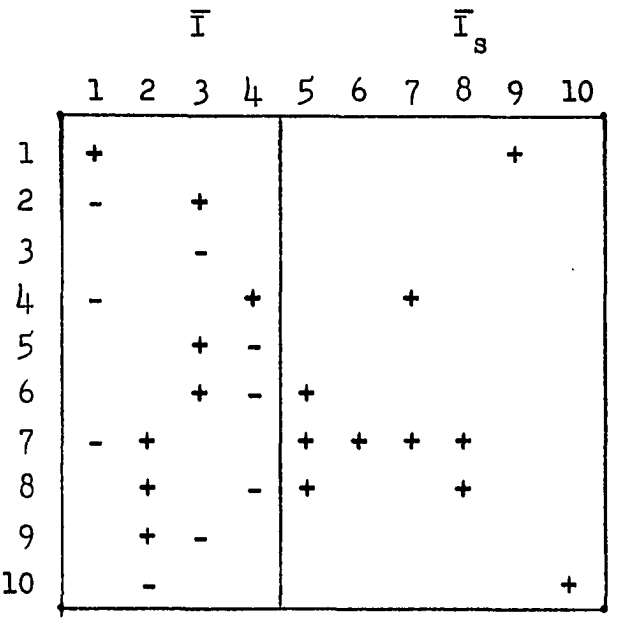
$$P_4^{-1} = \frac{1}{99}\begin{vmatrix} 55 & 44 & 44 & 22 & 22 \\ 44 & 73 & 37 & 32 & 23 \\ 44 & 37 & 73 & 23 & 32 \\ 22 & 32 & 23 & 52 & 25 \\ 22 & 23 & 32 & 25 & 52 \end{vmatrix}$$

The network of the previous example will be slightly rearranged to form the network for another example, as follows:

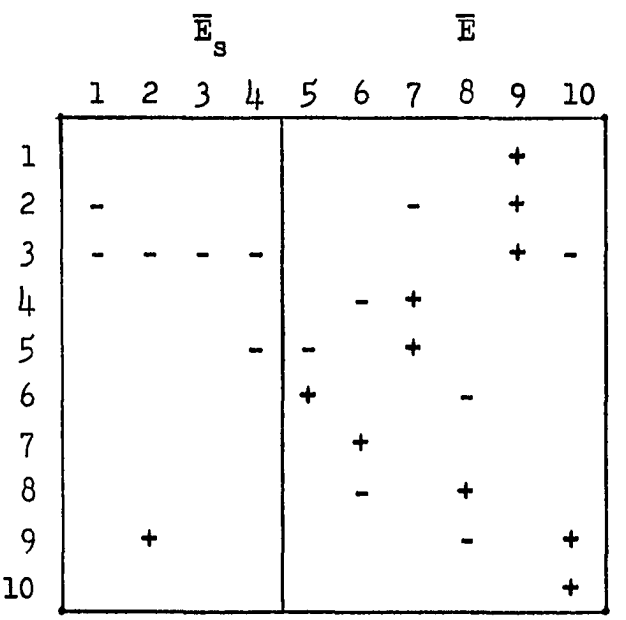


To avoid confusion with the first example all symbols for this example will be overlined. The same matrices as before may be calculated, and will be found to be as follows:

$\overline{H} = \overline{J}$



$\overline{H}_t^{-1} = \overline{V}$



$\bar{L}$ =

		$\bar{I}$				$\bar{I}_s$					
		1	2	3	4	5	6	7	8	9	10
$\bar{E}_s$	1	4	-	-	-	-	-	$\bar{2}$	-	+	
	2	-	4	-	-	2	+	+	2		-
	3	-	-	5	$\bar{2}$	+					
	4	-	-	$\bar{2}$	4	$\bar{2}$		+	-		
$\bar{E}$	5	-	2	+	$\bar{2}$	3	+	+	2		
	6	-	+			+	+	+	+		
	7	$\bar{2}$	+		+	+	+	2	+		
	8	-	2		-	2	+	+	2		
	9	+								+	
	10		-								+

$\bar{P}$ =

		$\bar{E}_s$				$\bar{E}$					
		1	2	3	4	5	6	7	8	9	10
$\bar{I}$	1	2	+	+	+			+		$\bar{2}$	+
	2	+	2	+	+				-	-	2
	3	+	+	+	+					-	+
	4	+	+	+	2	+		-		-	+
$\bar{I}_s$	5				+	2		-	-		
	6						3	-	-		
	7	+			-	-	-	3		-	
	8		-			-	-		3		-
	9	2	-	-	-			-		3	-
	10	+	2	+	+				-	-	3

$$\bar{L}_1^{-1} = \frac{1}{110} \begin{bmatrix} 51 & 29 & 30 & 35 \\ 29 & 51 & 30 & 35 \\ 30 & 30 & 50 & 40 \\ 35 & 35 & 40 & 65 \end{bmatrix}$$

$$\overline{P}_4^{-1} = \frac{1}{110}\begin{vmatrix} 121 & 44 & 66 & 66 & 33 & 33 \\ 44 & 66 & 44 & 44 & 22 & 22 \\ 66 & 44 & 86 & 46 & 38 & 28 \\ 66 & 44 & 46 & 86 & 28 & 39 \\ 33 & 22 & 38 & 28 & 59 & 29 \\ 33 & 22 & 28 & 38 & 29 & 59 \end{vmatrix}$$

Bases for Spaces

Many of the most interesting (and possibly confusing) aspects of network analysis fall in the general area of determining appropriate bases for the various vector spaces discussed above. The following comments are intended as an indication of some of these aspects.

Some standard results may be stated here for convenient reference. It is first recalled that given some map F between two spaces and the adjoint map F' between the dual spaces in the opposite direction, if dual bases are used in both pairs of spaces, the matrix of F' is the transpose of the matrix of F.

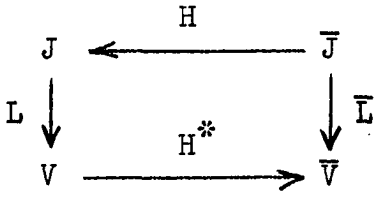
Let X and Y represent two bases for some space, and X' and Y' the corresponding dual bases for the dual space. Suppose the coordinates of some vector to be given with respect to base X, and it is desired to find the coordinates of the same vector with respect to base Y. Let a matrix A be constructed whose i'th column is made up of the coordinates of the i'th element of the base X with respect to the base Y.

If the coordinates of any vector with respect to X are now arranged as a column matrix and multiplied from the left by A, the result will be the column matrix of the coordinates of the vector with respect to Y. In the dual space, the transpose of A serves in the same way to transform the coordinates of a vector of the dual space with respect to Y' into the coordinates of the vector with respect to X'. Note that (as with adjoint maps) there is a reversal of direction in the dual space.

The above statements suggest that the use of dual bases will simplify the various matrix expressions encountered in analyses of vector space mappings, but it should be realized that there is no logical necessity for using dual bases. The use of dual bases can introduce other problems. Consider the examples of the preceding section. The split bases for V and J were obtained by first choosing the basis for J, which determined the matrix H, and then H was inverted and transposed to find the corresponding dual basis for V. This would seem to imply that the use of dual bases means that in only one of the two spaces is there any control over the basis elements, the other basis being determined by the dual requirement. This is essentially correct, although it will be a major point of the following discussion to show that at least those parts of the dual bases from E and I_s more or less naturally go together and likewise for I and E_s. However, the initial part of the discussion will be in as general a context as feasible.

Attention may be immediately directed to one part of the preceding discussion in which a dual basis requirement may be eliminated. The identification of I and J_o, from the matrix point of view, was accomplished by using the images under F of the basis elements of I for a basis of J_o, and therefore part of the basis for J. The identification of E_s and V_s then followed from the dual basis stipulation. This same type of identification will result from the milder stipulation that the images under F' of the basis elements of V from V_s be used as the basis for E_s. While J_o is fixed, being the image space of F, any V_s complementary to V_o may be used. Similar comments apply to the node end of the diagram.

Suppose L to be the matrix of a map between a space J and its dual space V with respect to some pair of bases (not necessarily dual), and it is desired to know the matrix $\bar{L}$ of the same map with respect to some different pair of bases (again not necessarily dual). By a slight abuse of notation, let J and V represent coordinates expressed with respect to the old bases, $\bar{J}$ and $\bar{V}$ coordinates with respect to the new bases, H the change of basis matrix from the new basis $\bar{J}$ to the old basis J, and H^* the change of basis matrix from the old basis V to the new basis $\bar{V}$. The relations may be conveniently indicated by the sketch:



The change of basis matrices H and H^* will be transposes only if it is specified that dual bases are to be used. The columns of H will be made up of the coordinates of elements of the new basis expressed with respect to the old basis elements, while the columns of H^* will be made up of the coordinates of the old basis elements expressed with respect to the new basis elements. The sketch may then be read in the same style as a map diagram to determine that $\overline{L} = H^*LH$ and $L = H^{*-1}\overline{L}H^{-1}$. If the basis is actually changed in only one of the two spaces, then the matrix corresponding to the space in which the basis stays the same becomes an identity matrix.

Any non-singular matrix may be used for H or H^* above and therefore determines a change of basis by implication. As a result of this it can be said that a number of theorems which have appeared in the electrical literature stating that, for instance, the determinant of a matrix of the loop equations of a network may only have certain values, can only be correct under additional assumptions about the bases under consideration.

As stated so far the change of basis calculations have not been directly concerned with the bases of most practical interest, which are changes of bases for the loop currents and the node-pair voltages. Since the loop and node equations are actually formed as parts of larger sets of equations, it will be necessary to approach the desired changes by way of changes of basis for the entire space. Suppose it

is desired to change from one set of split bases to another. (It will be recalled that the purpose of a split base is to allow the matrix identification of I and J_o, and E_s and V_s, or at the other end of the diagram J_s and I_s, and V_o and E. In this case the split base identifications are to be accomplished by using images of basis elements as basis elements as described earlier in this section, rather than by the use of dual bases.) To extend the previous abuse of notation a little further, let I and I_s represent coordinates with respect to the old split basis, $\bar{I}$ and $\bar{I}_s$ coordinates with respect to the new split basis, and likewise for E and E_s, $\bar{E}$ and $\bar{E}_s$. Because of the split basis restriction I and $\bar{I}$ refer to different bases for the same space, and likewise for E and $\bar{E}$. However, since the complementary spaces may change, I_s and $\bar{I}_s$ do not need to refer to the same space, nor E_s and $\bar{E}_s$.

If the change of basis equation is now written out, with the change of basis matrices H and H^* partitioned to agree with the partitioning established earlier for the L matrices, the result is

$$\begin{bmatrix} \bar{L}_1 & \bar{L}_2 \\ \bar{L}_3 & \bar{L}_4 \end{bmatrix} = \begin{bmatrix} H_1^* & H_2^* \\ H_3^* & H_4^* \end{bmatrix} \begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} \begin{bmatrix} H_1 & H_2 \\ H_3 & H_4 \end{bmatrix} .$$

Multiplying out the partitioned matrices, the expression of interest is $\bar{L}_1 = H_1^* L_1 H_1 + H_1^* L_2 H_3 + H_2^* L_3 H_1 + H_2^* L_4 H_3$. Now the fact that I and $\bar{I}$ refer to bases for the same space guarantees that H_3 is zero, and likewise the condition that E and $\bar{E}$

refer to bases of the same space gives that H_2^* is zero. H_2 and H_3^* will not be zero if the respective complementary spaces are changed. However, the two zero matrices are enough to reduce the expression to $\bar{L}_1 = H_1^* L_1 H_1$, and so it is only necessary to determine two matrices to find the new loop matrix. If the further restriction is made that dual bases be used, then only one matrix must be found, the change of basis matrix H_1 in I, whose columns are the coordinates of the new basis elements of I in terms of the old basis elements.

An exactly similar calculation can be made at the node end of the diagram. Letting $K = H^{-1}$, and $K^* = H^{*-1}$, the matrix equation becomes

$$\begin{bmatrix} \bar{P}_1 & \bar{P}_2 \\ \bar{P}_3 & \bar{P}_4 \end{bmatrix} = \begin{bmatrix} K_1 & K_2 \\ K_3 & K_4 \end{bmatrix} \begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} \begin{bmatrix} K_1^* & K_2^* \\ K_3^* & K_4^* \end{bmatrix}.$$

In this case K_3 and K_2^* are zero, and $\bar{P}_4 = K_4 P_4 K_4^*$. If dual bases are used, it is only necessary to determine K_4^*, whose columns are the coordinates of the new basis of E in terms of the old basis.

