HOMOLOGICAL STRUCTURE OF OPTIMAL SYSTEMS

Keith G Bowden

THESIS FOR PhD 1982

Dept of Control Engineering, Sheffield University

TEXT BOUND CLOSE TO THE SPINE IN THE ORIGINAL THESIS

Table of Contents

Diagrams

Glossary

CHAPTER II The Homology Theorem: Roth's Diagram

CHAPTER III Singular Homology Theory: Network Theory

CHAPTER IV De Rham Cohomology Theory: Field Theory

CHAPTER V Physical Structure in General Systems

CHAPTER VI Optimal Control Theory

(25) 6.1 The Partitioned System
(26) 6.2 The Torn System
(27) 6.3 Discrete Optimal Control as a Multidimensional Space Filter
(28) 6.4 The Forward System
(29) 6.5 The Continuous System

CHAPTER VII Example: Leontief's Economy

(30) 7 The Structure of published Input-Output Tables

CHAPTER VIII General Homology Theory

(31) 8.1 Construction of Exact Sequences from Complexes
(32) 8.2 Substructure of Discrete Optimal Control as a Long Exact Sequence
(33) 8.3 Substructure of Discrete Optimal Control as a Multidimensional Space Filter
(34) 8.4 Discrete Optimal Control as a Multidimensional Space Filter showing Homology Sequences and Scattering Structure

CHAPTER IX Computer Programming

(35) 9 Historical Evolution of High Level Languages

CHAPTER X Conclusions

(36) 10 Conceptual Diagram

Appendix I Topological Aspects of Invariant System Zeros

(37) 11.1 Internal Structure of Orthogonalisation Procedure
(38) 11.2 Internal Structure of Extended Orthogonalisation

Appendix III Amari's Generalised Diakoptics

(39) 13 Amari's Algebraic Diagram for Diakoptics

Abstract

Pure mathematics is often classified as continuous or discrete, that is into topology and combinatorics. Classical topology is the study of spaces in the small, modern topology or homology theory is the study of their large scale structure. The latter and its applications to General Systems Theory and implications on computer programming are the subject of our investigations.

A general homology theory includes boundary and adjoint operators defined over a graded category. Singular homology theory describes the structure of high dimensional simplicial complexes, and is the basis of Kron's tearing of electrical networks. De Rham Cohomology Theory describes the structure of exterior differential forms used to analyse distributed fields in high dimensional spaces. Likewise optimal control problems can be described by abstract homology theories. Ideas from tensor theory are used to identify the homological structure of Leontief's economic model as a real example of an optimal control system. The common property of each of the above systems is that of optimisation or equivalently the mapping of an error to zero. The criterion may be a metric in space, or energy in an electrical or mechanical network or system, or an abstract cost function in state space or money in an economic system and is always the product of a covariant and a contravariant variable.

The axiomatic nature of General Homology Theory depends on the definition of an admissable category, be it group, ring or module structure. Similarly real systems are analysed in terms of mutually recursive algebras, vector, matrix or polynomial. Further the group morphisms or mode operators are defined recursively. An orthogonal computer language, Algol82, is proposed which is capable of manipulating the objects described by homological systems theory, thus alleviating the tedium and insecurity incurred in implementing computer programs to analyse engineering systems.

Glossary

This work was written in parallel rather than from beginning to end - it just grew - and as such it should be read, to a certain extent, in the same way. It is not necessary to follow all the mathematical sections through in detail, many have probably been dealt with better elsewhere, eg in the references: it is the ideas behind them that are important. It is felt however that it is worthwhile introducing some mathematical concepts, all of which appear in the thesis, at this stage. An object will often be denoted by the same character regardless of the category we view it from. We will in fact consider an object to be in whichever category is convenient at a given time. The categories of objects used in this thesis are described below and shown pictorially in Fig(0.1). The following is (conventionally) written in terms of multiplicative groups. For additive groups, group multiplication must be replaced by addition, inversion by negation and unity by zero.

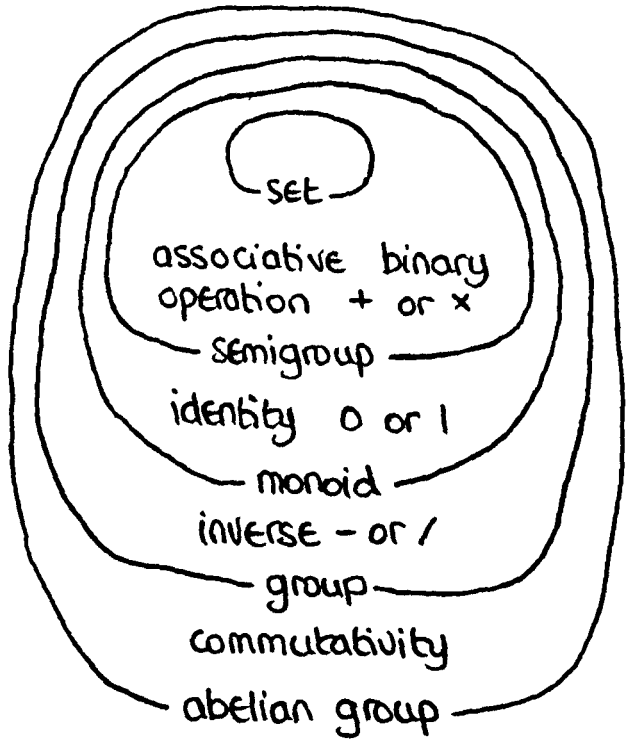
A semigroup is a set G together with an associative binary operation GxG->G, ie for any x,y,z in G, (xy)z=x(yz).

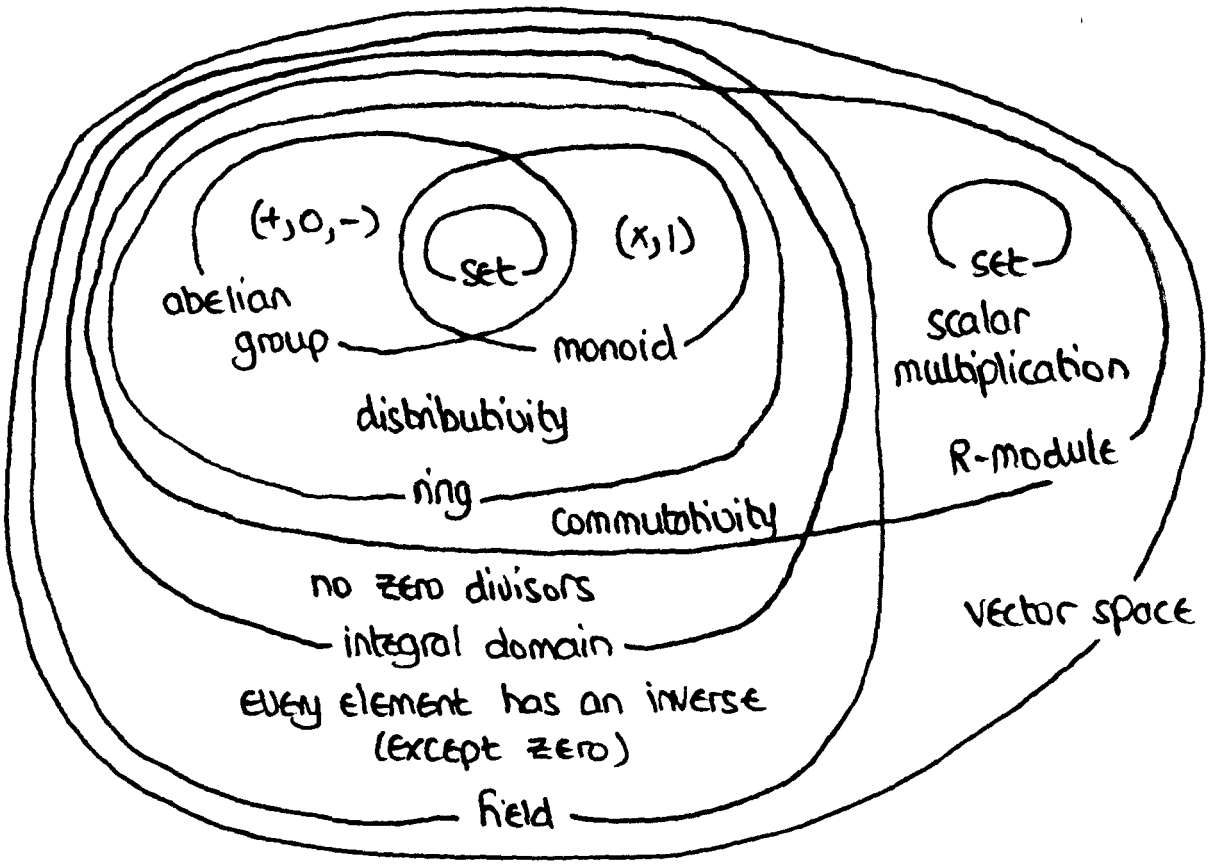
A monoid is a semigroup with an identity, e st, for any x<G, xe=ex=x.

A group is a monoid st for any x<G there exists an inverse $/x=x^{-1}$ st x/x=/xx=e.

An abelian group is a commutative group ie xy=yx. For instance the set of integers under addition is an abelian group as is the set of reals under addition or multiplication.

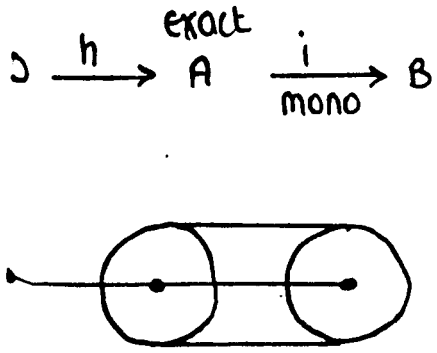
Fig (0.1) Categories

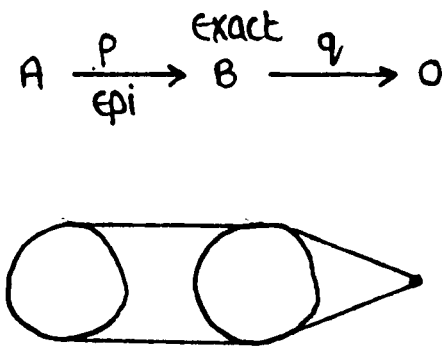


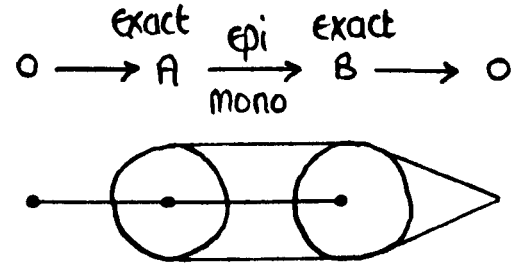


A subgroup is a subset preserving the group operation. If H is a subgroup of G, (or isomorphic to a subgroup of G) a coset of G by H is a set of elements gH, the quotient group G\H is the class of cosets of G by H with the law of composition (gH)(g'H)=(gg')H. (G modulo H.) Eg. the quotient set of the integers by the set of even numbers under addition is the class {odds, evens}. This is one of the most important concepts in the geometric theory. For example if $G=(R^3,+)$ and $H=(R,+)=\{h,0,0\}$ then G\H consists of all lines parallel to the x-axis (each line is a coset) and has two degrees of freedom - we can write $R^3\backslash R=R^2$ which explains the term quotient. Similarly if $H=(R^2,+)$ then G\H consists of all the planes perpendicular to the z-axis. G\H is a disjoint partition of G where the elements of G\H are copies of H.

A homomorphism or structure preserving map h:A->B between abelian groups is a function satisfying (ah)(a'h)=aa'h, then 0h=0 and (/a)h=/ah. The kernel (null-space) of h is the subgroup $h^{-1}(0)$ of A, the image (range) of h is the subgroup Ah of B, the cokernel of h is the quotient group B\Ah of B and to complete the duality we may define the coimage of h as the quotient group $A\backslash h^{-1}(0)$. Then h is a monomorphism (monic or 1:1) if its kernel is zero, see Fig(0.2), an epimorphism (epic or onto) if its cokernel is zero (equivalently if B is the image of h), see Fig(0.3), and an isomorphism or information preserving map A~B if it is monomorphic and epimorphic, see Fig(0.4). An endomorphism of A is a homomorphism A->A and an automorphism of A is an isomorphism A->A. If B is a subgroup of A then the monomorphism i:B->A given by ai=a, a<B is called the inclusion map or injection, the epimorphism p:A->A\B which sends each element of A into its coset is called the projection or surjection. For completeness homeomorphisms, diffeomorphisms and holomorphisms are continuous, differentiable and integrable maps respectively.







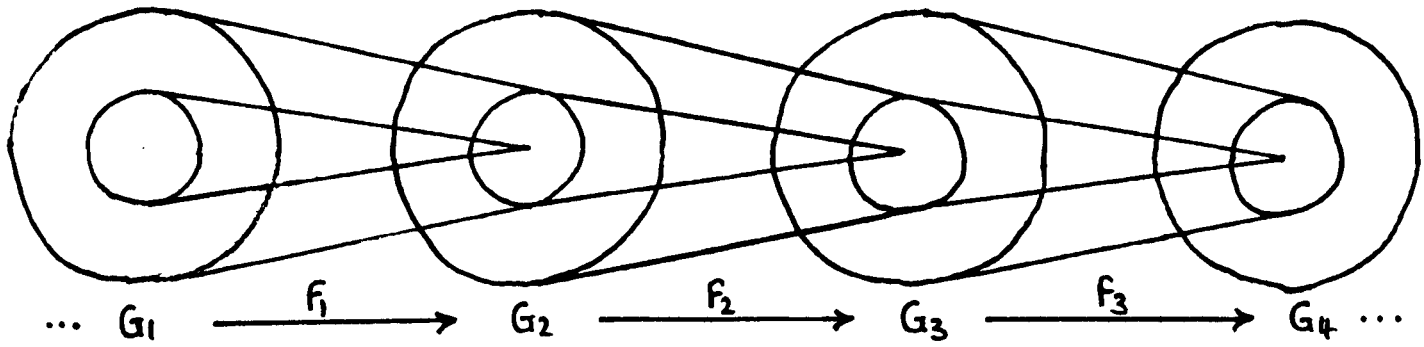
$im(h) = Ker(i) = 0$

Fig(0·2) Monomorphism

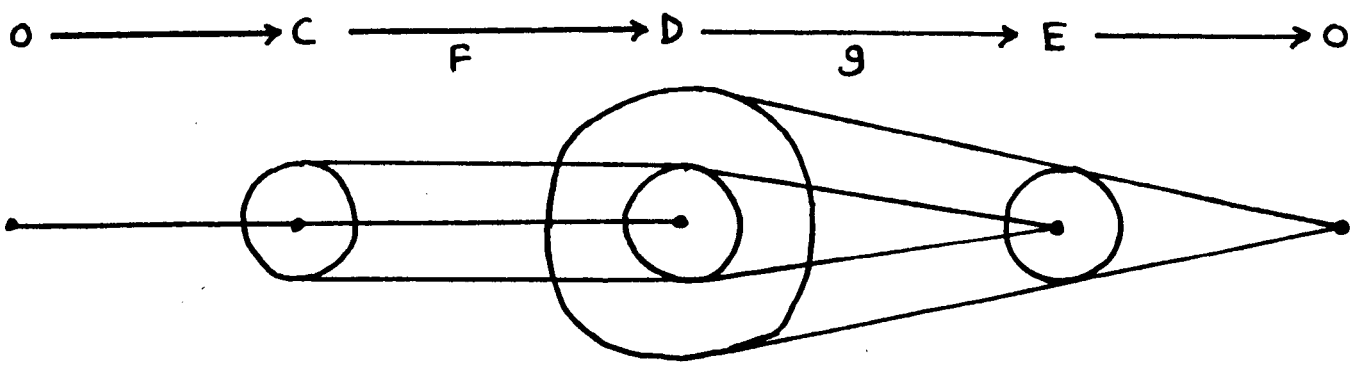
$im(p) = Ker(q)$

Fig(0·3) Epimorphism

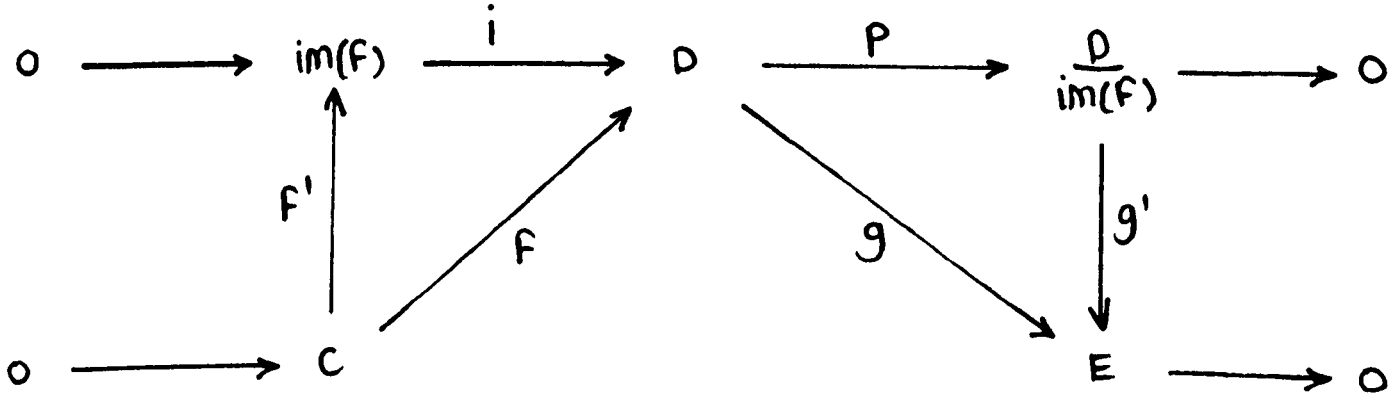
Fig(0·4) Isomorphism



Fig(0·5) Long exact sequence



Fig(0·6) Short Exact sequence



Now consider matrix multiplication as a homomorphism. $\underline{H}:R^m->R^n$ is a homomorphism (into) iff $\underline{H}(\underline{0})=\underline{0}$, ie $\underline{H}$ is full rank. $\underline{H}$ is a monomorphism (1:1) iff $\underline{H}^{-1}(\underline{0})=\underline{0}$, ie $\underline{Ha}=\underline{0}$ implies $\underline{a}=\underline{0}$, ie $\underline{H}$ is full rank and n>=m. $\underline{H}$ is an epimorphism (onto) iff $\underline{b}\backslash\underline{Ha}=\underline{0}$, ie for any $\underline{b}$ there exists an $\underline{a}$ such that $\underline{b}=\underline{Ha}$, ie $\underline{H}$ is full rank and m>=n. $\underline{H}$ is an isomorphism iff it is 1:1 and onto, ie full rank and square.

The behavior of tensor products of monomorphisms (eg torsion products in homological algebra) is often described using the terminology of exact sequences. There are such exact sequences of vector spaces, abelian groups, or more generally modules over any ring R, commutative or not. A triple

$$C\xrightarrow{f}D\xrightarrow{g}E$$

of abelian groups and homomorphisms is exact at D iff image(f)=kernel(g) and g.f=0. A sequence of abelian groups and homomorphisms

$$\ldots->G_1\xrightarrow{f_1}G_2\xrightarrow{f_2}G_3\xrightarrow{f_3}\ldots->G_n\xrightarrow{f_n}\ldots$$

is (long) exact iff each triple is exact, see Fig(0.5). An exact sequence

$$0->C\xrightarrow{f}D\xrightarrow{g}E->0$$

is called short exact, see Fig(0.6). This sequence is exact if and only if f is a monomorphism, g is an epimorphism, and it is exact at D. D is an <u>extension</u> of E by C. This is a generalisation of the concept of isomorphism in the sense that $h:G_1->G_2$ is an isomorphism iff

$$0->G_1\xrightarrow{h}G_2->0$$

is exact.

Note that in a short exact sequence as above, f is a monomorphism and identifies C with a subgroup $C' \leq D$. Thus up to isomorphism any short exact sequence can take the form

$0 \to C' \to D \to D \setminus C' \to 0$

where $D \to D \setminus C'$ is an epimorphism because taking $C'=im(f)$ from Fig(0.6), f' and g' are both isomorphisms.

A ring R is a set together with two identities {1,0} corresponding to two binary operations (.,+) st R is an abelian group under addition and a monoid under multiplication and the law of distributivity holds, ie $x(y+z)=xy+xz$. The set of reals under addition and multiplication is a ring. A commutative ring is a ring which is commutative under multiplication.

An integral domain is a commutative ring without zero divisors, ie no pair x,y exists st xy=0 where 0 is the identity under addition.

An ideal is a subset S of a ring R iff S is stable ie $x,y \in S \Rightarrow x-y \in S$ and $r \in R, s \in S \Rightarrow rs \in S$. A principle ideal S=Rr is an ideal generated by a single element $r \in R$. A principle ideal ring is a ring in which every ideal is principle. A principle ideal domain is a principle ideal ring which is also an integral domain.

A field,R is an integral domain st R-{0} is an abelian group under multiplication ie every element is a unit (element with a multiplicative inverse) except zero so R does not necessarily have to have an infinite element. Eg. the ring of polynomials K[z], over a field K is an integral domain. The units of K[z] are the polynomials of degree 0.

A (left) vector space, X over a field K is a set X and a map called scalar multiplication, st KxX->X:(a,x)->ax and the usual laws of distributivity, etc hold. This is a generalisation of the conventional vector space in which K=R. The outer or direct product of two vectors (multivalued objects) is essentially the set of products of their elements, $\{x_i,y_j\}$. The inner product is the contracted set or scalar produced by summing over certain of these products, eg as in the ordinary matrix product. The exterior product is an asymmetric sum, usually the contracted set $\{x_iy_j-x_jy_i\}$ as in the conventional cross product of vectors.

An R-module is a generalisation of a K-vector space where the field, K is replaced by a ring R, eg the space of polynomials K[z]. An R-module is said to be finitely generated iff there exists a finite set of generators or basis, and is called a free R-module if this basis is unique. A free R-module thus is nearly the same as a vector space. Since we are only interested in commutative rings we do not distinguish between left and right R-modules.

A graded module G is a module representable as a direct sum of a finite or denumerable number of modules.

An (associative) algebra over the commutative ring R with unit 1 is a left (right) module over R possibly lacking a unit, the multiplication satisfies (ra)(r'a')=rr'aa' where r,r' are in R. A ring is an algebra over R.

A category, C is a class of objects obj(C), together with:

(1) A function assigning to each pair (X,Y) of objects in C a set mor(X,Y). An element f<mor(X,Y) is called a morphism f:X->Y of C with domain X and codomain Y.

(2) A law of composition assigning to each triple of objects (X,Y,Z) in C a function mor(Y,Z)xmor(X,Y)->mor(X,Z). For morphisms g:Y->Z and f:X->Y, this function is written as gf:X->Z such that the following axioms hold:

(1) Associativity. If h:Z->W, g:Y->Z, f:X->Y then h(gf)=(hg)f.

(2) Identity. For each Y of C there exists I:Y->Y st If=f for f:X->Y and gI=g for g:Y->X.

If C and C' are two categories a functor F:C->C' is a pair of functions:

(1) An object function which assigns to each object A of C an object F(A) of C'.

(2) A mapping function assigning to each morphism f:X->Y of C a morphism F(f):F(X)->F(Y) of C' satisfying F(I(X))=I(F(X)) for any I<C and F(gf)=F(g)F(f) for any gf<C.

CHAPTER I Introduction: History

The recent integration of methods formerly peculiar to particular disciplines has led to satisfactory new techniques. This thesis attempts to unify various branches of engineering and scientific programming methodology within a very general branch of continuous mathematics known as differential topology (homology theory). The idea was originally suggested by Kron[1], as a justification of his work on network theory and tearing, based on an insight by Roth[11], and later extended by Branin[12] and Nicholson[6].

We give an account of Kron's systems theory, in the light of recent work and extend it into a unified theory with emphasis on the inherent physical structure of abstract general systems. Kron, in fact, is directly or indirectly responsible for many of the scientific computational methods used today. He continually emphasised that there must be an underlying justification for the proliferation of mechanical and electrical network analogies used as an aid to solve problems in widely varying scientific fields, engineering and sociology. Homology theory[5] was developed by pure mathematicians as an abstraction of certain classical branches of mathematics: the derivation and meaning of the word 'homology' is similar to that of 'analogy'. We take the view here that the role of the applied mathematician should be to interface the work of the pure mathematician to reality in a natural and useful way. In particular we attempt an integration of General Homology Theory, Optimal Systems Theory and Orthogonal Programming Languages.

This Chapter presents the historical background to Kron's network theory, the relevance of homology theory and the parallel developments in optimal control. Chapter II introduces the basic tools of least squares matrix algebra and presents a new theorem giving necessary and sufficient conditions for a very wide class of

systems to be optimal. Commutative diagrams are introduced and it is demonstrated that the solutions can be read off Roth's diagram. This idea originated in network theory and was extended to the distributed case by Branin and to optimal control by Nicholson. We present the general case. A hypothesis giving the set of sufficient categories for the homology theorem to hold is stated. The equations of Pontryagin's maximum principle are easily obtained as a special case. Chapter III reviews the electrical network as a chain complex structure with the current and voltage vectors as additive groups[10]. Nicholson's scattering theory is applied to the orthogonal network. Chapter IV reviews electromagnetic field theory in terms of exterior differential forms[2] and describes Kron's network model of Maxwell's equations and Branin's algebraic diagram. The finite element method is described and justified using de Rham's theorem. Chapter V gives some of the properties of physical structure of general systems. Well known analogies between very different physical systems can be shown to be based in their common mathematical structure. The difference between tensors-in-the-small and tensors-in-the-large, covariance, contravariance and scattering theory are explained. It is emphasised that one should always be aware of the tensorial structure even when working in matrix notation.

Chapter VI applies Kron's ideas to optimal control theory. The multistage optimal control problem is presented in partitioned matrix form and the solution obtained from the Homology Theorem. Roth's diagram for each torn stage is shown to model that for the overall problem and the Riccati equation is read straight off the diagram. The system is then reconnected into a form consistent with Kron's algebraic diagram for the 'multidimensional space filter' thus optimal control theory is presented as a chain complex in matrix form. This geometric analysis is seen to have much in common with that of Wonham[9] et al. The scattering structure of the orthogonal form of these equations is shown to lead to the Chandrasekar equations. The analysis is repeated for continuous optimal control. Chapter VII applies these ideas to Leontief's input-output model of the economy. Prices and commodity flows are shown to

be the co- and contravariant variables and the model shown to constitute a generalised network with 'money' as the utility function. The Walras law is the topological constraint equivalent of the optimisation problem. Some interesting results are obtained for the continuous and discrete dynamic cases.

In Chapter VIII we give the formal definition of a chain complex and state the Eilenberg-Steenrod axioms for a general homology theory in an admissable category. The de Rham cohomology theory of differential forms and the singular homology theory of simplicial (chain) complexes are shown to be special cases and the de Rham theorem is stated. Chapter IX gives a brief history of the development of high level computer languages and a computer language, Algol 68, which is capable of treating matrices, tensors, groups and algebraic diagrams as objects and manipulating them directly is described. The Conclusion reviews the results in the text and proposes some topics for future research. It is explained why this structure (homology theory) is thought to be the generalised network that Kron and Branin were looking for. Applications to general systems theory in terms of K[z] modules and to Artificial Intelligence with regard to the 'core' of a program are discussed. Appendix I shows the relevance of algebraic diagrams to the analysis of transmission zeros. Appendix II derives de Rham's theorem, this is mainly included to demonstrate the way mathematicians handle abstract objects. An abstract homology theory is described in terms of general chain complexes. Appendix III describes Amari's generalised diakoptics in operator form.

1.1 Network theory

Gabriel Kron was born in Transylvania in 1901, obtained a degree in electrical engineering at Michigan University in 1924, and rather than continue on to postgraduate work went on a lone walking tour around the world, taking copies of Weatherburn's 'Vector Analysis' and Forsythe's 'Differential Equations'. From his return in 1928 he worked for GEC until his retirement in 1966: he died suddenly in 1968. He received an honorary doctorate from Nottingham University in 1961. Over this period of time he developed, published and practiced a general theory of electrical machines, networks and general systems, with emphasis on both the physical structure of abstract systems and on his method of 'Diakoptics' or solution by tearing (decomposing) the network or system into smaller subsystems, solving each part separately and recombining to give the overall solution which, at least in the linear case, is exact. One great advantage of this method is that if any change were required to be made to any part of the system only that part need be solved again and the system recombined. Kron, in 1959, on an IBM 'card program calculator', inverted a 256 by 256 matrix by tearing into 16 subdivisions each taking about an hour to solve. Most of Kron's work appeared in the 'Electrical Journal' as a series of articles (1957 to 1959), later published as 'Diakoptics' (1963). Happ systemised Kron's network theory taking care of many special cases such as singular subdivisions.

Kron's electrical network analysis can be classified into two dual formulations: the mesh method and the node method. These are more general than the classical notions of a dual network as they include the non-realisable dual of a planar network. Thevenin's and Norton's theorems and Kirchoff's laws are all taken into account. The mesh method essentially involves defining a 'spanning tree' over the network. The currents in the branches of this tree are then independent. A

rectangular connection matrix of 1's, -1's and 0's is set up relating the directed currents in these branches to notional directed mesh (closed loop) currents. The unknown currents and voltages can be found in terms of the current and voltage generators by simple matrix manipulation (vectors in the large), including one matrix inversion of the order of the number of meshes involved. The connection matrices of the dual node method are annihilators of the mesh connection matrices. Kron retaliated to criticism of his method by showing that his singular connection matrices were part of a larger nonsingular matrix by including the link (non-tree) branches into an 'orthogonal network of solenoidal and lamellar currents'. He further insisted that his matrices were in fact second rank tensors (he called them tensors-in-the-large), the voltage and current vectors exhibiting the covariant and contravariant properties of 'across and through variables'. Diakoptics is an extension of this method which allows subsections of the network to be solved separately and the solutions combined with the remaining 'intersection network' using - from a matrix manipulation point of view - the Householder inversion lemma. Kron saw the intersection network as a 'miniature model of the original system'. A large number of small matrices has to be inverted resulting in greater computational efficiency. Kron continually emphasised that the network and not the equations should be torn as there is an actual loss of information in going from one to the other. He applied tearing to many practical cases including mechanical networks, linear programming models of transportation and load flow problems, (he is responsible for most methods of economic dispatch used today), network models of Poisson and diffusion equations and Schrodinger's equation, plastic and elastic fields, molecular models and to the solution of 'divided difference' equations. Despite the fact that he visualised the method of tearing in a number of interconnected spaces, Kron never used higher rank tensors in his publications.

A number of other groups have contributed towards this work, most particularly the 'Research Association for Applied Geometry' in Japan and the 'Tensor Society of Great Britain'. Kron became known in Japan in 1953. His work was applied by the RAAG to numerous practical problems, including tearing of plastic and elastic fields, electrodynamic, aerodynamic and hydrodynamic problems and notably tearing of Shannon's information space. In a similar way the TSGB publishes a quarterly journal extending Kron's work.

1.2 Homology theory

Roth (1959) showed how Kron's network theory was based in homological algebra and produced an algebraic diagram of exact sequences, in which consecutive pairs of maps annihilate each other, involving Kron's connection matrices. This was later extended by Branin (1966) to electromagnetic field theory and by Amari (1962) to the diakoptical case. Amari also dealt with diakoptical eigenvalue analysis. Onodera (1960) classified tearing into two dual cases, diakoptics or open circuiting being the cohomological case and codiakoptics or closed circuiting the homological case. Kondo and Iri (1958) dealt with the homology groups in detail in 'Theory of Trees, Cotrees, Multitrees, and Multicotrees'. Before Kron died he stated that he felt the correct algebraic structure for an excited electrical network was a 'fibre bundle': a statement which appears to have been taken seriously by the topologist Steenrod, but never extended further. Kron however took Roth's algebraic diagram and extended it into a 'multidimensional space filter' (1959) involving most of the basic concepts involved in his work including Gauss' and Stokes' theorems. Kron pointed out that De Rham's theorem, which he interpreted as stating that there exists an isomorphism between the homological, topological structure of underlying 'dead' electrical networks and the

cohomological, algebraic structure of the superimposed 'live' electromagnetic fields justified his network equations. De Rham's theorem also can be taken to give an isomorphism between electrical network theory of simplicial complexes and electromagnetic field theory of differential forms (Grassman algebra), justifying the methods of finite differences and finite elements where discretisation of fields is carried out by modelling the field with a mesh and taking the limiting case as the mesh gets finer. Kron's First Generalisation Postulate states that 'The n algebraic equations of a physical system with n degrees of freedom may be replaced by a single equation having the same form as that of a single unit of the system, if each variable is replaced by the appropriate n-matrix.' Rothman discusses the 'Philosophical Meaning of Tensor Theory', pointing out that Kron was a 'dignified follower of the Greek philosophers' in his 'continuous aim of finding a general principle' in Nature. Further that 'There is still a gap between the most advanced concepts of differential manifolds and Kron's representation. Cartan, De Rham, Hodge etc. take into consideration only two spaces being neighbours and not a network of spaces.'

Electromagnetic theory was treated by Maxwell in terms of vector analysis (in the small). This was found unsatisfactory by Einstein who used tensors (in the small) to develop Relativity Theory. Misnor and Wheeler progressed one stage further in their remarkable book 'Gravitation' and worked (partly) in terms of Eli Cartan's exterior 'differential forms': these are the objects whose topology is described by the de Rham cohomology theory. Differential forms were popularised by Harley Flanders (1963) and used by Lynn[13] et al in their book 'Differential Forms on Electromagnetic Circuits' to describe Kron's network models of electromagnetic fields. The operators grad, div and curl of electromagnetic theory can be generalised both to tensors and differential forms when just the boundary operator, d and the Hodge star operator, * cover all cases. A coboundary operator and a generalised Laplacian can be defined in terms of these. The Hodge theorem

generalises Helmholtz' theorem in higher dimensional spaces. Network theoretic versions of these formulae are also available, eg. the Poincare lemma becomes Kron's lemma.

Most of the work carried out in homological algebra by topologists disregarded practical applications other than to pure mathematics. A number of 'classical' homology theories arose based on (1) the structure of simplicial complexes (high dimensional polytopes) called singular homology (2) the structure of differential forms (high dimensional fields) called de Rham cohomology theory and (3) the structure of chain complexes, based on module theory eg. rings of polynomials, see Fig(10). The most important concept in singular homology theory is that of the short exact sequence ie. that the boundary of a simplex does not itself have a boundary. This is equivalent in de Rham cohomology theory to a generalisation of the statements curl(grad(.))=0 and div(curl(.))=0. These concepts both indicate some kind of minimisation - in the electromagnetic case that of energy. A ring consists of a set and two operations known as addition and multiplication subject to certain restrictions. Sets, groups, rings and modules etc. are known collectively as categories. Eilenberg and Steenrod defined a general homology theory over an arbitrary category subject to certain admissible conditions and gave the necessary structure and axioms for the structure to include the classical homology theories. Warner bases his general cohomology theory on axiomatic sheaf theory. Hilton noted that homology theory describes covariant objects and cohomology theory describes contravariant objects and instead used the terms cohomology and contrahomology respectively. Bourgin uses the generic term omology.

1.3 Control theory

The idea of a General Systems Theory was proposed by Ludwig von Bertalanffy and extended by Preston Hammer, Wayne Wymore, Norbert Weiner, Herbert Simon, G J Klir, David Wismer et al. After the definition of dynamical systems by Kalman, Falb and Arbib[4] (1969) in terms of semigroups, mathematical systems theory of optimal control based on the use of commutative diagrams soon followed. This was fully axiomised by Kalman whose work was based in module theory, the special case being a ring of polynomials in the inverse z (shift or discrete Laplace) operator. Mesarovic and Takahara[3] (1975) concentrate on the structural aspects of an abstract input-output system, rather than optimality and have classified different realisations of systems and their interrelationships in terms of categories and functors. Kalman's work originates partly in Arbib's treatment of automata theory and it is pointed out that a formal system of natural language is an example of a general system in this sense. Work in Artificial Intelligence: natural language and image processing - and in Computer Science - has recently been progressing towards this formulation. Indeed Kron saw his multidimensional space filter as a foundation for a self organising intelligent automaton. Computer languages like Algol 68 and Lisp allow rich algebraic structures to be defined and manipulated.

Classical control theory originated from the frequency domain techniques of Bode, Nyquist, Nichols, Evans (root locus) and Routh and Hurwitz and reappears in the neoclassical techniques of Macfarlane, Rosenbrock et al. In the USA, Bellman and Kalman dealt with dynamic programming and Gaussian estimation theory, respectively, leading to matrix Riccati and Chandrasekhar type differential equations, the separation (certainty equivalence) principle and Liapunov stability theory. In the USSR Pontryagin and his coworkers used the more general calculus of variations of Lagrange and Hamilton to formulate the celebrated Maximum Principle. Recently Wonham, Denham, Macfarlane et al have provided a geometric basis for control theory

similar to that of functional analysis. Owens has applied this to the analysis of transmission and decoupling zeros linking the frequency response and state space control theories. These are the fixed modes which are invariant under feedback and duality. Multilevel decomposition and coordination has been dealt with by Pearson for LQG systems, and Mesarovic in the set theoretic case (with a preoccupation for existence theorems). Nicholson (1971 - 1979) took Kron's electrical network theory and showed how it was analogous to the general and sequential least squares problem (as applied to tearing of a multimachine system), to multivariable control, to discrete and continuous optimal control theory and Kalman-Bucy filtering (also duals in a sense) and to the discrete and continuous smoothing problems by showing how each theory fitted into the mathematical structure of scattering theory using Redheffer's star operator. He also suggested implementing specialised computer hardware to realise this operation. Quittard-Pinon[16] (1981) has used a Roth type diagram in Time Series Analysis. Bellman's original work on invariant imbedding involved using a scattering formulation. Kailath , Ljung and Friedlander used this work to show that the Chandrasekhar equations could replace the Riccati equation of dynamic programming with more computational efficiency in many cases. Flanders treats Lagrangian dynamics in the context of differential forms, deriving the Euler-Lagrange equation and the Hamiltonian equations which are the basis of the Pontryagin maximum (minimum) principle (usually proved using fixed point theorems). Vanacek extended optimal control and estimation theory into the realms of higher dimensional tensors, a natural formulation for the representation of multilevel systems. Vanacek mentions that his solution should apply to the Toeplitz equations which have since been solved in the general case by Khabie-Zeitoune (1980).

Our work will show how optimal control theory and general systems theory fit into the homological scheme of things, filling in many gaps in current knowledge and deriving some new and efficient algorithms in the process. Our approach is that any optimisation problem, whether it be energy minimisation, cost function

minimisation or minimisation of a metric is equivalent to mapping something to zero. This something may simply be the system error, or the boundary of a boundary or it may be something of the form curl(grad) or div(curl) or it may be a metric on a manifold eg vector space. The induced short exact sequence(s) involves the boundary operators of homology theory. Indeed Wonham's algebraic systems theory implicitly uses many of these concepts without explicitly naming them. It is hoped that the reader will be able to see any particular problem in the context of the whole scheme and thus be able to find the most efficient solution in a systematic manner. In each chapter we deal with certain fixed topics eg. optimality, orthogonality, duality, tearing, scattering, 'live' and 'dead' structures, co- and contravariance, discretisation, analogies, de Rham's theorem, tensors and categories. The presentation is informal, rigour being left to the more mathematically inclined. The work is largely self contained though the Glossary and References [1] and [6] at the end of this chapter should be considered as essential.

The author has only recently discovered the work of Hermann[8] who has also used algebraic geometry as a tool in systems theory. His work concentrates on frequency domain concepts rather than optimality thus complementing this thesis. We find it relevant however to quote freely from a review of his book by Wonham[7]. 'Let A:nxn and B:nxm be fixed matrices with elements in a number field K... Let F:mxn and T:nxn be matrices over K representing, respectively, state feedback and change of basis (so T is nonsingular). Thus the set of all matrices we can hit, up to similarity, by means of feedback, is just the image of the map $f:K^{mxn}+K^{nxn}\to K^{nxn}:(F,T)\to T^{-1}(A+BF)T$. Notice the minor abuse of notation: as a rational map, f is actually defined only on a subset of its indicated domain, in this case where det(T)#0. In algebraic geometry f is said to be defined 'almost everywhere'... Generally speaking most of the formal structural synthesis problems that have been studied in linear multivariable control could be posed as questions

about the rational map f or suitably extended versions of it... Now comes the remarkable algebro-geometric fact: if the field K is C, you can conclude that f itself is 'almost onto', namely, the image of f includes almost every point of the codomain. So you can either hit or get as close as you like to every point of the codomain! In our example it turns out that (the derivative of f) Df is onto just when (A,B) is controllable. Then (K=C), for almost any $A_o \in C^{nxn}$ there will exist a feedback map $F \in C^{mxn}$ such that A+BF is similar to A_o. The catch is that you may need F to be complex even when A, B, and A_o are all real: and complex F may not have any systemic interpretation, and so may not be admissable in the original problem context. In the case K=R (where system theorists feel more at home), the implication 'Df onto => f almost onto' is no longer true. The foregoing approach, of using systemic properties to infer Df onto, to infer f almost onto..., therefore breaks down.

Now there may be other ways, as yet undiscovered, in which algebraic geometry might help. Hermann is very enthusiastic about this... The scenario is dramatic. The bad guys are the 'modern algebraic geometers, (who are not only) uninterested in applications-orientated material' (p. 105) but inflict on the unsuspecting public their 'gobbledy-gook and Bourbaki-style generalised nonsense... (being evidently) afflicted with a Death Wish, fatally smothering a beautiful, classical subject...' (p. 15). The good guys are the nineteenth century founders of the subject (Kronecker figures prominently), who are asserted to be 'much closer in spirit to the need (sic) of <u>modern</u> applied mathematics... (because) they were often much more concerned about progress in general science than are todays mathematicians!' (p. 74). Even if unwilling to take sides, one cannot help but recognize an authentic cri de coeur...' (Wonham[7]). These are this author's sentiments exactly. He would number Kron amongst the good guys... When reference was made to the TSGB the Head of a certain Mathematics Department declared 'Oh, We don't talk about that.'

One of the most important groups of results in mathematics that can be derived from homology theory are the 'fixed point theorems' of Brouwer, Kakutani, etc. These were used profusely by the mathematical economists to prove the existence of general equilibria in abstract economic systems. The general linear case of these models is the Leontief dynamic input-output model, as treated by Livesey. This is taken as a major example of an abstract control system with a wealth of inherent structure. Franksen (1969) discusses an electric network analog of Leontief's economy in detail in 'Mathematical Programming in Economics by Physical Analogs'. All the usual properties crop up. Prices are taken as covariant and commodity flows as contravariant variables. The objective function is thus money! The attempts of Franksen, Bott (a topologist, who used Grassman algebra) and Whitin to produce network models of the Leontief economy and Franksen's orthogonalisation of the input-output model and introduction of negprices are critically discussed. Teldahl's peculiar distributed economy, the natural extension of which is a relativistic economy with a finite time horizon, is shown to be a natural result. Walras' law is shown to be the economic minimum energy criterion.

1.4 References

(1) Kron, Diakopics: The Piecewise Solution of Large-Scale Systems, London, MacDonald, 1963.

(2) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(3) Mesarovic and Takahara, General Systems Theory: Mathematical Foundations, Academic Press, 1975.

(4) Kalman, Falb and Arbib, Topics in Mathematical Systems Theory, McGraw-Hill, 1969.

(5) Bourgin, Modern Algebraic Topology, Macmillan, 1963.

(6) Nicholson, Structure of Interconnected Systems, IEE Control Engineering Monograph 5, Peter Perigrinus, 1978.

(7) Wonham, Book Review of Ref[8], IEEE Trans AC-24, No 3, June 1979.

(8) Robert Hermann, Algebro-Geometric and Lie-Theoretic Techniques in Systems Theory, Part A: Interdisciplinary Mathematics, Vol 13, Brookline MA: Math Sci Press 1977. See also Vols 1-12.

(9) Wonham, Linear Multivariable Control: A Geometric Approach, Berlin, Springer-Verlag, 1974.

(10) Kondo and Iri, On the Theory of Trees, Cotrees, Multi-Trees and Multicotrees, RAAG Memoirs, Vol 2, 1958.

(11) Roth, An application of algebraic topology: Kron's method of tearing, Q. Applied Math, April 1959.

(12) Branin, The Algebraic-Topological Basis for Network Analogies and the Vector Calculus, Symposium on Generalised Networks, Poly. Inst. of Brooklyn, April 1966.

(13) Lynn, Balasubramanian and Sen Gupta, Differential forms on electromagnetic networks, Butterworth, 1970.

(14) Reid, Riccati Differential Equations, Academic Press, 1972.

(15) MacLane and Birkoff, Algebra, Second Edition, Collier MacMillan, 1979.

(16) Quittard-Pinon, (unknown), 4th ITSM, Valencia, June 1981.

(17) Martin and Hermann, A new application of Algebraic Geometry to Systems Theory, IEEE Conf Decision and Control, pp 821-5, 1976.

(18) Hermann, Algebro-Geometric and Lie-Theoretic Techniques in Systems Theory, Part A, (Interdisciplinary Mathematics Vol XIII), Maths Sci Press, 1977.

CHAPTER II The Homology Theorem: Roth's diagram

The various manifestations of tearing in orthogonal systems are realised by a set of related algorithms on partitioned matrices. Nicholson has shown how the ordinary least squares problem can be represented on Roth's diagram and thus has a dual pair of solutions, given by Kron's lemma. The Homology Theorem gives conditions on a topological system such that Roth's diagram commutes if and only if the system is optimal. Roth's diagram is constructed from an exact sequence related to the chain complex containing the system matrix and one of its annihilators. The adjoint system is constructed from a set of homomorphisms from this sequence to the optimisation criteria or metric. Finally the Hamiltonian equations of Pontryagin's maximum principle are shown to be implicit in Roth's diagram for the general nonlinear optimal control problem and the Euler-Lagrange equation can be read off the diagram.

2.1 Tearing in orthogonal systems

Kron introduced the celebrated orthogonal or complete network concept in reply to criticism of the 'non-physical' form of the 'singular connection tensors' in his circuit theory. The symmetry of the system is achieved by filling in the missing parts of the rectangular arrays - often with zero or unit submatrices - producing square matrices, by identifying the missing variables. Though this method overspecifies the system it has many computational and descriptional advantages. It will be found in the following chapters that the method is also important in general systems theory, though we more often use the singular form for derivations because of its nilpotent properties.

Kron used the orthogonal system as a basis for his method of tearing. In the 'finite element method' the connected system is a distributed field and the torn system is a lumped parameter representation. In optimal control theory the connected system will be assumed to be continuous and the torn system discrete. Tearing in fact is a ubiquitous concept. It may occur in space, as in the finite element method, in time as in optimal control, across a network, as in Kron's original method, between states in a decomposed system, or levels in a hierarchical system, or between spatial dimensions, as in Kron's space filter. In the same way that the homological structure of Kron's network theory gives rise to the optimal solution, it is the homology of the torn network which forces the exact solution in Diakoptics. Kron's method of tearing as applied to networks in fact has one particular peculiarity in that as well as producing a number of subnetworks a further network called the intersection network, consisting of a collection of single unconnected network elements is left. This is due to the fact that to separate the subnetworks, rather than just cutting branches Kron actually removes RLC elements from the network. No other decomposition method uses this technique and it is in fact not really necessary for network tearing.

Various algorithms are used as the basis of tearing in different problems. Kron's Diakoptics is simplistically based on Householder's formula

$(A+BCD)=/A-/AB/(/C+D/AB)D/A$ or in more conventional notation

$$(A+BCD)^{-1}=A^{-1}-A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1},$$

whereas in the orthogonal case use is made of Schur's lemma

$$/\begin{bmatrix} X_a & X_b \\ X_c & X_d \end{bmatrix}=\begin{bmatrix} /(X_a-X_b/X_dX_c) & -/X_aX_b/(X_d-X_c/X_aX_b) \\ -/X_dX_c/(X_a-X_b/X_dX_c) & /(X_d-X_c/X_aX_b) \end{bmatrix}$$

We will continue to use the slash notation to emphasise the ease of programming matrix algorithms. There is no ambiguity in its use as the / operator is monadic and therefore takes a higher priority than a diadic operator. Confusion may however arise when differential operators are involved. We take /(zP)=z/P. Chapter IX describes extensions to the language Algol68 in which matrix equations in the above form can be programmed directly.

The derivation of the discrete Riccati equation for multistage optimal control is based on the partitioned matrix version of the recent (196-) Thomas algorithm for the inversion of tridiagonal matrices.

For
$$\begin{aligned} B_1x_1+C_1x_2&=d_1\\ A_2x_1+B_2x_2+C_2x_3&=d_2\\ A_3x_2+B_3x_3+C_3X_4&=d_3\\ &\cdots\\ A_ix_{i-1}+B_ix_i+C_ix_{i+1}&=d_i\\ &\cdots\\ A_nx_{n-1}+B_nx_n&=d_n \end{aligned}$$

we can write the new system $A^2_{i-1}x_{i-1}+C_{i-1}x_i=b^2_{i-1}$

and hence the recursive algorithm

(1) $A^2_i=B_i-A_iA^2_{i-1}{}^{-1}C_{i-1}$, $i=2,\ldots,n$,

(2) $b^2_i=d_i-A_iA^2_{i-1}{}^{-1}b^2_{i-1}$, $i=2,\ldots,n$,

(3) $x_n=A^2_n{}^{-1}b^2_n$,

(4) $x_i=A^2_i{}^{-1}(b^2_i-C_ix_{i+1})$, $i=n-1,\ldots,1$.

This is also the basis of sequential least squares estimation and can be shown to be a special case of the recursive 'staircase method' of inversion for general block matrices involving successive application of Schur's lemma to a square submatrix partitioned such that the upper left and lower right hand blocks are square.

2.2 Ordinary least squares

This is the basis of all linear quadratic methodology. Paraphrasing Nicholson[4]: 'A least squares estimate for the parameter vector x in the measurement equation

$$y=Ax+v$$

with known mean and covariance properties

$$E(v)=0,\ E(vv')=R,\ E(xx')=S,\ E(vx')=0,$$

is obtained by minimising the function

$$J(x)=(y-Ax)'/R(y-Ax)+x'/Sx.'$$

We will now demonstrate how Roth's diagram may be used in general least squares problems by rewriting the optimisation in standard form using partitioned matrices

$$\min \begin{bmatrix} v \\ x \end{bmatrix}' \begin{bmatrix} /R & \\ & /Q \end{bmatrix} \begin{bmatrix} v \\ x \end{bmatrix}$$

subject to

$$\begin{bmatrix} I & A \end{bmatrix} \begin{bmatrix} v \\ x \end{bmatrix} = y$$

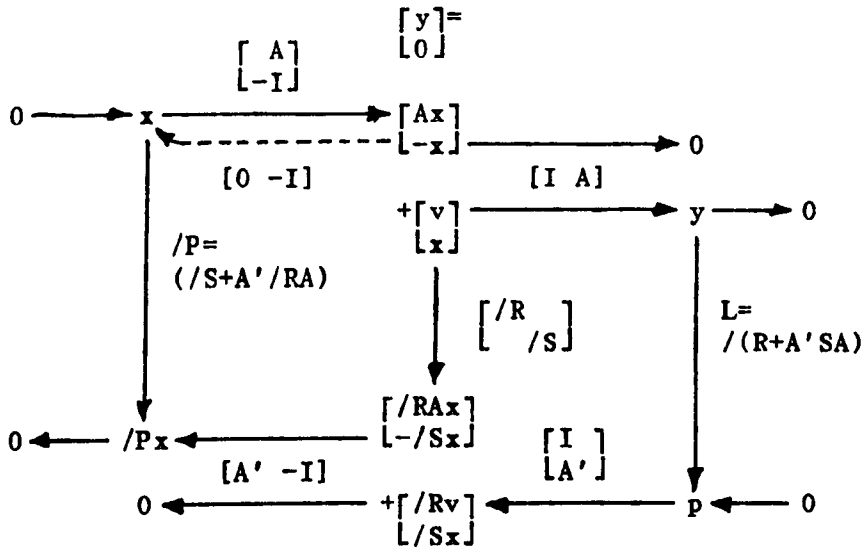
and putting

$$\begin{bmatrix} y \\ 0 \end{bmatrix} = \begin{bmatrix} Ax \\ -x \end{bmatrix} + \begin{bmatrix} v \\ x \end{bmatrix}.$$

The partitioned vector in the cost function must appear explicitly in the constraints. It is then placed top centre in Roth's diagram[2] and is mapped to its adjoint by the weighting matrix of the cost function. The partitioned coefficient matrix in the constant equation appears top right, the other three

positions are held by its adjoint (transpose) and annihilators so that the horizontal rows are exact and the diagram commutes. The solution can then be read directly off the diagram remembering that isomorphisms may be reversed by inverting the associated matrix.

By analogy with the electric circuit problem Roth's diagram takes the following form

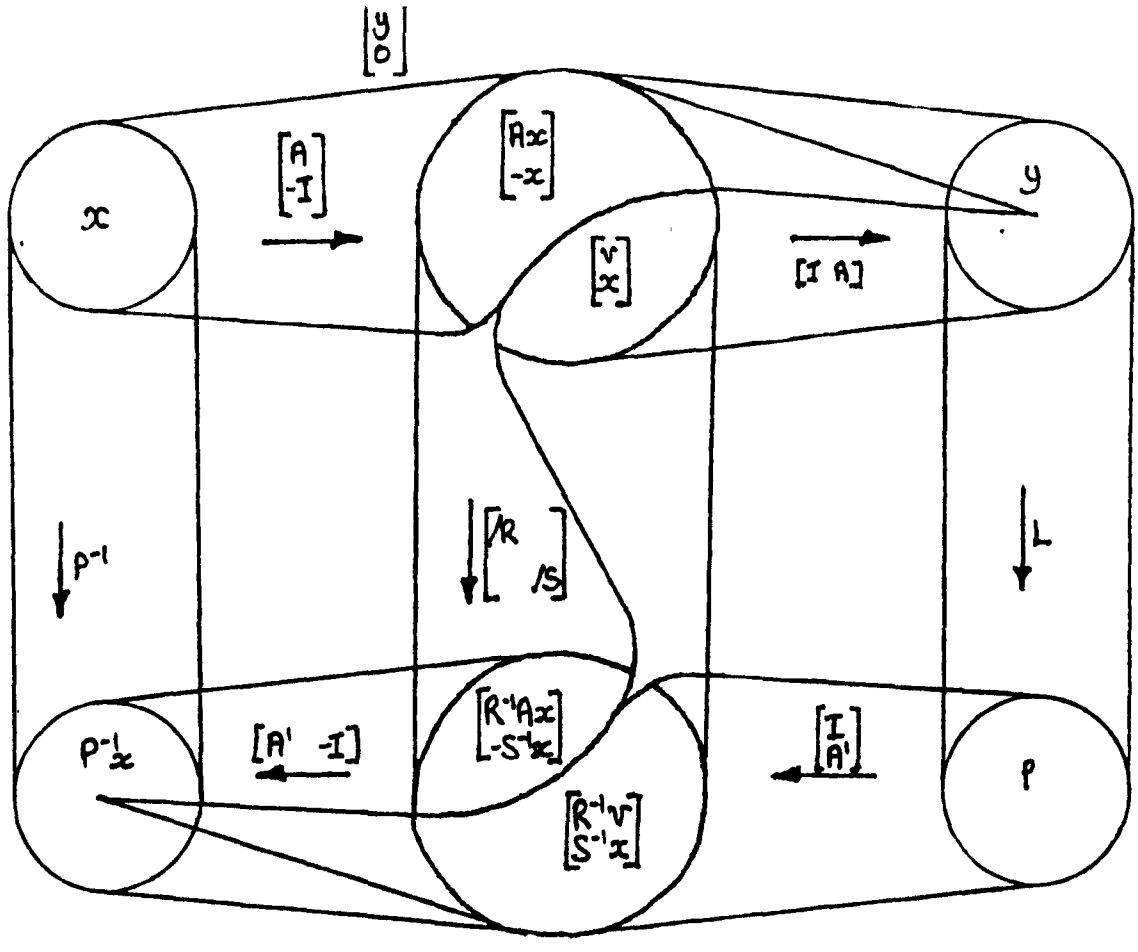


and we can find the optimal estimate x^0 from the left hand square

$$x^0=P[A' \ -I]\begin{bmatrix}/R\\ /S\end{bmatrix}\begin{bmatrix}y\\ x\end{bmatrix}=PA'/Ry$$

Any inverse of $\begin{bmatrix}A\\ -I\end{bmatrix}$ will give x^0 from $\begin{bmatrix}Ax\\ -x\end{bmatrix}$, eg $[0 \ -I]$.

This diagram requires some explanation. It is an extension of the idea of a commutative diagram in pure mathematics. The latter is a diagram of objects and homomorphisms such that following the arrows round from A to B gives the same result whichever path is taken, in the sense that the composition of the functions labelling the arrows is the same. Only if an arrow is labelled with an isomorphism may it be followed backwards and the inverse function taken. Our version of Roth's diagram above, is shown in a more pictorial form in Fig(2.1). The object



Fig(2·1) Ordinary least squares

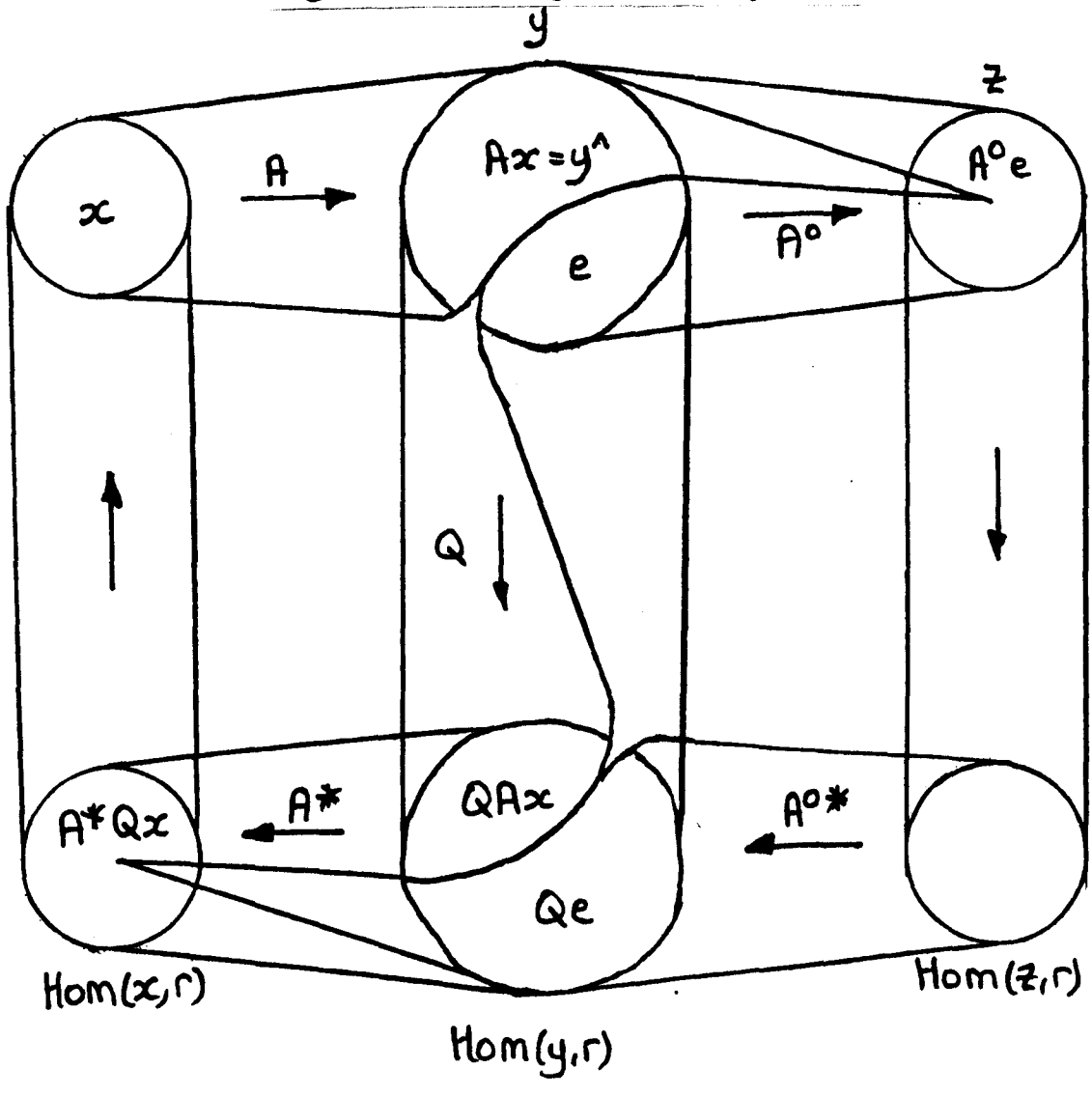
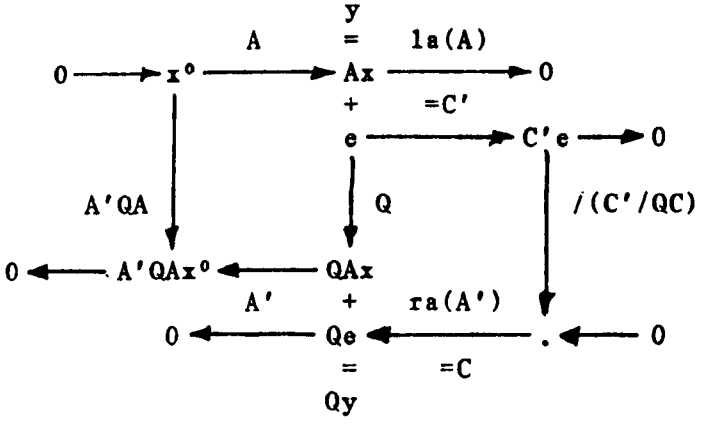


Fig (2·2) Roth type Venn diagram

$$\left\{\begin{bmatrix} y \\ 0 \end{bmatrix}\right\} = \left\{\begin{bmatrix} Ax \\ -x \end{bmatrix}\right\} \oplus \left\{\begin{bmatrix} v \\ x \end{bmatrix}\right\} = \mathrm{im}\begin{bmatrix} A \\ -I \end{bmatrix} \oplus \ker[I \; A]$$

considered as the group of all values of [y' 0']' and is split into the direct sum of its subgroups such that the substructure of the problem can be clearly seen, based on the technique used by Branin. The category of the objects in the diagram depends on the problem. In ordinary least squares we deal with vector spaces, in network theory with the singular homology groups, in control theory with K[s] or K[z] modules and in field theory with differential forms. Commutative diagrams may be considered as dual, in a sense, to signal flow diagrams as used in control theory, in that mappings in the former compose to the identity round a closed loop whereas objects in the latter sum to zero at a summing point.

Roth's diagram in general gives two forms of solution. For instance given the overspecified equation set y=Ax+e where A is nxm and n>m we may wish to minimise the weighted sum of squares of the error e=y−Ax with covariance matrix /Q with la = left annihilator and ra = right annihilator (see Fig(2.2))



min(y−Ax)'Q(y−Ax) subject to y=Ax+e

gives from the diagram $x^{o}=/(A'QA)A'Qy$ (round the left hand side)

and $y^0=Ax^0=A/(A'QA)A'Qy$

therefore $e=y-y^0=(I-A/(A'QA)A'Q)y$ ... $(A'Qe=0)$

$=/QC/(C'/QC)C'y$ (round the right hand side)

which must be true for all values of y so we can state

Theorem (Kron's lemma) If A and C are rectangular nonzero matrices such that $\ker(A')\geq\mathrm{im}(C)$ and Q is nonsingular and of commutative order then

$A/(\underline{A'QA})\underline{A'Q}+\underline{/QC}/(C'\underline{/QC})C'=I$

where the underscore just groups the weighted coboundary elements together. This virtually unknown formula completes the set available for naive matrix manipulation and we will call it Kron's lemma - as he seems to have had little else named after him. It is the analog of the definition of the Laplacian in electromagnetic theory, and finds important application in Bowden's algorithm for fitting data to models[10].

Finally Kron emphasised in his later papers, particularly 'Four Abstract Reference Frames of an Electrical Network' that the mathematical structure of an electrical network is that of the 'fibre bundle'. According to Kron 'The most frequently cited example of a complex is the so-called polyhedron, in which each p-network forms the boundaries of a (p+1)-network ... A special case of the polyhedron is the graph. The topological complex associated with a conventional electric network is however quite a different structure. The p-networks happen to be coincident instead of bounding each other. This rather unusual type of complex is called a fiber bundle.' We will expand somewhat on this by demonstrating the connection

between fibre bundles and ordinary least squares. A fibre bundle is a pentuple (B,X,Y,p,g) where B, X and Y are topological spaces, B is the bundle space, X is the base space, Y is the fibre and g is the bundle group, a group of homeomorphisms on the Y. p is a continuous map p:B->X. The space $Y_x=p^{-1}(x)$ is called the fibre over x and each Y_x is homeomorphic to Y. In matrix notation the projection may be written $\underline{P}\underline{b}=\underline{x}$ where the fibre over $\underline{x}$ is $\underline{P}^{-1}\underline{x}$, the coimage of $\underline{P}$. A cross section of B is a continuous map f:X->B such that pf(x)=x. This is equivalent to taking a pseudoinverse of $\underline{P}$, $\underline{P}^{+}:\underline{x}->\underline{b}$ such that $\underline{P}\underline{P}^{+}\underline{x}=\underline{x}$ where $\underline{P}^{+}=/(\underline{P}'\underline{Q}\underline{P})\underline{P}'\underline{Q}$ for any $\underline{Q}$. The mathematics of fibre bundle theory is well developed and closely interwoven with homology theory. There is ample scope for future research here.

2.3 The Homology Theorem

The notation y=Ax+e will be used in this section to preserve continuity with the least squares theory. Let C be a 'sufficient' category (as will be explained below) then if x,y<C and

$$x \xrightarrow{A} y$$

is such that y>Ax then the monomorphism

$$Ax \xrightarrow{id} y$$

is called the (natural) inclusion map or injection, the epimorphism

$$y \xrightarrow{p} y\backslash Ax$$

which sends each element of y into its coset is called the projection or surjection, hence the sequence

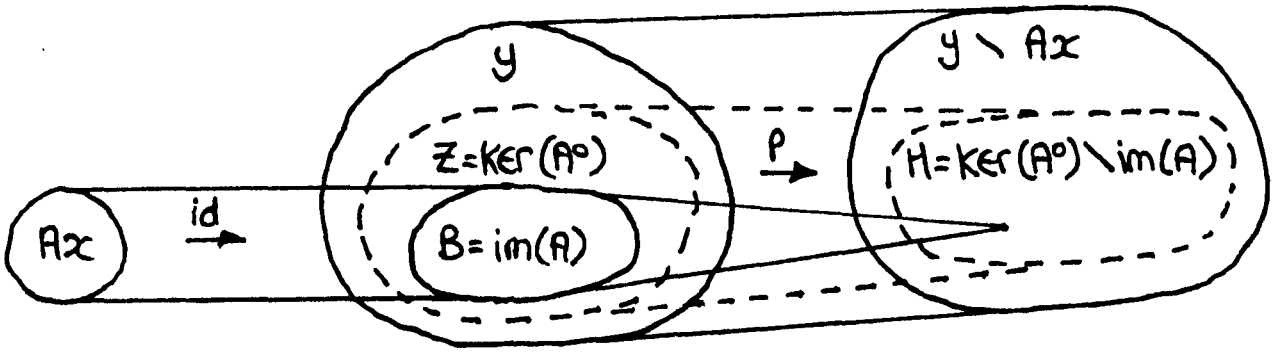
$$0 \xrightarrow{id} Ax \xrightarrow{p} y \xrightarrow{} y\backslash Ax \xrightarrow{} 0$$

is short exact. The 'zero' of y\Ax is Ax. This is shown more clearly in the Venn type diagram, in Fig(2.3), (ignore the dashed lines for now).

For example if A=[1 1]' we have [1 1]'R->R^2->R^2\[1 1]'R. R^2\[1 1]'R is the set of straight lines in the 'xy plane' parallel to x=y, [1 1]'R is simply the set of points {x,x} (common notation in geometric control theory). Obviously the line x=y is the 'zero' in the quotient set as 'addition of lines' {x,y:x=y+c}+{x,y:x=y} = {x,y:x=y+c} shows.

We wish to choose x given y to minimise some cost function r(x,y). We can write x^o={x:y=Ax, dr(x,Ax)/dx=0}=f(y). Therefore $y^o=Ax^o=Af(y)$ which is a contraction mapping. If Af obeys certain continuity properties then there is a fixed point of the function at which $y=y^o$. For example using the Kakutani fixed point theorem (proved using homology theory!) Af must be upper semicontinuous.

Now consider the homology group H=ker(A^o)\im(A)=Z(A^o)\B(A), where A^o is any left annihilatorof A. We have the sequence Ax=B(A)<Z(A^o)<y. In the diagram below all rows and columns are short exact. We shall see a similar diagram for electric circuit theory in which H^1=0, and the dual cohomology group H_1=Z(A')\B($A^{o'}$).