The formation of the node and loop matrices, as expressed by the sketch on page 58, can be treated in an almost identical fashion. Letting $K = H^{-1}$, etc., as before, the basic equations are $L = H^* Z H$ and $P = KYK^*$. If the H and K matrices are conformably partitioned into <u>two</u> rectangular submatrices each, the expressions take the forms:

$$\begin{bmatrix} L_1 & L_2 \\ L_3 & L_4 \end{bmatrix} = \begin{bmatrix} H_a^* \\ H_b^* \end{bmatrix} Z \begin{bmatrix} H_a & H_b \end{bmatrix} \qquad \text{and}$$

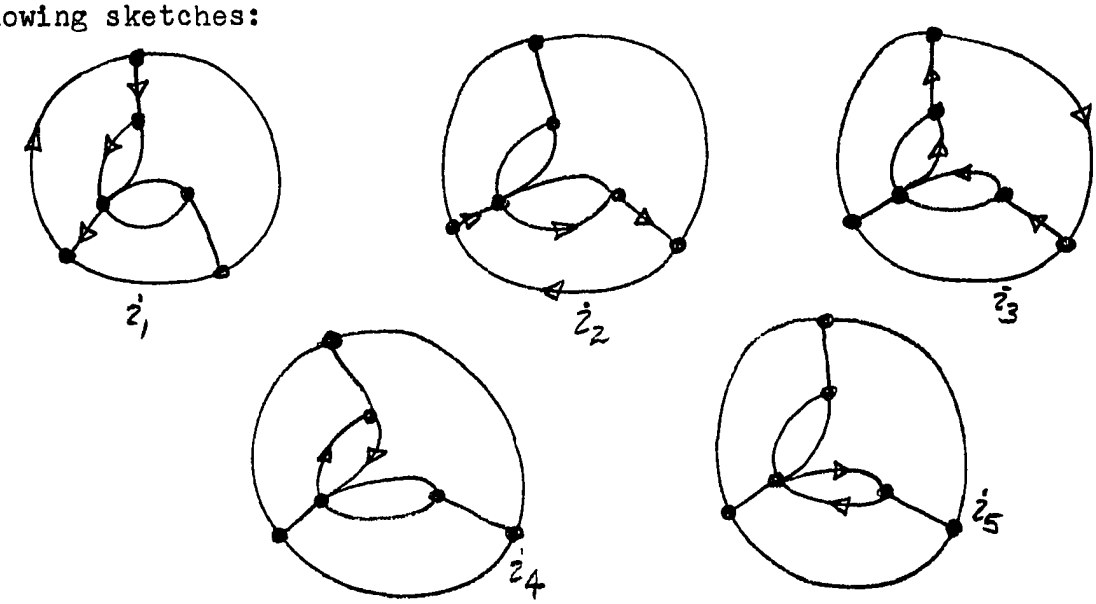
$$\begin{bmatrix} P_1 & P_2 \\ P_3 & P_4 \end{bmatrix} = \begin{bmatrix} K_a \\ K_b \end{bmatrix} Y \begin{bmatrix} K_a^* & K_b^* \end{bmatrix} .$$

The final expressions are $L_1 = H_a^* Z H_a$ and $P_4 = K_b Y K_b^*$. With respect to dual bases, to form the loop matrix it is only necessary to know H_a, whose columns are the coordinates of the loop current basis elements expressed with respect to the natural basis elements. Likewise in forming the node matrix with respect to dual bases, it is only necessary to know K_b^*, whose columns are the coordinates of the node-pair voltage basis elements expressed with respect to the natural basis elements.

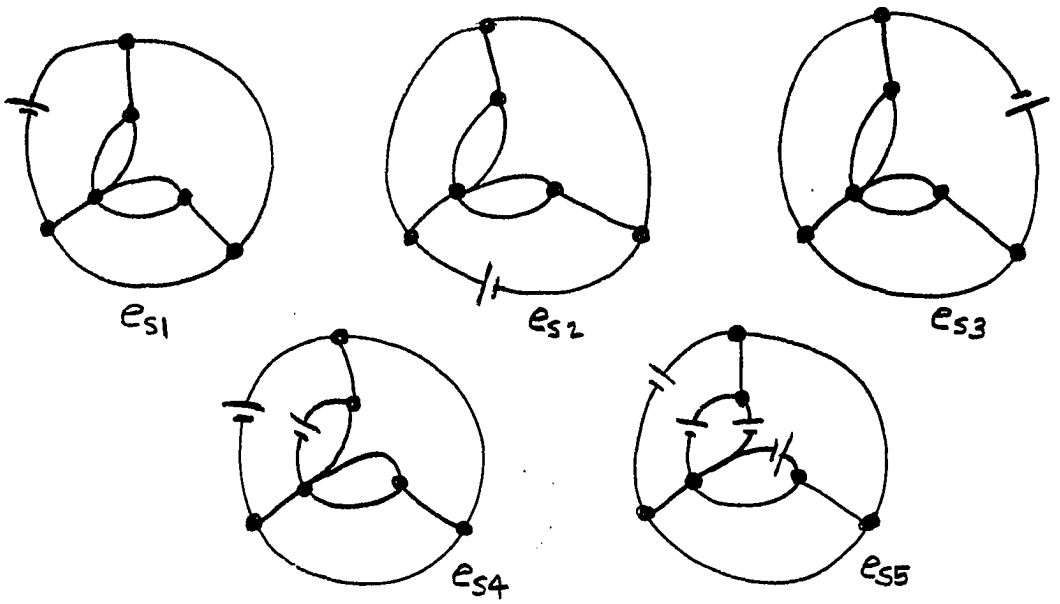
It has been amply demonstrated that the use of dual bases may bring about an appreciable simplification in the preceding equations, so the remainder of this chapter will be devoted to discussion and examples of dual bases. In particular, it will be useful to discuss the dual basis of E_s corresponding to a given basis of I and likewise the dual basis of I_s corresponding to a given basis of E.

The identification of a pair of dual bases for I and E_s was an essential part of the demonstration that the Roth Diagram represented the network problem. The precise

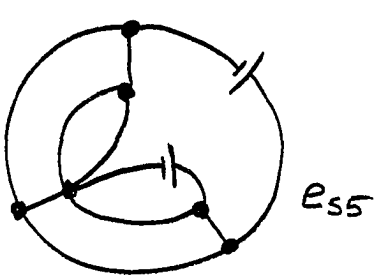
procedure may be restated: The selection of a basis for I is equivalent to the selection of a basis for the tiesets of the network, and the dual basis element to a given basis element i_s is to be a net plus one volt of source potential in the branches of the same tieset, and a net zero volts around each of the other tiesets of the basis. The first complete example considered will be used as a source of examples. The basis for I chosen there may be conveniently indicated by the following sketches:



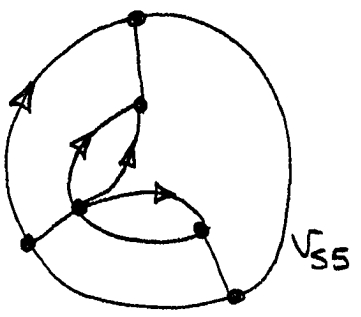
where, as before, an arrow on a branch indicates one ampere in the direction indicated, and the unmarked branches are to have zero current. The following sketches indicate a possible selection of actual physical one volt voltage sources to set up each of the dual basis elements of E_s.



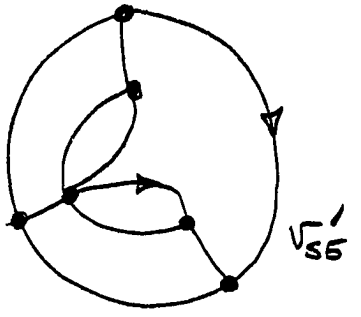
An alternative and somewhat more correct interpretation of these sketches is that they represent a basis for a possible V_s, i.e., v_{s4} would be one volt across each of the two branches in the direction shown, and its image under F' is e_{s4}. Note that the selection of a basis for I immediately determines the unique dual basis for E_s, but there will normally be many possible v_s's such that the images of its basis elements are the basis elements of E_s. For instance, if the physical sources to set up e_{s5} were taken in the following manner



the resulting implied branch voltage would change from



to

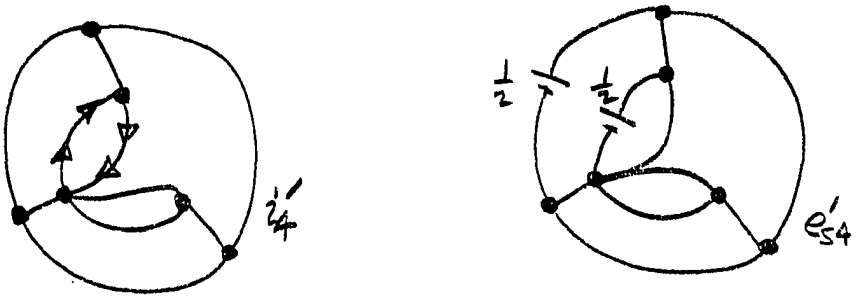


where the arrows represent plus one volt across the branches in the direction indicated. Inspection will show that v'_{s5} cannot be expressed in terms of the original implied v'_ss, so its use in place of v_{s5} would definitely correspond to a new subspace V_s, and not just a change of basis in the old subspace.

It may be helpful to keep in mind that it is not the actual physical voltage sources themselves which enter into the equations, so to speak, but abstract symbols representing the net source voltage around a loop. In the case of current sources the same type of statement may be made, but matters can remain on a fairly concrete level since the current sources from the very start are closely tied to the network nodes. If it were assumed that only planar networks would be considered, the loops of the network could be tied to a specific set of planar meshes, and the same conveniences enjoyed as in the case of current sources.

The check that a given representation of an e_s properly induces the correct voltages in the loops is of course just a matter of checking around each loop in turn. In general,

if some rather complicated new basis were chosen for I, and it was desired to find the implied new dual basis for E_s, it would likely be most efficient to find the transposed inverse of the change of basis matrix for I, which becomes the change of basis matrix for E_s. In simple cases, direct inspection is adequate. Suppose that a change of basis were made in which all the basis elements for I remained fixed, except that the new $i'_4 = 2i_4$. The significant corresponding change in the dual basis is that $e'_{s4} = (1/2)e_{s4}$, which may be sketched as

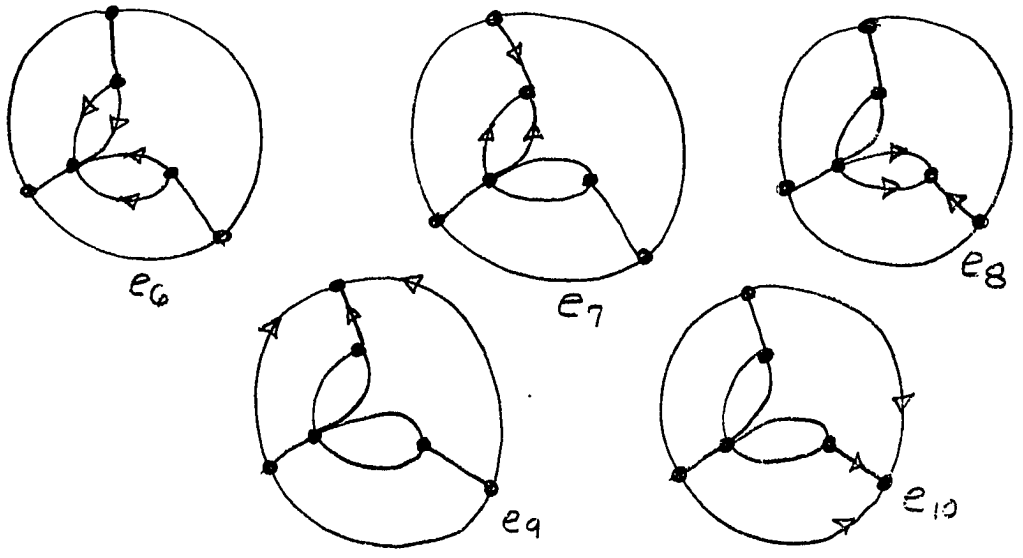


Checking on the induced voltages in the loops, there will still be zero voltage induced in all but the fourth loop, and plus one volt in the fourth loop, since it must be traversed twice.

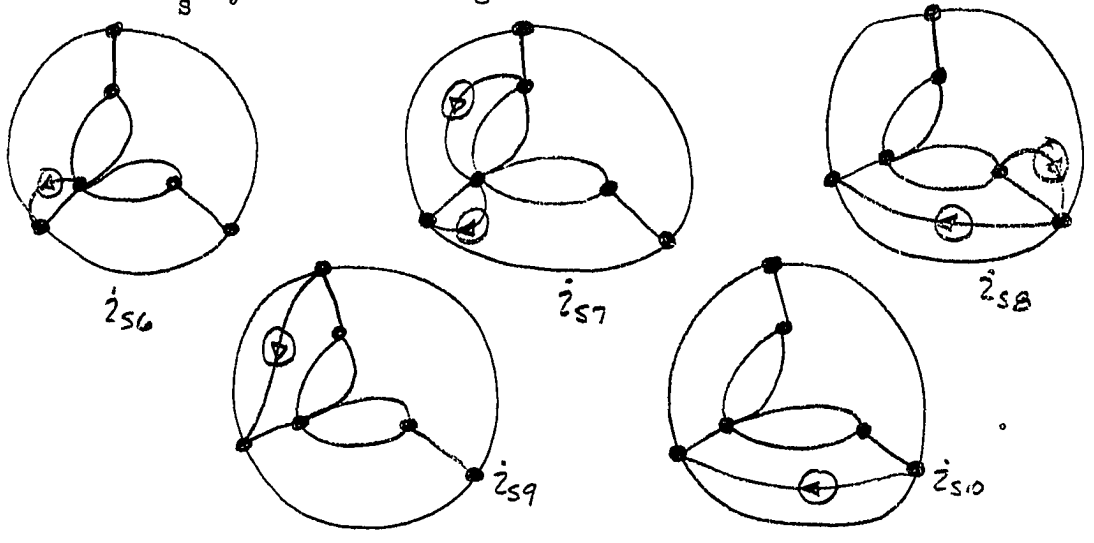
When a given branch appears more than once in a tieset, i.e., the corresponding loop current is set up by a current of other than one ampere in the particular branch, an alternative point of view is sometimes helpful. The basic condition for a set of dual bases is that $e_{sm}(i_n)$ is to be plus one if m = n, and zero otherwise. For any e_{sm} there is at least one v_{sm} such that $e_{sm} = F'v_{sm}$ so the expression can be written as $F'v_{sm}(i_n) = v_{sm}(Fi_n)$, and since F is the injection of I into J this may just as well be written as $v_{sm}(i_n)$. The

determination of the dual basis of E_s corresponding to a given basis of I may thus be referred back to the branch spaces, and essentially involves an expression of the form v(j). Now an earlier discussion showed that in at least some cases this is the expression for the power dissipated in the branches, so the determination of some e_{sm} is then a matter of finding a branch voltage which gives one watt of power dissipation if loop current i_m is the network current, and zero watts if the network current is any other loop current of the basis. It is now obvious why the voltage source of the immediately preceding example had to be reduced by the same factor the dual loop current was increased.