Fig(2·3) Venn diagrams for the short exact sequence

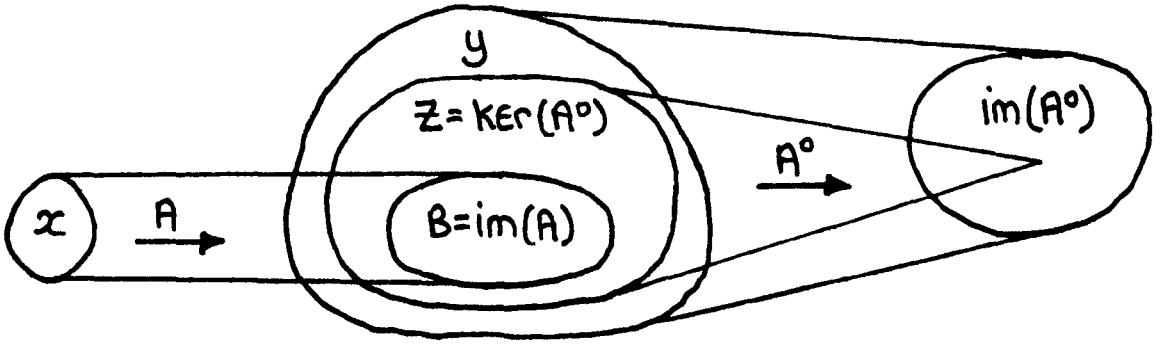
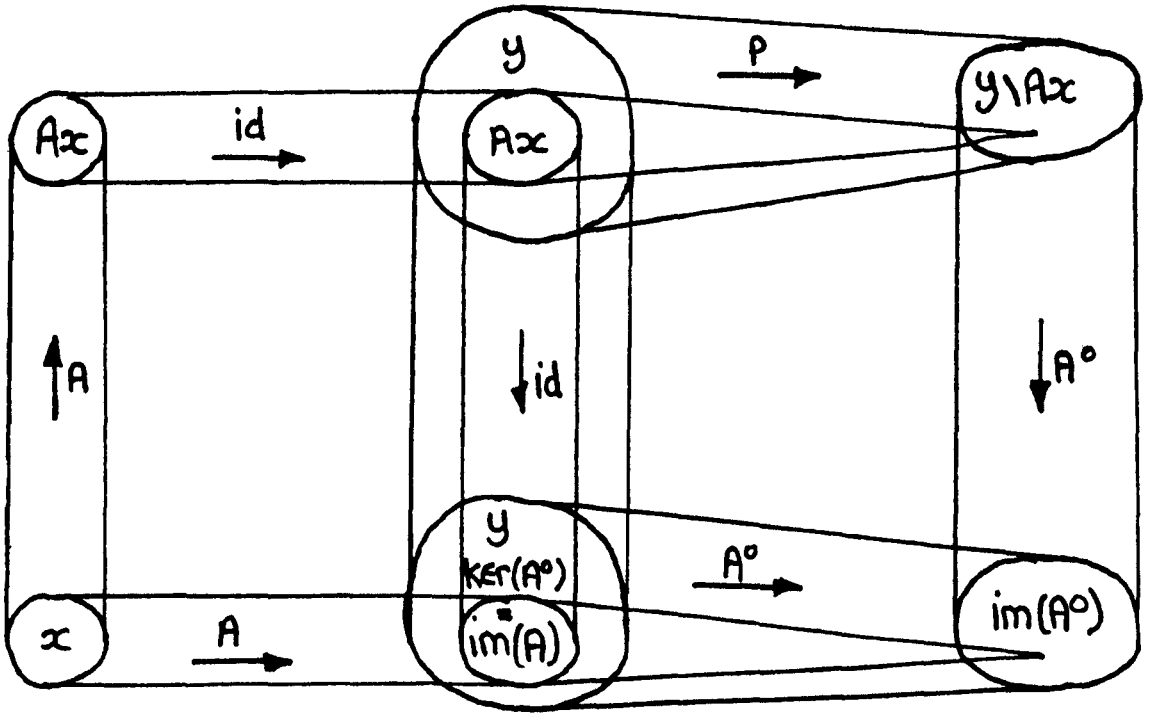
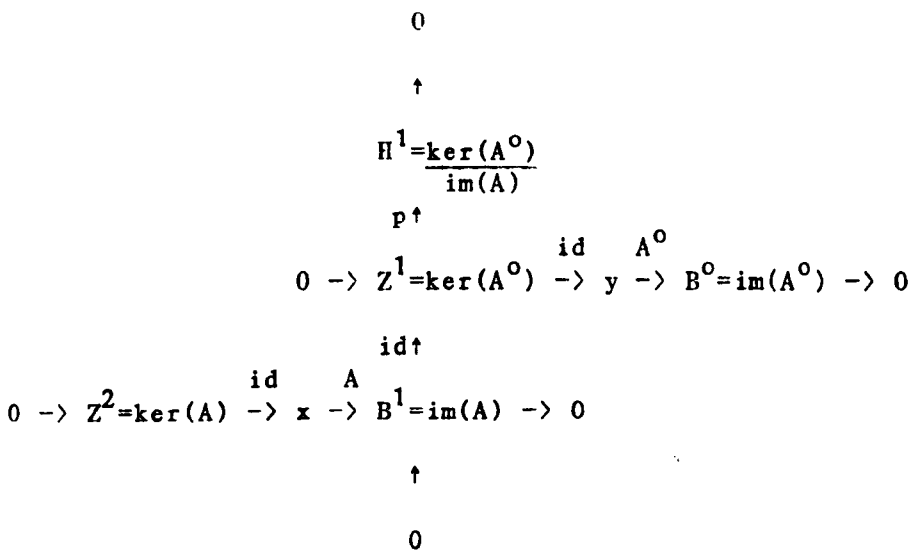


Fig (2·4) Venn diagrams for the chain complex



Fig(2·5) Construction of A°



The short exact sequence $0\to B^1\to Z^1\to H^1\to 0$ is contained in the sequence $0\to Ax\to y\to y\backslash Ax\to 0$, see Fig(2.3) again. In general the sequence

$$0\to x\xrightarrow{A} y\xrightarrow{A^o} \mathrm{im}(A^o)\to 0$$

is not short exact (see Fig(2.4)). For example if $A=[1\ 1\ 0]'$ and $A^o=[1\ -1\ 1]$ then $A^o[0\ 1\ 1]'=0$ but $[0\ 1\ 1]'\neq Ax$ for any value of x. However A^o can always be chosen such that the sequence is isomorphic to

$$0\to Ax\xrightarrow{id} y\xrightarrow{p} y\backslash Ax\to 0$$

hence $\ker(A^o)=\mathrm{im}(A)$ and $H=0$ (see Fig(2.5)). For instance in the matrix case if $x\in R^m$, $y\in R^n$, $n<m$ then $y\backslash Ax\in R^{n-m}=R^n\backslash R^m$ and A^o must be a $(n-m)$ x m matrix, $A^o\in \mathrm{Hom}(R^n,R^{n-m})$, the set of all homomorphisms from R^n to R^{n-m}, a generalisation of the idea of dual space. For a more detailed description of Hom, see the beginning of Ch VIII.

Next consider the contravariant mappings shown by dashed arrows in Figs (2.6). For the diagrams to commute define the adjoints $A^*:f\to fA$ and $A^{o*}:g\to gA^o$, with $fA\in \mathrm{Hom}(x,r(x,y))$, $f\in \mathrm{Hom}(y,r(x,y))$, $gA^o\in \mathrm{Hom}(y,r(x,y))$ and $g\in \mathrm{Hom}(A^o y,r(x,y))$ where $r(x,y)\in R$, some coefficient group. The * operator will be found to transpose

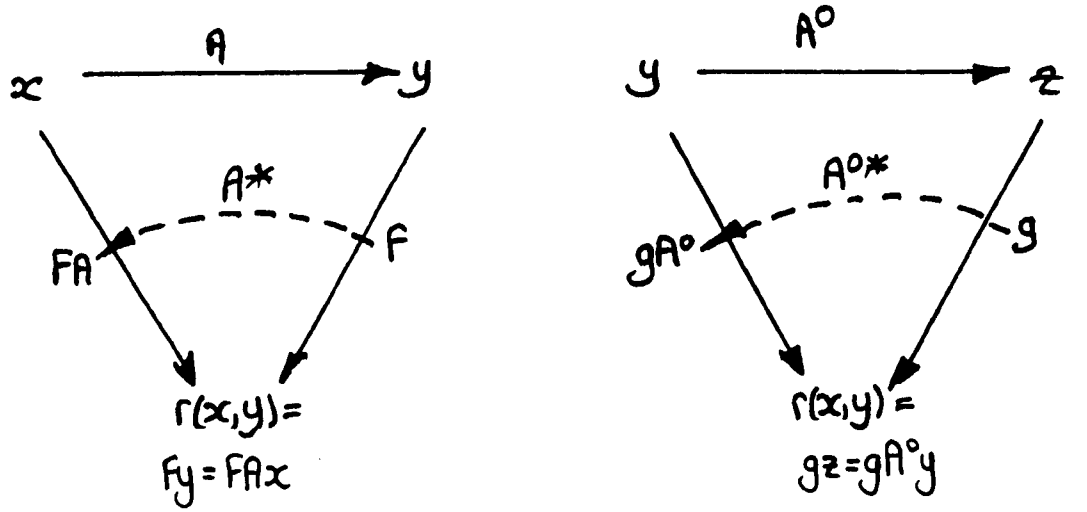
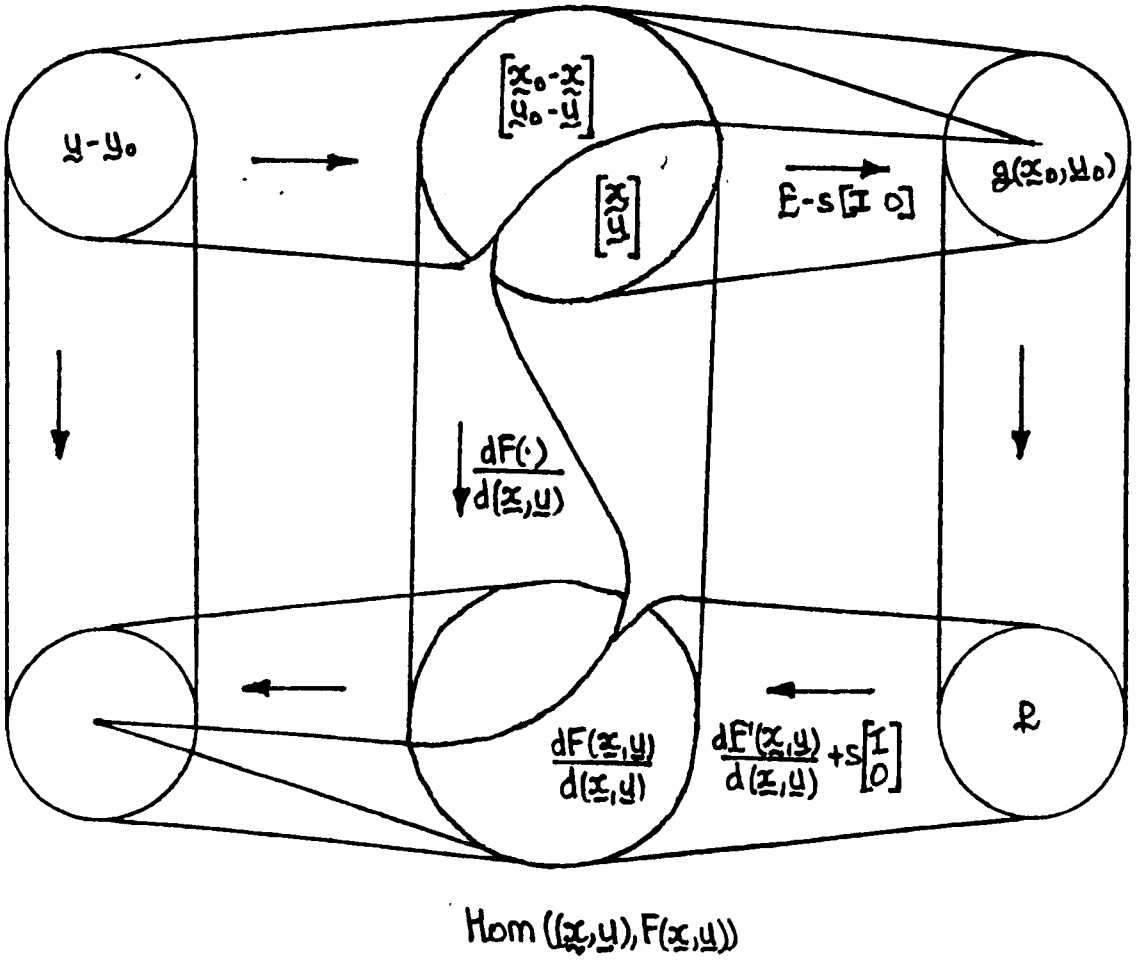


Fig (2·6) Construction of the adjoints

Fig (2·7) The Euler-Lagrange equation



matrices, conjugate complex numbers, negate the Laplace operator and invert the lag operator of time series analysis. It also relates curl, grad and div and higher order operators of partial differential equations.

The following diagram is short exact (by symmetry)

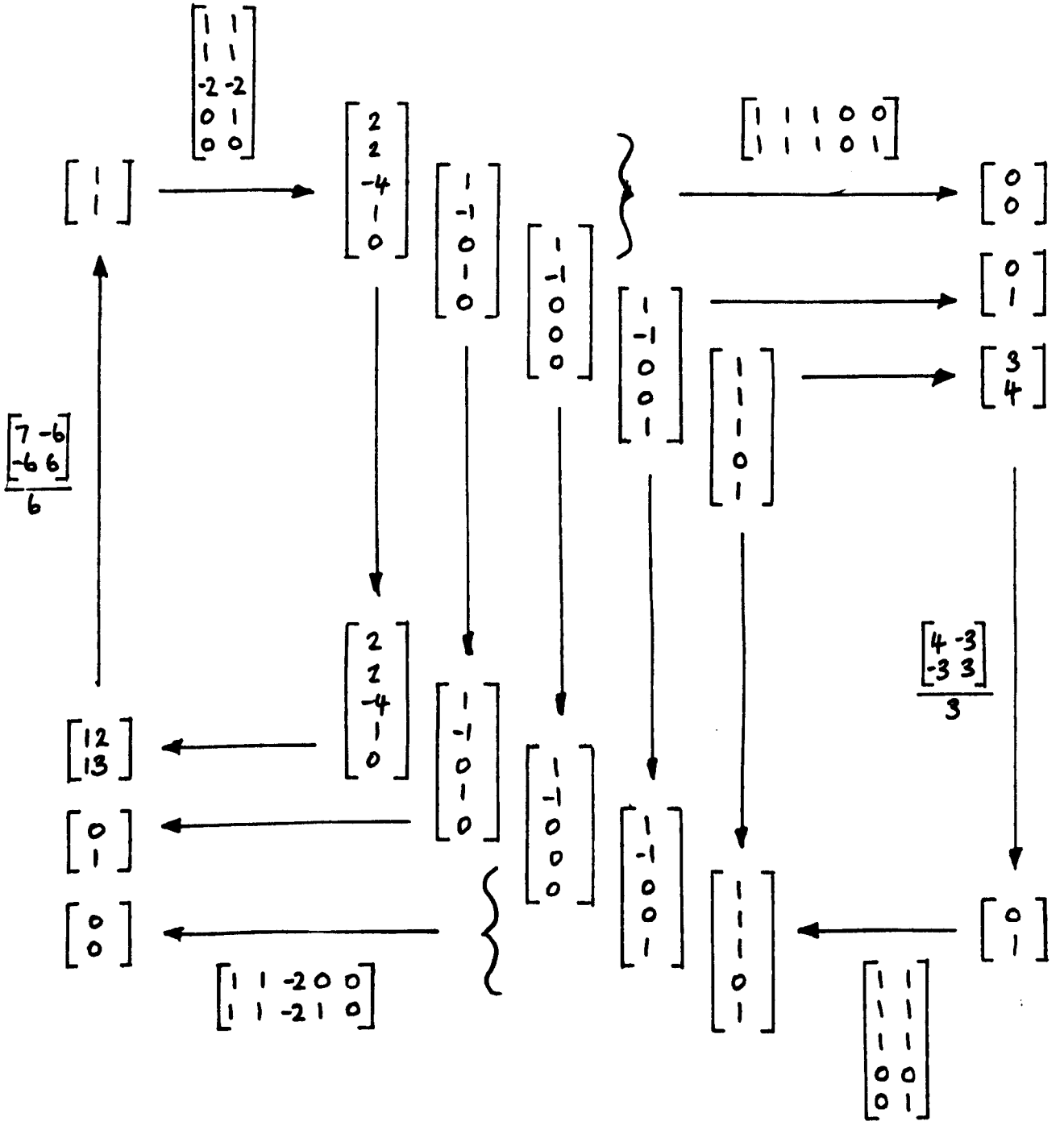
$$0 \leftarrow -\mathrm{Hom}(x, r(x,y)) \overset{A^*}{\leftarrow} -\mathrm{Hom}(y, r(x,y)) \overset{A^{o}*}{\leftarrow} -\mathrm{Hom}(A^{o}y, r(x,y)) \leftarrow -0$$

and is related to the original sequence by three isomorphisms depending on the structure of the coefficient group R. We choose these isomorphisms such that the diagram commutes. Following the Venn diagram, Fig(2.2) clockwise round the left hand side, we can see that there is a contraction mapping from y back into itself. Further convergence is obtained in one iteration. x^o can thus be found from y. We have again borrowed the notation from least squares theory. Note that though $x \# x^o$, normally $\{x\}=\{x^o\}$. This may be confusing as we are using the notation x as an abbreviation for $\{x\}$. Strictly we should also consider the case where $\{x^o\} < \{x\}$, though this follows naturally. Another form of solution may be found by following the mappings clockwise around the right hand side of the diagram, giving e^o from y. Note that the diagram is rotationally symmetrical, this is precise as can be seen by the duality in the theory. Fig(2.8) shows the five types of chain that occur when the sequences are not exact. This is important in network theory.

So for at least some categories the following is true.

<u>The Homology Theorem</u> For a given sufficient category C and spaces E,F,G in C and a suitable monomorphism, f (or epimorphism g) there exists an epimorphism g (monomorphism f) such that the following diagram is exact

$$0 \rightarrow E \overset{f}{\rightarrow} F \overset{g}{\rightarrow} G \rightarrow 0.$$



Fig(2·8) Examples of the five basic types of 'chain' vector in ordinary least squares (Roth's diagram)

There exists a short exact sequence relative to a scalar field R with the arrows reversed

$$0 \leftarrow \mathrm{Hom}(E,R) \xleftarrow{f^*} \mathrm{Hom}(F,R) \xleftarrow{g^*} \mathrm{Hom}(G,R) \leftarrow 0$$

where Hom(E,R) is the contravariant set of all homomorphisms E->R and f* is the adjoint function *:f->f*. Further there exist natural isomorphisms

$$\begin{array}{ccccccccc} 0 & \rightarrow & E & \xrightarrow{f} & F & \xrightarrow{g} & G & \rightarrow & 0 \\ & & \simeq & & \simeq & & \simeq & & \\ 0 & \leftarrow & \mathrm{Hom}(E,R) & \xleftarrow{f^*} & \mathrm{Hom}(F,R) & \xleftarrow{g^*} & \mathrm{Hom}(G,R) & \leftarrow & 0 \end{array}$$

then the solution of the optimisation problem min r(x) with respect to e<E with r<R, x<F is given if and only if the diagram commutes. That is the diagram commutes for all optimal E^0, F^0, G^0, f^0, g^0, etc with respect to some r.

Commutation of Roth's diagram is equivalent to optimality. This appears to be true at least for continuous problems, eg it may not hold in linear programming due to the inequality constraints, despite the inherent duality in the problem. Nevertheless the Kuhn-Tucker conditions may well have their analogs within homology theory.

Hypothesis C is a sufficient category for the Homology Theorem to hold if and only if it is an admissable category for homology theory (see Chapter VIII), and the axioms for homology theory hold. A system of the form described above will be termed a homological system.

2.4 The nonlinear case

Consider the scalar case given $y=f(x)+e$. We want

$$\min_x \|e\| \text{ or } \frac{d}{dx}(y-f(x))^2=0$$

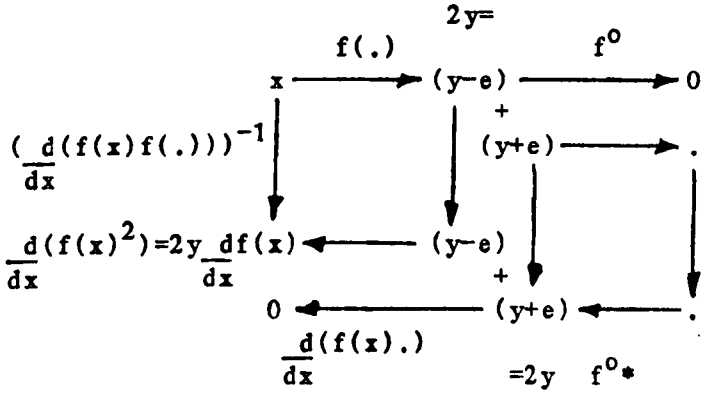
and dropping the constant $dy^2/dx=0$,

$$d(f(x)^2-2yf(x))/dx=0$$

or $d((e-y)(e+y))/dx=d(e^2-y^2)/dx=0$

giving $d(f(x)^2)/dx=2ydf(x)/dx$.

Alternatively we can construct Roth's diagram as shown below.



The general term in $\text{Hom}(x,f(x)^2)$ is $df(x)^2/dx$. Reading around the lhs we have

$$x^o=(d(f(x)f(.))/dx)^{-1}2(df(x)/dx)y.$$

In fact the scalar case, though an interesting example, can be seen to have little practical value as it is about equivalent to setting $df(x)/dx=0$. However it does lead us on to the time varying case.

The general time varying problem[9] can be written in state space form

$$\min_{\underline{u}} \int F(\underline{x},\underline{u})dt \text{ subject to } \underline{g}(\underline{x},\underline{u})=s\underline{x}-\underline{f}(\underline{x},\underline{u})=\underline{0},$$

with one of either $\underline{x}(0)=\underline{x}_0$ or $\underline{u}(0)=\underline{u}_0$ usually set to zero. The solution is given by the Euler-Lagrange equation from the calculus of variations

$$\left(\frac{\partial}{\partial(\underline{x},\underline{u})} - s\frac{\partial}{\partial(s\underline{x},s\underline{u})}\right)L(\underline{x},\underline{u},\underline{p})=0,$$

where the Lagrangian $L(\underline{x},\underline{u},\underline{p})=F(\underline{x},\underline{u})+\underline{p}'\underline{g}(\underline{x},\underline{u})=\underline{p}'s\underline{x}-H(\underline{x},\underline{u},\underline{p})$,

where $\underline{p}'$ is called the Lagrange multiplier or costate variable and the Hamiltonian $H(\underline{x},\underline{u},\underline{p})$ is the difference between kinetic and potential energy in the mechanical case. These are the equations of a conservative holonomic (integrable) system. Flanders[8] treats them in terms of differential forms.

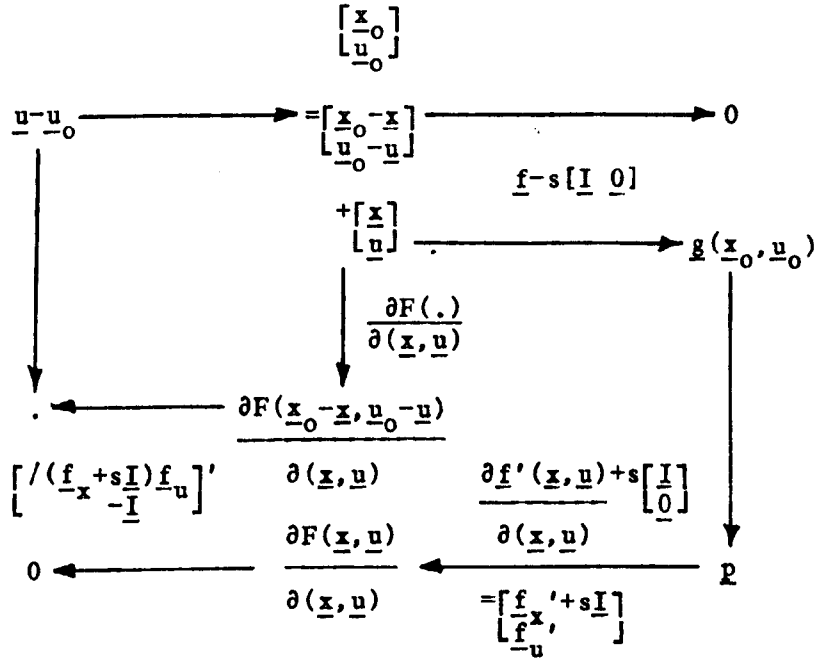
Performing the (partial) differentiations as in the following table

	$F(\underline{x},\underline{u})$	$\underline{p}'s\underline{x}$	$-\underline{p}'\underline{f}(\underline{x},\underline{u})$
$\frac{\partial}{\partial(\underline{x},\underline{u})}$	$\frac{\partial F(\underline{x},\underline{u})}{\partial(\underline{x},\underline{u})}$	$\underline{0}$	$-\underline{p}'\frac{\partial \underline{f}(\underline{x},\underline{u})}{\partial(\underline{x},\underline{u})}$
$-s\frac{\partial}{\partial(s\underline{x},s\underline{u})}$	$\underline{0}$	$(-s\underline{p}',\underline{0})$	$\underline{0}$

the solution is $\frac{\partial F(\underline{x},\underline{u})}{\partial(\underline{x},\underline{u})}-s\underline{p}'(\underline{I},\underline{0})-\underline{p}'\frac{\partial \underline{f}(\underline{x},\underline{u})}{\partial(\underline{x},\underline{u})}=\underline{0}$.

Note that all of this is achieved without reference to $(\underline{x},\underline{u})_o$.

The alternative approach is to construct Roth's diagram (see Fig(2.7)). Writing $\underline{f}_x=\partial\underline{f}(\underline{x},\underline{u})/\partial\underline{x}$ we can split $(\underline{x},\underline{u})_o$ into two parts $(\underline{x},\underline{u})+(\underline{x}_o-\underline{x},\underline{u}_o-\underline{u})$ and construct the top short exact sequence below.



Following Chang we can drop the integral in the s-plane (see Chapter VI for a discussion of this) and take $F(\underline{x},\underline{u})$ as the coefficient group. The general term in $\mathrm{Hom}((\underline{x},\underline{u}),F(\underline{x},\underline{u}))$ is $\partial F(\underline{x},\underline{u})/\partial(\underline{x},\underline{u})$ hence the central isomorphism is $\partial F(.)/\partial(\underline{x},\underline{u})$ giving the lower sequence. The Lagrange multiplier appears at the bottom right and the initial conditions at the top right. The Euler-Lagrange equation is implicit in the mapping at the bottom right. The lack of symmetry between the upper and lower short exact sequences $(*:\underline{f}\to\partial\underline{f}'(\underline{x},\underline{u})/\partial(\underline{x},\underline{u}))$ can be removed by considering small changes in the diagram then

$$\partial\underline{f}(\underline{x},\underline{u})=\frac{\partial\underline{f}(\underline{x},\underline{u})}{\partial\underline{x}}\partial\underline{x}+\frac{\partial\underline{f}(\underline{x},\underline{u})}{\partial\underline{u}}\partial\underline{u}$$

and *:$\partial \underline{f}(\underline{x},\underline{u})/\partial(\underline{x},\underline{u}) \to \partial \underline{f}'(\underline{x},\underline{u})/\partial(\underline{x},\underline{u})$ as expected.

Finally we have the Hamiltonian

$$H(\underline{x},\underline{u},\underline{p})=\underline{p}'\underline{f}(\underline{x},\underline{u})-F(\underline{x},\underline{u})=\underline{p}'s\underline{x}-F(\underline{x},\underline{u}),$$

therefore $s\underline{x}=\partial H(\underline{x},\underline{u},\underline{p})/\partial \underline{p}$ ... (1).

Further $(\partial/\partial \underline{x}-s\partial/(s\underline{x}))(\underline{p}'s\underline{x}-H(\underline{x},\underline{u},\underline{p}))=0$

therefore $s\underline{p}=-\partial H(\underline{x},\underline{u},\underline{p})/\partial \underline{x}$ ... (2).

(1) and (2) are the equations of Pontryagin's maximum principle. Other equations can be derived by partially differentiating F, L or H with respect to $\underline{x}$, $\underline{u}$, $\underline{p}$ or t. For example expanding

$$\frac{dH(\underline{x},\underline{u},\underline{p})}{dt} = \frac{\partial H(\underline{x},\underline{u},\underline{p})}{\partial \underline{p}} \frac{d\underline{p}}{dt} + \frac{\partial H(\underline{x},\underline{u},\underline{p})}{\partial \underline{x}} \frac{d\underline{x}}{dt}$$

$$=\dot{\underline{x}}'\dot{\underline{p}}-\dot{\underline{p}}'\dot{\underline{x}}=0,$$

which will be found to be related to the Walras law in an economic context.

The foregoing analysis can easily be extended to the discrete case following Ref(5). The gradient matrices may be calculated using the techniques in Ref(1).

2.5 References

(1) Bernussou and Geromel, An Easy Way to Find Gradient Matrix of Composite Matricial Functions, IEEE Trans AC-26, No 2, April 1981.

(2) Roth, An Application of Algebraic Topology: Kron's Method of Tearing, Quarterly of Applied Maths, April 1959, No 1.

(3) Kondo and Iri, On the Theory of Trees, Cotrees, Multitrees and Multicotrees, RAAG Memoirs, Vol 2, 1958.

(4) Nicholson, Structure of Interconnected Systems, IEE Control Engineering Monograph 5, Peter Perigrinus, 1978.

(5) Lahudde, Discrete Hamiltonian Mechanics, I J General Systems, Vol 6, No 1, 1980.

(6) Chern, Some new viewpoints in Differential Geometry in the Large, Bull Amer Math Soc 52, 1946, pp 1-30.

(7) Hermann and Martin, A new application of Algebraic Geometry to Systems Theory, IEEE Conf Decision and Control, Dec 1976.

(8) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(9) Athens and Falb, Optimal Control, McGraw-Hill, 1966.

(10) Bowden, Fitting Data to Models, to be published in Matrix and Tensor Quarterly.

(11) Ben-Israel and Charnes, Generalised Inverses and the Bott-Duffin Network Analysis, Journal of Mathematical Analysis and Applications, Vol 7, pp 428-35, 1963.

(12) Iri et al, Algebraic and Topological Theory and Methods of Linear Programming with Weak Graphical Representation, RAAG Memoirs, Vol 4, 1968.

CHAPTER III Singular Homology Theory: Network Theory

We review the application of the topological structure of chain complexes to the analysis of the electrical network as an optimal system in the light of the results of the last chapter. Kron's orthogonal network and the associated algebraic diagram are introduced. We briefly mention scattering theory and network tearing[3-16].

3.1 Topology

Point set topology is the study of spaces in the small, that is the way in which adjacent points of a space are connected. It is concerned with topological spaces, with real fields, rational fields, integer fields, compact spaces, Hausdorff spaces, Banach spaces, Hilbert spaces. It does not matter much what the dimension of the space is.

Algebraic topology is the study of spaces in the large. It is concerned with objects and their relationships in n-dimensional space. It does not matter so much what the point set topology of the space is. It is the study of the effects of dimensionality upon structure.

One of the implications of homology theory is that the algebraic topological structure of different n-dimensional spaces are isomorphic regardless of the differential structure of the underlying manifold[17]. Imagine for instance two interlinked toroids in Euclidean 3-space. These can be approximated by covering the surfaces with small plane faces or facets. The faces can be represented by a number of points in our space, with lines joining appropriate points being

boundaries of the faces. That these simplicial approximations are homotopic with our original toroids, that is that they can be continuously deformed until they are identical, and that the topological properties of the linkage has not changed due to the approximation seems obvious - though is non trivial mathematically especially in the general case. That the homological structure is independent of the triangulation used can also be seen. If we now replace our real n-space with an integer valued discrete n-space, it can be seen that if the scale is made small enough with respect to the size of the toroids then the points can be approximated by integer valued n-tuples. Again the simplicial approximations generated by these points are homotopic to the original toroids and their homological structure remains unchanged. It can be seen that the point set topology is irrelevant.

Studying the properties of simplicial approximations to general objects in n-space is tantamount to studying the properties of n-space itself. Other isomorphisms are that between the homological structure of a space and the cohomological structure of its dual: and between a space and superimposed functions (de Rham's theorem).

3.2 Chain complexes

The notation in this chapter conforms with that of the network theorists, Kondo, Iri, Roth etc. Homology groups are shown superscripted and cohomology groups subscripted. An abstract cell E_r is an entity bearing an integer dimension, degree or grade, r. A cell complex is a collection of cells together with

(1) an incidence relation on pairs of cells $A_r \langle B_{r-1}$, relating eg. a surface to its edges,

(2) a set of integer valued incidence numbers $C_{ab}(r)$ on A_r and B_{r-1} such that $C_{ab}(r) \# 0 \Rightarrow B_{r-1} \langle A_r$ and

(3) $\sum_b C_{ab}(r) C_{bc}(r-1) = 0$ where it is assumed that for all but a finite number of b values either $C_{ab}(r)$ or $C_{bc}(r-1)$ is 0.

If $B_{r-1} \langle A_r$ then B_{r-1} is a face of A_r but this does not guarantee $C_{ab}(r) \# 0$.

An (orientated) r-simplex S_r is an ordered set of r+1 points or vertices $(P_0, P_1, \ldots, P_r)$. A <u>simplicial complex</u> is a cell complex where the r-cells are r-simplexes. The boundary dS of a simplex S is a formal sum of simplices of one lower dimension with integer coefficients

$$d(P_0, \ldots, P_n) = \sum_{i=0}^{n} (-1)^i (P_0, \ldots, P_i', \ldots, P_n)$$

where the prime means that P_i is omitted. In this case (3) is equivalent to the statement that the boundary of a boundary is zero ie $dd(P_0, \ldots, P_r)$ gives two terms in $(P_0, \ldots, P_i', \ldots, P_j', \ldots, P_r)$ of opposite signs so that everything cancels. The 0-vertices or faces determine the complex uniquely.

A <u>chain complex</u> is a cell complex where the r-cells are r-chains ie formal sums $C = \sum A_i S_i$ and the A_i may be real constants or elements of K-modules, etc. The boundary of a chain is defined by $dC = \sum A_i dS_i$ and by linearity $ddC = 0$. Dual cell and cochain complexes can be defined similarly. This is the basis of Kron's network theory. The A_i are the 'live' voltages or currents superimposed on the 'dead' network of simplexes S_i.

We follow Kondo and Iri[4]. 'The group of r-chains (cochains) C_r (C^r) is defined as the additive group with the S_r (S^r) as its basis, then the boundary and coboundary operators $d:C^r->C^{r-1}$ and $d:C_r->C_{r+1}$ also $d:C^o->0$ and $d:C_n->0$. The set of those elements of C_r which are mapped to 0 by d forms the group of r-cycles Z_r. The set of those elements of C_r which are images of elements of C_{r+1} forms the subgroup of r-boundaries F_r. From dd=0 we have the subsets $C_r>Z_r>F_r$. Since the groups now under consideration are commutative, every subgroup is a normal subgroup. Therefore the factor groups $C_r\backslash Z_r$ and $Z_r\backslash F_r$ are of interest. Since by definition Z_r is the kernel of the homomorphism $d:C_r->F_{r-1}$ the following isomorphism may easily be observed: $C_r\backslash Z_r \simeq F_{r-1}$ in connection with the first of the factor groups. As for the latter it is well known that $H_r=Z_r\backslash F_r \simeq B_r+T_r$ where B_r is a free additive group of rank R_r and T_r is a group isomorphic with the direct sum of a number of cyclic groups the ith of which is of order $t_r(i)$ such that $t_r(i+1)$ divides $t_r(i)$. B_r is called the rth Betti group, R_r the rth Betti number, T_r the r-dimensional torsion group, the $t_r(i)$'s the r-dimensional torsion coefficients and H_r, the r-dimensional homology group. Finally C_r is isomorphic to the sum $C_r \simeq Z_r+V_r$ where $V_r \simeq F_{r-1}$ represents an r-tree. All the above extends to the dual case simply by reversing the order of the indices and swapping subscripts and superscripts. Therefore the index r+1 must be replaced by r-1 and vice versa. Furthermore the First Duality Theorem states that $B^r \simeq B_r$ and $T^r \simeq T_{r+1}$ or equivalently that $R^r=R_r$ and $t^r=t_{r+1}$. In particular since obviously $F^n=0$ and $F_o=0$ we have $H^n \simeq Z^n \simeq B^n$ and $H_o \simeq Z_o \simeq B_o$ and $T^n=0$, $T_o=0$ also $C^o=Z^o$ and $C^n=Z_n$. From the fact that a one cell is incident to one and only one 0-cell with positive sign and one and only one 0-cell with negative sign, it follows that $T^o=T_1=0$ and $R^o=R_o$ is equal to the number of connected components of the complex. A complex whose r-dimensional homology group vanishes is called acyclic in the dimension r. Thus the chain group C_r is a free group and its basis may be sorted into five types of basis chains distinguished by their behaviour with respect to the boundary and coboundary operators.' The interrelationships of the various groups[4] are

summarised in Fig(3.1).

The above description is also applicable to the 2-dimensional complex of a conventional network in which every r-dimensional cycle (r>0) can be made to bound. The 1-dimensional homology group vanishes because by definition there exists no non-bounding 1-cycle. The diagram representing the homological structure of a network is as follows

$$
\begin{array}{ccccc}
 & & 0 & & \\
 & & \uparrow & & \\
 & & H^0=B^0 & & \\
 & & \uparrow & & \\
0 & 0 & 0\to Z^0 & \to C^0 & \to 0 \\
\uparrow & \uparrow & \uparrow & & d \\
H^2=B^2 & 0\to Z^1 & \to C^1 & \to F^0 & \to 0 \\
\uparrow & \uparrow & d\uparrow & & \\
0\to Z^2 & \to C^2 & \to F^1 & \to 0 & 0 \\
\uparrow & d\uparrow & & & \\
0 & 0 & & &
\end{array}
$$

All rows as well as all columns are exact. The relations between topological and electrical network terminology are defined as follows

Fig(3·1) The interrelations of various groups such as have been expounded in the above [4.] can be summarized in the following diagrams.

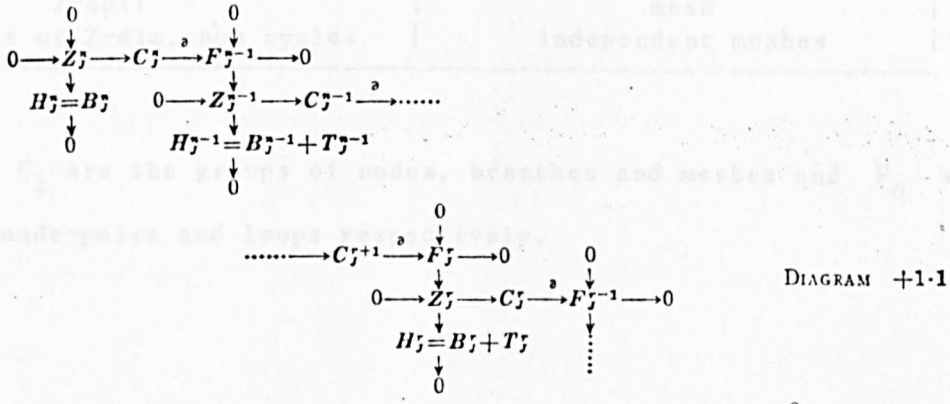


DIAGRAM +1·1

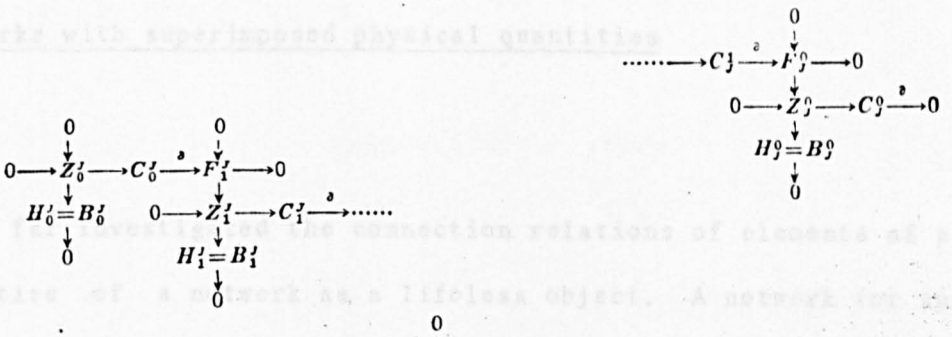
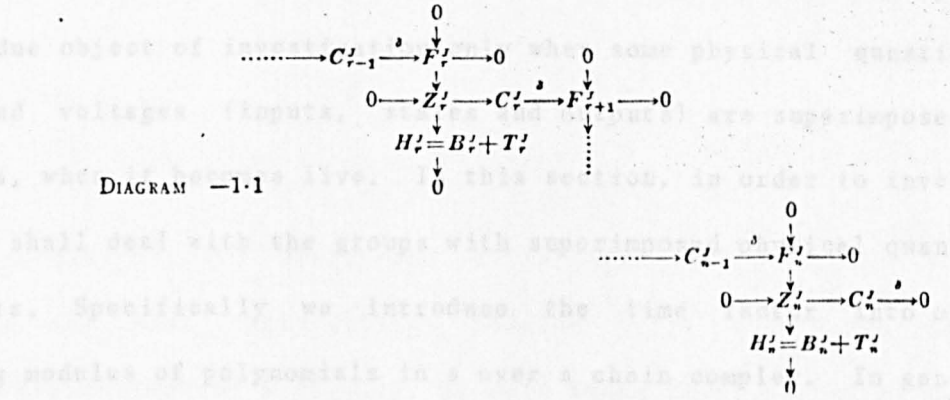


DIAGRAM −1·1



In the above diagrams all rows as well as all columns are exact.[1)]

1) A sequence of homomorphisms $\cdots\to A_{r-1}\to A_r\to A_{r+1}\to\cdots$ is said to be exact if, for each integer r, the image of $\phi_{r-1}: A_{r-1}\to A_r$ coincides with the kernel of $\phi_r: A_r\to A_{r+1}$. Especially, $0\to A\to B$ means that A is a subgroup of B, $A\to B\to 0$ means that A is mapped onto B; therefore $0\to A\to B\to C\to 0$ means that $C\cong B/A$, and $0\to A\to B\to 0$ means that $A\cong B$; etc.

topological terminology	network terminology
0-cell	node
bounding 0-cycle	node-pair
basis of 0-dim. non cocycles	independent nodes
1-cell	branch
1-cycle or boundary	loop
1-cocycle or coboundary	cut-set or star
2-cell	mesh
basis of 2-dim. non cycles	independent meshes

C_o, C_1 and C_2 are the groups of nodes, branches and meshes and F_o and $F_1=Z_1$ the groups of node-pairs and loops respectively.

3.3 Networks with superimposed physical quantities

We have so far investigated the connection relations of elements of a network ie. the properties of a network as a lifeless object. A network (or system) however, becomes a due object of investigation only when some physical quantities such as currents and voltages (inputs, states and outputs) are superimposed on it, or in other words, when it becomes live. In this section, in order to investigate a live network we shall deal with the groups with superimposed physical quantities as the coefficients. Specifically we introduce the time factor into our analysis by considering modules of polynomials in s over a chain complex. In general, physical quantities, such as current and voltage, form an additive group: when their instantaneous values only are of importance, they can be considered to belong to the field of real numbers: when they are regarded as functions of time, or more generally as distributions with respect to time, they should be considered to belong to a certain topological group of the respective functions or distributions: when the steady state analysis of a linear alternating current network is our concern, they can be regarded as belonging to the field of complex numbers, etc.

In any case we denote by K the group to which the respective physical quantities belong. In the most general case K may be a principle ideal domain and our manifold is a chain complex.

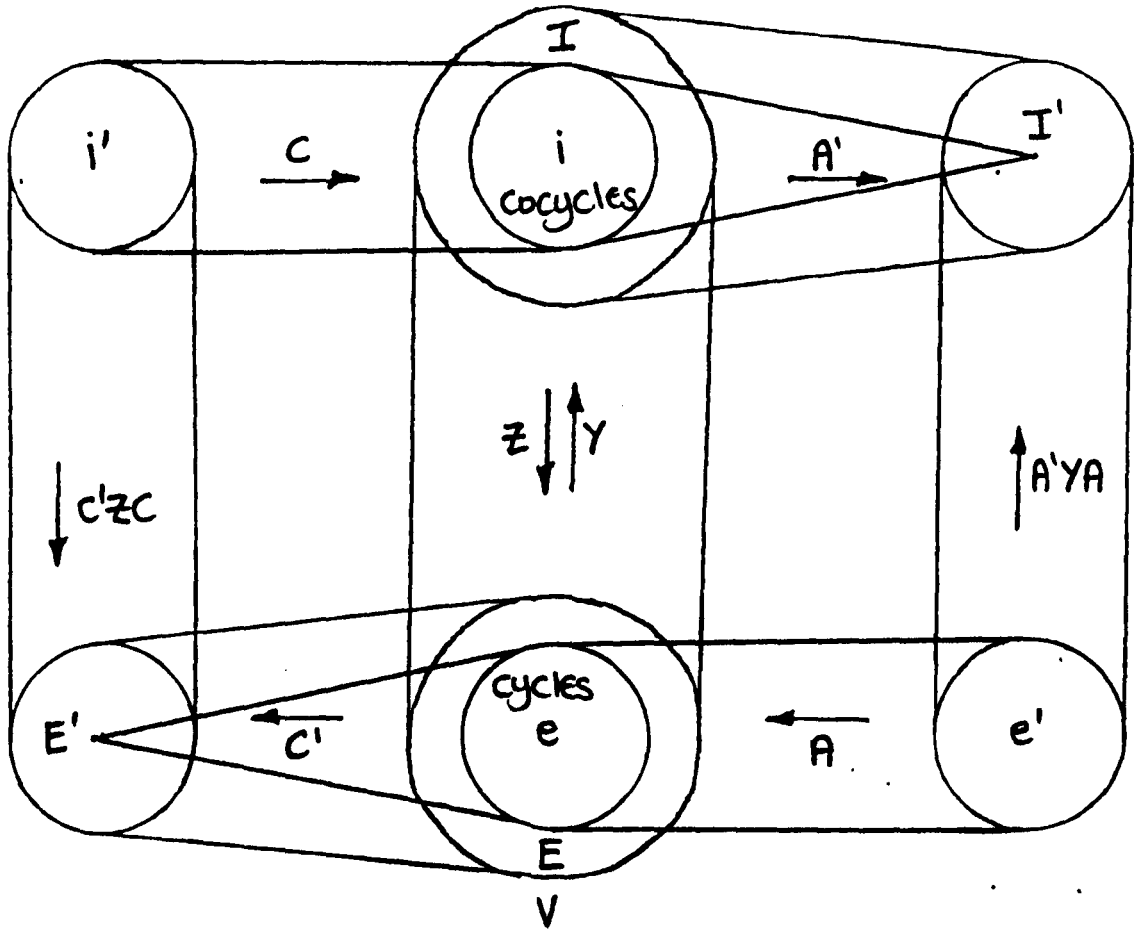
In his topological analysis of Kron's method of tearing, Roth[7] studied the isomorphism between the middle short exact sequence of the above diagram and its dual (see also Fig(3.2) showing the substructure of the problem in network theory notation)

$$\begin{array}{ccccccccc} 0 & \to & H^1(K) & \to & C^1(K) & \xrightarrow{d} & P^o(K) & \to & 0 \\ & & \simeq & & \simeq & & \simeq & & \\ 0 & \leftarrow & H_1(K) & \leftarrow & C_1(K) & \xleftarrow{d} & P_o(K) & \leftarrow & 0 \end{array}$$

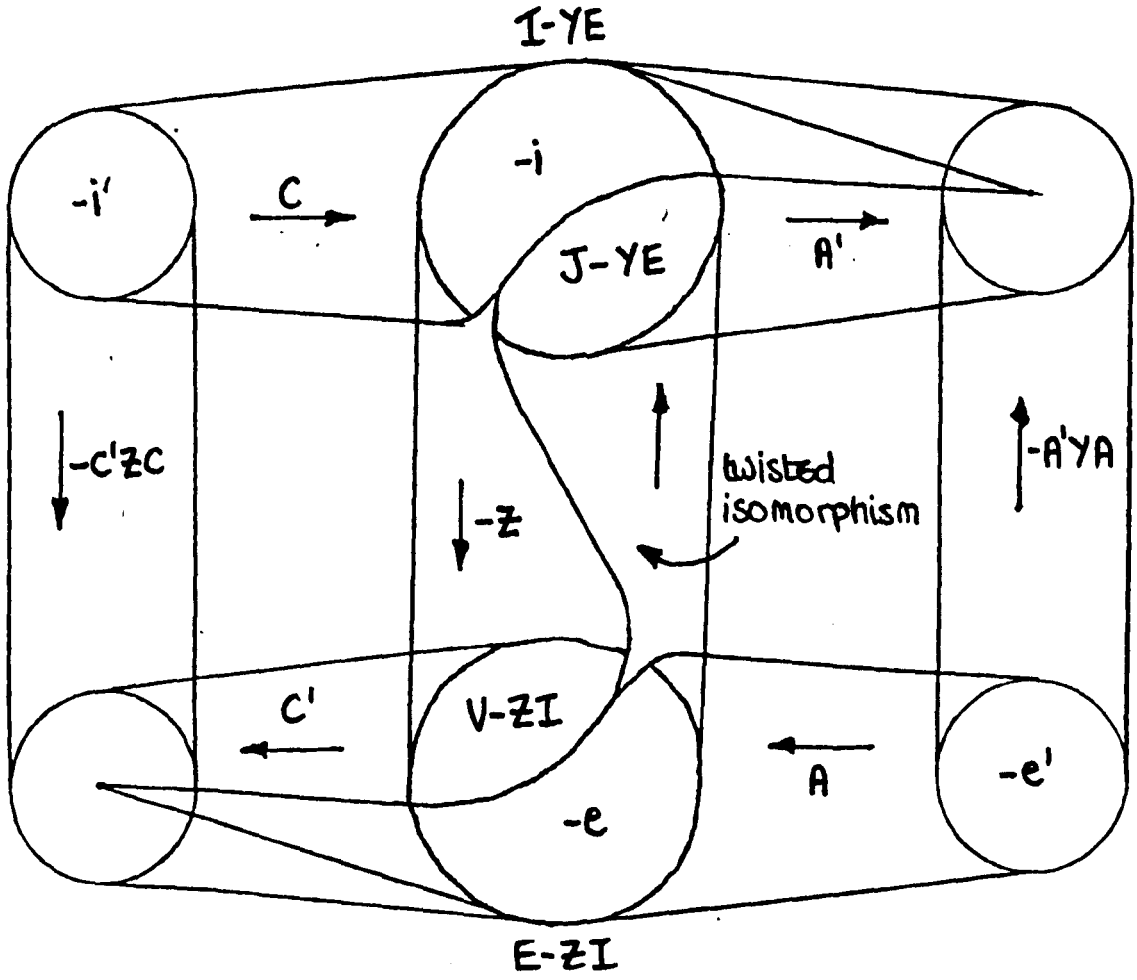
	Hu	Kondo	Roth
cycles	$H_1(X)$	$Z^1=F^1$	$H^1(K)$
chains	$H_1(X,A)$	C^1	$C^1(K)=H^1(K,K_o)$
boundaries	$H_o(A)$	F^o	$P^o(K)$

Comparison of notation

The table should clarify notational differences which appear to have caused problems in the past, particularly because Roth calls $H_r(K)$ the r-dimensional homology group instead of $H_r(K,K_o)$. It only shows correspondences between the way the authors use the variables, not the groups themselves - ie it depends whether one looks at the homology sequence or the chain complex. K_o is the 0-skeleton of points. The topologist Steenrod defined the property of 'ohmicness' (generalised power definiteness) of the isomorphism $\underline{H} \simeq \underline{H}^*$ for the general network problem to have a unique solution.



Fig(3·2) Chain groups in the electrical network problem



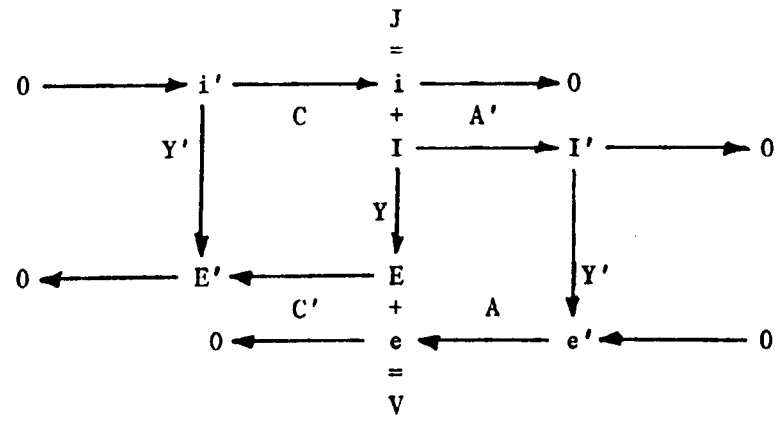
Fig(3·3) Roth's diagram for the electrical network problem

Roth describes the process of solving the network as 'an untangling of a twisted isomorphism between the homology and cohomology sequences of K modulo its zero skeleton' (see Fig(3.3) showing the substructure of the group homomorphisms). Note how Branin's diagram has to be redrawn to show Roth's twisted isomorphism. Branin ignored the underlying topological structure of the network and considered mappings between the spaces of the superimposed algebraic structure ie the coefficients of the homology groups.

Now define the vectors

i of b branch currents and
I of corresponding generators,
i' of m mesh currents,
e of b branch voltages and
E of corresponding generators and
e' of n node-pair voltages.

then Roth's isomorphism becomes Ohm's law E+e=Z(I+i) or I+i=Y(E+e) where Z is the diagonal matrix of branch impedances. Kirchoff's voltage and current laws are represented by the top and bottom sequences of Roth's diagram. We define a bxm matrix C relating mesh and branch variables where the incidence numbers $c_{ij}=(1,-1,0)$ if the ith orientated branch is (positively, negatively, not) incident to the jth orientated mesh. Each branch or 'primitive network' consists of one impedance (R,L,C) and one source (I,E). In a similar way define the nxm matrix relating the mesh and node variables then the following diagram commutes (ie. we get the same answers whichever way round we follow the arrows) and the rows are short exact hence A'C=0 where A' is the transpose of A.



This is Branin's form of the diagram (see also Fig(3.3)) actually due to Roth. By following the arrows around we obtain

C A'	V=ZJ	J=YV
0->i'->J->I'->0	E+e=Z(I+i)	I+i=Y(E+e)
≃ ≃ ≃	C'e=0 Kirchoff	A'i=0
0<-E'<-V<-e'<-0	C'E=E'	A'I=I'
C' A	i=Ci'	e=Ae'
mesh branch node	i'=/(C'ZC)C'(E-ZI)	e'=/(A'YA)A'(I-YE)

As can be seen the solution for i' (e') can be written down directly by following the arrows round the right (left) hand side of the diagram.

These solutions are a result of the inherent energy minimisation

$$\min_{e'} ||i||_Z = \min_{i'} ||e||_Y \qquad \text{or} \qquad \frac{d}{de'} i'Zi=0 \quad \frac{d}{di'} e'Ye=0$$

The dual of this in the sense of de Rham's isomorphism is distance minimisation over the underlying manifold considered as a metric space. Another equation we can of course invoke is that of power invariance under coordinate transformation, that is ie=i'e' (cf energy methods in electromagnetic theory).

3.4 The orthogonal network

Kron[14] retaliated to criticism of his method with regard to the use of singular connection matrices, by including extra 'open meshes' in the branch to mesh transformation resulting in a square C. Consider the mesh method and define (1) a spanning tree of t branches with independent currents over the network: the b-t cotree branches are called links, (2) a set of b paths through the network consisting of m closed paths or meshes and b-m open (lamellar) paths, which give enough degrees of freedom to take all the current into account. We perform a non-singular transformation from tree and link branches to open and closed meshes. We make a minor change in notation from Branin's to Kron's in that we now show lamellar variables in large letters and solenoidal variables in small letters. This only means that e and E are interchanged. Kron also showed current related variables superscripted and voltage related variables subscripted.

Nicholson[6] showed as follows how the solution to the orthogonal network problem could be exhibited as a scattering product involving two consecutive obstacles. 'The general electrical network problem includes the connection of b-primitive branches specified by E+e=Z(I+i) or V=ZJ where Z is a symmetrical impedance matrix for the primitive network and e,I represent branch voltage and source current vectors respectively. In the orthogonal formulation the branch connections are defined in terms of square non-singular connection matrices $C=[C_c\ C_o]$ and $A=[A^c\ A^o]=/C'$ related to specified closed and open paths containing variables $i^c{}',e_c{}'$ and $I^o{}',E_o{}'$ respectively. The open meshes can always be chosen to pass through the tree branches resulting in the form of connection matrix shown below. In the connected network

$$J=[C_c\ C_o]\begin{bmatrix} i^c{}' \\ I^o{}' \end{bmatrix} \text{ and } V=[A^c\ A^o]\begin{bmatrix} e_c{}' \\ E_o^c{}' \end{bmatrix}$$

and with specified trees and links

$$C=\begin{bmatrix} C_{tc} & C_{to} \\ C_{lc} & C_{lo} \end{bmatrix}=\begin{bmatrix} C_t & 1 \\ 1^t & 0 \end{bmatrix},\quad A=\begin{bmatrix} A^{tc} & A^{to} \\ A^{lc} & A^{lo} \end{bmatrix}=\begin{bmatrix} 0 & 1_t \\ 1_l & A^l \end{bmatrix}$$

where 1 is the unit matrix. The branch variables E, i and the tree-branch voltages E_o' and mesh currents $i^{c'}$ are related by

$$E=\begin{bmatrix} E_t \\ E_l \end{bmatrix}=A^o E_o',\quad i=\begin{bmatrix} i_t \\ i^l \end{bmatrix}=C_c i^{c'}$$

The branch variables are also constrained by the Kirchoff laws

$C_c'E=0$, $A^{o'}i=0$, and $E'i=0$.

New equivalent sources e_l, I^t are then referred to the links and open path or tree branches respectively, with

$$e=\begin{bmatrix} e_t \\ e_l \end{bmatrix}=A^c e_c'=\begin{bmatrix} 0 \\ C_c'e \end{bmatrix},\quad I=\begin{bmatrix} I_t \\ I^l \end{bmatrix}=C_o I^{o'}=\begin{bmatrix} A^{o'}I \\ 0 \end{bmatrix}$$

where $e_c'=C_c'e$ and $I^{o'}=A^{o'}I$ represent equivalent induced mesh-voltage and nodal current sources respectively with arbitrary sources e, I. Combining the equations and rearranging gives the solutions

$$\begin{bmatrix} i^{c'} \\ E_o' \end{bmatrix}=\begin{bmatrix} -/Z_a Z_b & /Z_a \\ Z_d-Z_c/Z_a Z_b & Z_c/Z_a \end{bmatrix}\begin{bmatrix} I^{o'} \\ e_c' \end{bmatrix}=\begin{bmatrix} Y^b/Y^d & Y^a-Y^b/Y^d Y^c \\ /Y^d & -/Y^d Y^c \end{bmatrix}\begin{bmatrix} I^{o'} \\ e_c' \end{bmatrix}$$

where

$$\begin{bmatrix} Z_a & Z_b \\ Z_c & Z_d \end{bmatrix}=\begin{bmatrix} C_t' & 1 \\ 1^t & 0 \end{bmatrix}\begin{bmatrix} Z_t & \\ & Z_l \end{bmatrix}\begin{bmatrix} C_t & 1 \\ 1^t & 0 \end{bmatrix}=\begin{bmatrix} C_t'Z_tC_t+Z_l & C_t'Z_t \\ Z_tC_t & Z_t \end{bmatrix}$$

$$\begin{bmatrix} Y^a & Y^b \\ Y^c & Y^d \end{bmatrix}=\begin{bmatrix} 0 & 1_t \\ 1_l & A^{l'} \end{bmatrix}\begin{bmatrix} Y^t & \\ & Y^l \end{bmatrix}\begin{bmatrix} 0 & 1_t \\ 1_l & A^l \end{bmatrix}=\begin{bmatrix} Y^l & Y^lA^l \\ A^{l'}Y^l & Y^t+A^{l'}Y^lA^l \end{bmatrix}$$

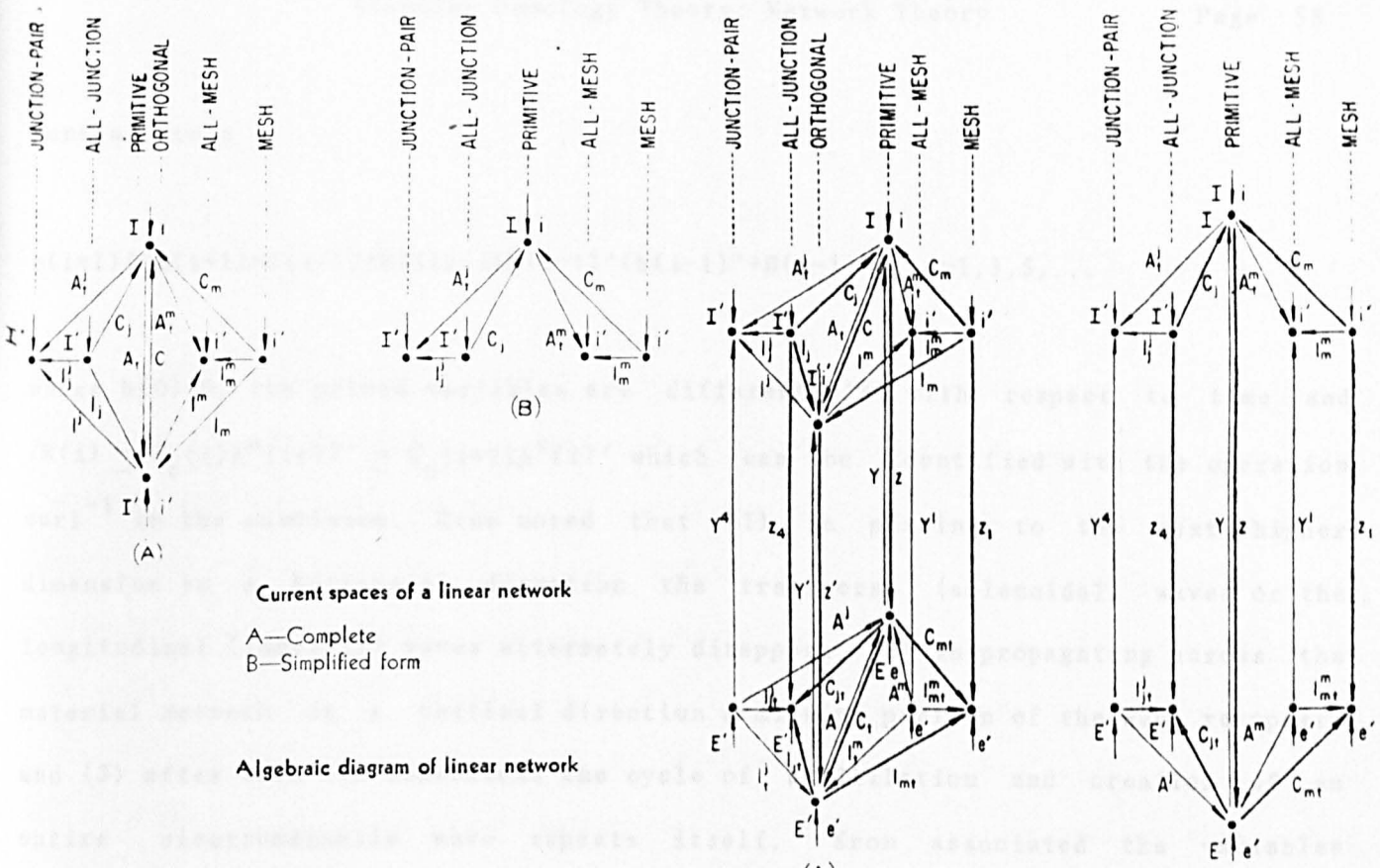
The usual mesh and node solutions are included with $i=L(e-ZI)$ and $E=M(I-Ye)$ where $L=C_c/(C_c'ZC_c)C_c'$ and $M=A^o/(A^{o'}YA^o)A^{o'}=Z-ZLZ$ represent the branch admittance and branch impedance matrices respectively.

Kron generalised the conventional electric network to a k-dimensional polyhedron or wave model. The electromagnetic structure formed by interconnecting two or more wave models into one network is called a 'multidimensional space filter'. The equations represent the orthogonal formulation of the electrical network problem, and a similar form is used to characterise each 'isolated' higher dimensional network. The square connection matrices as well as their four submatrices (boundary operators) can be arranged into an algebraic diagram as illustrated in Fig (3.4). Rectangular incidence matrices M(i) of -1,0 and 1 elements interconnect the spatial elements in adjacent networks, such as branches with nodes or planes, and are related to the boundary operators or partial connection matrices C_c, A^o for the ith and i+1th networks with

$$C_c(i)A^o(i+1)'=M(i),\ C_c(i)'A(i)=0,\ M(i-1)M(i)=0,\ L(i)M(i-1)'=0.$$

It is emphasised that with an isolated q-network, no incidence matrix M is associated, only a connection matrix C or A. The transformation diagram includes residual type operators directed horizontally between similar closed and open path variables, with impedance - admittance type projection operators projected vertically between the closed and open path dual variables.

Kron introduced a concept of wave propagation into the polyhedron model, with a transverse electromagnetic wave associated with closed path dielectric and magnetic variables (d(i)',e(i)) and (h(i),b(i)') respectively, and a longitudinal wave associated with open path dielectric and magnetic variables (D(i)',E(i)) and (H(i),B(i)') respectively where the primed variables are differentials with respect to time. Propagation proceeds from the 0-dimensional points to the higher dimensional network elements, and induces electrical and electromagnetic variables, consistent with the form of Maxwell's field equations. A cycle of 'open-circuit' wave propagation repeats after each two dimensions and is represented by the



Current spaces of a linear network

A—Complete
B—Simplified form

Algebraic diagram of linear network

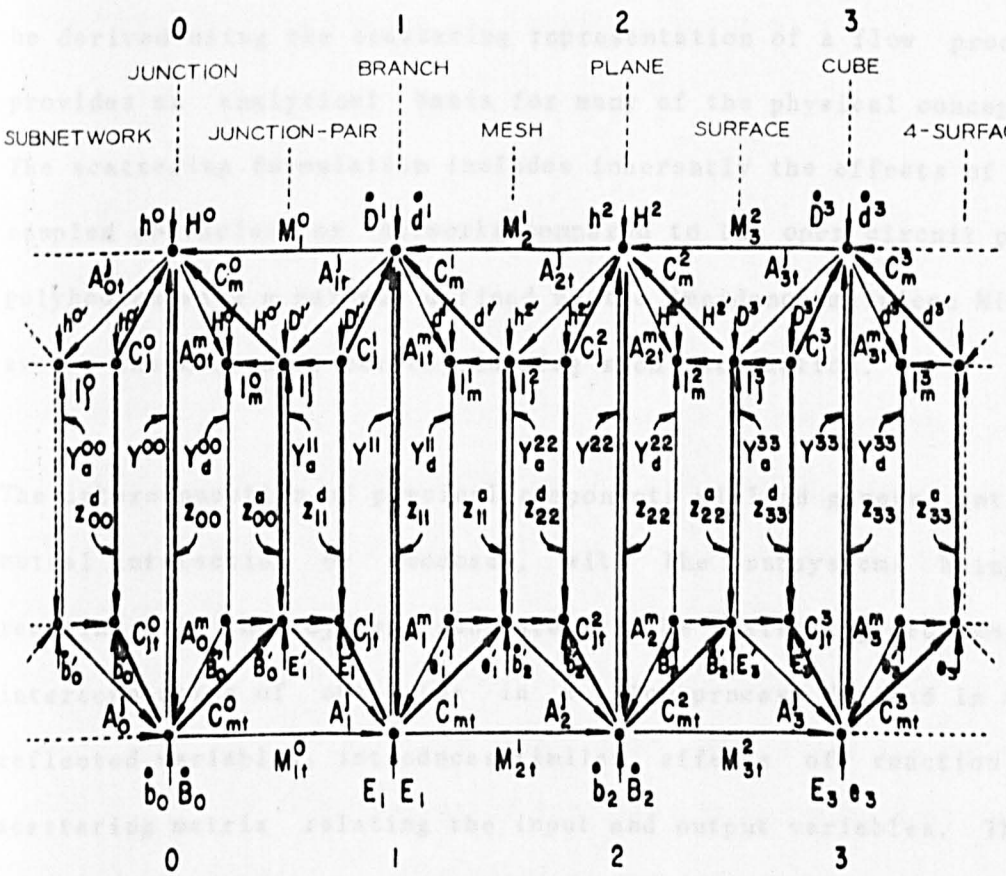


Fig (3·4) Simplified algebraic diagram of polyhedron [18]

general steps

$$b(i+1)'+B(i+1)=Z(i+1)/M(i)Y(i)M(i-1)'(b(i-1)'+B(i-1)'),\ i=1,3,5,...$$

where $b(0)=0$, the primed variables are differentials with respect to time and $/M(i) \pm C_c(i)A^o(i+1)' \pm C_o(i+1)A^c(i)'$ which can be identified with the operation $curl^{-1}$ in the continuum. Kron noted that (1) in passing to the next higher dimension in a horizontal direction the transverse (solenoidal) waves or the longitudinal (lamellar) waves alternately disappear, (2) in propagating across the material network in a vertical direction a missing portion of the wave reappears and (3) after each two dimensions the cycle of annihilation and creation of an entire electromagnetic wave repeats itself. Kron associated the variables $d(i)'+D(i)'$ and $b(i+1)'+B(i+1)'$ of the polyhedron model with the ith and i+1th order divided differences $(1-z^{-1})^i$ and $(1-z^{-1})^{i+1}$ of the estimation problem: the minimisation of error being associated with that of energy. Similar properties may be derived using the scattering representation of a flow process, and this also provides an analytical basis for many of the physical concepts discussed by Kron. The scattering formulation includes inherently the effects of interaction between coupled obstacles or networks compared to the open circuit propagation across the polyhedron with a pathway defined by the incidence matrices M(i) which apparently avoids the necessity for considering such interaction.