Now consider current sources. A basis for E is equivalent to selecting a basis for the cutsets. The corresponding dual basis element of I_s to a given basis element of E is to be a net plus one ampere of source current across the branches of the corresponding cutset, and a net zero amperes across each of the other cutsets of the basis. It was for convenience in making this calculation that an earlier stipulation was made that current sources were always to be bridged across single branches. By way of example, the basis for E of the network used before may be indicated by the following sketches:

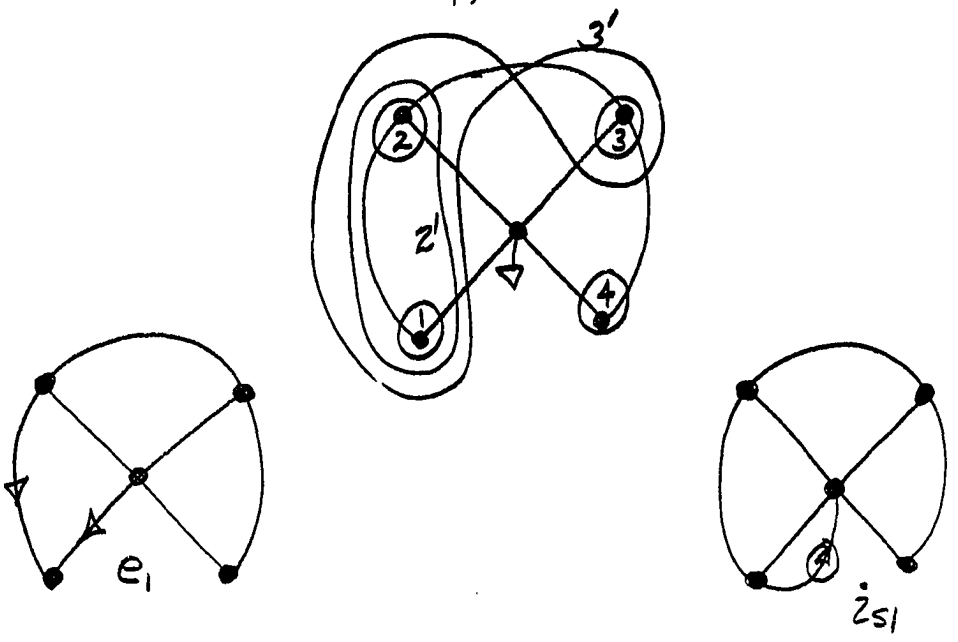


and a possible physical realization of the corresponding dual basis of I_s by the following sketches:

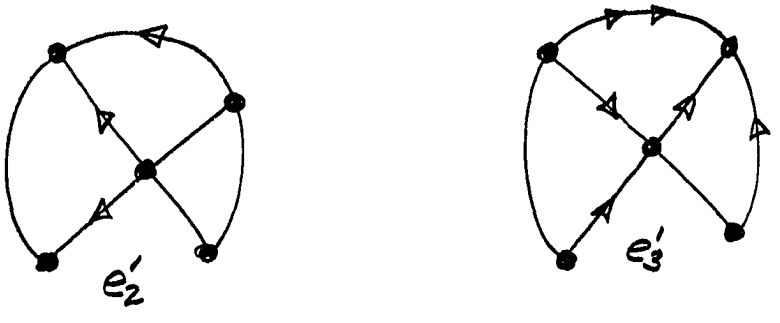


Comments corresponding to those made about the voltage sources apply here.

As a final example of current source manipulation, consider the following network. The small circles identified by unprimed numbers may be taken to indicate a node to datum set of variables. For instance, e_1 and i_{s1} are as indicated by the sketches below the network.



Suppose it is now desired to change to a new set of variables in which e_1 and e_4 are to be unchanged, but e_2 and e_3 are to be changed to the following:



and it is desired to determine the corresponding new dual basis for the current sources.

Let the new variables be indicated by primes. The precise changes of variable may then be indicated by the following two matrices, the second being obtained by transposing the first:

	e'_1	e'_2	e'_3	e'_4
e_1	+	+	-	
e_2		+	-	
e_3			+	
e_4				+

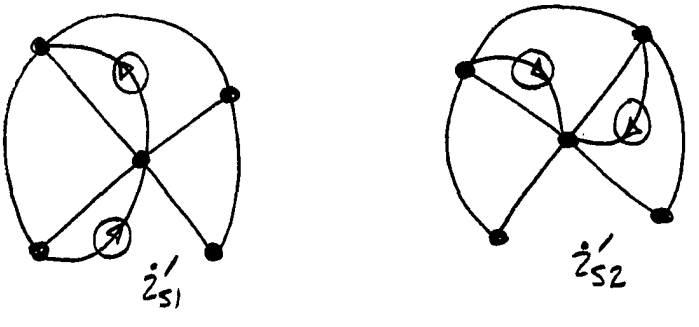
	i_{s1}	i_{s2}	i_{s3}	i_{s4}
i'_{s1}	+			
i'_{s2}	+	+		
i'_{s3}	-	-	+	
i'_{s4}				+

The other two sets of relations are obtained by inversion, giving:

	e_1	e_2	e_3	e_4
e'_1	+	-		
e'_2		+	+	
e'_3			+	
e'_4				+

	i'_{s1}	i'_{s2}	i'_{s3}	i'_{s4}
i_{s1}	+			
i_{s2}	-	+		
i_{s3}		+	+	
i_{s4}				+

The first two current sources of the new basis are essentially different from those of the original basis, and may be indicated as follows:



An additional interpretation of a dual basis element suggests itself from these sketches. The basis for E was indicated in each case by drawing closed curves. If these curves are regarded as defining closed surfaces, each element of the basis of E may be associated with a closed surface, and the

corresponding element of I_s is to be one ampere across the surface in the positive direction, and zero amperes across the other surfaces.

An additional example of a large-scale change of basis may be found in the final section of the dissertation.

CHAPTER III

THE REARRANGEMENT OF NETWORKS

It seems quite logical that if a network whose solution is known is rearranged slightly, the solution of the rearranged network should be obtainable from the original solution by an amount of calculation more or less proportional to the amount of rearrangement. This chapter will demonstrate that this is indeed the case, and will give the specific formulas whereby the solution of either network may be used to aid in the solution of the other, in a sense to be made precise.

The concept of finding the solution of a network by utilizing the solution of a "torn" version seems to have been emphasized first by Kron (8). However he does not seem to have published any consistent method whose range of applicability can be determined, much less proofs that the necessary steps can be performed. Several authors have commented upon the difficulties which many readers have in understanding some of Kron's publications (9, 10). His desire to develop methods to handle large networks is understandable if it is recalled that numerical solutions to large-scale physical

problems are often determined by consideration of an electric network model of the physical system, and a good representation often requires many branches in the network model. Kron has published a number of such network models of differential equations (11-13). In the technical discussion to follow, explicit formulas will be given for the inverses of the loop and node matrices of the original network in terms of the corresponding inverses of the torn network, and vice-versa. As might be expected from the discussions of the previous chapter, the calculations are closely related to those of inverting a matrix by bordering or partitioning, and in fact take the same numerical form in some cases. The use of these techniques allows known network solutions to be used to aid in the solution of rearranged versions of the network, or conversely, a network and a rearranged version can both be solved for the same cost as that of solving the original network plus a cost which is related to the difference between the original network and the rearranged version.

The development of the formulas will be based upon another diagram due to Roth, the so-called Squirrel Cage Diagram, to be given shortly. The Squirrel Cage Diagram is essentially a representation of the fact that the diagrams of the two networks, as given in the previous chapter, may be embedded in a larger commutative diagram if the rearrangement takes either of the forms of "tearing" or "pinching", to be described. Roth's abstract analysis of the Squirrel

Cage Diagram (7) is inadequate, his discussion merely stating that there is some solution of the "torn" network which will give the desired solution of the original network, which is obvious from the fact that the two networks are required to have the same branches and has nothing to do with "tearing" or the diagram. In particular he does not show how to specify the network problem for the torn version in a useful fashion such that the solution of the torn network will give the desired solution of the original network, which will be seen to be a central point in the discussion to follow, and to require that the diagram as given by Roth be augmented with additional spaces and maps. Roth's matrix analysis of the Squirrel Cage Diagram (3) is similarly inadequate. For instance, if Equation 7.5 of (3) is to be correct, it is necessary to make the stipulation that (in his notation) $e_o = e^{\#} = 0$, the meaning of which is not at all clear, but seems to imply that voltage sources will be allowed in only part of the network. Likewise, he gives no analysis of the conditions under which the matrix inversion (L_1^{-1}, in his notation) can be performed, and his examples are incomplete.

The Squirrel Cage Diagram

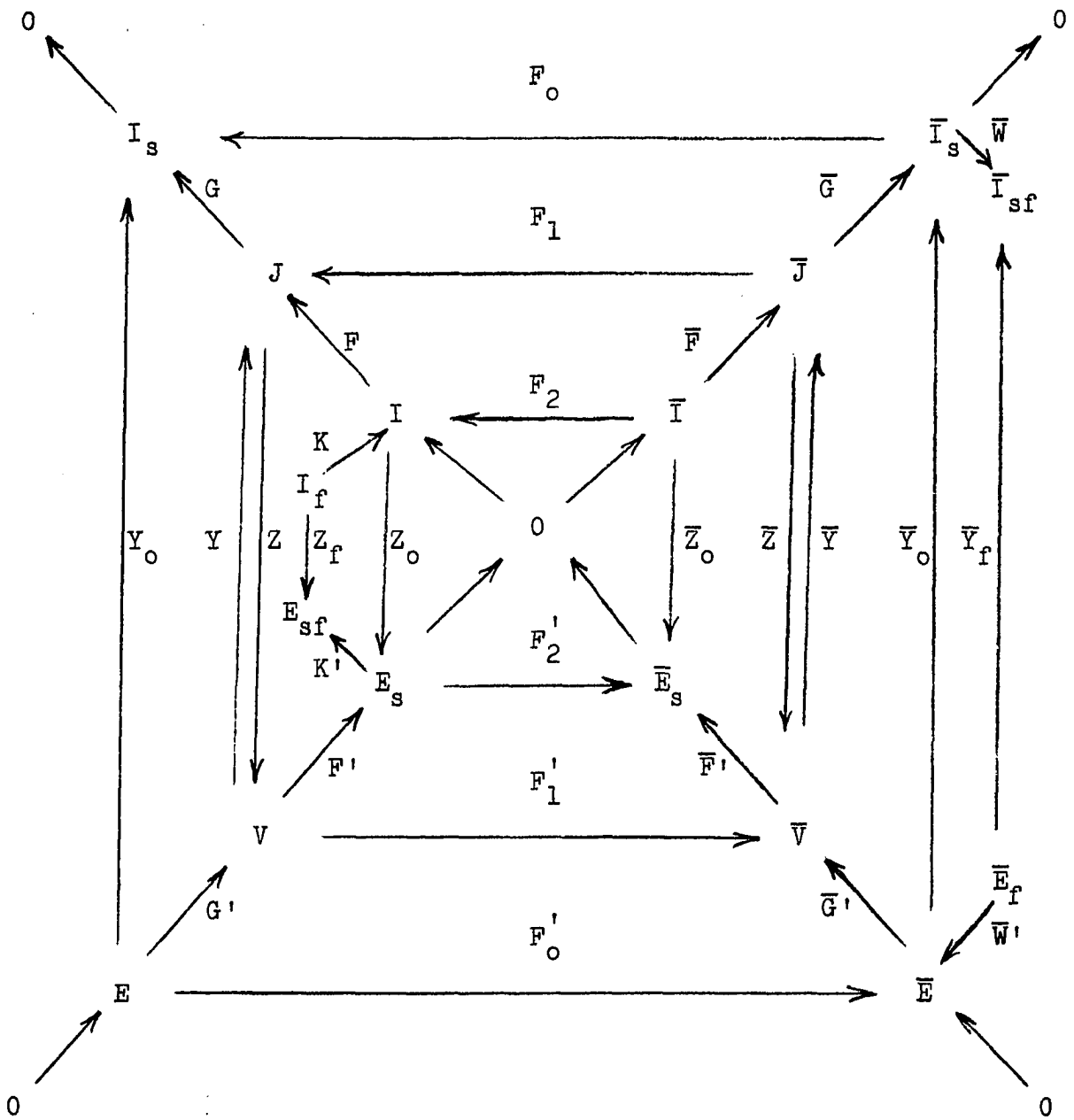
Suppose the two networks of interest to be each furnished with the complete apparatus of spaces and maps as described in the preceding chapter, the symbols of one of the networks being distinguished by overlining: J will represent

the branch current space of one network and $\overline{J}$ the branch current space of the other, for instance. The networks will be related in a fashion to be described, and the one whose symbols are overlined will be called the torn version, and the other the pinched version.

The essential relations between the two networks are now established by the stipulation of the existence of two maps satisfying conditions as follows: Let F_1 be an invertible map from $\overline{J}$ to J, and F_0 a map from $\overline{I}_s$ to I_s, such that $F_0\overline{G} = GF_1$ (commutative) and $\overline{Z} = F_1^{\prime}ZF$, where as usual $F_1^{\prime}$ is the adjoint of F_1.

These two maps induce a map F_2 from $\overline{I}$ to I. For any element of $\overline{I}$, $0 = \overline{GFi} = GF_1\overline{Fi}$, so $F_1\overline{Fi} = Fi$ for some unique element of i of I, since the final maps are all monomorphisms. The image of $\overline{i}$ under F_2 is then defined to be this i.

With the maps defined above and their adjoints, the result is the Squirrel Cage Diagram as given by Roth:


0
0
F_o
I_s
$\overline{I}_s$
$\overline{W}$
G
$\overline{G}$
$\overline{I}_{sf}$
F_1
J
$\overline{J}$
F
$\overline{F}$
F_2
I
$\overline{I}$
K
I_f
Y_o
Y
Z
Z_f
Z_o
0
$\overline{Z}_o$
$\overline{Z}$
$\overline{Y}$
$\overline{Y}_o$
$\overline{Y}_f$
E_{sf}
K'
F_2'
E_s
$\overline{E}_s$
F'
F_1'
$\overline{F}'$
V
$\overline{V}$
G'
$\overline{G}'$
$\overline{E}_f$
$\overline{W}'$
F_o'
E
$\overline{E}$
0
0

The remainder of the spaces and maps will be partly defined at this time. E_{sf} is defined as the null space of F_2' and K' is a projection of E_s onto E_{sf} along some complementary subspace. I_f and K are the corresponding dual space and adjoint map. Strictly speaking, I_f is not a subspace of I, as implied by the notation, but it is isomorphic to the image space of K, which is a subspace of I. The map from I_f to E_{sf} will be given later, but it is worth noting that it will <u>not</u> be taken as $K'Z_oK$, as might be expected by comparison with the definition of Z_o. The definitions at the other end of the diagram follow the same pattern, $\overline{I}_{sf}$ is defined to be the null space of F_o and $\overline{W}$ as a projection of $\overline{I}_s$ onto $\overline{I}_{sf}$ along some complementary subspace, and $\overline{E}_f$ and $\overline{W}'$ are the corresponding dual space and adjoint map, and finally, the map from $\overline{E}_f$ to $\overline{I}_{sf}$ will <u>not</u> be defined as $\overline{W}Y_o\overline{W}'$. A comment similar to that about I_f applies to $\overline{E}_f$.