The interconnection of physical components will in general introduce an effect of mutual interaction or feedback, with the subsystems being influenced by and reacting with the adjacent subsystems. The scattering problem, concerned with the interconnection of obstacles in a flow process defined in terms of incident and reflected variables, introduces similar effects of reaction within a combined scattering matrix relating the input and output variables. The solution matrix of the orthogonal network can be decomposed into a Redheffer star product

$$\begin{bmatrix} & -/Z_a Z_b & /Z_a \\ Z_d & -Z_c/Z_a Z_b & Z_c/Z_a \end{bmatrix} = \begin{bmatrix} Z_b & 0 \\ Z_d & Z_c \end{bmatrix} * \begin{bmatrix} -/Z_a & /Z_a \\ -/Z_a & /Z_a \end{bmatrix} = \begin{bmatrix} C_t' Z_t & 0 \\ Z_t & Z_t C_t \end{bmatrix} * \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} o / Z_a$$

$$\begin{bmatrix} Y^b/Y^d & Y^a - Y^b/Y^d Y^c \\ /Y^d & -/Y^d Y^c \end{bmatrix} = \begin{bmatrix} /Y^d & -/Y^d \\ /Y^d & -/Y^d \end{bmatrix} * \begin{bmatrix} Y^b & Y^a \\ 0 & Y^c \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} o / Y^d * \begin{bmatrix} Y^1 A^1 & Y^1 \\ 0 & A^1, Y^1 \end{bmatrix}$$

where o is the Kronecker or tensor product. The star product can thus represent the form of an orthogonal network solution and illustrates a basis for connecting network tree and link elements. It may appear that the solution does not introduce inherently the properties of a return difference operator. This would be supported by Kron's reference to the polyhedron model as an open circuit structure of networks, however the orthogonal network solution can be considered to contain a return difference effect within the off diagonal elements and thus includes the ability to incorporate a priori information. In the scattering problem this effect and the return difference operator are introduced into each successive stage of the multistage process by the effects of the connecting zone between adjacent obstacles, in contrast to the connection of adjacent orthogonal networks in the polyhedron model based on the incidence matrices M(i).'

3.5 Tearing

The process of tearing involves the removal of just enough complete branches to split the network up into a number of unconnected subnetworks. This collection of removed branches is known as the intersection network and should preferably be taken from the most loosely coupled branches of the original network. Kron emphasised that mere partitioning of the network equations is a far more difficult task to perform efficiently than that of tearing the network as information is actually lost in formulating the equations from the network. For the mesh method with voltage sources[6], Ohm's law of the intersection network is

$Zi=e$

where Z is a diagonal matrix of known nodal impedances and i and e are unknown vectors of equivalent current sources due to subdivision, and nodal voltages across the removed branches, respectively. For the subnetworks

$\underline{Z}\underline{i}=\underline{E}+\underline{e}$

where $\underline{Z}$ is the block diagonal matrix of known mesh impedances, one block for each subnetwork, and $\underline{E}$, $\underline{i}$ and $\underline{e}$ are partitioned vectors of known voltage sources, unknown mesh currents and unknown equivalent voltage sources of the torn branches respectively. Also

$i=\underline{C}\underline{i}$ and

$\underline{e}=-\underline{C}'e$

where $\underline{C}$ is a partitioned connection matrix relating the intersection network to the subnetworks. The fundamental equations of diakoptics can thus be written

$$\begin{bmatrix} Y & -\underline{C} \\ \underline{C}' & \underline{Z} \end{bmatrix}\begin{bmatrix} e \\ \underline{i} \end{bmatrix}=\begin{bmatrix} 0 \\ \underline{E} \end{bmatrix}$$

Substituting and using Householders formula the unknown mesh currents are

$\underline{i}=/(\underline{Z}+\underline{C}'Z\underline{C})\underline{E}$ ...(1)

$=(/\underline{Z}-/\underline{Z}\underline{C}'/(/Z+\underline{C}/\underline{Z}\underline{C}')\underline{C}/\underline{Z})\underline{E}$ ...(2)

Now computing equation (1) involves inversion of $\underline{Z}+\underline{C}'\underline{Z}\underline{C}$ which is of the dimension of the total number of meshes of the original network. Computing (2) involves the inversion of $\underline{Z}$, a block diagonal matrix in which each block is of the dimension of the number of meshes in the associated subnetwork, and of Z, the diagonal matrix of intersection impedances and $/Z+\underline{C}/\underline{Z}\underline{C}'$ of the dimension of the number of branches of the intersection network. Thus equation (2), the basic equation of Diakoptics results in a considerable saving of computational effort. In a similar way Onodera developed the dual node method with current sources or codiakoptics which effectively allows the removal of nodes rather than branches. The fundamental equations of codiakoptics are

$$\begin{bmatrix} \underline{Y} & -\underline{C} \\ \underline{C}' & \underline{Z} \end{bmatrix} \begin{bmatrix} \underline{e} \\ \underline{i} \end{bmatrix} = \begin{bmatrix} \underline{I} \\ \underline{0} \end{bmatrix}$$

Kondo's generalised diakoptics allows each subnetwork to be solved by either the mesh or the node method, whichever is the most convenient.

3.6 References

(1) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresmann, 1971.

(2) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(3) Amari, Topological Foundations of Kron's Tearing of Electrical Networks, RAAG Memoirs Vol 3, 1962.

(4) Kondo and Iri, On the Theory of Trees, Cotrees, Multi-Trees and Multicotrees, RAAG Memoirs, Vol 2, 1958.

(5) Mizoo, Iri and Kondo, On the Torsion and Linkage Characteristics and the Duality of Electric, Magnetic and Dielectric Networks, RAAG Memoirs, Vol 2, 1958.

(6) Nicholson, System Concepts in Kron's Polyhedron Model and the Scattering Problem, Int. J. Control, Vol 20, No 4, 1974.

(7) Roth, An application of algebraic topology: Kron's method of tearing, Q. Applied Math, April 1959.

(8) Branin, The Algebraic-Topological Basis for Network Analogies and the Vector Calculus, Symposium on Generalised Networks, Poly. Inst. of Brooklyn, April 1966.

(9) Bramellar, Practical Diakoptics for Electrical Networks, Chapman and Hall, London, 1969.

(10) Happ, Diakoptics and Networks, Academic Press, 1971.

(11) Iri, Supplements to the Theory of Trees, Cotrees, Multitrees and Multicotrees, RAAG Memoirs 3, A-X, 1962, 4-22.

(12) Kron, Four abstract reference frames of an Electric Network, IEEE Trans PAS 87, No 3, March 1968, p 815.

(13) Kron, Invisible Dual (n-1) Networks induced by Electric 1-Networks, IEEE Trans CT 12, No 4, Dec 1965, p 464.

(14) Kron, The Misapplication of Graph Theory to Electric Networks, AIEE Trans Pt 1, Vol 81, Sept 1962.

(15) Onodera, Topological Synthesis of Non-Reciprocal Resistance Networks, IEEE Monograph No 428E, Feb 1961.

(16) Onodera, A new approach to Kron's method of analysing Large Systems, IEE Monograph No 403E, Oct 1960.

(17) Bourgin, Modern Algebraic Topology, Macmillan, 1963.

(18) Kron, Basic Concepts of Multidimensional Space Filters, Proc IEE, Nov 1959.

(19) Hoffman, Kron's Method of Subspaces, Quart Applied Maths, Vol 2, pp 218-31, 1944.

(20) Synge, The Fundamental Theorem of Electrical Circuits, Quart Applied Maths, Vol 9, pp 113-27, 1951.

(21) Roth, Kron's Method, Pts I, II, III and IV, Quart Applied Maths, Vol 61, pp 83, 1955 and Vol 62, pp 32 and 45, 1956.

(22) Roth, An Application of Algebraic Geometry to Numerical Analysis, Proc Nat Acad Sci, Vol 41, pp 518-21, 1955.

(23) Roth, The Validity of Kron's Method of Tearing, Ibid, pp 599-600.

(24) Iri, Supplements to the Theory of Trees, Cotrees, Multitrees and Multicotrees, RAAG Memoirs, Vol 3, pp 310-28, 1962.

CHAPTER IV De Rham Cohomology Theory: Field Theory

This Chapter looks at the space of exterior differential forms or de Rham Cohomology Theory and its relationship to vector and tensor algebra in the small. Boundary and coboundary operators and the Hodge * operator are introduced, putting the analysis into the context of General Homology Theory. Maxwell's equations for electromagnetic waves are displayed on an extended form of Roth's diagram and Branin's network model is shown to be the basis of the Finite Element Method.

4.1 Exterior differential structures

The study of differential forms[1 - 4] generalises the concepts of gradient, divergence and curl of a vector or tensor field to higher dimensional spaces. Let R be the field of real numbers and L an n-dimensional vector space over R. For each p=0,1,...,n we shall construct a new vector space V^pL, the space of differential or exterior p-forms on L. We begin with $V^oL=R$, $V^1L=L$ and $\dim(V^pL)={}^nC_p=n!/(p!(n-p)!)$, $0\leq p\leq n$. Exterior n-forms are objects that occur under an integral sign or as the result of a differentiation. For example a line integral takes a 1-form in 3-space $\underline{w}=Adx+Bdy+Cdz$, a surface integral takes a 2-form $\underline{w}=Pdydx+Qdzdx+Rdxdy$ and a volume integral takes a 3-form $\underline{w}=Hdxdydz$. The absence of terms dzdy, etc suggests some kind of symmetry. The absence of terms dxdx, etc suggests skew-symmetry.

The exterior product of two one forms is a generalisation of the cross product of two vectors

$$\underline{w}_1 \uparrow \underline{w}_2 = (A_1B_2 - B_1A_2)dx\uparrow dy + (B_1C_2 - C_1B_2)dy\uparrow dz + (C_1A_2 - A_1C_2)dz\uparrow dx$$

In general $\uparrow : V^pL \times V^qL \to V^{p+q}L$. The wedge product obeys the distributive law and the alternation rules $dx\uparrow dz = -dz\uparrow dx$ etc and $dz\uparrow dz = 0$ etc. It is not commutative. Note that the $\uparrow$ is often ommitted so that $dx\uparrow dy = dxdy$. The exterior product thus represents an element of area with a rotational orientation. The cross product is a vector mutually at right angles to both $\underline{w}_1$ and $\underline{w}_2$. We write $\underline{w}_1 \uparrow \underline{w}_2 = *(\underline{w}_1 x \underline{w}_2)$. Misnor and Wheeler[1] visualise the exterior product of two 1-forms as a honeycomb or eggcrate structure, independent of the z-axis, with a rotational arrow in each tube produced. Imagine a stationary bundle of straws each rotating in the same direction. The tubes are narrow and their cross section is irrelevant, but their number gives the magnitude of the 2-form.

A p-form can be expressed in terms of the components of a skew-symmetric tensor of rank p in a general coordinate system

$$\underline{w} = \frac{1}{p!} \sum A_{ab\ldots p} dx_a \uparrow \ldots \uparrow dx_p$$

An alternative formulation as the ordered sum over a set of p-tuples can be made. Exterior multiplication of forms must satisfy the associative law, unlike cross multiplication of vectors. Further for a p-form $\underline{w}$ and a q-form $\underline{v}$

$\underline{w}\uparrow\underline{v} = (-1)^{pq}\underline{v}\uparrow\underline{w}$ if $p+q<n$ (anticommutation)

$= 0$ if $p+q>n$.

The inner product of two 1-forms in 3-space $\underline{w}_1.\underline{w}_2=A_1A_2+B_1B_2+C_1C_2$, often written $(\underline{w}_1,\underline{w}_2)$, with heavy or double brackets. In general $.:V^pLxV^pL->R$. For two p-forms $\underline{w}=\underline{w}_1\uparrow\underline{w}_2\ldots\uparrow\underline{w}_p$ and $\underline{v}=\underline{v}_1\uparrow\underline{v}_2\ldots\uparrow\underline{v}_p$ the dot product is given by the determinant of the matrix whose ijth element is $\underline{w}_i.\underline{v}_j$. It follows that the inner product of a p-form and a q-form is nonexistent and further that the inner product is commutative and distributive.

The 1-form $\underline{w}x\underline{v}$ is known as the dual of the 2-form $\underline{w}\uparrow\underline{v}$, as above. Hodge defined the star operator

$$\underline{w}\uparrow\underline{v}=*\underline{w}.\underline{v}ds,\quad *:V^pL->V^{n-p}L$$

where * being monadic operates first, ie only on $\underline{w}$, and ds is an n-dimensional volume. The star operation on $\underline{w}$ gives the dual n-p form $*\underline{w}$. Its value can be found by equating coefficients. Wheeler has shown that for the general case

$$**\underline{w}=(-1)^{np+p+s}\underline{w}$$

where $\underline{w}$ is a p-form in n-space and s is the number of minus signs in the inner product of the base vectors.

The gradient of a scalar is an example of an exterior derivative. This can be generalised using the rules

$$\underline{d}(\underline{w}+\underline{v})=\underline{dw}+\underline{dv},\quad \underline{d}(\underline{w}\uparrow\underline{v})=\underline{dw}\uparrow\underline{v}+(-1)^p\underline{w}\uparrow\underline{dv}\text{ and }\underline{d}(dx)=0.$$

We have $\underline{d}:V^pL\to V^{p+1}L$. Note that the gradient of a scalar is the generalised curl of a 0-form $\underline{dw}$, the conventional curl of a vector is the dual of the generalised curl of a 1-form $*\underline{dv}$. A p-form $\underline{w}$ is called closed if $\underline{dw}=0$ and exact if there exists a p-1 form $\underline{v}$ such that $\underline{w}=\underline{dv}$.

The generalised divergence takes a p-form into a p-1 form. The conventional divergence of a vector gives a scalar. The extension of this concept to a p-form is in three stages (1) the dual of $\underline{w}$ gives a n-p form $*\underline{w}$ (2) the exterior derivative $\underline{d}*\underline{w}$ takes the n-p form into an n-p+1 form (3) a dual of this n-p+1 form gives the required p-1 form $*\underline{d}*\underline{w}$. Wheeler defines the divergence or codifferential operation as

$$\underline{bw}=(-1)^{np+n+s+1}*\underline{d}*\underline{w}.$$

If $\underline{w}$ is a p-form and $\underline{v}$ is a p+1 form then $(\underline{dw}).\underline{v}=\underline{w}.(\underline{dv})$. This may be interpreted as saying $\underline{b}$ and $\underline{d}$ are adjoint.

The Laplacian of a vector $\underline{w}$ is written in vector notation

$\mathrm{grad}(\mathrm{div}\underline{w})-\mathrm{curl}(\mathrm{curl}\underline{w}))$ or $\underline{d}(\underline{d}.\underline{w})-\underline{d}\mathrm{x}(\underline{d}\mathrm{x}\underline{w})$.

In the notation used in differential forms, the Laplacian of a p-form in n-space is

$\underline{L}=\underline{dbw}+\underline{bdw}$. We have $\underline{L}^*=*\underline{L}$.

$\underline{L}$ is self adjoint ie $\underline{Lw}.\underline{v}=\underline{w}.\underline{Lv}$. The form is called harmonic if the Laplacian of $\underline{w}$ is zero, then $\underline{dw}=0$ and $\underline{bw}=0$.

If $\underline{w}$ is a p-form then the second differential

$\underline{d}\underline{d}w=0$.

In Euclidean 3-space the application of this equation to a scalar a and a vector $\underline{v}$ gives

$*\underline{d}\underline{d}a=0$ and $*\underline{d}*(*\underline{d}v)=*\underline{d}(\underline{d}v)=0$.

This is a way of expressing the well known equations curl(gradient)=0 and div(curl)=0.

Hodges theorem If $\underline{w}$ is any p-form then there is a p-1 form $\underline{v}$, a p+1 form $\underline{u}$ and a harmonic form $\underline{y}$ such that $\underline{w}=\underline{d}v+\underline{b}u+\underline{y}$, where $\underline{d}v$, $\underline{b}u$ and $\underline{y}$ are unique. This is the generalisation of Helmholtz theorem in 3-space. Euclidean space can be decomposed as the direct sum

$$E^p=\underline{L}E^p+H^p=\underline{d}\underline{b}E^p+\underline{b}\underline{d}E^p+H^p=\underline{d}E^{p-1}+\underline{b}E^{p+1}+H^p$$

where H^p is the space of harmonic p-forms, the space orthogonal to $\underline{L}E^p$. We define the Green's operator $\underline{G}:E^p\rightarrow\underline{L}E^p$ by setting $\underline{G}a$ equal to the unique solution of $\underline{L}w=\underline{a}-\underline{H}a$ in $\underline{L}E^p$. $\underline{G}$ commutes with $\underline{d}$, $\underline{b}$ and $\underline{L}$ or any operator which commutes with $\underline{L}$. From the Hodge theorem $\underline{L}\underline{G}+\underline{H}=1$.

The generalised Stokes theorem

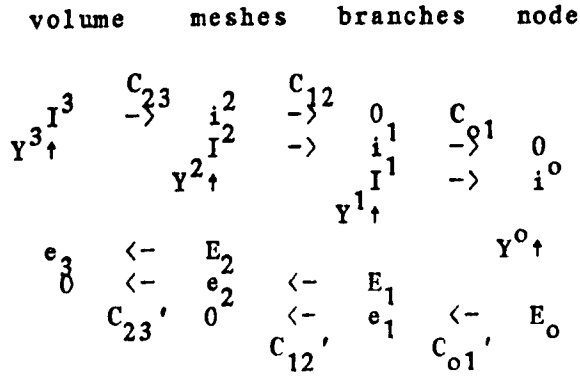
$$\oint_{dS} \underline{w} = \int_{S} \underline{dw}$$

where $\underline{w}$ is a p-form, to be integrated over the p+1 dimensional boundary of S and $\underline{dw}$ is a p+1 form to be integrated over the p-dimensional region S.

As yet we have made no mention of the time factor in our analysis. Again the most general case is that of modules of polynomials in s over the chain complex of exterior differential forms. We consider Maxwell's equations for electromagnetic fields.

4.2 Network model for Maxwell's equations

Branin[8] has generalised Kron's 2-network analysis to that of a 3-network of nodes, branches, meshes and volumes. This completely describes the 2-network problem with inductive linkages in the sense of Kondo and Iri and provides a useful introduction to Kron's network model of Maxwell's field equations which it describes. Branin gives the transformation diagram for a simply connected 3-network as follows



The middle square does not commute. The basic theorems regarding boundary and coboundary operators $C_{o1}C_{12}=0$ and $C_{12}C_{23}=0$ hold, also $Y^3=/(C_{23}'Z_2C_{23})$ and $Y^o=C_{o1}Y^1C_{o1}'$. As the Betti numbers of a simply connected complex are all unity, that is the homology groups all vanish, the dimensions of the subspaces are

Subspace	Dimension
i^o, E_o, e_1	$p_1=n_o-1=\text{rank}(C_{o1})$
i^1, e_2	$p_2=n_1-p_1=\text{rank}(C_{12})$
i^2, I^3, e_3	$p_3=n_2-p_2=n_3=\text{rank}(C_{23})$

where the n_j are the numbers of j-cells in the 3-complex. Corresponding to the two Kirchoff laws for a linear graph are four such laws for a 3-complex

$$C_{o1}i^1=0 \text{ or } i^1=C_{12}I^2,\ C_{12}i^2=0 \text{ or } i^2=C_{23}I^3,$$

$$C_{12}'e_1=0 \text{ or } e_1=C_{o1}'E_1,\ C_{23}'e_2=0 \text{ or } e_2=C_{12}'E_2.$$

Similarly, there are two expressions corresponding to Ohm's law

$$i^1+I^1=Y^1(e_1+E_1) \text{ or } i^2+I^2=Y^2(e_2+E_2).$$

Branin states that he 'originally expected to find two versions of the 3-network problem analogous to Roth's electrical network problem for a linear graph. However every attempt to set up and solve a Roth type problem, with i^o and e_3 specified has failed - apparently because such a problem is underdetermined. The only 3-network problem which the author has been able to define and solve corresponds roughly, to the electrical network problem - but with notable differences. Specifically one cannot assume arbitrary vectors I^1, E_1, I^2, E_2 since E_1 and I^2 are determined by I^1 and E_2. The 3-network problem, then is

given (1) a simply connected 3-complex and its connection matrices C_{o1}, C_{12} and C_{23}, (2) the isomorphisms Y^1 and Y^2 or Z_1 and Z_2 and (3) the arbitrary vectors I^1 and E_2,

find the vectors i^1, i^2, I^2, I^3 and E_o, E_1, e_1, e_2 such that Kirchoff's and Ohm's laws hold.'

Branin derives the solutions

$$E_1=/(C_{12}(/(C_{12}'Z_1C_{12})-Y^2)C_{12}')C_{12}(Y^2E_2+/(C_{12}'Z_1C_{12})C_{12}'Z_1I^1),$$

$$I^2=/(C_{12}'(/(C_{12}Y^2C_{12}')-Z_1)C_{12})C_{12}'(Z_1I^1+/(C_{12}Y^2C_{12}')C_{12}Y^2E_2),$$

$$E_o=/(C_{o1}Y^1C_{o1}')C_{o1}(I^1-Y^1E_1),$$

$$I^3=/(C_{23}'Z_2C_{23})C_{23}'(E_2-Z_2I^2),$$

and obtains i^1, i^2, e_1, e_2 from Kirchoff's laws. These equations are related to the discrete Riccati equation.

Maxwell's equations summarise the macroscopic electromagnetic field theory. In vector notation

$$\mathrm{curl}\underline{E}=-d\underline{B}/dt,\quad \mathrm{curl}\underline{H}=\underline{J}+d\underline{D}/dt,$$

$$\mathrm{div}\underline{B}=0,\quad \mathrm{div}\underline{D}=\rho,$$

$\underline{B}=\mu\underline{H}$, $\underline{D}=\varepsilon\underline{E}$ and $\underline{J}=\sigma\underline{E}$

where $\underline{E}$ is electric field, $\underline{H}$ is magnetic field, $\underline{B}$ is magnetic flux, $\underline{D}$ is electric flux and $\underline{J}$ is electric current density. ρ is electric charge density. μ, ε and σ are the permeability, permittivity and impedance of the medium respectively (equivalent to L, C and R.) From $\underline{B}=\mathrm{curl}\underline{A}$ we have $\mathrm{curl}(\underline{E}+d\underline{A}/dt)=0$ with $\mathrm{div}\underline{A}=0$ (or $\mathrm{div}\underline{A}=-\varepsilon\mu dv/dt$ which is the Lorentz gauge and is equivalent to the continuity equation) hence $\underline{E}+d\underline{A}/dt=-\mathrm{grad}(v)$, where $\underline{A}$ is the vector magnetic potential and v is the scalar electric potential. Under the Lorentz gauge from

$$\mathrm{curl}(\mathrm{curl}\underline{A})=\mathrm{grad}(\mathrm{div}\underline{A})-\mathrm{div}(\mathrm{grad}\underline{A})$$

we obtain $\mu\varepsilon d^2\underline{A}/dt^2-\Delta\underline{A}=\mu\underline{J}$

and similarly $\mu\varepsilon d^2v/dt^2-\Delta v=\rho/\varepsilon$,

the wave equations, where $\Delta=\mathrm{div}(\mathrm{grad})$. An extended version of Branin's diagram is shown in Fig (4.2). Further we can find from $\mathrm{div}(\underline{J}+d\underline{D}/dt) = \mathrm{div}(\sigma+d\varepsilon/dt)\underline{E} = (\sigma+d\varepsilon/dt)\rho/\varepsilon = 0$ that $\sigma\rho+\varepsilon d\rho/dt=0$. The Poynting vector or power flow $\underline{P}=\underline{E}\times\underline{H}$, the Poynting theorem is $\mathrm{div}\underline{P}=s\underline{B}.\underline{H}+\underline{E}.(s\underline{D}+\underline{J})$. By defining the natural set of 1-forms Maxwell's equations become

$$*d\underline{E}=\underline{b}*\underline{E}=-s\underline{B},\quad *d\underline{H}=\underline{b}*\underline{H}=\underline{J}+s\underline{D},$$

$*\underline{d}*\underline{B}=\underline{b}\underline{B}=0$, $*\underline{d}*\underline{D}=\underline{b}\underline{D}=\rho$. Maxwell's equations can be further condensed as one spinor equation.

From Stoke's theorem and Gauss' theorem

$$\oint_{dS}\underline{E}=-\iint_{S}{}^{*}s\underline{B},\quad \oint_{dS}\underline{H}=\iint_{S}{}^{*}(\underline{J}+s\underline{D}),$$

$$\oint\!\!\iint_{dV}{}^{*}\underline{B}=0,\quad \oint\!\!\iint_{dV}{}^{*}\underline{D}=\iiint_{V}{}^{*}\rho,$$

where dS is the boundary of a surface S and dV is the boundary of a volume V. These are Maxwell's equations in exterior integral form. Lynn[4] goes on to derive Maxwell's equations in space time coordinates and as tensor densities:

$$dE_y/dx-dE_x/dy=-\varepsilon_{pxy}dB^p/dt,$$

$$dH_y/dx-dH_x/dy=\varepsilon_{pxy}(J^p+dD^p/dt),$$

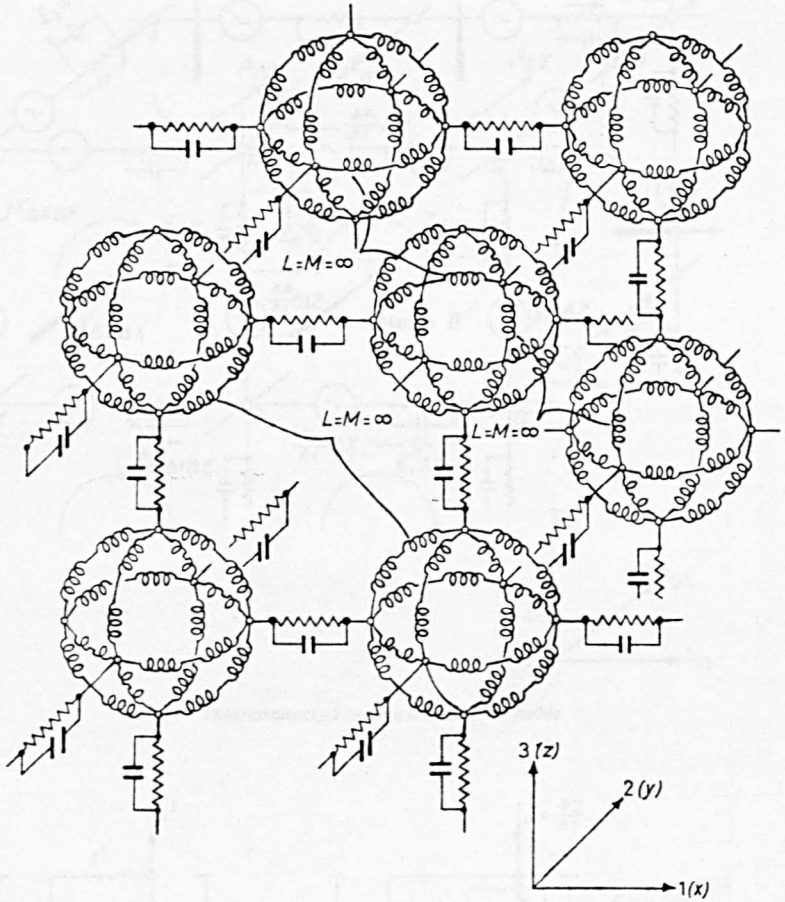
$$dB^p/dp=0 \text{ and } dD^p/dp=\rho,$$

in Cartesian coordinates where all differentials are partials and p is an arbitrary direction. Again the metric tensor appears in arbitrary coordinate systems.

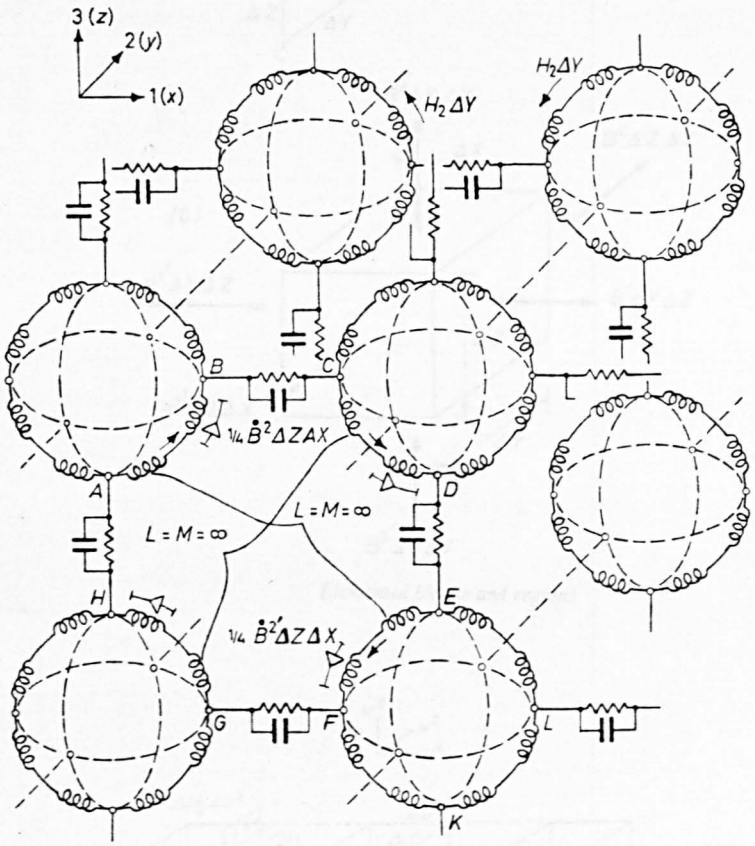
Maxwell's equations can be used to construct a network model of fields in free space. This is equivalent to the method of finite differences in which the space is discretised with a regular Cartesian mesh. The region is first subdivided and then integrations performed in the subregions presented as network quantities. When Maxwell's equations are expressed in exterior derivative form the network model will be independent of the choice of coordinates. Kirchoff's voltage and current laws for the network are interpreted by the relation $\underline{ddw}=0$ for differential forms. Stoke's theorem reduces some of the surface and volume integral's to line and surface integrals. In this form they interpret Faraday's and Ampere's laws. The model developed by Kron, see Fig(4.1.1), makes use of ideal transformers in

addition to resistors, inductors and capacitors. However it is difficult to display Maxwell's equations in an algebraic diagram based on Kron's model, therefore Lynn introduced another network model, see Fig(4.1.3), in which the integral's of the differential forms of electric field are lumped into the branches representing the edges of each block of the discretisation. Each edge of each block in the Cartesian mesh is replaced by a primitive electric circuit branch consisting of a voltage source in series with a resistor and capacitor in parallel. This in fact is a standard procedure in network analysis and design. Considering a branch parallel to the x-axis and taking the x component, or the dual dydz component out of each term in Maxwell's equations the voltage across the RC pair is $E_x dx$. The current through the resistor is $J^x dydz$ and through the capacitor is $\sigma D^x dydz$. The conductance and capacitance respectively are $\sigma dydz/dx$ and $j\omega\varepsilon dydz/dx$ using complex algebra where ω is the frequency in radians. From $\underline{d}x\underline{A}=\underline{B}$ the voltage across the ideal source is $\sigma A_x dx$ due to the mesh current $B^x dydz$. The voltage across a branch is thus $\sigma A_x dx+E_x dx$. The total impedance of the branch is $1/(\sigma+j\omega\varepsilon)$. Maxwell's equations can be reconstructed from the network in the obvious way. In a similar way the relationship between the surface and volume variables (2-network) can be developed. Branin's and Lynn's algebraic diagrams for the network field model are shown below. A Venn type diagram is shown in Fig(4.2). This should be compared with Branin's diagram for a 3-network.

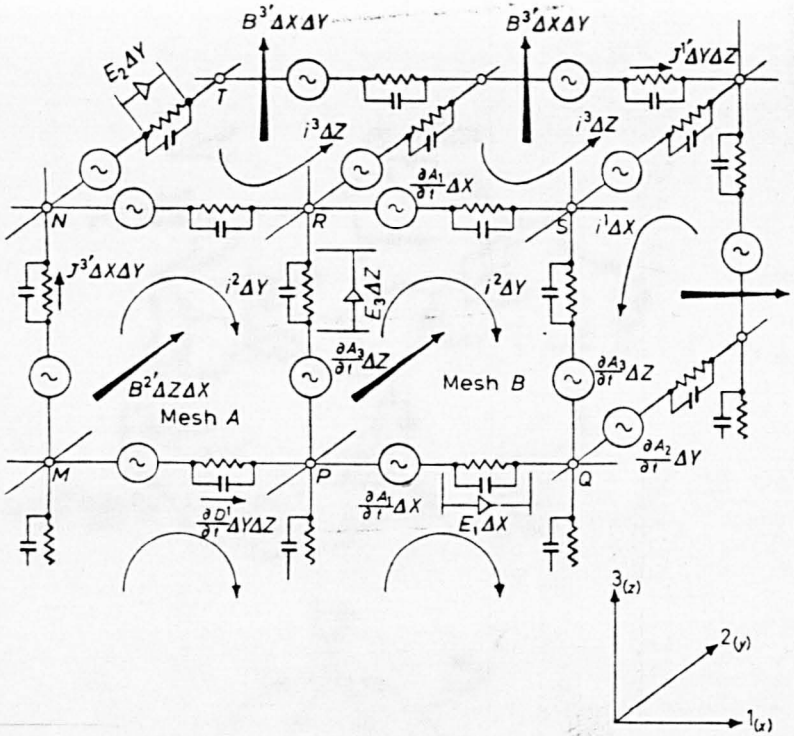
The most important question left open on the subject of partial differential equations is the insertion of the boundary conditions into Roth's diagram for the Homology Theorem which is of course essential for solving real problems!



Kron's network model

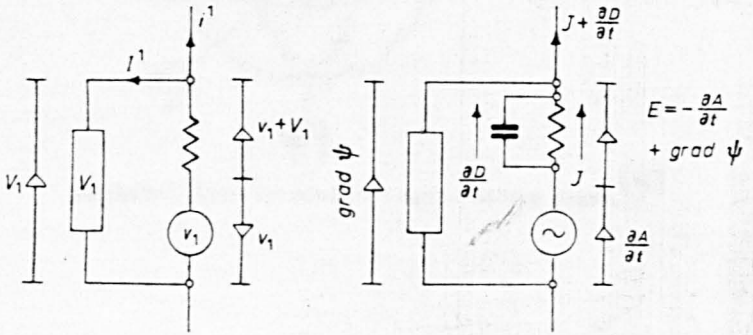


Kron's network model in x–z planes

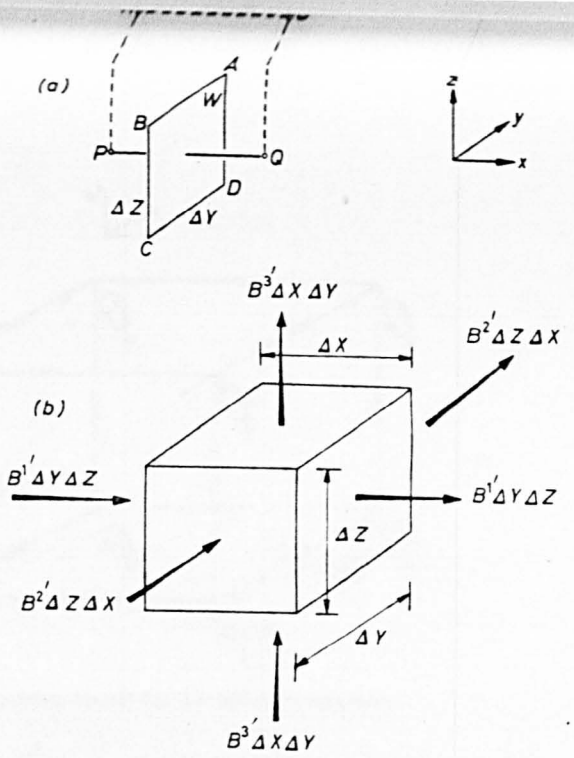


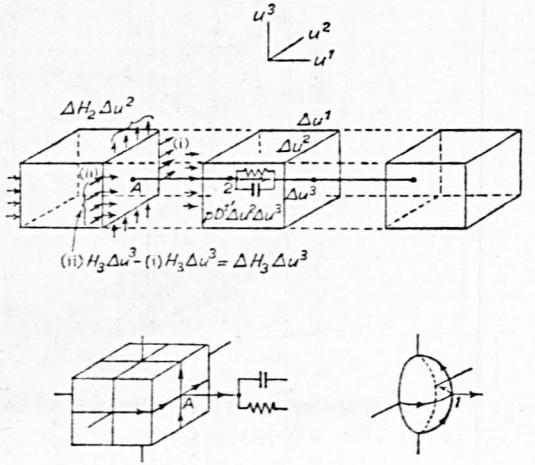
Fig(4·1·2)

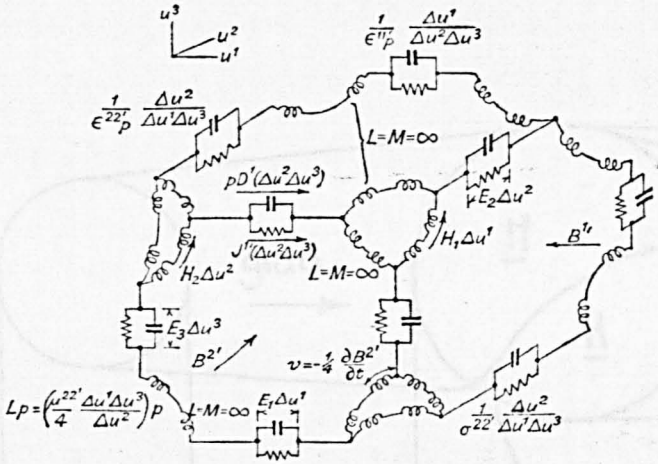
Interconnected network model for fields

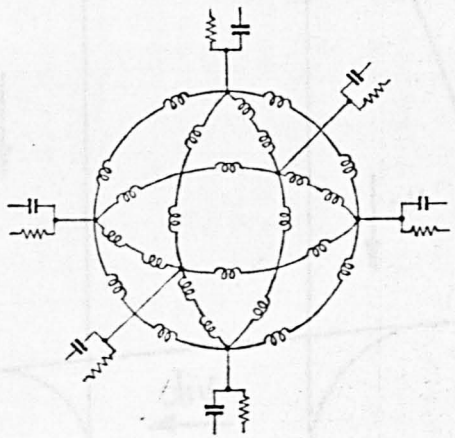


Primitive 1-network (interconnected network model for fields)



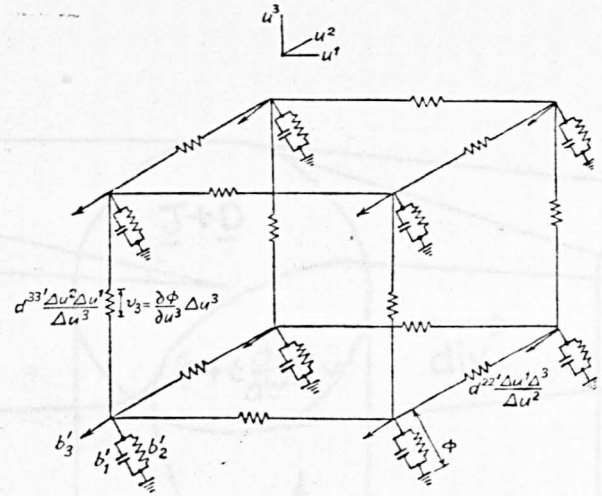




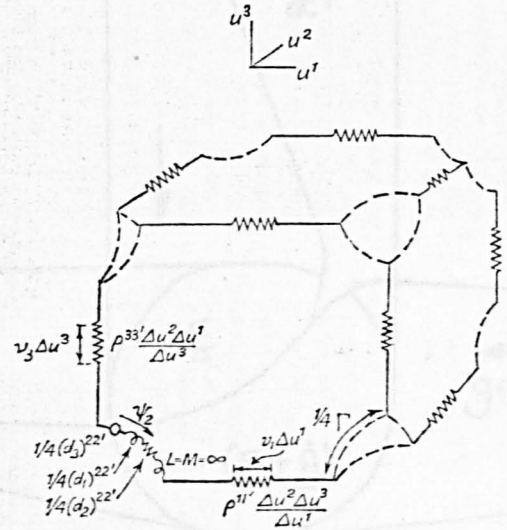


Fig(4·1·3)

Equivalent circuit for Maxwell's equations (one block). [4]



Equivalent circuit for the diffusion equation.



Equivalent circuit for solenoidal fluid flow (one block).

Fig(4·2) Extended Roth's diagram for Maxwell's equations

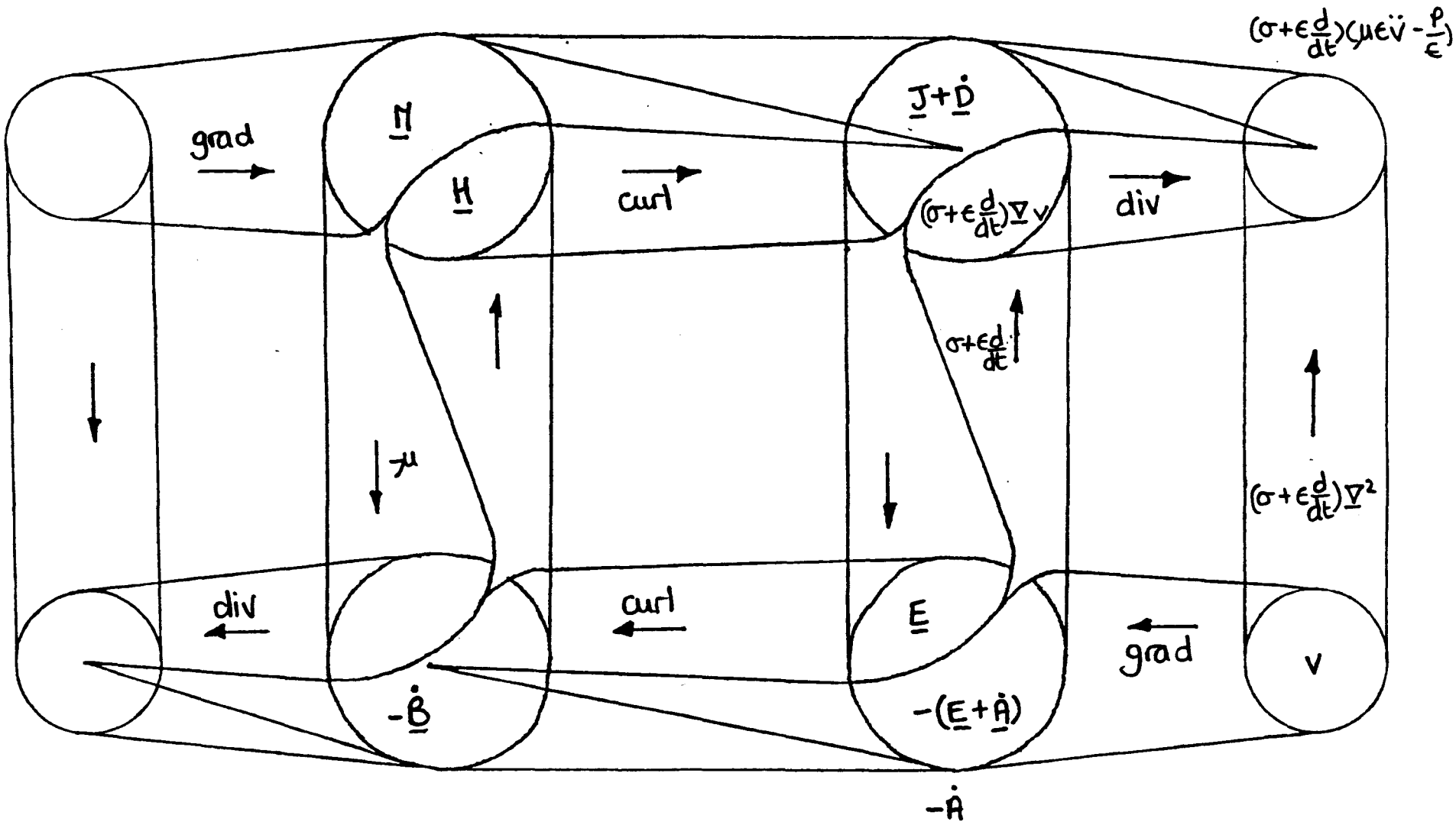


Fig (4·3) An example of the confusion surrounding duality (IEEESpectrum)

Help wanted

Duality is a useful concept in circuit theory. Using it, one may take a circuit and a true statement regarding it, perform the dual transformation on both, and obtain another true statement regarding another circuit. Circuit theory is based on Maxwell's equations. Therefore, duality must be inherent in them. One would guess that two of the equations must be the duals of the other two. However, this is not obvious from their usual formulation. I have made local inquiries concerning ways of formulating Maxwell's equations such that the duality principle is immediately obvious from them. I have met with little success. Therefore, I now call on the larger community for assistance.

The word "duality" is ambiguous. I am concerned with the principle that regards as dual pairs: mesh and node; KVL and KCL; voltage and current; series and parallel; etc. I am specifically not concerned with the principle which regards as dual pairs: B and H and E.

Any information on this equation will be appreciated.

John A. Baldwin, Jr.
Dept. of Electrical Engineering
University of California
Santa Barbara, Calif.

Maxwell or Kirchhoff?

I agree with John A. Baldwin, Jr. (Oct. 1975, p. 26) that duality is a very useful concept in circuit theory. On the contrary, it is very nebulously defined in electromagnetic theory. This perhaps can support the thesis that "circuit theory is based on Maxwell's equations" is a miconception. Rather, circuit theory is based on graph theory and Ohm's law and Kirchhoff's laws. The word "duality" in circuit theory is very unambiguous.

The parallelism between mesh and node, KVL and KCL, voltage and current, etc., is very striking. Two circuits are called dual if the corresponding graphs are dual. The purpose here is not to discuss the abstract properties of dual graphs but simply to state the result, which is: The necessary and sufficient condition for a network to have a geometrical dual is that it is a planar network.[1] Rules for finding the dual of a network, once it is determined that it has a dual, are fairly well known to circuit theorists.[2] In matrix form, the duality condition between two networks N_1 and N_2 can also be stated as $A_1 = B_2$ where A_1 is the incident matrix of N_1 and B_2 is the loop matrix of N_2. Furthermore, the number of branches of N_1 and N_2 must be equal and the rank of A_1 must equal the rank of B_2. Thus, the correspondence between the nodes of N_1 (rows of A_1) and the loops of N_2 (rows of B_2) coupled with the correspondence between the branches of N_1 and N_2 constitute the duality conditions.

The reason why Maxwell's equations cannot be formulated to reflect the principle of duality lies, in my opinion, in the fact that the "medium" of electromagnetic radiation is continuous and cannot be construed as a planar graph.

Timothy Jordanides
California State University
Long Beach, Calif.

1. Whitney, H., "Nonseparable and planar graphs," *Trans. Amer. Math. Soc.*, vol. 34, no. 2, pp. 339-362, 1932.

2 ...ling, H., *Electric Networks*. New York: Wiley, 1... pp. 198-201.

Circuit theory is not based on Maxwell's equation. It is founded on the definition of lumped elements (resistance, capacitance, etc.) and of two rules of connection (Kirchhoff's laws). If one writes this set of equations, it is directly apparent that the interchange of voltage and current yields the same set although some equations are transformed into another equation of the set. This is the only basis of duality that is a property of Kirchhoff's model for electric circuits.

The Maxwell model is totally different. For instance, it relies on the existence of a three-dimensional space that is totally ignored by Kirchhoff's model. As a result of these different axioms, one model yields partial differential equations and the other ordinary differential equations. This stresses that the two mathematical models, although applied to the same physical phenomenon, enjoy different properties. One should not expect duality to be a property of Maxwell's model because it is a property of Kirchhoff's model.

The letter raising this interesting problem points to a basic flaw in engineers' education. Much too often, there is a total confusion between the models and the physical reality. They are so well identified that one expects the same properties to belong to the one end and to the other. Then, it is quite natural to expect different models to enjoy the same property.

Jacques Neirynck
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

I would take exception to Prof. Neirynck's response, particularly the statements, "Circuit theory is not based on Maxwell's equations. It is founded on the definition of lumped elements (resistance, capacitance, etc.) and of two rules of connection (Kirchhoff's laws) . . . The Maxwell model is totally different."

All five of these lumped-constant equations are derivable from Maxwell's equations. Kirchhoff's voltage law (KVL) follows from

$$\text{curl}\left(\mathbf{E} + \frac{\partial \mathbf{A}}{\partial t}\right) = 0 \qquad (1)$$

$$V_{ab} = -\int_b^a \mathbf{E} \cdot d\mathbf{l} \qquad (2)$$

and the application of Stokes' theorem. Kirchhoff's current law (KCL) follows from

$$\text{div}\left(\mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\right) = 0 \qquad (3)$$

$$I = \int_s \mathbf{J} \cdot d\mathbf{s} \qquad (4)$$

and the application of Gauss's theorem. The relations between the voltage registered by a voltmeter connected in parallel across a circuit element (L, R, or C) and the current registered by an ammeter placed in series with it are derived from Maxwell's equations in many texts (see, for example, W. H. Hayt, Jr., *Engineering Electromagnetics*). This contradicts the statements cited above.

Equations (1) and (3) suggest the following dual pairs in Maxwell's equations.

E: Electric field	**J:** current density
A: Magnetic vector potential curl	**D:** displacement div
Stokes' theorem closed path	Gauss's theorem closed surface
B: magnetic flux density	ρ: electric charge density
λ: magnetic flux linkage	Ω: electric charge

However, there are difficulties involved with going further. For example, the equation $D = \epsilon E$ has the dual $A = ?J$, which appears to be nonphysical.

Although the existence of duality in lumped-constant electromagnetic theory does not guarantee its existence in the continuum version of the same science, there is, on the other hand, no reason to rule it out categorically.

Duality is also a property of the lumped-constant science of mechanics: mass, dashpot, spring, Newton's force, and velocity laws. However, local inquiries have come up with no evidence of duality in continuum mechanics. There are too many coincidences for duality to be an accidental property of lumped-constant sciences.

I thank those who responded to my original inquiry. However, the question remains unanswered. Briefly put, it is this. The five equations of linear lumped-constant circuit theory are

$$v_L = L\frac{di_L}{dt}$$

$$v_R = i_R R - i_R/G$$

$$i_C = C\frac{dv_C}{dt}$$

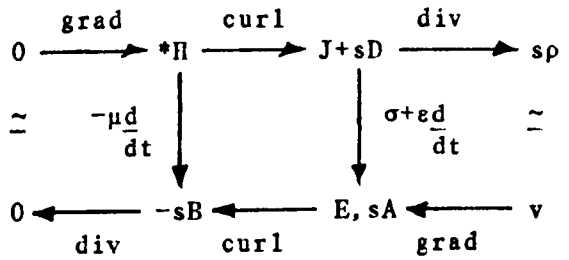
$$\sum_{\text{node}} i_n = 0 \quad \text{(KCL)}$$

$$\sum_{\text{mesh}} v_n = 0 \quad \text{(KVL)}$$

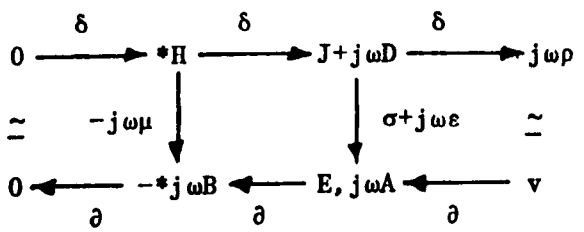
If one performs the transformation $i \leftrightarrow v$, $L \leftrightarrow C$, $R \leftrightarrow G$, KVL $\leftrightarrow$ KCL, mesh $\leftrightarrow$ node, one ends up with the same equations but in a different order. Is there a similar set of transform pairs that changes the equations of continuum electromagnetic theory into themselves?

John A. Baldwin, Jr.
University of California
Santa Barbara, Calif.

volume meshes branches nodes



Branin



Lynn

4.3 The finite difference method

An efficient method is given for solving the finite difference equations of a Laplacian field over a finite region. The area is split up into rectangular subregions and a new algorithm for solving each subregion is presented.

Lynn[4] states that the network model for the region of space studied is established by first dividing the region into subregions small enough to give the desired degree of accuracy. The subregions are blocks formed with edges dx_1, dx_2 and dx_3. In fact the limit as $dx_i \to 0$ does not have to be taken to produce an exact solution for each point. This is the method of finite differences[12] or finite elements. For example consider Laplace's equation in two dimensions for a scalar potential, $(\partial\delta+\delta\partial)a=0$, with given boundary conditions. De Rham's theorem not only justifies the finite element method but also differentiation by limits.

Laplace's equation over a region in two dimensional space is usually solved on a computer using the finite difference method. This involves covering the region with a regular Cartesian mesh. The potential of an interior point is given by $a_o=(a_N+a_S+a_E+a_W)/4$, the adjacent potentials at each point of the compass, with similar formulae for points adjacent to the boundary. The equations are exact and involve no discretisation error. They are usually solved by relaxation - an iterative Gauss-Seidel process which scans through the points in turn in some predefined order updating the current estimate of the potential at a point in terms of the values at the surrounding points. Convergence can be improved by using Carre's method of successive overrelaxation which overestimates the error by an optimal factor.

The method is quicker than the finite element method which relies on direct matrix inversion by ordinary methods, but does not result in an explicit matrix inverse hence the problem must be resolved for every new set of boundary conditions. Much more efficient for this kind of problem are methods of partitioned matrix inversion, or tearing[1]. Systems of this nature typically result in the inversion of a tridiagonal block matrix which can be solved for a particular right hand side (set of boundary conditions) by the partitioned matrix version of the Thompson algorithm - or inverted by the block staircase method. With a further level of partitioning we can prove what we hope to be an even more efficient inversion algorithm for subregions of the problem. This method has the advantages that (1) it is efficient. The conventional way to estimate the speed of an algorithm is to calculate the number of multiplications necessary to achieve a desired degree of accuracy. Assume we are working on a machine with a wordlength that is long compared to the accuracy required. Then the amount of time it takes our recursive algorithm to solve the problem to different degrees of accuracy is fixed, whereas convergent iteration takes longer and longer for higher levels of accuracy. (2) It gives an exact inverse so that we can change our boundary conditions without

resolving the problem. (3) It solves the problem in pieces so that we can actually change the shape of the boundary in one particular section without resolving any of the other subregions. This is one of the main advantages of Kron's method.

Assuming the region can reasonably be split up into a small number of rectangular subregions plus a number of nonrectangular subregions near the boundary which must be solved by direct inversion, then for any such mxn region with assumed boundary conditions (found by the usual diakoptical methods) to find the potential at each point we must invert the well known matrix

$$\begin{bmatrix} Z & I & & & & \\ I & Z & I & & & \\ & I & Z & I & & \\ & & \cdots & \cdots & & \\ & & & I & Z & I \\ & & & & I & Z \end{bmatrix} \text{ where } Z = \begin{bmatrix} -4 & 1 & & & & \\ 1 & -4 & 1 & & & \\ & 1 & -4 & 1 & & \\ & & \cdots & \cdots & & \\ & & & 1 & -4 & 1 \\ & & & & 1 & -4 \end{bmatrix}$$

where Z is an mxm matrix and the partitioned matrix has n rows and n columns. I is the unit matrix.

We now present a new algorithm for inverting this matrix by tearing the region into n columns. Assume the inverse is a partitioned matrix with n rows and n columns where each element is an mxm matrix X_{ij}. Then

$$-X_{i-1,j}+ZX_{i,j}-X_{i+1,j}=\delta_{ij}I, \quad i,j=1,\ldots,n$$

where δ_{ij} is the Kronecker delta function (0 if i=j and 1 otherwise)

and $X_{oj}=X_{n+1,j}=0$.

Define $Z_{i+1}=ZZ_i-Z_{i-1}$, a polynomial in Z, where

$Z_0=I$, $Z_{-1}=0$, $Z_{-2}=-I$ etc.

We can now write an orthogonal type equation

$$\begin{bmatrix} Z_{i+1} \\ Z_i \end{bmatrix}=\begin{bmatrix} Z & -I \\ I & 0 \end{bmatrix}\begin{bmatrix} Z_i \\ Z_{i-1} \end{bmatrix}$$

In a natural way this can be written in state space form $\underline{Z}_{i+1}=\underline{\underline{Z}}\underline{Z}_i$.

We also have

$$\underline{\underline{Z}}^i=\begin{bmatrix} Z_i & -Z_{i-1} \\ Z_{i-1} & -Z_{i-2} \end{bmatrix}$$

Theorem $X_{ij}=Z_{min(i,j)-1}Z_{n-max(i,j)}/Z_n$

where $/A=A^{-1}$ the ordinary matrix inverse (this notation is preferred when implementing matrix algorithms in Algol68).

Proof Substituting in $-X_{i-1,j}+ZX_{ij}-X_{i+1,j}=\delta_{ij}$

(1) For i=j, $X_{i-1,j}=Z_{i-2}Z_{n-i}/Z_n$, $X_{ij}=Z_{i-1}Z_{n-i}/Zn$, $X_{i+1,j}=Z_{i-1}Z_{n-i-1}/Z_n$ and $\delta_{ij}=I$.

Therefore $-X_{i-1,j}+ZX_{ij}-X_{i+1,j}=$

$-Z_{i-2}Z_{n-i}/Z_n+ZZ_{i-1}Z_{n-i}/Z_n-Z_{i-1}Z_{n-i-1}/Z_n=$

$(-Z_{i-2}Z_{n-i}+(Z_{i-2}+Z_i)Z_{n-i}-Z_{i-1}Z_{n-i-1})/Z_n=$

$(Z_iZ_{n-i}-Z_{i-1}Z_{n-i-1})/Z_n=I=\delta_{ij}$. QED

(2) For $i<j$ and hence $i>j$ by symmetry $X_{i-1,j}=Z_{i-2}Z_{n-j}/Z_n$, $X_{ij}=Z_{i-1}Z_{n-j}/Z_n$, $X_{i+1,j}=Z_iZ_{n-j}/Z_n$ and $\delta_{ij}=0$.

We have $-X_{i-1,j}+ZX_{i,j}-X_{i+1,j}=$

$-Z_{i-2}Z_{n-j}/Z_n+ZZ_{i-1}Z_{n-j}/Z_n-Z_iZ_{n-j}/Z_n=0$. QED

The programming overheads for awkwardly shaped regions in the general case may prove unwieldy, but for specific problems this would seem to be an ideal method. The use of Algol68 as the programming medium would lessen the burden considerably. An attempt to solve the equations using Kron's network analysis, with electric field as the dual variable, resulted in a less efficient algorithm. This is in fact the basis of the boundary element method. An alternative approach to the analysis of distributed systems is in terms of the directional Laplace and z-transforms,

$(s_x^2+s_y^2)\phi=0$ and it is easy to show that

$(z_x-2+z_x^{-1}+z_y-2+z_y^{-1})\phi=0$ as before.

A similar approach can be taken in cylindrical coordinates.

4.4 References

(1) Misner, Thorne and Wheeler, Gravitation, Freeman, San Francisco, 1973.

(2) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresman, 1971.

(3) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(4) Lynn, Balasubramanian and Sen Gupta, Differential forms on electromagnetic networks, Butterworth, 1970.

(5) Mizoo, Iri and Kondo, On the Torsion and Linkage Characteristics and the Duality of Electric, Magnetic and Dielectric Networks, RAAG Memoirs, Vol 2, 1958.

(6) Nicholson, System concepts in Kron's polyhedron model and the scattering problem, Int. J. Control, Vol 20, No 4, 1974.

(7) Roth, An application of algebraic topology: Kron's method of tearing, Q. Applied Math, April 1959.

(8) Branin, The Algebraic-Topological Basis for Network Analogies and the Vector Calculus, Symposium on Generalised Networks, Poly. Inst. of Brooklyn, April 1966.

(9) Bramellar, Practical Diakoptics for Electrical Engineers, Chapman and Hall, 1969.

(10) Isodori et al, Nonlinear decoupling via Feedback: A Differential Geometric Approach, IEEE Trans AC-26 No 2, 1981.

(11) Orner et al, Least Squares Simulation of Distributed Systems, IEEE Trans AC-20, No 1, Feb 1975.

(12) Bowden, Laplacian Arrays: Part I, A direct solution of the discretised form of Laplace's equation, Matrix and Tensor Quarterly, March 1982.

(13) Kron, Diakoptics: The piecewise solution of large scale systems, McDonald, 1963.

(14) Bowden, Computer aided design of pole pieces for an X-band magnetron, Dept. of Electrical Engineering, Sheffield University, 1974.

(15) Bowden, Matrix Manipulation in Fortran, Basic, Pascal and Algol68, NELP Computer Centre Document, No 448.

(16) Plonsey and Collins, Principles and Applications of Electromagnetic Fields, McGraw-Hill, 1961.

(17) Davies and Radley, Electromagnetic Theory, Vols I and II, Oliver and Boyd, 1969 and 1972.

(18) Chern, Some new viewpoints on Differential Geometry in the Large, Bulletin of the American Mathematical Society, Vol 52, pp 1-30, 1946.

CHAPTER V Physical Structure in General Systems

Tensor space is introduced and a comparison of the use of tensors in the small (ie vector algebra) with the exterior algebra is made. Contra and covariance in general systems is discussed and leads to the definition of tensors in the large (ie matrix algebra). Our discussion of scattering theory is continued.

5.1 Tensors-in-the-small

The tensor product of two finite real vector spaces is the quotient space $JoK=(JxK)\backslash(J.K)$ where the inner product space $J.K$ is the subspace of the Cartesian product JxK generated by elements of the form $(j_1+j_2,k)-(j_1,k)-(j_2,k)$ and $(j,ak)-a(j,k)$ etc $a\in R$, $j\in J$, $k\in K$, that is $(j_1+j_2)ok=j_1ok+j_2ok$ and $a(jok)=joak$ etc. Let J^* be the dual space consisting of all real linear functions on J then the tensor space J_{rs} associated with J, r times covariant and s times contravariant is

$$\underbrace{J \text{ o } \dots \text{ o } J}_{\text{r-copies}} \text{ o } \underbrace{J^* \text{ o } \dots \text{ o } J^*}_{\text{s-copies}}.$$

The direct sum $T(J)=\oint_{r,s} J_{rs}$, where $J_{oo}=R$ is the tensor algebra of J. It is a noncommutative, associative, graded algebra under o-multiplication. Let $C(J)$ be the subalgebra $\oint_p J_{po}$ of $T(J)$ and $I(J)$ be the two sided ideal in $C(J)$ generated by the set of elements of the form joj for $j\in J$ and set $I_pJ=I(J) \cap J_{po}$. It follows that $I(J)=\oint_p I_pJ$ and is a graded ideal in $C(J)$. The exterior algebra $V(J)$ of J is the graded algebra $C(J)\backslash I(J)$. If we set $V^oJ=R$, $V^1J=J$ and $V^pJ=J_{po}\backslash I_pJ$, $k\geq 2$ as before then $V(J)=\oint_p V^pJ$. In particular the residue class containing $j_1o\dots oj_p$ is $v_1\wedge\dots\wedge v_p$. These definitions should be compared with those in Section(4.1), describing the space of differential p-forms.

We now look at the exterior derivative, product and the Hodge star operator in tensor notation. A differential p-form, in terms of a skewsymmetric tensor of rank p is:

$$\underline{w}=\frac{1}{p!}A_{ab\ldots p}dx^a\uparrow dx^b\uparrow\ldots\uparrow dx^p$$

where we use Einstein's summation convention, ie. summation is assumed over repeated (dummy, umbral or running) contra-covariant pairs of indices on the same side of an equality. If the tensor $A_{ab\ldots p}$ transforms between coordinate systems using the same transformation tensor as the base vectors it is called covariant (subscripts). If the transformation involves the inverse of the base transformation tensor it is called contravariant (superscripts).

The exterior product of two 1-forms

$$\underline{w}=A_a dx^a \qquad \underline{v}=B_b dx^b$$

is

$$\underline{w}\uparrow\underline{v}=\frac{1}{2}(A_aB_b-A_bB_a)dx^a\uparrow dx^b$$

and the skew-symmetric tensor $A_aB_b-A_bB_a$ is thus seen to give the exterior product. This can be generalised to higher forms by use of the generalised Kronecker delta. The simplest form is well known:

$$\delta^i_a=0,\ i\#a \text{ and } =1 \text{ otherwise}$$

The general form is given by the determinant of a matrix whose ijth element is δ_{ij}. For example

$$\delta^{ij}_{ab}=\begin{vmatrix}\delta^i_a & \delta^i_b\\ \delta^j_a & \delta^j_b\end{vmatrix} \qquad \delta^{ijk}_{abc}=\begin{vmatrix}\delta^i_a & \delta^i_b & \delta^i_c\\ \delta^j_a & \delta^j_b & \delta^j_c\\ \delta^k_a & \delta^k_b & \delta^k_c\end{vmatrix}$$

$$=(\delta^i_a\delta^j_b-\delta^i_b\delta^j_a)$$

=1 (-1) if the {ij...} are distinct and constitute an even (odd) permutation of the {ab...} and =0 otherwise and we can write

$$(A_aB_b-A_bB_a)=\delta^{ij}_{ab}A_iB_j$$

In general the exterior product of a p-form $\underline{w}$ and a q-form $\underline{v}$ is given by

$$H_{abc...r}=\frac{1}{p!q!}\,d^{mn...pq...s}_{ab...r}\,W_{mn...p}\,V_{q...s}$$

The inner product of two p-forms in terms of coefficients V and W is

$$\frac{1}{p!}V_{ab...p}W^{ab...p}$$

The repeated indices sum the components of V and W. This process is known as contraction in tensor terminology. The skew-symmetric tensor W is expressed in terms of contravariant components. This in fact is a special case of contraction in which the tensor is not merely reduced in rank but in fact results in a scalar.

A special skew-symmetric tensor ε is now introduced whose terms are 0, 1 and -1. In n dimensions the Levi-Civita tensor

$$\varepsilon^{ab...p}=1,\ -1 \text{ or } 0$$

depending on whether (1) an even permutation of a,b,...,p will restore the sequence 1,2,...,n or (2) an odd permutation will restore it or (3) any index is repeated, respectively. The tensor ε is useful for dealing with determinants and dual tensors. If a tensor W of rank p has V as its dual tensor then

$$V^{ij...k} = \frac{1}{p!} W_{ab...c} \varepsilon^{ab...cij...k}$$

This agrees with the definition of the Hodge star operator. If non Cartesian coordinates are used a more general form involving the metric tensor must be used.

The generalised Kronecker delta can be used to find the tensor form of the exterior derivative. If $W_{bc...e}$ is a skewsymmetric tensor of rank p, the generalised curl is

$$\frac{1}{p!} \delta^{abc...e}_{mn...q} \partial W_{bc...e} / \partial x^a$$

The divergence is obtained by taking the dual of the curl of the dual of a tensor.

It is seen that the maze of indices often makes it difficult to observe the differences between different types of quantities. Flanders[13] compares tensor analysis with the exterior algebra and concludes that each has its own natural area of application particularly with regard to symmetry. Tensor analysis only consists of techniques for handling indexed quantities and lacks an established substantial body of deep results comparable to the exterior calculus. In tensor analysis it is often difficult to see the range of application as everything seems to work in a coordinate patch. Tensor fields do not behave themselves under mappings whereas there is always a naturally induced differential form due to a mapping on a space. It is often difficult to discover the deeper invariants in physical situations using tensors, whereas they appear to arise naturally in the exterior algebra. Tensor analysis of geometric situations restricts one to the use of natural frames

associated with a local coordinate system. The exterior algebra allows the use of Cartan's moving frames. Often a combination of techniques is in order. This is particularly illustrated in Misner and Wheeler's 'Gravitation'.

5.2 Covariance and contravariance

According to Branin[2] 'In devising network models for dynamical systems the traditional approach has been to make a term by term comparison between the different equations describing the dynamical system and those describing a cognate electrical system. Although this practice usually works, it is by no means infallible and may even fail to lead to an analogy when one actually exists.

The inherent weakness in this approach is the fact that, in comparing the equations of performance of the dynamical system, with either the mesh or the nodal equations of the electrical system, one is completely unable to recognise the differing topological character of the two types of variable involved. In other words, after the mesh or nodal equations have been compiled, no trace remains to indicate which variables sum to zero at a point and which sum to zero around a closed path. Consequently either the mass-inductance or the mass-capacitance analogy may emerge, depending on whether the mesh or nodal equations were taken as a standard of comparison.

As long as the mechanical system being modelled can be represented as a planar graph, no practical difficulty arises from using either of these two analogies. This is because of the well known theorem that any planar graph has a dual in which the roles of the two types of network variable may be interchanged. When the graph representing the mechanical system is nonplanar, however, ... the mass-inductance

analogy ... fails, since it implies that force, which sums to zero at a point, is analogous to voltage which sums to zero around a closed path. The mass-capacitance analogy, on the other hand, always applies because it is topologically consistent in making force the analogy of current and velocity the analogy of voltage.'