Although not indicated, it will also be necessary to use two reverse maps: let F_{or} be any reverse of F_o, and F_{2r}' any reverse of F_2'.

Some properties of the various maps will now be determined. (1) F_o is an epimorphism. For any i_s, $i_s = Gj = GF_1\overline{j} = F_o\overline{Gj}$ for some $\overline{j}$, since F_1 establishes an isomorphism between J and $\overline{J}$. (2) $F_1\overline{F} = FF_2$. This follows immediately from the definition of F_2 as a map induced by F_1. (3) F_2 is a monomorphism. Suppose $\overline{i} \neq 0$. Then since F, $\overline{F}$ and F_1 are all monomorphisms, $0 \neq F_1\overline{Fi} = FF_2\overline{i}$, so

$F_2\bar{I} \neq 0$. (4) Since F_1 is invertible J and $\bar{J}$ must have the same dimension, so the two networks must have the same number of branches, although this implies nothing about the interconnections of the branches. (5) The torn version may have more nodes than the pinched version. It cannot have fewer, since then F_2 could not be an epimorphism. If it does have more nodes, it must have fewer loops than the pinched version, since the number of branches is the same. (6) The commutative requirement $F_o\bar{G} = GF_1$ guarantees that a given branch of the torn version ends up with the same bounding nodes, so to speak, at I_s, whether the calculation is <u>via</u> J or $\bar{I}_s$. (7) The condition $\bar{Z} = F_1'ZF$, or the equivalent condition on the Y maps, is easily satisfied if each network is constructed from identical branches. (8) $\bar{Z}_o = F_2'Z_oF_2$ and $Y_o = F_o\bar{Y}_oF_o'$, which follows immediately from the various commutativities.

Several of the above properties may be combined to justify the statement: a torn version of a network may be obtained by opening up some of the loops of the original network. Conversely pinched versions may be obtained by joining nodes. These methods may reduce the complexity of the network by making the relationships involving some branches trivial by shorting or opening them. This would not represent the most general modification, since an example will be given later of a torn version in which all the branches of the pinched version enter into the relations of the torn

version in a non-trivial fashion. The definition of F_2 indicates a necessary condition: every element of $\bar{I}$ must have an image in I. Since F_2 was induced by F_1 (which assigns every branch to itself, so to speak), this may be interpreted to say that every tieset of the torn version must also be a tieset of the pinched version. Likewise on the other end of the diagram, every cutset of the pinched version must also be a cutset of the torn version.

Analysis of the Squirrel Cage Diagram

The situation to be considered is as follows: it is desired to find the solution of a given network problem for the pinched version, i.e., given e_s and i_s, what are the corresponding v and j? Since the map F_1 establishes an isomorphism between the two networks, for any v and j there must be some corresponding $\bar{v}$ and $\bar{j}$ such that $j = F_1\bar{j}$ and $v = F_1'^{-1}\bar{v}$. The essence of the calculation is to specify a network problem for the torn version, that is some pair of sources $\bar{e}_s$ and $\bar{i}_s$, such that the solutions $\bar{j}$ and $\bar{v}$ may be used to find the corresponding j and v of the pinched version. The information known initially about the pinched network problem is the sources, i.e., e_s and i_s, so it will be necessary to determine from these the corresponding $\bar{e}_s$ and $\bar{i}_s$. Once these are known the solution of the torn version follows as described in Chapter II, and then F_1 and $F_1'^{-1}$ may be used to find the desired solution of the pinched version.

In terms of matrix calculations, the solution of a network problem is equivalent to finding the inverse of either the loop or node matrix of the network, since when one is known the other may be found without additional inversions, as described by the formulas of Chapter II. Specific formulas will be given for the inverses of both the loop and node matrices of the pinched version in terms of the inverses of the loop and node matrices of the torn version.

Conversely, the calculation of the solution of the torn network in terms of a solution of the pinched network will follow a precisely dual procedure.

The determination of $\bar{e}_s$ and $\bar{i}_s$ as functions of e_s and i_s requires a certain amount of discussion, but certain points may be settled immediately. First, e_s and i_s determine unique corresponding values of $\bar{e}_s$ and $\bar{i}_s$ due to the various isomorphisms established by the maps, so any set of sources for the torn network which satisfy the proper conditions must be the correct, unique, sources needed. Likewise $\bar{j}$ and $\bar{v}$ are uniquely determined by e_s and i_s. The image of v under F' is e_s, so $F_2'F'v$ is the image of e_s in $\bar{E}_s$ under F_2'. Now $\bar{v} = F_1'v$, and $\bar{e}_s = \overline{F'v} = \bar{F}'F_1v$. The relation $F_1\bar{F} = FF_2$ implies the relation $\bar{F}'F_1' = F_2'F'$, so $\bar{e}_s = F_2'F'v = F_2'e_s$, and thus depends only upon e_s, and not at all upon i_s. Since F_2' is an epimorphism, it will in general have a non-trivial null space, and elements of this subspace will not influence $\bar{e}_s$. It will be seen that in fact these elements influence elements

of the current source space $\bar{I}_s$.

By a commutativity argument similar to the one above it is easily seen that the relation $i_s = F_o \bar{i}_s$ must hold, and so it might be loosely said that elements in the null space of F_o (an epimorphism) are not influenced by i_s, and as a matter of fact it will be seen that it is precisely these elements which are influenced by the elements of the null space of F_2'.

The determination of the correct element of $\bar{I}_s$ therefore represents the major problem in the analysis of the diagram. That this is the major problem seems to have been first recognized by J. B. Giever (personal communication).

Given an $\bar{e}_s$ and an $\bar{i}_s$, the check that they are the correct ones is to see if the solution they determine in the pinched network correctly maps into the given e_s and i_s, which is to say that the following two equations must be satisfied:

$$i_s = GF_1\bar{j} = GF_1(\bar{F}\bar{Z}_o^{-1}\bar{e}_s + \bar{Y}\bar{G}'\bar{Y}_o^{-1}\bar{i}_s) \qquad \text{and}$$

$$e_s = F'ZF_1 j = F'ZF_1(\bar{F}\bar{Z}_o^{-1}\bar{e}_s + \bar{Y}\bar{G}'\bar{Y}_o^{-1}\bar{i}_s) \quad .$$

It is already known that $\bar{e}_s = F_2' e_s$ and $i_s = F_o \bar{i}_s$. Since all the maps are linear, the correct $\bar{i}_s$ must be some linear function of e_s plus some linear function of i_s. The calculation will therefore be done in two parts, to simplify the expressions: (1) Given the sources of the pinched version to be the original i_s, and $e_s = 0$, find the corresponding $\bar{e}_{s1}$ and $\bar{i}_{s1}$. (2) Given the sources of the pinched version to be the original e_s, and $i_s = 0$, find the corresponding

$\bar{e}_{s2}$ and $\bar{i}_{s2}$. The necessary sources in the general case will then be $\bar{e}_s = \bar{e}_{s1} + \bar{e}_{s2}$ and $\bar{i}_s = \bar{i}_{s1} + \bar{i}_{s2}$.

The calculations are based upon the following two comments. First, for any $\bar{i}_s$, $F_2'F'F_1'^{-1}\bar{G}'\bar{Y}_o^{-1}\bar{i}_s = 0$ by application of the commutation rules, which says that a solution determined by any element of $\bar{I}_s$ corresponds to a voltage source in the pinched version which lies in the null space E_{sf} of F_2'. Second, if any element of I is mapped into $\bar{I}_s$ by way of J and $\bar{J}$, its image lies in the null space $\bar{I}_{sf}$ of F_o, which follows immediately by application of the commutation rules as follows: $F_o\bar{G}F_1^{-1}Fi = GF_1F_1^{-1}Fi = GFi = 0$, by exactness of the sequence.

(1) Since the voltage sources in the pinched version are zero for this calculation, $\bar{e}_{s1} = 0$, and the voltage equation above becomes $0 = F'ZF_1\overline{YG}'\bar{Y}_o^{-1}\bar{i}_{s1}$, and $i_s = F_o\bar{i}_{s1}$. One of these equations can be immediately satisfied by the use of some reverse map of F_o, namely if $\bar{i}_{st} = F_{or}i_s$ then $i_s = F_o\bar{i}_{st}$. Since reverses are not unique, it is of course highly unlikely that such a "trial" value is the correct one so the voltage equation will not be satisfied. However, it is now shown that it is possible to take such a trial value and calculate the necessary correction so that both equations will be satisfied.

It was shown above that the image in E_s of any element of $\bar{I}_s$ lies in the null space E_{sf} of F_2', so any projection K' will give this same image, considered as an element of E_{sf}.

The term to be cancelled may then be explicitly indicated as $e_{sf} = K'F'ZF_1\overline{YG}'\overline{Y}_o^{-1}\overline{i}_{st} = K'F'F_1'^{-1}\overline{G}'\overline{Y}_o^{-1}F_{or}i_s$. Now the correction $\overline{i}_{sc}$ to $\overline{i}_{st}$ must not disturb the image of the F_o calculation, so it must lie in the null space of F_o. It was shown above that a sure way to get such elements is to start with any element of I and map it to $\overline{I}_s$, so in particular elements of I_f have this property. The procedure is then to determine some i_f such that its image in $\overline{I}_s$, and consequently in E_{sf}, cancels the error term, so the equation to be solved is $-e_{sf} = (K'F'F_1'^{-1}\overline{G}'\overline{Y}_o^{-1}GF_1^{-1}FK)i_f$, or $-e_{sf} = Z_f i_f$. Assuming the existence of an inverse to Z_f (which will be discussed later) the correction term is then found to be

$$\overline{i}_{sc} = -\overline{G}F_1^{-1}FKZ_f^{-1}K'F'F_1'^{-1}\overline{G}'\overline{Y}_o^{-1}F_{or}i_s, \text{ and } \overline{i}_{s1} = \overline{i}_{st} + \overline{i}_{sc}.$$

(2) In this case it is immediate that $\overline{e}_{s2} = F_2'e_s$ and that $\overline{i}_{s2}$ must lie in $\overline{I}_{sf}$, the null space of F_o, since $i_s = 0$. Now the image in E_s of the solution determined by $\overline{e}_{s2}$ by itself is immediately calculated to be $Z_oF_2\overline{Z}_o^{-1}F_2'e_s$, so the image in E_s of the $\overline{I}_s$ component must be the difference between e_s and this term: $e_s - Z_oF_2\overline{Z}_o^{-1}F_2'e_s$. But this is precisely the problem treated in (1), so the result of applying the same technique gives the $\overline{I}_s$ component as:

$$\overline{i}_{s2} = \overline{G}F_1^{-1}FKZ_f^{-1}K'(e_s - Z_oF_2\overline{Z}_o^{-1}F_2'e_s).$$

The final result for the sources of the torn network is now obtained by adding together the two contributions to each source, giving $\overline{e}_s = F_2'e_s$ and:

$$\bar{i}_s = F_{or} i_s - (\bar{G}F_1^{-1}FKZ_f^{-1}K')F'F_1'^{-1}\bar{G}'\bar{Y}_o^{-1}F_{or}i_s$$
$$+ (\bar{G}F_1^{-1}FKZ_f^{-1}K')\,(e_s - Z_oF_2\bar{Z}_o^{-1}F_2'e_s) \quad .$$

There remains the necessity of showing that Z_f is invertible under reasonable conditions. It is now shown that the map $\bar{G}F_1^{-1}FK$ is a monomorphism, since then the discussions of Chapter I may be applied, and in particular if the network is ohmic, so to speak, $\bar{Y}_o^{-1}$ will be ohmic and invertibility will be guaranteed. The map will be a monomorphism if the equation $\bar{G}F_1^{-1}FKi_f = 0$ implies that $i_f = 0$. By exactness the equation does imply that $F_1^{-1}FKi_f = \overline{Fi}$, for some $\bar{i}$, or $FKi_f = F_1\overline{Fi} = FF_2\bar{i}$, and since F is a monomorphism $Ki_f = F_2\bar{i}$. Now for any e_{sf}, $e_{sf}(F_2\bar{i}) = F_2'e_{sf}(\bar{i}) = 0$, and so also $0 = e_{sf}(Ki_f) = K'e_{sf}(i_f) = e_{sf}(i_f)$. The only element of I_f which annihilates every element of E_{sf} is the zero element, so $i_f = 0$.

It has therefore been demonstrated that a solution of the pinched version can be found in terms of the solution of the torn version, at the primary cost of an auxiliary inversion of a map between spaces whose dimension equals the difference in the dimensions of the I and $\bar{I}$ spaces. Note that the same difference exists between the dimensions of the E and $\bar{E}$ spaces, except that the dimension of the $\bar{E}$ space is larger than that of E, while the dimension of $\bar{I}$ is smaller than that of I. This difference in dimension will be called the <u>degree of rearrangement</u> of the two networks.

The process also requires that the inter-space maps

F_o, F_1 and F_2 be found, as well as F_1^{-1}. As will be illustrated by the matrix interpretation which follows, this is essentially a matter of finding a change of basis matrix and its inverse matrix, which is usually a fairly simple process, at least for small degrees of rearrangement.

The case of complete tearing may be conveniently considered here, i.e., the torn version has all its branches open-circuited. In this case $\bar{I}$ is a null space since there are no loops and the degree of rearrangement equals the dimension of the I space, so the inversion of Z_f requires the inversion of a matrix the same size as the loop matrix, which seems to hold small promise.