These ideas are inherent in traditional distributed field theory using 'tensors-in-the-small'. Kron extended them to 'tensors-in-the-large' in his network analysis. The best introduction to the subject is Franksen's excellent series of papers on the 'economic network' concept. Franksen draws an analogy between economics and engineering, working on the basis that they both make similar assumptions, constructs an electrical network analogy of Leontief's input-output analysis and solves it using quadratic programming. He assumes pure elasticity of demand and plasticity of supply and solves the problem in a similar way to Kron using the simplex method. Starting with Weyl's ideas of symmetry he introduces generalised versions of the First and Second Laws of Thermodynamics. The First Law is based on the observation, by Leibnitz in 1693, that 'In any isolated system there must be an invariant entity changeable in form but indestructable.' This is equivalent to conservation of energy in physics or Walras' law in economics (see Chapter VII). This entity may be written as the product of two quantities of the form flow and potential. The Second Law may be stated 'A flow will only occur from a higher potential to a lower potential.' This gives rise to the concept of an orientated graph in network theory and is responsible for Franksen's definition of 'negprices' or negative prices in economics in that commodity flow will only occur from a sector which has a low value for that commodity to a sector which has a high value. He also notes that analogies may be made between systems of this sort by means of homomorphisms (actually functors) ie mappings that retain the structure of organisation. Generalisations of the classical concepts of covariance and contravariance may be made in terms of transvariables (across variables) that sum to zero around a closed loop and are measured on the interval scale between two

points of a system without cutting the interconnections (one point may be chosen as reference) and intervariables (through variables) that sum to zero at a node and are measured on the ratio scale by cutting the interconnections of the system and inserting the measurement device in between the cut points. Franksen further classifies the variables of a system into the intensive and extensive reference frames and defines the economic equivalents of content, cocontent, energy and coenergy. In mechanics the statements of Kirchoff's laws are replaced by D'Alembert's principle - the well known technique of summing forces to zero at a point. The table below compares the classifications in various physical systems. Note that quadratic cost function weighting matrices and impedance tensors are doubly contravariant. Any vector in real space has dimensions $M^m T^t L^l$ where the vector is contravariant if l is positive, covariant if negative and invariant if zero. It is intensive or extensive depending on t.

	covariance	contravariance	minimise
traditional tensors i-t-1 definition	transforms with coordinate system	transforms against coordinate system	
Franksen's, Kron's definition	transvariables (across vars) $\sum$ closed loop = 0	intervariables (through vars) $\sum$ node = 0	
in general	potential	flow, flux	
tensors	subsript	superscript	
vectors	row	column	
measured on	interval scale without cutting	ratio scale with cutting	
metric space	space	dual space	metric
economics	negative prices	commodity flows	(-) money
optimal control	costates	states	cost fn
electric circuits	voltage	current	watts
electromagnetics	electric field	magnetic flux	energy
mechanics	displacement	force	ditto
rotational	rotation	torque	ditto
thermodynamics	temperature	heat transfer	calories
liquid flow	head	flow	ditto
gas flow	pressure	mass flow	ditto

5.3 Tensors-in-the-large

Kron[1] was insistent that the objects he was dealing with were actually 'tensors-in-the-large', with all the properties of conventional tensors except those associated with a coordinate system, and not just matrices which are simply arrays of numbers. This was (1) because the variables they transformed possessed properties of co- and contravariance and (2) by analogy with electromagnetic theory (tensors-in-the-small). For this reason he (followed by many others) used matrices with indices representing their dimensions. There is little loss in simplicity in using tensor notation in control and systems theory - one should at least always be aware of it in a similar way to the systems approach to engineering currently popular - and there are a number of advantages. The Einstein summation convention comes to the rescue as far as matrix multiplication is concerned. Tensors with ranks higher than 2 can be used when representing multivariable or multilevel systems. Multiple indices can be handled just as easily in most programming languages - and the summation convention could even be implemented in Algol 68! The structure of tensor equations is much richer than that of matrix equations. More nonlinear equations can be represented (though this could be considered a disadvantage when formulating a model) and the positions of the indices point to contravariant (superscript) or covariant (subsript) variables giving a closer representation of physical reality than matrices. Tensors are commutative under multiplication. Further outer products (and sums), similar to the Kronecker product of matrix notation, are directly represented. These methods make way for more efficient computational algorithms. Thus we have both an improved and a more general philosophy. In practice tensors are only normally used in field theory and at the level of abstract mathematics.

Various mathematical structures are available for the study of physical systems. The operations of addition or multiplication in a particular field structure can can be used as a means of classification. If associative the object is a semigroup. If also there exists an inverse, an identity element and the operator commutes we have an abelian group. If both operations exist and are distributive the whole structure is called a ring. The classifications along with their associated operations are called categories. Transformations between categories are called functors eg the 'forgetful functor' always takes a category to a more simple one by omitting part of its structure. This is a concept often used intrinsically in proofs eg fixed point theorems. Inevitably the generality of theory and the specificness of results have an inverse relationship depending on the mathematical structure used to define the system. According to Vanacek 'Let us refer to one area in which the introduction of richer structures has proved useful. The Maxwell theory of electromagnetic field in vacuum can be described either by 24 scalar equations or by 8 vector equations or by 3 tensor equations or by one spinor equation. (It is to be understood that this is not just through the trivial direct sum.) The mentioned reduction at the same time contributed to the knowledge and the deepening of the Maxwell theory.'

A tensor (in-the-small) A, of rank 1, is defined as a column or row vector in R^n and as contravariant or covariant according to whether its components transform under a change in coordinate system, with i as free index, as

$$A^i=A^j\frac{dx^i}{dy^j} \text{ or } A_i=A_j\frac{dx^j}{dy^i}$$

respectively, with summation assumed from 1 to n.

A tensor (in-the-large)

$$A^{i_1,i_2,\ldots,i_n}_{j_1,j_2,\ldots,j_m} \quad \text{or} \quad A_{i_1,i_2,\ldots,i_n;j_1,j_2,\ldots,j_m}$$

of rank n+m is said to be n times contravariant and m times covariant and exists in n+m dual coupled linear subspaces written as

$$R^{i_1} x R^{i_2} x \ldots x R^{i_n} x R^*_{j_1} x R^*_{j_2} x \ldots x R^*_{j_m}$$

where some of the R_j are dual to some of the R^i. In a problem where n#m, or some of the spaces are not dual, then we simply have not included all the variables in the real problem (orthogonal formulation). The tensor can be visualised as an n+m dimensional array of elements and written as a $\sum i_k$ by $\sum j_k$ partitioned matrix where the covariant indices are allowed to vary across the matrix and the contravariant ones down. Normal addition adds corresponding elements in tensors of the same rank and dimensions. Inner (matrix) multiplication occurs when contra-covariant pairs of indices from different elements in an outer product are equal with summation by contraction over those pairs. Inverses are defined in terms of the unit tensor. Scalar products are allowed. When inverting a square tensor (rank 2) the appropriate indices are raised or lowered, ie dy^{-1} above is covariant. Normally we make no distinction between a tensor x and its component x^j. Contravariant and covariant indices should always balance across an equation (in the same way as powers of MLT in physics units). The Kronecker delta (unit tensor) is a special case of the generalised unit tensor

$$\Delta^{i_1 i_2 \ldots i_n}_{j_1 j_2 \ldots j_n} = \delta^{i_1}_{j_1} \delta^{i_2}_{j_2} \ldots \delta^{i_n}_{j_n}$$

=1 if every contra-covariant pair of indices are equal and =0 otherwise, as well as the generalised Kronecker delta.

The transposes of

$$A^{i_1,i_2,\ldots,i_n}_{j_1,j_2,\ldots,j_m} \text{ are } A^{perm_1\{i_1,i_2,\ldots,i_n\}}_{perm_2\{j_1,j_2,\ldots,j_m\}}$$

ie any permutations of the indices. Symmetric and skewsymmetric parts of the tensor are respectively

$$A_{(i_1\ldots i_p)}=\sum_{perm} A_{perm\{i_1\ldots i_p\}}/p! \text{ and}$$

$$A_{[i_1\ldots i_p]}=\sum_{perm} A_{perm\{i_1\ldots i_p\}}(-1)^P/p!$$

where $P=\varepsilon_{perm\{i_1\ldots i_p\}}$

The dual of a tensor

$$*A^{i_1 i_2\ldots i_p}{}_{i_{(p+1)}\ldots i_n}=\varepsilon_{i_1 i_2\ldots i_n} A^{i_1 i_2\ldots i_p}/p!$$

The exterior product (basis of the Grassman algebra) is the skewsymmetric part of the outer product, for instance the vector cross product is related to the exterior product by the Hodge star operator, $*:\underline{x}\underline{x}\underline{y}\rightarrow\underline{x}\uparrow\underline{y}$

$$\underline{x}\underline{x}\underline{y}=\begin{bmatrix} x_j y_k - x_k y_j \\ x_k y_i - x_i y_k \\ x_i y_j - x_j y_i \end{bmatrix},\quad 2\underline{x}\uparrow\underline{y}=\underline{x}\underline{y}'-\underline{y}\underline{x}'=\begin{bmatrix} 0 & x_j y_i - x_i y_j & x_k y_i - x_i y_k \\ x_i y_j - x_j y_i & 0 & x_k y_j - x_j y_k \\ x_i y_k - x_k y_i & x_j y_k - x_k y_j & 0 \end{bmatrix}$$

Examples[8] (1) The Luenberger canonical form for systems of the form $A(s)x=Bu$ can be written in tensor notation

$$A=E_{m1}\times A_1+\sum_{j=2}^{m}((E_{j-1,j}\times I_n)+(E_{mj}\times A_j))$$

where $(E_{ij})_{gh}=I$ (if $i=g$ and $j=h$)

$=0$ (otherwise),

and $A(s)=A_1+A_2 s+A_3 s^2+\ldots-s^m$

and the A_i are nxn matrices and we can write $s\underline{x}=\underline{A}\underline{x}+\underline{B}\underline{u}$. This will be seen in the next chapter. Use may be made of the Kronecker delta to represent such systems.

(2) Multilevel or decomposed systems may be represented as higher order tensors. Vanacek[7] has derived new higher order Riccati equations for estimation and control in this way, using adjoints in tensor space. These problems may be represented within Roth's diagram. Vanacek concludes by saying 'we are convinced that (the LQG problem) ... has not been closed yet. This is because of not fully digged algebraic structure (!)'

(3) The Lyapunov stability matrix equation can be written

$$A^i_{.k}Q_{il}+A^j_{.l}Q_{kj}=(A^i_{.k}\delta^j_{.l}+\delta^i_{.k}A^j_{.l})Q_{ij}=-\delta_{kl}$$

Now if Q_{ij} is symmetrical we have

$$(A^i_k\delta^j_l+\delta^i_kA^j_l+(A^i_l\delta^j_k+\delta^i_lA^j_k)(1-\Delta^{ij}_{ji}))Q_{ij}=-\delta_{kl}$$

and Q_{ij} can be found from inversion. This algorithm can be programmed directly with the help of a delta function, with less effort than say Macfarlane's algorithm.

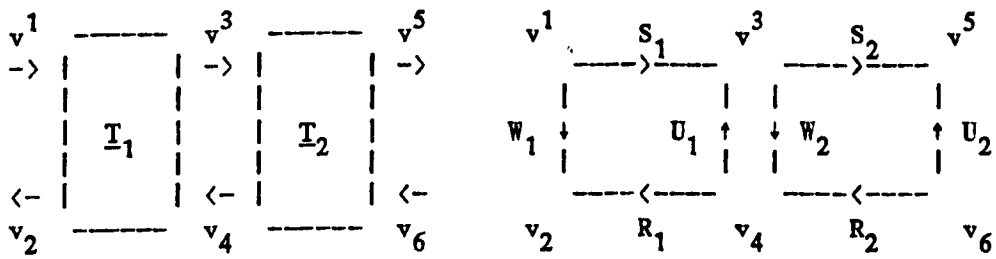
5.4 Scattering theory

Paraphrasing Nicholson[14]: 'The interconnection of subsystems in many physical and socioeconomic systems introduces effects of feedback with each unit being influenced by its neighbours. The general scattering problem is concerned with the introduction of such obstacles into, or distributed parameters in, flow processes in which disturbances are reflected, transmitted or absorbed at a subsystem or obstacle boundary. In a serial flow process the interaction between coupled stages

is defined in terms of a combined scattering matrix.' Nicholson has used this as an analytical basis upon which Kron's polyhedron model could be investigated. With two adjacent obstacles the reflected and incident waves are represented by

$$\begin{bmatrix} v^3 \\ v_2 \end{bmatrix} = \underline{T}_1 \begin{bmatrix} v^1 \\ v_4 \end{bmatrix} \text{ and } \begin{bmatrix} v^5 \\ v_4 \end{bmatrix} = \underline{T}_2 \begin{bmatrix} v^3 \\ v_6 \end{bmatrix} \text{ where } \underline{T}_i = \begin{bmatrix} S_i & U_i \\ W_i & R_i \end{bmatrix}$$

The cascaded process may also be represented by a transformation or signal flow diagrams



then for the combined structure

$$\begin{bmatrix} v^5 \\ v_2 \end{bmatrix} = \underline{T}_1 * \underline{T}_2 \begin{bmatrix} v^1 \\ v_6 \end{bmatrix} = \begin{bmatrix} S_2/(I-U_1W_2)S_1 & U_2+S_2U_1/(I-W_2U_1)R_2 \\ W_1+R_1W_2/(I-U_1W_2)S_1 & R_1/(I-W_2U_1)R_2 \end{bmatrix} \begin{bmatrix} v^1 \\ v_6 \end{bmatrix}$$

where $\underline{T}_1 * \underline{T}_2$ defines the star or scattering product of two partitioned matrices. The star product has a higher priority than normal matrix multiplication. It is associative and distributive but not commutative, takes the ordinary matrix inverse and obeys the transposed conjugate relationship.

Nicholson has represented the internal structure of the combined scattering matrix in Fig (5.1) and as an interconnected lattice type structure Fig (5.2) having much in common with the double helix model for the DNA molecule due to Watson. Half of this molecule serves as a template for the reproduction of further molecules and also for the production of RNA and various enzymes. It is worth noting the growing interest in the AI field in the use of the DNA molecule as a model for the production of intelligent programs. There is recent interest in representing recursive or ladder algorithms as lattice structures [9]. Nicholson has further investigated scattering in distributed media, electrical networks, invariant

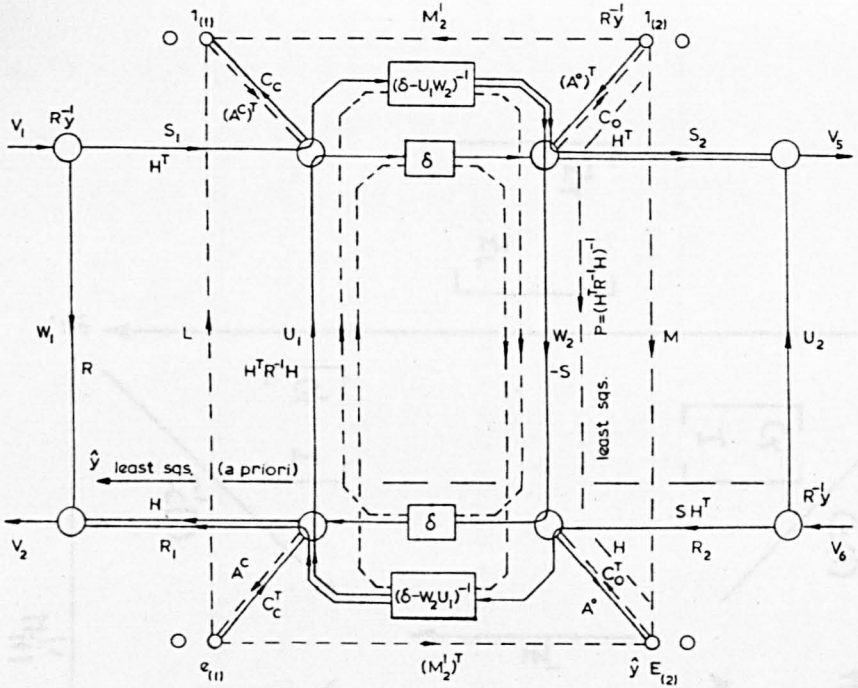


Figure 5.1 Internal structure of combined scattering matrix. [24]

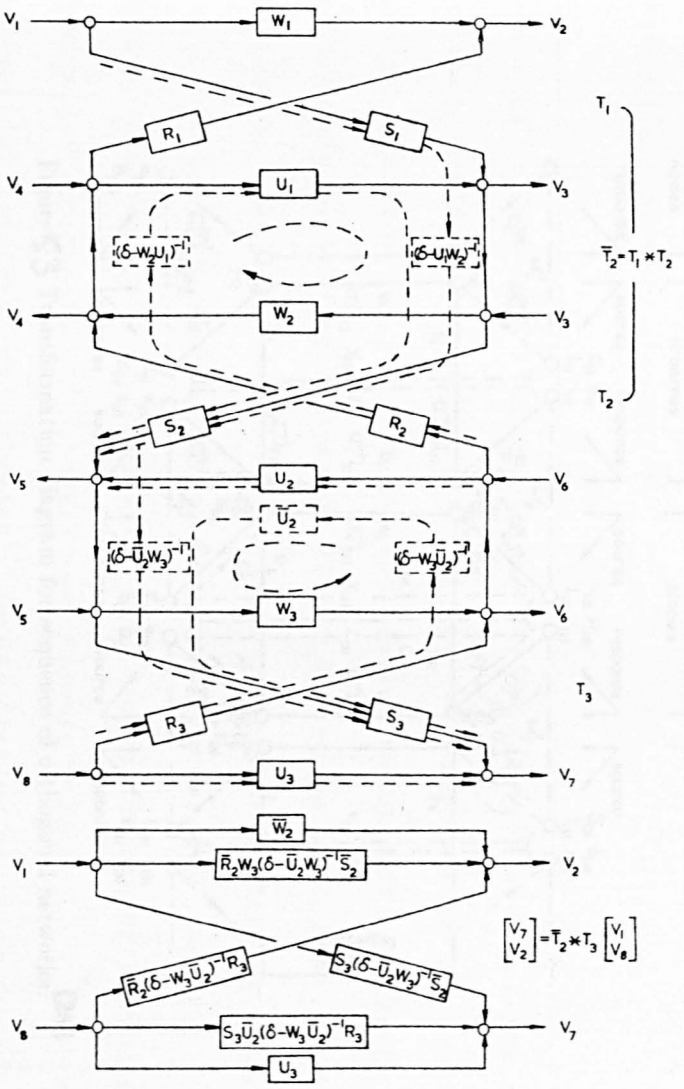


Figure 5.2 Interconnected lattice-type sequence. [24]

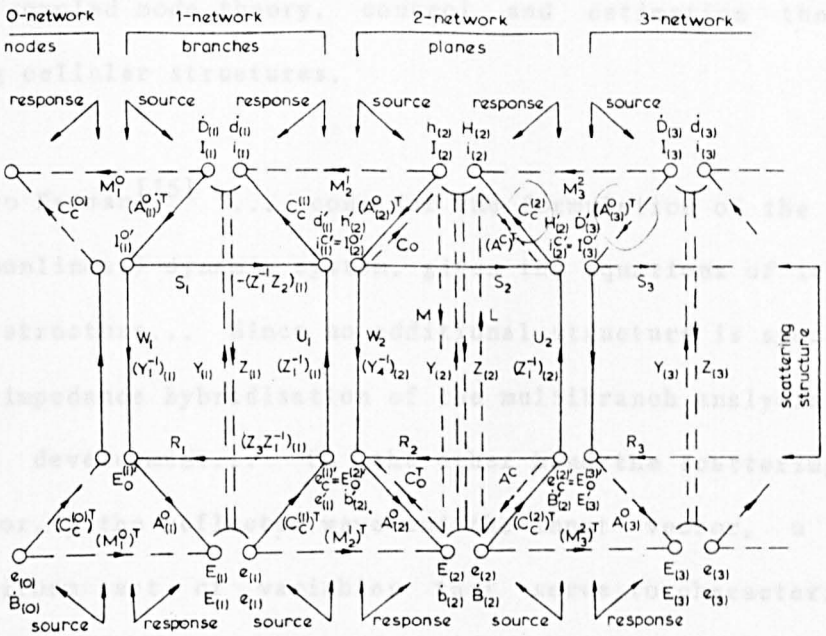
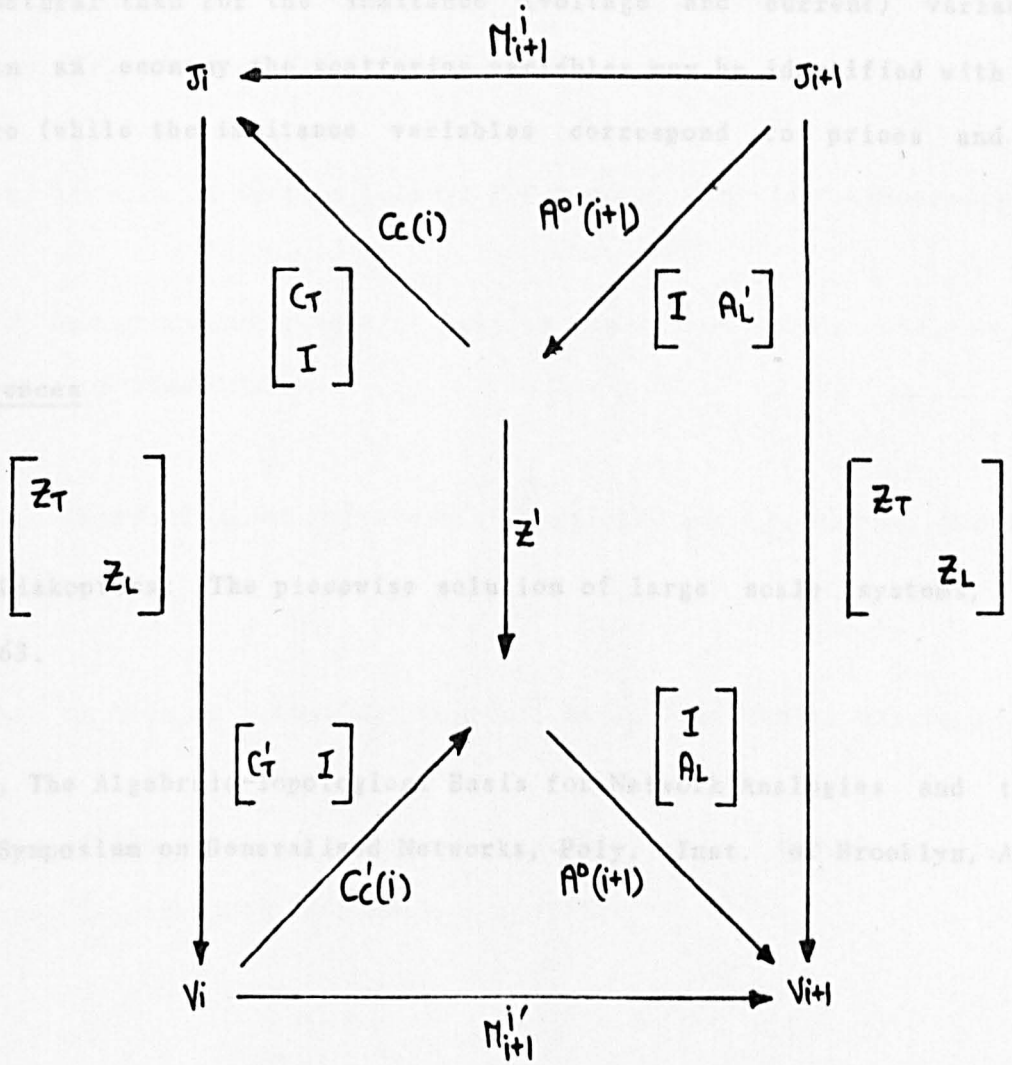


Figure 5.3 Transformation diagram for sequence of orthogonal networks. [24]



embedding, coupled mode theory, control and estimation theory and generalised interacting cellular structures.

According to Kalman[15] '... consider the formulation of the state equations for a (possibly nonlinear) dynamic system, given the equations of its components and the connection structure... Since no additional structure is specified, the convenient admittance-impedance hybridisation of the multibranch analysis is not applicable to the present development... On the other hand the scattering variables, with the output vector, y the reflected wave and the input vector, u the incident wave, yield a uniform set of variables that serve to characterise the most commonly encountered components of network and system theory. In fact, in many nonelectrical systems the physical interpretation of the scattering variables is far more natural than for the immitance (voltage and current) variables. For instance in an economy the scattering variables may be identified with income and expenditure (while the immitance variables correspond to prices and commodity flows).'

5.5 References

(1) Kron, Diakoptics: The piecewise solution of large scale systems, MacDonald, London, 1963.

(2) Branin, The Algebraic-Topological Basis for Network Analogies and the Vector Calculus, Symposium on Generalised Networks, Poly. Inst. of Brooklyn, April 1966.

(3) Spillers, On Diakoptics: Tearing an arbitrary system, (unknown), Vol XXIII, No 2, 1965.

(4) Pereya and Scherer, Efficient computer manipulation of tensor products with applications to multidimensional approximation, Mathematics of Computation, Vol 27, No 123, July 1973.

(5) Powers, Solving AX+XB=C by Control and Tearing, (unknown).

(6) Wonham, Linear Multivariable Control: A Geometric Approach, Berlin, Springer Verlag, 1974.

(7) Vanacek, LQG Problem of Estimation and Control: Evolution in the discrete, finite time, Kybernetica, Vol 11 (1975) No 4.

(8) Bowden, Tensors in Systems Theory, PhD Report, Sheffield University, 1977.

(9) Shensa, Recursive Least Squares Lattice Algorithms - A Geometric Approach, IEEE Trans AC-26 No 3 June 1981.

(10) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresmann, 1971.

(11) Bowden, On Inverse Suboptimal Control, Matrix and Tensor Quarterly, Sept 1981.

(12) Journal of the Franklin Institute, Special Issue: Modern Aspects of Large Scale Systems Science, Vol 286, No 6, Dec 1968.

(13) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(14) Nicholson, Structure of Interconnected Systems, IEE Control Engineering Monograph 5, Peter Perigrinus, 1978.

(15) Kalman, Aspects of Network and Systems Theory, (unknown), 19 .

CHAPTER VI Optimal Control Theory

This chapter applies Kron's ideas to optimal control theory. It takes much from the work of Nicholson but, rather than concentrating on scattering theory, uses Roth's diagram to quickly get results and to show the analogies between different problems, and different forms of the same problem. A sequence of mappings associated with multistage optimal control is introduced and shown to constitute a chain complex, thus for the first time giving the direct connection with homology theory. Wonham's geometric theory is closely paralleled, but from a far more structural point of view. According to Wonham[9] '...the geometry was first brought in out of revulsion against the orgy of matrix multiplication which linear control theory mainly consisted of, not so long ago. But secondly and of greater interest, the geometric setting rather quickly suggested new methods of attacking synthesis which have proved to be intuitive and economical: they are also easily reduced to matrix arithmetic as soon as you want to compute.'

6.1 The partitioned system

Consider the multistage linear optimal control problem, in a similar form to that dealt with by Nicholson[1].

$$\min \sum_{t=1}^{n} y(t)'Q(t)y(t)+u(t-1)'R(t-1)u(t-1)$$

where

$$zx(t)=x(t+1)=A(t)x(t)+B(t)u(t),\ t=0,\ldots,n-1$$

with

$$y(t)=C(t)x(t) \text{ and } x(o)=x_o.$$

This can be rewritten as an ordinary least squares problem as follows:

$$\min_{\underline{u}} \begin{bmatrix} x(1) \\ x(2) \\ \dots \\ x(n) \end{bmatrix}' \begin{bmatrix} C(1)'Q(1)C(1) & & & \\ & C(2)'Q(2)C(2) & & \\ & & \dots & \\ & & & C(n)'Q(n)C(n) \end{bmatrix} \begin{bmatrix} x(1) \\ x(2) \\ \dots \\ x(n) \end{bmatrix} + \begin{bmatrix} u(o) \\ u(1) \\ \dots \\ u(n-1) \end{bmatrix}' \begin{bmatrix} R(o) & & & \\ & R(1) & & \\ & & \dots & \\ & & & R(n-1) \end{bmatrix} \begin{bmatrix} u(o) \\ u(1) \\ \dots \\ u(n-1) \end{bmatrix}$$

where

$$\begin{bmatrix} I & & & \\ -A(1) & I & & \\ & \dots & & \\ & & -A(n-1) & I \end{bmatrix} \begin{bmatrix} x(1) \\ x(2) \\ \dots \\ x(n) \end{bmatrix} = \begin{bmatrix} A(o)x(o) \\ 0 \\ \dots \\ 0 \end{bmatrix} + \begin{bmatrix} B(o) & & & \\ & B(1) & & \\ & & \dots & \\ & & & B(n-1) \end{bmatrix} \begin{bmatrix} u(o) \\ u(1) \\ \dots \\ u(n-1) \end{bmatrix}$$

or

$$\min_{\underline{u}} \underline{x}'\underline{Q}\underline{x}+\underline{u}'\underline{R}\underline{u}$$

where

$\underline{A}\underline{x}=\underline{x}_o+\underline{B}\underline{u}$ given $\underline{x}_o$.

Substituting for $\underline{x}$

$$\frac{d}{d\underline{u}'}((\underline{x}_o-\underline{B}\underline{u})'/\underline{A}'\underline{Q}/\underline{A}(\underline{x}_o-\underline{B}\underline{u})+\underline{u}'\underline{R}\underline{u})=0$$

and it is easy to show that the optimal control

$$\underline{u}^o=/(\underline{B}'/\underline{A}'\underline{Q}/\underline{A}\underline{B}+\underline{R})\underline{B}'/\underline{A}'\underline{Q}/\underline{A}\underline{x}_o$$

$$=(/\underline{R}-/\underline{R}\underline{B}'/(\underline{A}/\underline{Q}\underline{A}'+\underline{B}/\underline{R}\underline{B}')\underline{B}/\underline{R})\underline{B}'/\underline{A}'\underline{Q}/\underline{A}\underline{x}_o$$

using Householder's formula. We have to invert a matrix of the form

$$\begin{bmatrix} /Q & -/QA' & & \\ -A/Q & A/QA'+/Q & -/QA' & \\ & \dots & \dots & \dots \\ & & -A/Q & -A/QA'+/Q \end{bmatrix} + \begin{bmatrix} B/RB' & & & \\ & B/RB' & & \\ & & \dots & \\ & & & B/RB' \end{bmatrix}$$

This is the block tridiagonal matrix inversion problem of multistage linear optimal control. The finite differences problem to which a direct solution was obtained at the end of Chapter IV is a special case.

6.2 Practical aspects

We have the following possible methods of solution

(1) Direct inversion. Does not make use of the special structure of the matrix. Very slow and inefficient. Similar methods such as LU decomposition used by many commercial packages.

(2) Sparse matrix techniques. Better, but using brute force techniques on structured matrices is never going to be very efficient.

(3) The Thomas recursive matrix algorithm given in Chapter II. Needs to be resolved for every new set of initial conditions.

(4) The discrete Riccati equation. This is virtually the same as (3). The usual problems mentioned are that the $\underline{P}$ matrix must either be stored at every reverse step or recalculated at every forward step, and must simultaneously be kept symmetrical. Infinite time optimisation of a time invariant system leads to the algebraic Riccati equation and a constant $\underline{P}$ matrix.

(5) Staircase inversion (using Schur's lemma). This is similar to (3) and (4) but actually produces the inverse.

(6) Explicit solution along the lines covered in Section (4.3). This appears to be feasible for the time invariant case. Define the sequence Z(i) such that

$$YZ(i-1)+ZZ(i)+Y'Z(i+1)=0$$

where $Y=B/RB'-A/Q$

and $Z=B/RB'+A/QA'+/Q$

so that

$$\begin{bmatrix} Z(i+1) \\ Z(i) \end{bmatrix}=\begin{bmatrix} -/Y'Z & -/Y'Y \\ I & 0 \end{bmatrix}\begin{bmatrix} Z(i) \\ Z(i-1) \end{bmatrix}$$

or

$$\underline{Z}(i+1)=\underline{\underline{Z}}\underline{Z}(i)$$

and

$$\underline{Z}(n)=\underline{Z}^{n}\underline{Z}(o)=\underline{Z}^{i}\underline{Z}(n-i).$$

Further defining the two series

$F(-1)=-/YY'$	$G(-2)=-/YZ$
$F(0)=0$	$G(-1)=I$
$F(1)=I$	$G(0)=0$
$F(2)=-/Y'Z$	$G(1)=-/Y'Y$
...	...

which both fit the equation for Z(i) and

$$\begin{bmatrix} F(n) & G(n-1) \\ F(n-1) & G(n-2) \end{bmatrix}=\begin{bmatrix} F(2) & G(1) \\ F(1) & G(0) \end{bmatrix}^{n}\begin{bmatrix} F(0) & G(-1) \\ F(-1) & G(-2) \end{bmatrix}$$

$$=\begin{bmatrix} F(i) & G(i-1) \\ F(i-1) & G(i-2) \end{bmatrix}\begin{bmatrix} F(n-i) & G(n-i-1) \\ F(n-i-1) & G(n-i-2) \end{bmatrix}$$

which produces the identities relating the F's and G's. It is possible to find explicit formulas for the ijth element of the inverse partitioned matrix, in terms of the F's and G's. These appear to be related to the Chandrasekar equations.

(7) When Bellman[32] first postulated the principle of optimality which led to the techniques of dynamic programming and Kalman filtering he was actually investigating scattering problems in layers, which he termed invariant embedding, and as such dealt not only with the Riccati equation but also Chandrasekar's equations. Largely ignored in control theory until recently these have now been shown by Kailath[2 - 6] et al to be more efficient than the Riccati equation in some time invariant cases.