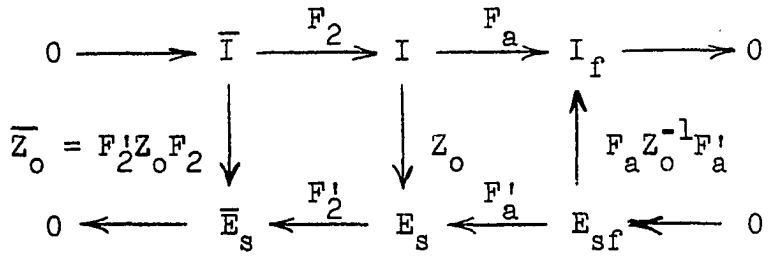
The procedure to use the pinched network solution to aid in the solution of the torn version is in every way the dual of the above procedure, and only the results will be given: $i_s = F_o \bar{i}_s$, $\bar{Y}_f = \bar{W}\bar{G}F_1^{-1}FZ_o^{-1}F'F_1'^{-1}\bar{G}'\bar{W}'$, and

$$e_s = F_{2r}'\bar{e}_s + F'F_1'^{-1}\bar{G}'\bar{W}'\bar{Y}_f^{-1}\bar{W}(\bar{i}_s - \bar{Y}_1F_o'Y_o^{-1}F_o\bar{i}_s - \bar{G}F_1^{-1}FZ_o^{-1}F_1'^{-1}\bar{e}_s).$$

The question of complete pinching, i.e., all branches short-circuited, suffers the same fate as complete tearing, since the inversion of Y_f in this case will require the inversion of a matrix the same size as the node matrix of the torn version.

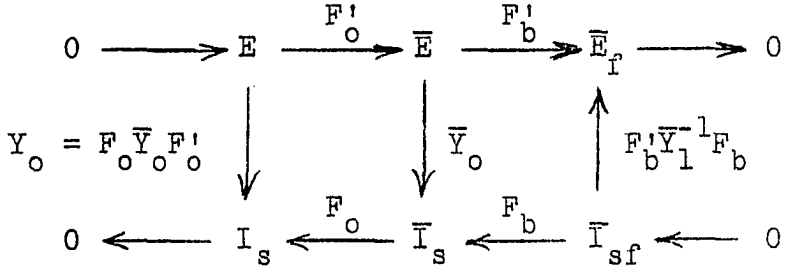
Comparison with Bordering and Partitioning

If the spaces and maps at the loop end of the diagram are isolated in a separate diagram, with E_{sf} injected into E_s by F_a', the following diagram results:



and comparison with the discussions of Chapter I will show that this is the proper diagram so that Z_o may be inverted by the partitioning technique, or $\bar{Z}_o$ by the bordering technique, and the necessary inversions are precisely of the same size matrices as must be inverted in the use of tearing or pinching.

Similarly at the node end of the diagram, the separated diagram takes the form:



and a similar comment is appropriate.

The question may then be fairly raised, is there any benefit to tearing or pinching that can not be similarly obtained by the techniques of bordering and partitioning? It may be noted here, in anticipation of the results of the next section on matrix interpretations, that in many cases the actual numerical work will be identical whether, for instance, Z_o is inverted by means of partitioning or by tearing, etc.

In general the answer is yes, in the following spirit: Suppose neither version has been previously solved. Then the use of tearing or pinching allows both networks to be solved simultaneously, in the sense of obtaining the inverses of both Z_o and $\overline{Z}_o$ or Y_o and $\overline{Y}_o$ at a cost of inversions not much greater than that required to solve either network by itself, assuming the degree of rearrangement is small, of course. The corresponding operation of arbitrary bordering or partitioning would also give two inverses at the same cost in calculation, but only one of them would be applicable to a known physical problem.

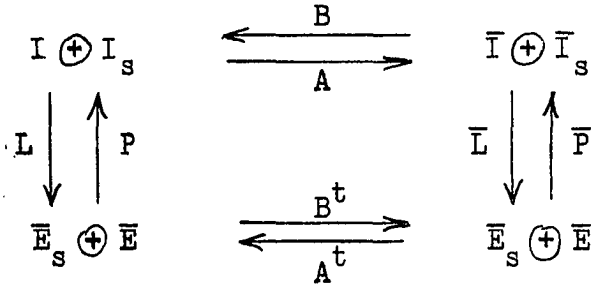
In the case where one of the networks has been previously solved, the advantage of tearing and pinching is quite pronounced. In the case of arbitrary bordering, for example, given a map whose inverse is desired, and another map between larger spaces whose inverse is known, it is not at all obvious how the smaller spaces are to be embedded in the larger spaces such that the necessary commutative conditions are satisfied.

Matrix Interpretations

Two matrix techniques will now be given to illustrate the abstract procedures developed above. In the first technique no assumptions are made about special choices of bases, special matrix forms, etc. In the second technique it is shown that with no restrictions on the networks it is possible to "match" the bases of the two such that the techniques

of matrix partitioning and bordering can be immediately applied. The symbol U will always stand for an appropriately sized identity matrix, necessarily square; while the symbol 0 will always stand for a completely zero matrix, square or rectangular as appropriate.

It will be assumed that dual, split bases have been selected for each network and therefore each of the center spaces of the network diagrams decomposed into subspaces: $J = I \oplus I_s$, $\bar{V} = \bar{E}_s \oplus \bar{E}$, etc. Let B represent the matrix of F_1 with respect to the bases selected, and A represent F_1^{-1}. Since dual bases are in use, the transpose of B, B^t, will then be the matrix of F_1', and A^t the matrix of $F_1'^{-1}$. The ways in which these matrices are related is indicated by the following sketch:



where, for instance, $L = A^t \bar{L} A$.

Any column matrix representing a set of coordinates of an element of one of these spaces can be conveniently partitioned into two matrices, each submatrix containing the coordinates pertaining to one of the subspaces of the direct sum representation of the space. If the A and B matrices are partioned conformably, the result in the case of the A matrix is:

$$\begin{bmatrix} \bar{i} \\ \bar{i}_s \end{bmatrix} = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} \begin{bmatrix} i \\ i_s \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} e_s \\ e \end{bmatrix} = \begin{bmatrix} A_1^t & A_3^t \\ A_2^t & A_4^t \end{bmatrix} \begin{bmatrix} \bar{e}_s \\ \bar{e} \end{bmatrix} .$$

The definition of F_2 above was a result of the fact that the image of any element of $\bar{I}$ was an element of I, and this same fact also guarantees that the B_3 partition of the B matrix below is zero as indicated:

$$\begin{bmatrix} i \\ i_s \end{bmatrix} = \begin{bmatrix} B_1 & B_2 \\ 0 & B_4 \end{bmatrix} \begin{bmatrix} \bar{i} \\ \bar{i}_s \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \bar{e}_s \\ \bar{e} \end{bmatrix} = \begin{bmatrix} B_1^t & 0 \\ B_2^t & B_4^t \end{bmatrix} \begin{bmatrix} e_s \\ e \end{bmatrix} .$$

The first calculation is to be that of the inverse of L_1, the loop matrix of the pinched version, so take the current sources to be zero: $i_s = 0$. Under this condition the expressions for the sources of the torn version become $\bar{e}_s = B_1^t e_s$ and $\bar{i}_s = A_3 i$, and it is this last quantity which must be determined by an auxiliary calculation. Writing out the solution of the torn version: $\bar{e} = \bar{P}_4^{-1}(A_3 i - \bar{P}_3 B_1^t e_s)$ and $\bar{i} = \bar{L}_1^{-1}(B_1^t e_s - \bar{L}_2 A_3 i)$, and transferring this solution back to the pinched version the result is:

$$i = B_1 \bar{L}_1^{-1}(B_1^t e_s - \bar{L}_2 A_3 i) + B_2 A_3 i, \text{ and}$$

$$e_s = A_1^t B_1^t e_s + A_3^t \bar{P}_4^{-1}(A_3 i - \bar{P}_3 B_1^t e_s) .$$

The second of these two expressions will be solved for the term $A_3 i$, and the result when substituted in the first will give the desired answer: $i = L_1^{-1} e_s$.

The second of the two equations may be rearranged as $(U - A_1^t B_1^t + A_3^t \bar{P}_4^{-1} \bar{P}_3 B_1^t)\, e_s = A_3^t \bar{P}_4^{-1} A_3 i$, and it will now be

shown that it is possible to construct a matrix Q such that $Q(A_3^t \bar{P}_4^{-1} A_3) = A_3$, so it is possible as a result to solve the equation for the term $A_3 i$. In general A_3 is not even square, much less non-singular, so to get some hold upon its properties it must be reduced to canonical form: for any matrix A_3 there exists at least one non-singular matrix C such that:

$$A_3 = C \begin{bmatrix} 0 & 0 \\ X & U \end{bmatrix} = \begin{bmatrix} C_o & C_a \end{bmatrix} \begin{bmatrix} 0 & 0 \\ X & U \end{bmatrix} ;$$

where the size of the identity matrix U is the same as the rank of A_3, and X has no special properties. Replacing A_3 in $A_3^t \bar{P}_4^{-1} A_3$ by this canonical form the result is:

$$A_3^t \bar{P}_4^{-1} A_3 = \begin{bmatrix} 0 & X^t \\ 0 & U \end{bmatrix} C^t \bar{P}_4^{-1} C \begin{bmatrix} 0 & 0 \\ X & U \end{bmatrix} ,$$

and defining $T = C^t \bar{P}_4^{-1} C$ and partitioning T so that T_4 below is the same size as the rank of A:

$$A_3^t \bar{P}_4^{-1} A_3 = \begin{bmatrix} 0 & X^t \\ 0 & U \end{bmatrix} \begin{bmatrix} T_1 & T_2 \\ T_3 & T_4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ X & U \end{bmatrix} = \begin{bmatrix} X^t T_4 X & X^t T_4 \\ T_4 X & T_4 \end{bmatrix} .$$

Finally define a matrix D as:

$$D = \begin{bmatrix} 0 & 0 \\ 0 & T_4^{-1} \end{bmatrix}$$

where the vertical dimension of D is the same as the dimension of the $\bar{I}_s$ space, and the horizontal dimension the same

as the I space.

A direct calculation will now show that if Q is defined by Q = CD, it will have the desired property. As for the invertibility of T_4, it may be calculated as $T_4 = C_a^t \bar{P}_4^{-1} C_a$. Now since C is non-singular, the rank of C_a must equal its horizontal dimension, and therefore C_a is a representation of a monomorphism so the considerations of Chapter I may be applied to the question of the existence of the inverse. In particular, if P is ohmic, then the inverse will exist.

If these manipulations and substitutions are made, the result for the inverse of the loop matrix is finally:

$$L_1^{-1} = B_1 \bar{L}_1^{-1} B_1^t + (B_2 - B_1 \bar{L}_1^{-1} \bar{L}_2)\, Q A_3^t\, (B_2^t - \bar{L}_3 \bar{L}_1^{-1} B_1^t) \quad .$$

The corresponding calculation of the inverse of the node matrix gives the result: $P_4^{-1} = A_4^t (\bar{P}_4^{-1} - \bar{P}_4^{-1} Q A_3^t \bar{P}_4^{-1}) A_4$.

The use of the pinched version to aid in the solution of the torn version may now be considered. The only essential difference is that the auxiliary inversion takes a slightly different form. The desired expression is $R(A_3 L_1^{-1} A_3^t) = A_3^t$, where R is to be found. The necessary expressions are as follows:

$$A_3^t = C \begin{bmatrix} U & X \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} S_1 & S_2 \\ S_3 & S_4 \end{bmatrix} = C^t L_1^{-1} C \quad , \quad D = \begin{bmatrix} S_1^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

and finally R = CD. The inverses are then found to be:

$$\bar{L}_1^{-1} = A_1 (L_1^{-1} - L_1^{-1} R A_3 L_1^{-1}) A_1^t, \text{ and}$$

$$\bar{P}_4^{-1} = B_4^t P_4^{-1} B_4 + (B_2^t - B_4^t P_4^{-1} P_3)\, R A_3\, (B_2 - P_2 P_4^{-1} B_4) \quad .$$

The forms of the above two results already show interesting similarities to the matrix bordering and partitioning formulas of Chapter I. The correspondence will now be made exact. Specifically, assume that the bases for the two spaces have been so chosen that the A and B matrices take the following special forms:

$$\begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix} = \left[\begin{array}{ccc} U & A_{12} & A_2 \\ 0 & A_{32} & A_{41} \\ 0 & 0 & U \end{array}\right] \text{ and } \begin{bmatrix} B_1 & B_2 \\ 0 & B_4 \end{bmatrix} = \left[\begin{array}{ccc} U & B_{21} & B_{22} \\ 0 & B_{23} & B_{24} \\ 0 & 0 & U \end{array}\right].$$

This places no additional restrictions on the networks at all, being merely a matter of amount of effort involved in matching up the bases in the two networks. Such a pair of bases will be called <u>matched</u>. The original partitions are indicated by the dotted lines. B_1 can be reduced to the form shown by the selection of whatever basis is used for $\bar{I}$ as part of the basis for I, and likewise the form of B_4 by the use of the basis for E as part of the basis for $\bar{E}$. The form of the A matrix then follows as a result of the form of the B matrix.

The primary result of using matched bases is that $\bar{L}_1$ is identical to a submatrix of L_1, and P_4 to a submatrix of $\bar{P}_4$. Specifically, let L_1 and $\bar{P}_4$ be partitioned:

$$L_1 = \begin{bmatrix} L_{11} & L_{12} \\ L_{13} & L_{14} \end{bmatrix} \qquad \bar{P}_4 = \begin{bmatrix} \bar{P}_{41} & \bar{P}_{42} \\ \bar{P}_{43} & \bar{P}_{44} \end{bmatrix}$$

where L_{11} is to be the same size as $\bar{L}_1$ and $\bar{P}_{44}$ the same size as P_4. If matched bases are used, then $L_{11} = \bar{L}_1$ and $P_4 = \bar{P}_{44}$.

The discussions of Chapter I on the use of matrix partitioning and bordering are all now applicable.

If only the node or loop matrix is of interest, it is not necessary to completely match bases. If it is desired to embed $\overline{L}_1$ in L_1, all that is necessary is to use the basis of $\overline{I}$ as part of the basis of I. To embed P_4 in $\overline{P}_4$ it is only necessary to use the basis of E as part of the basis for $\overline{E}$.

Examples

The two networks used as examples in Chapter II will also serve as the first example here. Using the same bases as before, the A and B matrices are:

A =

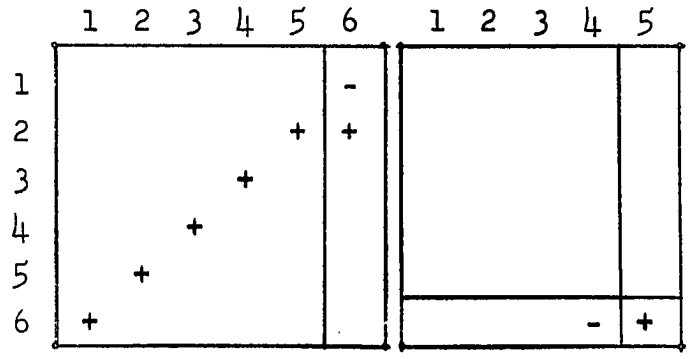
	1	2	3	4	5	6	7	8	9	10
1	+									
2		+								
3			+							
4				+						
5				+	-					
6				-	+	+				
7							+			
8								+		
9									+	
10										+

B =

	1	2	3	4	5	6	7	8	9	10
1	+									
2		+								
3			+							
4				+						
5				+	-					
6					+	+				
7							+			
8								+		
9									+	
10										+

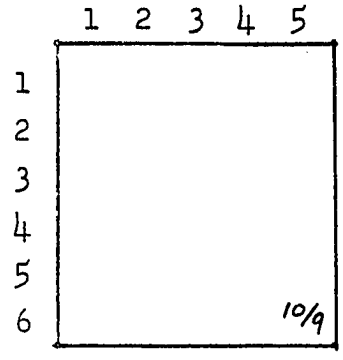
Selecting a C matrix as shown, the other matrices then follow:

$$A_3 = \begin{bmatrix} C_o & C_a \end{bmatrix} \begin{bmatrix} 0 & 0 \\ X & U \end{bmatrix} =$$

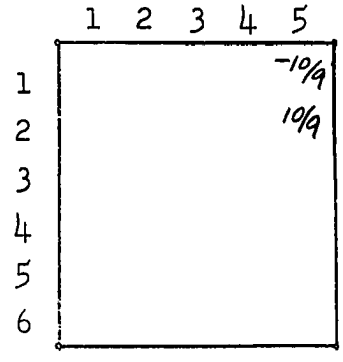


$$T_4 = \begin{bmatrix} 9/10 \end{bmatrix} \qquad T_4^{-1} = \begin{bmatrix} 10/9 \end{bmatrix}$$

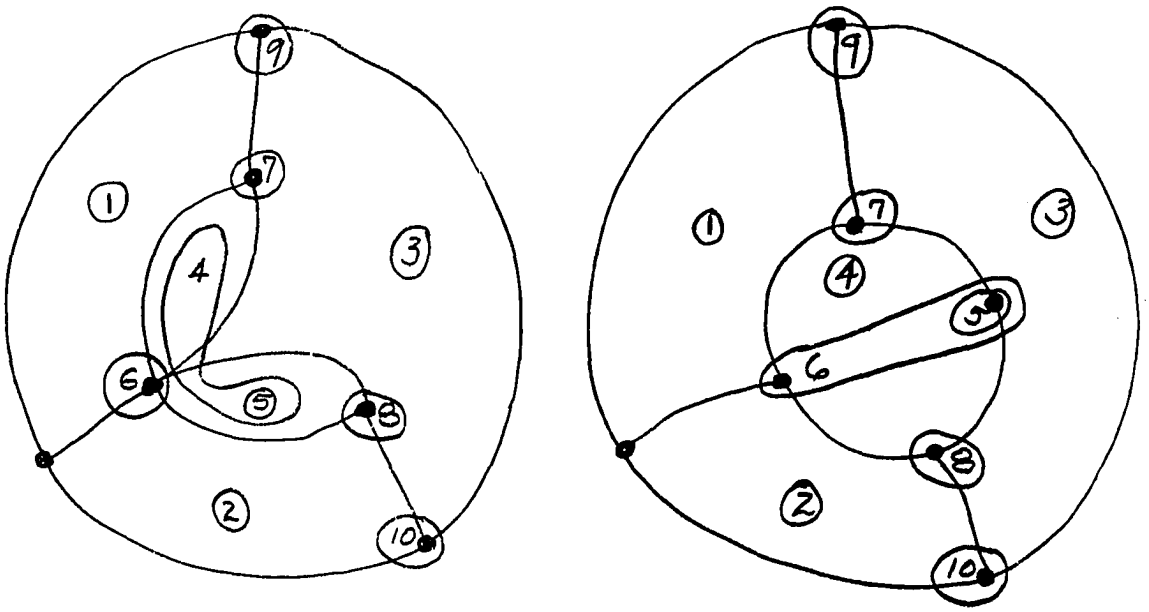
$$D =$$



$$Q = CD =$$



To illustrate the use of matched bases, let the bases of the two networks be chosen as indicated by the following sketches, in the same style as in Chapter II.



The loop and node matrices may then be determined as

$L_1 =$

	1	2	3	4	5
1	4	-	-	-	
2	-	4	-	-	-
3	-	-	5	$\bar{2}$	-
4	-	-	$\bar{2}$	4	2
5		-	-	2	2

$\bar{L}_1 =$

	1	2	3	4
1	4	-	-	-
2	-	4	-	-
3	-	-	5	$\bar{2}$
4	-	-	$\bar{2}$	4

$P_4 =$

	1	2	3	4	5
1	5	$\bar{2}$	$\bar{2}$		
2	$\bar{2}$	3		-	
3	$\bar{2}$		3		-
4		-		3	-
5			-	-	3

$\bar{P}_4 =$

	1	2	3	4	5	6
1	2	2	-	-		
2	2	5	$\bar{2}$	$\bar{2}$		
3	-	$\bar{2}$	3		-	
4	-	$\bar{2}$		3		-
5			-		3	-
6				-	-	3

and it is seen that in each case the smaller matrices appear as submatrices of the larger matrices.

CHAPTER IV

ANALYSIS OF A NETWORK DERIVED FROM A LINEAR HEAT FLOW PROBLEM

The following discussion is for the purpose of presenting an example of the use of the techniques given in Chapter III in a practical context. It would not be feasible to discuss all aspects of the solution of a complicated problem, so the following examples have been arranged to illustrate the major advantages and disadvantages of the techniques of Chapter III. A two-dimensional heat flow problem is presented for the sake of simplicity, although, of course, the techniques are not limited to this situation.

Complicated heat flow problems are usually solved on digital computers by means of finite difference equations. The given continuous body is replaced by a discrete, lumped system, in which the mass is collected into lumps at the points of a grid, and it is assumed that the grid points are connected by strips of thermally conducting material. The equations for the discrete system are finite in number and will be seen to have an excellent interpretation as the equations of an electric network. When boundary conditions are specified the

equations for the discrete system may be solved to yield the distributions of temperature and heat flow over the grid points. If the grid, or net, spacing is taken small enough, the results of the calculation are usually a good approximation to the distributions in the continuous body from which the approximation was derived. If the boundary conditions are changed, it is necessary to repeat the entire calculation.

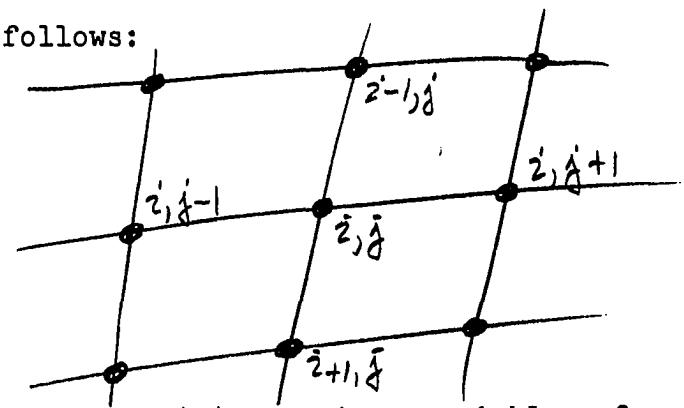
The network model procedure to be given envisages a somewhat more ambitious program. Essentially, the techniques of operational calculus will be used to obtain the solution of the problem for any boundary conditions, in the sense that no significant additional calculation will be necessary to find the distributions once the specific boundary conditions are given. Even further, the boundary conditions may be quite arbitrary functions of time. The network model approach also allows the use of the rearrangement techniques of Chapter III where appropriate.

On the other hand, the network manipulations are based upon the assumption that the relations among the variables are linear. Therefore this network approach cannot be applied to the many heat flow problems which are significantly non-linear. The network approach will also usually require more calculation than a single direct solution for fixed boundary conditions.

In order to establish the network representation, one proceeds as in the derivation of a set of finite difference

equations. Let K represent the thermal conductivity of the body, M the mass density, and c the specific heat. These may all be functions of position in the body. Let the thickness of the body be represented by b, and suppose a square net of points laid out with h as the net spacing.

Let all of the mass (and, therefore, heat storage capacity) with h/2 each side of a net point be lumped at the net point, and all of the heat transmitting material lumped into rods connecting the net points. The area of each rod will be bh and the mass at each point will be Mh^2b, where M is to be evaluated at the net point. A typical set of net points might then be indicated as follows:



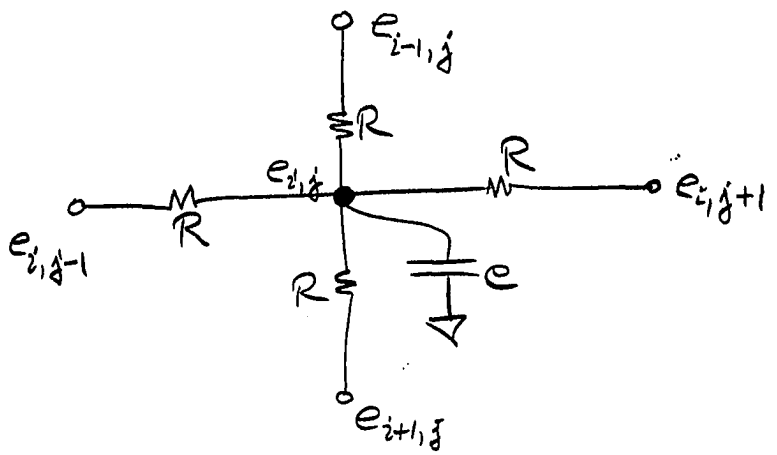
Let T be used to represent temperature variables, Q heat flow variables, and t the time variable. The result of writing the energy balance equation at the i,j net point is then:

$$dT_{i,j}/dt = \frac{Q_{i,j}}{SMbh^2} = \frac{K(T_{i+1,j}+T_{i-1,j}+T_{i,j+1}+T_{i,j-1}-4T_{i,j})}{SMh^2} .$$

If such an equation is written at each net point, the result is a set of ordinary differential equations. The usual finite difference derivation would now also replace the time derivative

by a difference approximation, but this is not appropriate for the present discussion.

Now consider the following segment of an electric network:



If currents are summed at the center node, the result is:

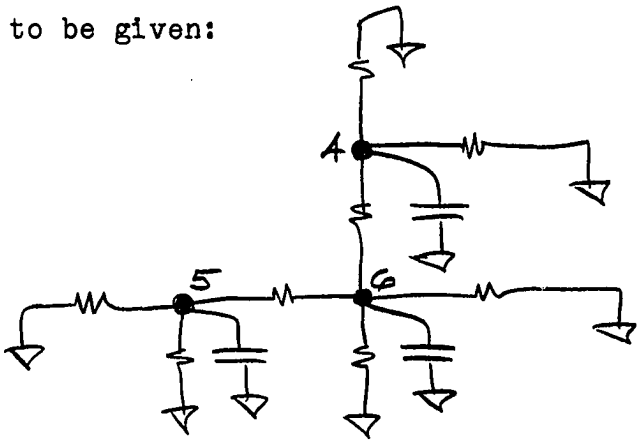
$$4e_{i,j} + RC(de_{i,j}/dt) - e_{i,j+1} - e_{i,j-1} - e_{i+1,j} - e_{i-1,j} = 0 ,$$

which is easily seen to be identical in form with the heat equation given above. If the parameters are now arranged so that $RC = SMh^2/K$, and voltages interpreted as temperatures and currents as heat flows, a network model for the heat flow approximation has been established. Since with linear networks the resistances and capacitances do not depend upon the voltages or currents, the model is only applicable to heat flow problems in which the specific heat and thermal conductivity do not depend upon the temperatures or heat flows.

The establishment of a heat flow boundary condition is accomplished by the specification of the appropriate Q variable, and corresponds to the specification of a current source in the network model. Likewise, the specification of a temperature boundary condition corresponds to the specification

of a voltage source in the network model. If only one type of source is desired in the network model, it is straightforward to convert voltage sources to equivalent current sources, and vice-versa. For networks derived from two-dimensional heat flow problems, it is likely that the node method of analysis will be the most convenient, so it will be taken that only current sources appear in the network, and it is desired to solve for the voltages. The remainder of the discussion will be entirely in terms of the solution of the electric network.

Suppose the following network of one ohm resistors and one farad capacitors to be given:



and $\bar{e}_4$, $\bar{e}_5$ and $\bar{e}_6$ are the voltages with respect to ground at the indicated points. This network is small enough that the solution procedures may be explicitly indicated, as well as the specific procedure whereby the solution to be found may be used to aid in the solution of a rearranged version of the network. A larger rearrangement problem will then be discussed to show that no change in procedure is necessary.

The node matrix of the network above may be immediately

written as:

	$\bar{e}_4$	$\bar{e}_5$	$\bar{e}_6$
$\bar{I}_{s4}$	s+3		-
$\bar{I}_{s5}$		s+3	-
$\bar{I}_{s6}$	-	-	s+4

where s is the usual complex variable of the Laplace transform. Inverting the matrix, the result is:

1/D

$s^2+7s+11$	+	s+3
+	$s^2+7s+11$	s+3
s+3	s+3	$(s+3)^2$

where D is the determinant of the node matrix, in this case $D(s) = (s+2)(s+3)(s+5)$.

For many electrical engineering calculations this inverse is the desired quantity, and the network is regarded as solved once the inverse is found. Some additional manipulations will now be given to clarify the meaning of these expressions. Consider, for instance, the 4,5 entry of the inverse, which relates the transform of the current source applied at node 5 to the transform of the resulting voltage at node 4. Let the entry be expanded in partial fractions:

$$\frac{1}{(s+2)(s+3)(s+5)} = \frac{1/3}{(s+2)} - \frac{1/2}{(s+3)} - \frac{1/6}{(s+5)} \quad ,$$

from which the inverse transform may be immediately written down as:

$$\bar{g}_{4,5}(t) = (1/3)\,\epsilon^{-2t} - (1/2)\,\epsilon^{-3t} + (1/6)\,\epsilon^{-5t} \quad ,$$

for t positive, and as zero for t negative. This function is called a weighting function, and gives the voltage developed at node 4 due to a unit impulse of current being applied at node 5.

Suppose now the current supplied to node 5 to be zero for t negative, and some specific $\bar{i}_{s5}(t)$ for t positive. The convolution integral may then be used to calculate the resulting voltage at node 4 as:

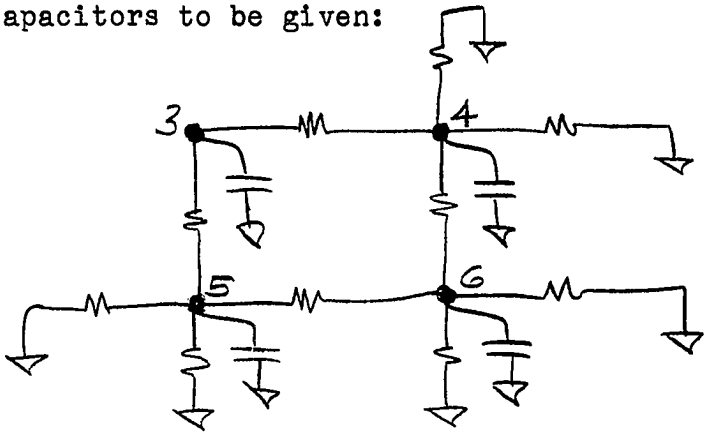
$$\bar{e}_4(t) = \int_0^\infty \bar{g}_{4,5}(x)\, \bar{i}_{s5}(t-x)\, dx$$

for t positive, and, of course, the voltage will be zero for t negative. Therefore, if the weighting functions are known, the voltage which results due to any current source may be found by a simple and inexpensive integral calculation. If there are current sources connected to more than one node, they are simply considered one at a time, since their effects may be superimposed, a property of linear networks.

Because of the simplicity of the network structure and the normalization of the element values, the above calculations were all quite easy. In larger networks, there are usually two points of difficulty. The inversion of the node matrix and the factoring of the denominator polynomial D(s) are both, usually, relatively expensive compared to the other manipulations. The present discussions are concerned only with the matter of the matrix inversion. Note that all nine entries in the inverse matrix above are non-zero, even though the original matrix had two zero entries. This is usually

the case, even though the original matrix be quite sparse, because it is only when special conditions of symmetry exist in the network that a given current source will not produce voltages at all the network nodes.

Suppose now the following network of one ohm resistors and one farad capacitors to be given:



and it is desired to find its solution, that is, the inverse of its node matrix P_4:

$$P_4 = \begin{array}{c|cccc} & 3 & 4 & 5 & 6 \\ \hline 3 & s+2 & - & - & \\ 4 & - & s+4 & & - \\ 5 & - & & s+4 & - \\ 6 & & - & - & s+4 \end{array}$$

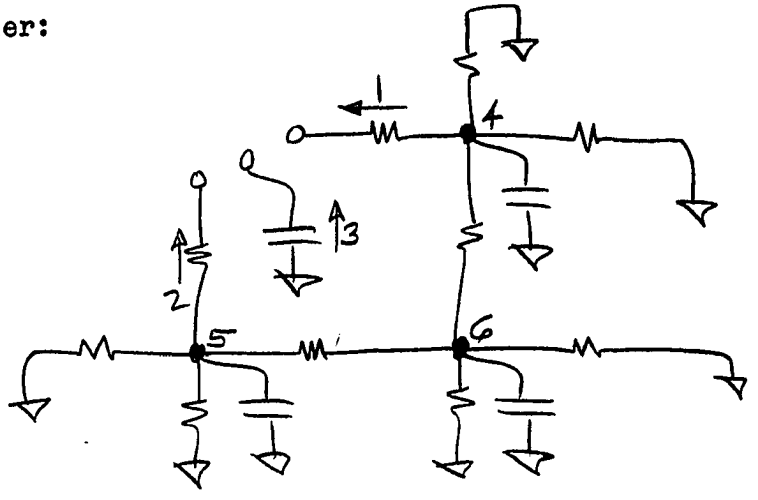
In this case it is not difficult to obtain P_4^{-1} by direct calculation, and it is found to be:

$$P_4^{-1} = 1/D_o$$

$s^2+8s+14$	$s+4$	$s+4$	2
$s+4$	$\frac{s^3+10s^2+30s+26}{s+4}$	$\frac{2(s+3)}{s+4}$	$s+2$
$s+4$	$\frac{2(s+3)}{s+4}$	$s^3+10s^2+30s+26$	$s+2$
2	$s+2$	$s+2$	s^2+6s+6

where $D_o(s) = s^3+10s^2+28s+20$. An alternative calculation of this inverse by the rearrangement method will now be given.

To use the techniques of Chapter III it is first necessary to have a pair of networks which are torn and pinched versions of each other, and a necessary part of this is that the two networks must have the same number of branches. It may now be noted that the first network above differs from the network of interest in that three branches are missing. If the three branches are added to the first network in the following manner:



it will be a torn version of the network of interest, and its node matrix will be:

$$\overline{P}_{4a} = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & + & & & & & \\ 2 & & + & & & & \\ 3 & & & s & & & \\ 4 & & & & s+3 & & - \\ 5 & & & & & s+3 & - \\ 6 & & & & - & - & s+4 \end{array}$$

using the additional voltage variables indicated. Note that $\bar{e}_4$ represents one volt across only three of the resistors connected to the node, the added resistor not being included, and likewise for $\bar{e}_5$. Now the lower submatrix has been previously inverted, so the inversion of $\bar{P}_{4a}$ is completely trivial, and this network may be taken as solved. In this case, inspection is adequate to determine that the two networks are torn and pinched versions of each other. In more complicated cases a matrix calculation can be made to settle the point. Let H be the change of basis matrix of Chapter II from the split basis to the natural basis of the pinched version, and $\bar{H}$ the corresponding matrix of the torn version. Then the matrix B of Chapter III may be found from $B = H^{-1}\bar{H}$, and, if the submatrix $B_3 = 0$, the two networks are torn and pinched versions of each other.

The discussions of Chapter III may be restated for present purposes as follows. It is desired to find P_4^{-1}. Suppose it were possible to find a matrix $\bar{P}_{4b}$ which contained P_4 as a submatrix in the lower right corner, and whose inverse is known. Indicating the inverse by

$$\bar{P}_{4b}^{-1} = \begin{array}{|c|c|}\hline L_{1\#} & L_{2\#} \\ \hline L_{3\#} & L_{4\#} \\ \hline \end{array} \quad ,$$

where $L_{4\#}$ is the same size as P_4, the desired inverse may be found from $P_4^{-1} = L_{4\#} - L_{3\#}L_{1\#}^{-1}L_{2\#}$, so the inversion of P_4 is replaced by the inversion of $L_{1\#}$ and some matrix multiplications. Further, in Chapter III the procedure is given for

changing the basis of the node-pair voltages of the torn network so that the node matrix with respect to the new basis contains P_4 as a submatrix. Let the new basis for the node-pair voltages of the torn version be indicated by primes on the symbols. Then the change of basis matrix $\overline{H}$ becomes:

$H =$

	$\bar{e}'_1$	$\bar{e}'_2$	$\bar{e}'_3$	$\bar{e}'_4$	$\bar{e}'_5$	$\bar{e}'_6$
$\bar{e}_1$	+		+	-		
$\bar{e}_2$		+	+		-	
$\bar{e}_3$			+			
$\bar{e}_4$				+		
$\bar{e}_5$					+	
$\bar{e}_6$						+

where the new basis elements have been chosen as specified in Chapter III. It should be noticed that due to the special form of $\overline{H}$ in this particular case, its inverse may be found by just changing the signs of all the off-diagonal terms. If the new node matrix for the torn network is now found by the formula $\overline{P}_{4b} = \overline{H}_t \overline{P}_{4a} \overline{H}$, the result is:

$\overline{P}_{4b} =$

+		+	-		
	+	+		-	
+	+	s+2	-	-	
-		-	s+4		-
	-	-		s+4	-
			-	-	s+4

and it is indeed seen that P_4 appears as a submatrix. The

inverse of $\bar{P}_{4b}$ follows immediately from the expression $\bar{P}_{4b}^{-1} = \bar{H}^{-1}\bar{P}_{4b}^{-1}\bar{H}_t^{-1}$. Letting $D(s) = (s+2)(s+3)(s+5)$, $Q(s) = s+3$, and $P(s) = s^2+7s+11$, the results may then be written in partitioned form as:

$$L_{1\#} = \begin{array}{|c|c|}\hline 1+(1/s)+(P/D) & (1/s)+(1/D) \\ \hline (1/s)+(1/D) & 1+(1/s)+(P/D) \\ \hline \end{array}$$

$$L_{2\#} = \begin{array}{|c|c|c|c|}\hline -(1/s) & (P/D) & (1/D) & (Q/D) \\ \hline -(1/s) & (1/D) & (P/D) & (Q/D) \\ \hline \end{array}$$

$$L_{3\#} = (L_{2\#})_t$$

$$L_{4\#} = \begin{array}{|c|c|c|c|}\hline (1/s) & & & \\ \hline & (P/D) & (1/D) & (Q/D) \\ \hline & (1/D) & (P/D) & (Q/D) \\ \hline & (Q/D) & (Q/D) & (Q^2/D) \\ \hline \end{array}$$

Inverting $L_{1\#}$, the result is:

$$L_{1\#}^{-1} = 1/R(s+4) \begin{array}{|c|c|}\hline s^4+12s^3+48s^2+72s+30 & s^3+10s^2+32s+30 \\ \hline s^3+10s^2+32s+30 & s^4+12s^3+48s^2+72s+30 \\ \hline \end{array}$$

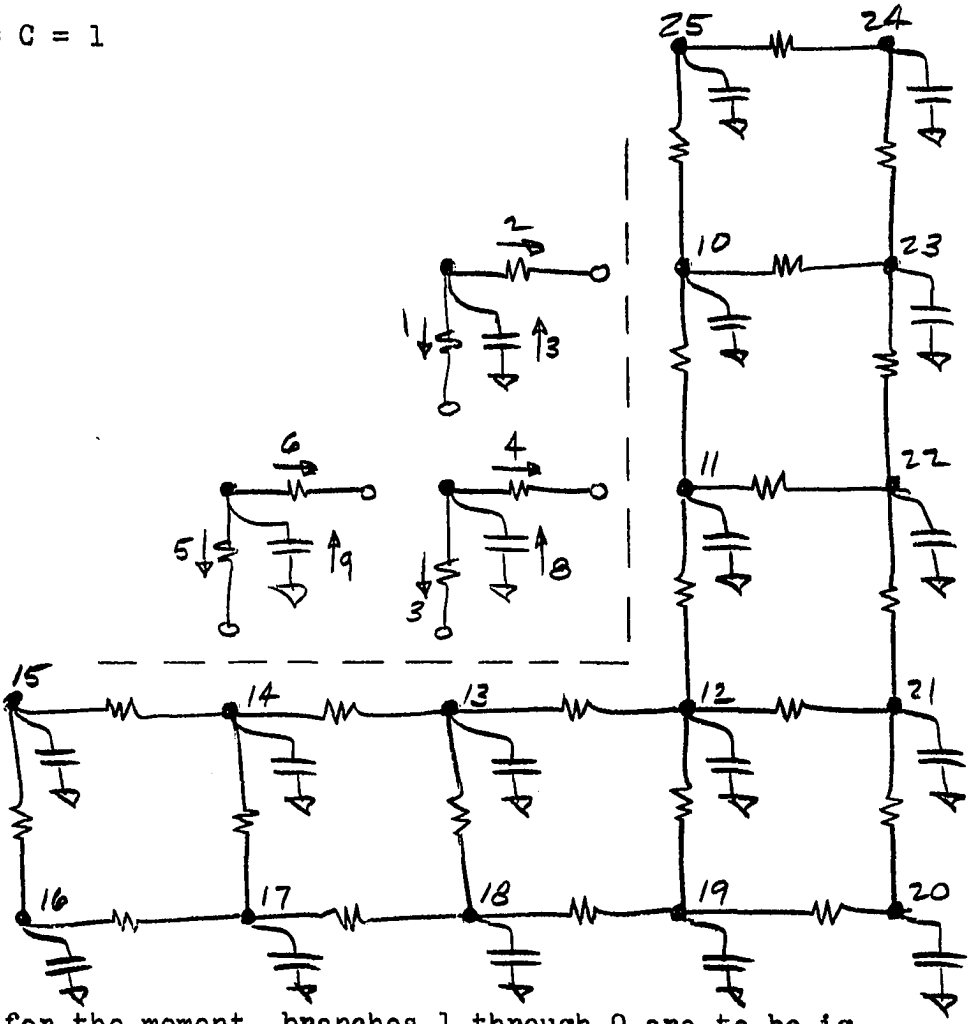
where $R(s) = s^3+10s^2+28s+20$. If the calculation for P_4^{-1} in terms of these variables is now carried through, the result given previously will be obtained.

Summarizing, the inversion of P_4, a 4 by 4 matrix, has been replaced by some matrix multiplications and the inversion of $L_{1\#}$, a 2 by 2 matrix. In this small case the indirect procedure is a more difficult way to perform the inversion, but

it should be obvious that in problems in which P_4 is considerably larger than $L_{1\#}$ the indirect procedure would likely be far the better of the two methods.

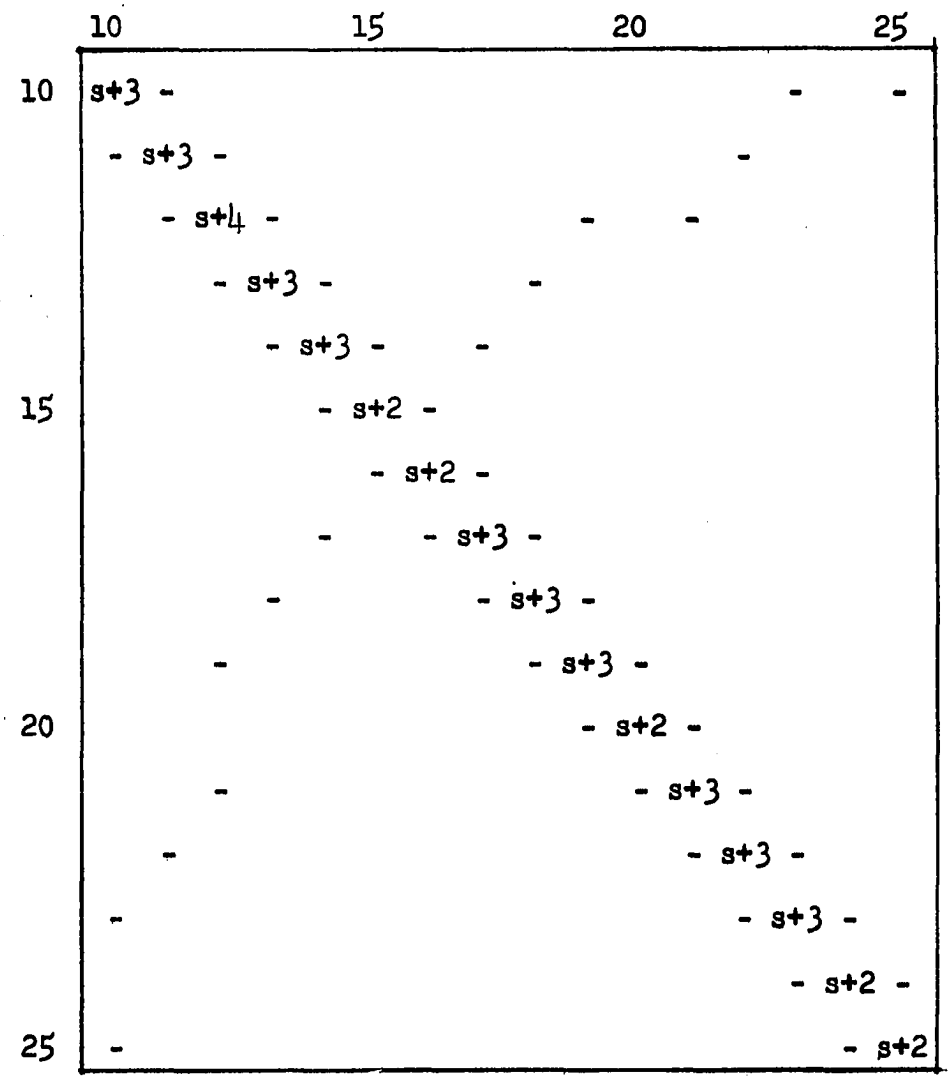
To give an indication of the appearances of the matrices for a larger problem, the $\overline{P}_{4a}$ matrix and the $\overline{H}$ matrix will be given for the following networks, which are simply the previous ones extended as far as typography will allow. The original network, which is solved directly, is the following:

R = C = 1



where, for the moment, branches 1 through 9 are to be ignored.

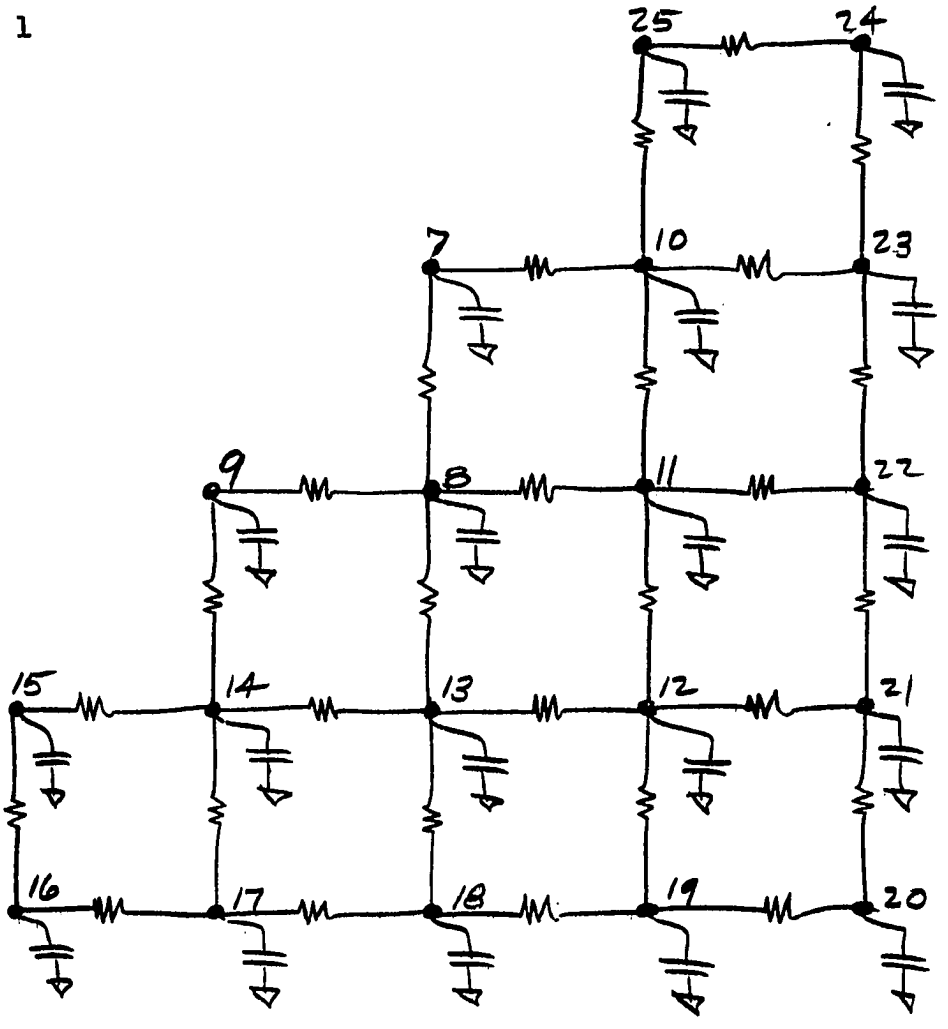
The node matrix is:



Although there are many zero entries, it is likely that every entry of the inverse will be non-zero in this case.

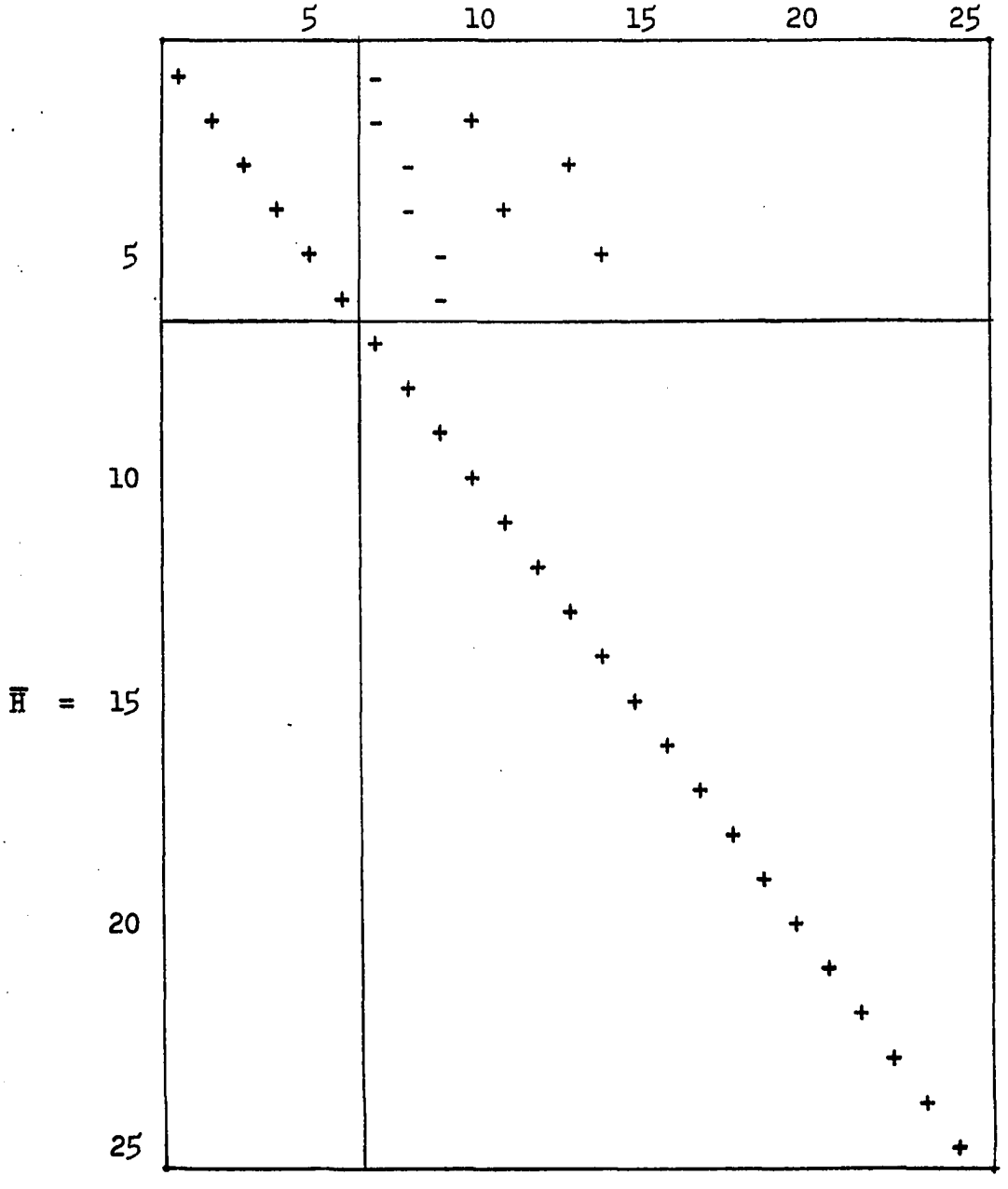
Suppose now the network whose solution is desired is:

R = C = 1



The previous network may be made into a torn version of this network by the addition of branches 1 through 9 as indicated in the previous diagram, and the node matrix of the modified network will be the original matrix augmented by main-diagonal terms only. Its inverse can therefore be immediately found in terms of the original inverse.

The remaining item needed to perform the manipulations is the change of basis matrix $\bar{H}$, which is found to be:



Note that again the inverse of this matrix may be obtained by just changing the signs of the off-diagonal terms.

Thus the inversion of the 19 by 19 node matrix of the pinched version may be accomplished by the inversion of a

9 by 9 matrix, and some matrix multiplications. The common estimate that the cost of inverting an n'th order matrix is proportional to n^3 then indicates an order of magnitude of reduction of cost of $(9/19)^3$, or a factor of about eight.

Some of the major features of the rearrangement technique will now be summarized. The use of the technique requires no trial and error calculations. The critical first step, of determining if two networks are pinched and torn versions of each other, requires essentially only the formation of the change of basis matrix for each network. Such matrices are simple to construct and easy to manipulate. The remaining steps then follow an explicit procedure, only standard matrix operations being necessary. Maximum use is made of known solutions of problems related to a given problem. The technique allows large networks to be solved more economically than before, or allows more alternatives to be investigated for the same cost.

LIST OF REFERENCES

(1) P. R. Halmos, Finite-Dimensional Vector Spaces, John Wiley & Sons, New York, 1953.

(2) J. P. Roth, An Application of Algebraic Topology to Numerical Analysis: On the Existence of a Solution to the Network Problem, Proc. Nat. Acad. Sci., vol. 41 (1955), pp. 518-521.

(3) J. P. Roth, An Application of Algebraic Topology: Kron's Method of Tearing, Quar. Appl. Math., vol. 17 (1959), pp. 1-24.

(4) N. Jacobson, Lectures in Abstract Algebra, Vol. II - Linear Algebra, Van Nostrand, New York, 1953.

(5) E. A. Guillemin, The Mathematics of Circuit Analysis, John Wiley & Sons, New York, 1949.

(6) E. A. Guillemin, Introductory Circuit Theory, John Wiley & Sons, New York, 1953.

(7) J. P. Roth, The Validity of Kron's Method of Tearing, Proc. Nat. Acad. Sci., vol. 41 (1955), pp. 599-600.

(8) G. Kron, A Set of Principles to Interconnect the Solutions of Physical Systems, Jour. Appl. Phys., vol. 24 (1953), pp. 965-980.

(9) P. Le Corbeiller, Matrix Analysis of Networks, John Wiley & Sons, New York, 1950.

(10) R. R. Sabroff and T. J. Higgens, A Critical Study of Kron's Method of Tearing, Part 1, Matrix and Tensor Quar., vol. 7 (1957), pp. 107-113.

(11) G. Kron, Equivalent Circuit of the Field Equations of Maxwell, Proc. I.R.E., vol. 32 (1944), pp. 289-299.

(12) G. Kron, Electric Circuit Models of the Schrodinger Equation, Phys. Rev., vol. 67 (1945), pp. 39-43.

(13) G. Kron, Equivalent Circuits of the Elastic Field, Jour. Appl. Mech., vol. 11 (1944), pp. A149-161.