(8) Direct calculation of p(o) from P(n) and, say, x(o) using scattering products and a doubling formula. This is the most efficient method of all though appears to be virtually unknown in practice. Like the FFT it is most useful when the number of time intervals can be written as an integer power of 2. Write

$$\begin{bmatrix} x(n) \\ p(n) \end{bmatrix} = \begin{bmatrix} E & F \\ G & H \end{bmatrix}^n \begin{bmatrix} x(o) \\ p(o) \end{bmatrix} = \begin{bmatrix} E_n & F_n \\ G_n & H_n \end{bmatrix} \begin{bmatrix} x(o) \\ p(o) \end{bmatrix} = \begin{bmatrix} x(n) \\ P(n)x(n) \end{bmatrix}$$

and

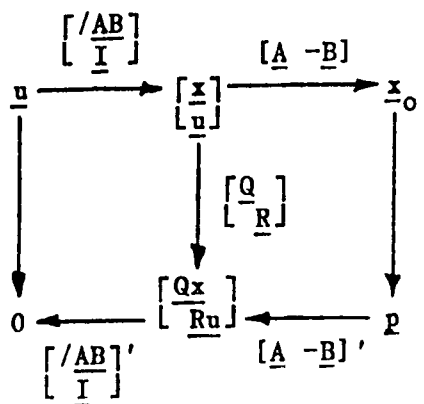
$$p(o) = /(P(n)F_n - H_n)(G_n - P(n)E_n)x(o).$$

6.3 A Kron type approach

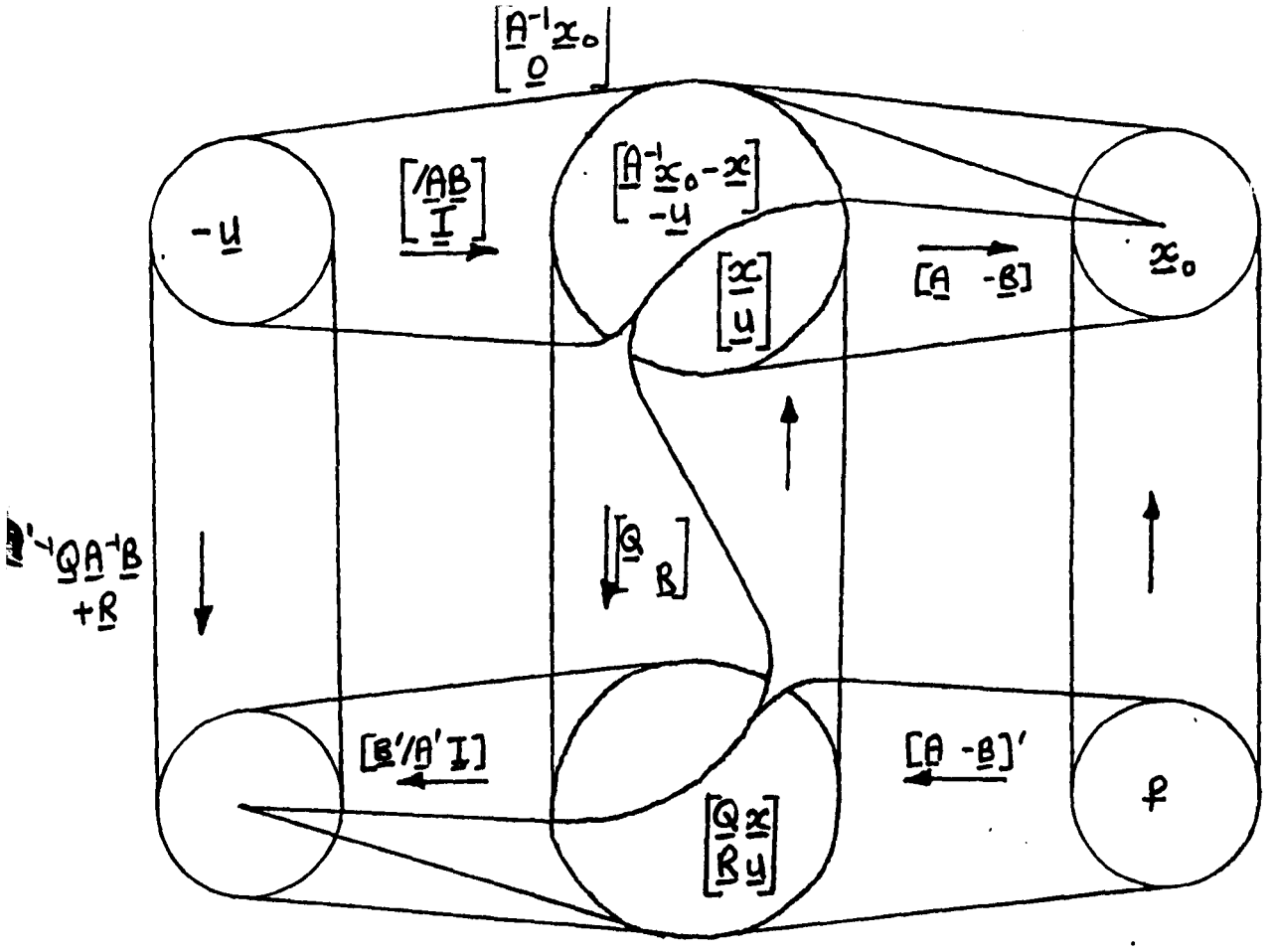
We again rewrite the multistage optimal control problem, this time in standard ordinary least squares form

$$\min \begin{bmatrix} \underline{x} \\ \underline{u} \end{bmatrix}' \begin{bmatrix} \underline{Q} & \\ & \underline{R} \end{bmatrix} \begin{bmatrix} \underline{x} \\ \underline{u} \end{bmatrix} \text{ given } [\underline{A} \ -\underline{B}] \begin{bmatrix} \underline{x} \\ \underline{u} \end{bmatrix} = \underline{x}_o$$

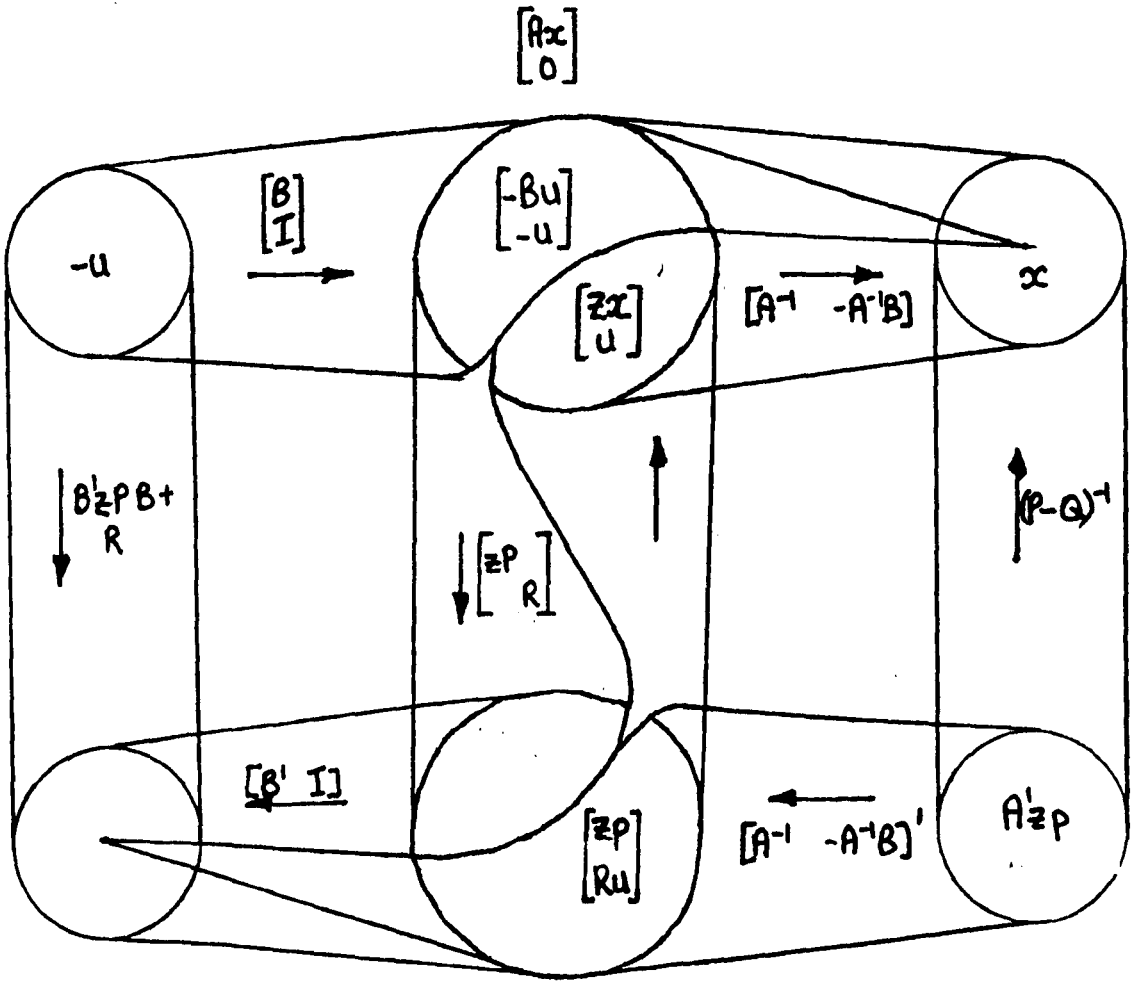
to which the results can be written down by inspection from Roth's diagram (see Fig(6.1))



giving



Fig(6·1) The partitioned system



Fig(6·2) The horn system

$$\begin{bmatrix}\underline{x}\\ \underline{u}\end{bmatrix}=/\begin{bmatrix}\underline{Q} & \\ & \underline{R}\end{bmatrix}[\underline{A}\ -\underline{B}]'/([\underline{A}\ -\underline{B}]/\begin{bmatrix}\underline{Q} & \\ & \underline{R}\end{bmatrix}[\underline{A}\ -\underline{B}]')[\underline{A}\ -\underline{B}]\underline{x}_o$$

$$=(\begin{bmatrix}\underline{I} & \\ & \underline{I}\end{bmatrix}-\begin{bmatrix}/\underline{AB}\\ \underline{I}\end{bmatrix}/(\begin{bmatrix}/\underline{AB}\\ \underline{I}\end{bmatrix}'/\begin{bmatrix}\underline{Q} & \\ & \underline{R}\end{bmatrix}\begin{bmatrix}/\underline{AB}\\ \underline{I}\end{bmatrix})\begin{bmatrix}/\underline{AB}\\ \underline{I}\end{bmatrix}'\begin{bmatrix}\underline{Q} & \\ & \underline{R}\end{bmatrix})\underline{x}_o$$

which are precisely the formula we derived earlier for $\underline{u}^o$. Note how the use of Householder's formula is replaced by the use of Kron's lemma. This approach bears much in common with the use of direct (topological) rather than energy methods in electromagnetic theory.

The costate vector $\underline{p}$ appears quite naturally in Roth's diagram giving as usual

$\underline{Q}\underline{x}=\underline{A}'\underline{p}$ and

$\underline{R}\underline{u}=-\underline{B}'\underline{p}$ or

$p(t)=C(t)'Q(t)C(t)x(t)+A(t)'p(t+1)$ where $p(n)=C(n)'Q(n)C(n)x(n)$

and $u(t)=-/R(t)B(t)'p(t+1)$.

The optimal trajectory and undetermined multipliers are thus defined by a two point boundary value problem and represented by a pair of coupled linear difference equations in scattering form

$$\begin{bmatrix}x(t+1)\\ p(t)\end{bmatrix}=\begin{bmatrix}A(t) & -B(t)/R(t)B(t)'\\ C(t)'Q(t)C(t) & A(t)'\end{bmatrix}\begin{bmatrix}x(t)\\ p(t+1)\end{bmatrix}$$

with

$$\begin{bmatrix}x(o)\\ p(n)\end{bmatrix}=\begin{bmatrix}x\\ C(n)'Q(n)C(n)x(n)\end{bmatrix}$$

We have investigated manipulation of the block diagonal equation $\underline{p}=\underline{P}\underline{x}$ but have yet found no useful results.

6.4 The torn system

Now let us tear the original problem back up into n stages. For the rest of this chapter we assume without loss of generality C(t)=I unless otherwise stated. We may minimise the zeroth Hamiltonian

$$\min H(o)$$

such that $x(t+1)=A(t)x(t)+B(t)u(t)$

where $2H(t)=x(t)'Q(t)x(t)+u(t)'R(t)u(t)+2H(t+1)$.

By Bellman's Principle of Optimality or Invariant Embedding we may minimise the tth Hamiltonian with respect to u(t) by assuming that the t+1th Hamiltonian is optimal with respect to u(t+1). Further taking the optimal value of H(t) to be quadratic with respect to x(t)

$$H(t)=x(t)'P(t)x(t)$$

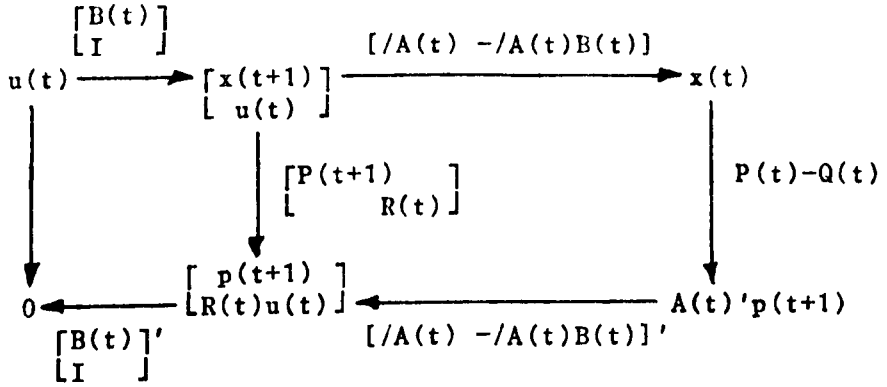
and augmenting the state space x(t+1) with u(t), the term in x(t) is invariant and the 2 drops out as the two remaining terms are both quadratic. We may again define a standard ordinary least squares problem

$$\min \begin{bmatrix} x(t+1) \\ u(t) \end{bmatrix}' \begin{bmatrix} P(t+1) & \\ & R(t) \end{bmatrix} \begin{bmatrix} x(t+1) \\ u(t) \end{bmatrix}$$

given

$$[/A(t) \; -/A(t)B(t)] \begin{bmatrix} x(t+1) \\ u(t) \end{bmatrix} = x(t)$$

from which the results may again be written down by inspection from Roth's diagram (see Fig(6.2))



where the costate vector

$$p(t)=P(t)x(t).$$

This transformation is the basis of the sweep method of solution. From the right hand square of the diagram we have the backward recursion relationship

$$P(t)-Q(t)=/([/A(t) -/A(t)B(t)]/\begin{bmatrix}P(t+1) & \\ & R(t)\end{bmatrix}\begin{bmatrix}/A(t)' & \\ -B(t)'/A(t)' & \end{bmatrix})$$

$$=A(t)'/(/P(t+1)+B(t)/R(t)B(t)')A(t) \quad \text{or from the lhs ...}$$

$$=A(t)'(P(t+1)-P(t+1)B(t)(R(t)+B(t)'P(t+1)B(t))^{-1}B(t)'P(t+1))A(t)$$

or Riccati equation, where $P(n)=Q(n)$. We will therefore in future draw this diagram reversed from left to right which turns out to be consistent with Kron's notation. Observe the similarity between Roth's diagram for the partitioned system and for the torn systems: $\underline{Q}$ is replaced by $P(t+1)$ which represents the contribution to the objective function of time increments from t+1 to n. The other differences are due to the inverted definition of $\underline{A}$ in the partitioned system. We see that we can write $H(t)=p(t)'x(t)$. This can be considered as an example of Kron's First Generalisation Postulate. The torn system is a model of the original system.

The diagram also gives Lewis' recent form for the solution of singular systems and can be extended to the decentralised and minimum time cases with Q=R=0 [11,12].

6.5 The reconnected system

We can now perform a bit of surgery on our commutative diagrams, following Kron's procedure in Ref[8] and join the sequence together into the algebraic diagram for a multistage linear optimal control problem in Fig(6.3).

The interface matrix E(t) is introduced in order to glue successive stages of the problem together. Kron had no problem with this as the consecutive stages of the space filter were assumed to be of compatible order and E(t) set to I. Working on the assumption that the top and bottom sequences are reverse complex conjugate (though this is not necessarily true) we have, for the diagram to commute

$$P(t+1)-Q(t+1)=E(t)*(B(t)'P(t+1)B(t)+R(t))E(t)$$

Performing an eigenvalue analysis on both sides

$$U(t)'\lambda_U(t)U(t)=E(t)*V(t)'\lambda_V(t)V(t)E(t)$$

and by abuse of notation

$$E(t)=/V(t)\lambda_V(t)^{-0.5}\lambda_U(t)^{0.5}U(t).$$

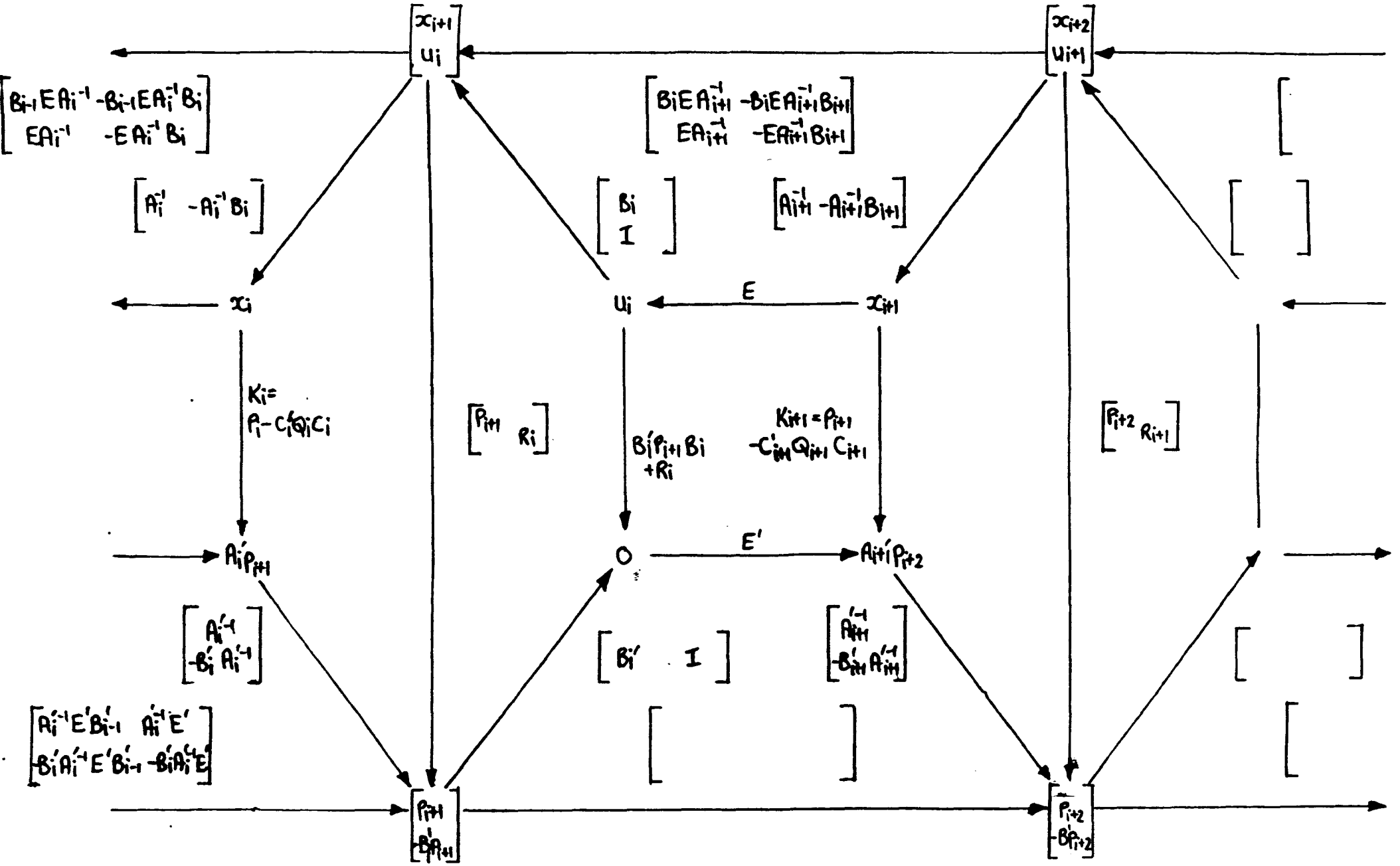


Fig (6·3) Discrete optimal control as a multidimensional space filter

where the U(t) $\{\lambda_U(t)\}$ and V(t) $\{\lambda_V(t)\}$ are the eigenvector matrices [diagonal eigenvalue matrices] of P(t+1)-Q(t+1) and B(t)'P(t+1)B(t)+R(t) respectively. There appears to be a relationship between the E matrices and the Chandrasekar equations in the time invariant case, as P(t+1)-Q=Y(t+1)/Y(t+2)P(t+2)A and B'P(t+1)B+R=/P(t+1)(I-/A'Y(t)/Y(t+1)) where P(t+1)-P(t+2)=Y(t+2)Y(t+1)'. It may also be of interest to investigate the annihilators of E(t).

Alternatively from the inverse (time invariant) scattering structure we may identify the interface matrices with the elements E=-/RB'zP and E*=zPB/R of the inverse interface matrix.

Following Kron and Nicholson we have a (reverse) homology sequence along the top of the diagram, a scattering sequence through the middle and a cohomology sequence along the bottom, see Fig(8.4). This is consistent with Kron's diagrams but not with Nicholson's optimal control analogy which is in forward sequence form. The optimisation criteria appear as vertical mappings as usual.

That the sequence of incidence matrices

$$M(t)=\begin{bmatrix} B(t)E(t)/A(t+1) & -B(t)E(t)/A(t+1)B(t+1) \\ E(t)/A(t+1) & -E(t)/A(t+1)B(t+1) \end{bmatrix}=\begin{bmatrix} -B/RB'zP/A & B/RB'zP/AB \\ -/RB'zP/A & /RB'zP/AB \end{bmatrix}$$

define a <u>chain complex</u> can be seen as follows:

(1) The K-module sequence [x(t+1)' u(t)']' is (intuitively) contravariant.

(2) M(t)M(t+1)=0

(3) $\mathrm{im}(M(t+1)) < \ker(M(t))$

The surprising fact that (3) is not an equality can be demonstrated thus

Example For a constant time system with

$A=I$, $B=[1\ 1]'$, $E=[1\ 1]$

we have $M=\begin{bmatrix}1 & 1 & -2\\ 1 & 1 & -2\\ 1 & 1 & -2\end{bmatrix}$ so $\begin{bmatrix}1 & 1 & -2\\ 1 & 1 & -2\\ 1 & 1 & -2\end{bmatrix}\begin{bmatrix}1 & 1 & -2\\ 1 & 1 & -2\\ 1 & 1 & -2\end{bmatrix}=0$

but given $x(t+1)=[2\ 0]'$, $u(t)=1$ then

$$\begin{bmatrix}1 & 1 & -2\\ 1 & 1 & -2\\ 1 & 1 & -2\end{bmatrix}\begin{bmatrix}2\\ 0\\ 1\end{bmatrix}=0 \text{ but } \begin{bmatrix}2\\ 0\\ 1\end{bmatrix}\#\begin{bmatrix}1 & 1 & -2\\ 1 & 1 & -2\\ 1 & 1 & -2\end{bmatrix}\begin{bmatrix}.\\ .\\ .\end{bmatrix}$$

Thus for the first time we have a sequence of matrices each of which annihilates its successor but that do not form an exact sequence. The matrices are singular (obviously), and nilpotent in the time invariant case. The reason this does not normally happen is that we always choose minimal annihilators. Matrices of this form do not exist with dimension less than 3x3.

The kernel of M in (x_1,x_2,u)-space is a plane $x_1+x_2=2u$. The image is the subspace $x_1=x_2=u$, a straight line. We are interested in the homology module

$H(M)=\ker(M)\backslash\mathrm{im}(M)=\{x_1+x_2=2u\}\backslash\{x_1=x_2=u\}=$

$\{x_1,x_2,(x_1+x_2)/2\}\backslash\{u,u,u\}=$

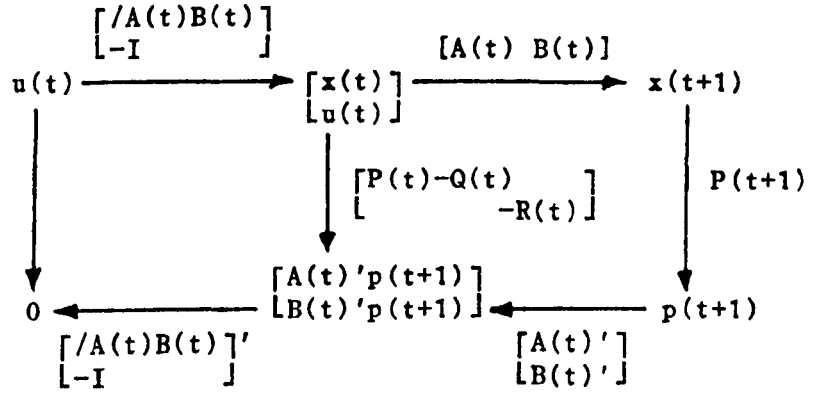
$\{(x_1,-x_1,0)+\{u,u,u\}\}$

which is a series of straight lines parallel to im(M) parameterised along a perpendicular line through the origin. There is one degree of freedom as $R^2\backslash R=R$.

We have $BE/A(x(t+1)-B(t)u(t))=0$. The im(B) is just the space $x(t)=0$, the ker(M) is $BEx(t)=0$.

6.6 The forward form

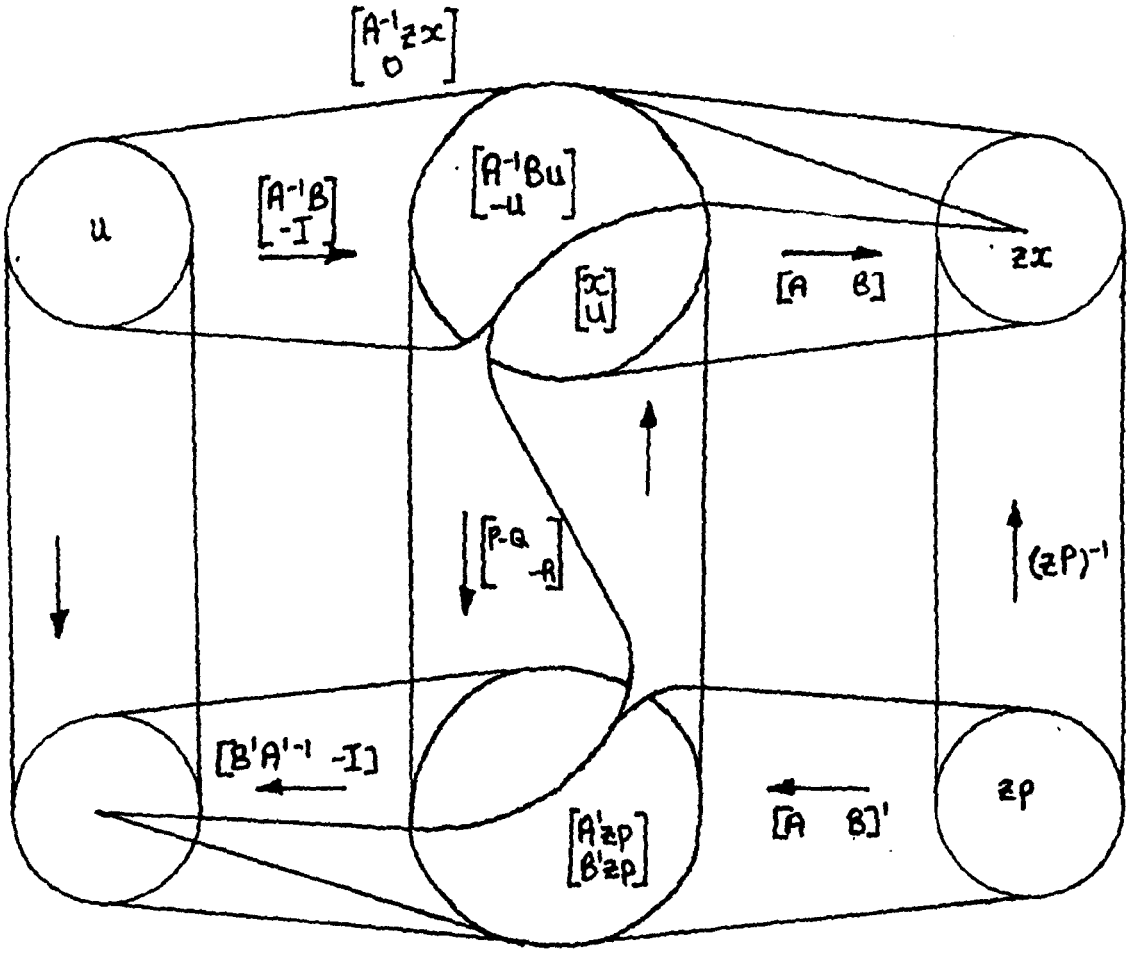
If we use $[x(t)'\ u(t)']'$ as the augmented state space then Roth's diagram (see Fig(6.4) for the substructure) becomes



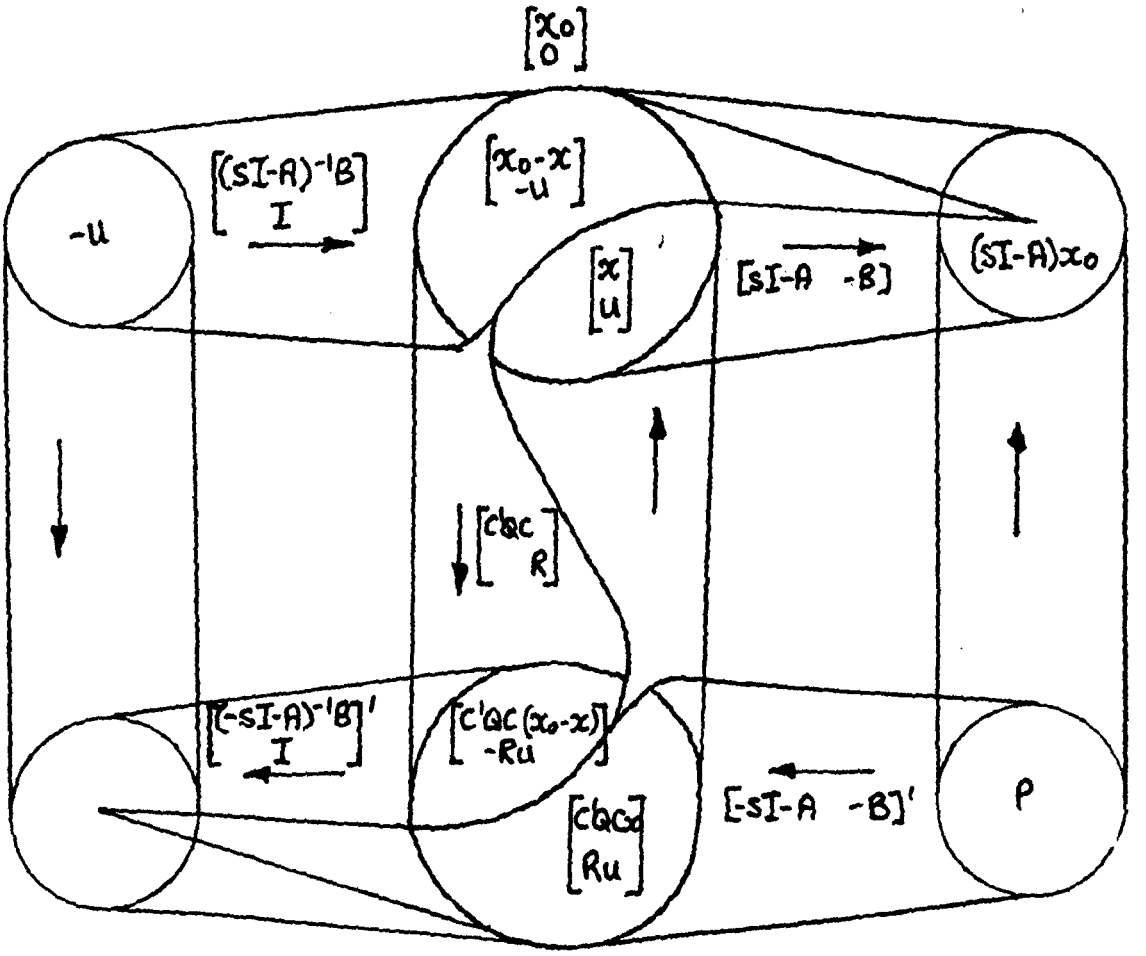
which is equivalent to

$$\min x(t)'Q(t)x(t)+u(t)'R(t)u(t)-p(t)'x(t)$$

and we have the forwards recursion relationship



Fig(6.4) The forward system



$$/P(t+1)=[A(t)\ B(t)]/\begin{bmatrix}P(t)-Q(t) & \\ & -R(t)\end{bmatrix}\begin{bmatrix}A(t)'\\ B(t)'\end{bmatrix}$$

which is of course the Riccati equation. The diagrams must be glued together in ascending order consistent with Nicholson's analogy. This can be demonstrated by investigating the forward scattering formulation.

7 The orthogonal system

	Networks	Backwards	Forwards
	$\begin{bmatrix}V_t\\ V_1^t\end{bmatrix}\begin{bmatrix}J^t\\ J^1\end{bmatrix}$	$\begin{bmatrix}zp\\ -B'zp\end{bmatrix}\begin{bmatrix}zx\\ u\end{bmatrix}$	$\begin{bmatrix}A'zp\\ B'zp\end{bmatrix}\begin{bmatrix}x\\ u\end{bmatrix}$
	$\begin{bmatrix}e^c\\ E_o^c\end{bmatrix}\begin{bmatrix}i^c\\ I^o\end{bmatrix}$	$\begin{bmatrix}0\\ A'zp\end{bmatrix}\ \begin{bmatrix}u\\ x\end{bmatrix}$	$\begin{bmatrix}0\\ zp\end{bmatrix}\ \begin{bmatrix}-u\\ zx\end{bmatrix}$
$C_c\ C_o$	$\begin{bmatrix}C_t:I\\ I^t:0\end{bmatrix}$	$\begin{bmatrix}B:A\\ I:0\end{bmatrix}$	$\begin{bmatrix}/AB:/A\\ -I\ :\ 0\end{bmatrix}$
	$\begin{bmatrix}Z_t\\ Z_1\end{bmatrix}$	$\begin{bmatrix}zP\\ R\end{bmatrix}$	$\begin{bmatrix}P-Q\\ -R\end{bmatrix}$
$Z_1\ Z_2$ $Z_3\ Z_4$	$\begin{bmatrix}C_t'Z_tC_t+Z_1 & C_t'Z_t\\ Z_tC_t & Z_t\end{bmatrix}$	$\begin{bmatrix}B'zPB+R\ B'zPA\\ A'zPB\quad A'zPA\end{bmatrix}$	$\begin{bmatrix}-R/(R+B'zPB)R\ B'/A'(P-Q)/A\\ /A'(P-Q)/AB\quad /A'(P-Q)/A\end{bmatrix}$
$A^c\ A^o$	$\begin{bmatrix}0:I\\ I:A^1\end{bmatrix}$	$\begin{bmatrix}0:\quad /A'\\ I:-B'/A'\end{bmatrix}$	$\begin{bmatrix}0\ :A'\\ -I:B'\end{bmatrix}$
	$\begin{bmatrix}Y^t\\ Y^1\end{bmatrix}$	$\begin{bmatrix}/(zP)\\ /R\end{bmatrix}$	$\begin{bmatrix}/(P-Q)\\ -/R\end{bmatrix}$
$Y^1\ Y^2$ $Y^3\ Y^4$	$\begin{bmatrix}Y^1 & Y^1A^1\\ A^1,Y^1\ Y^t+A^1,Y^1A^1\end{bmatrix}$	$\begin{bmatrix}/R\ -/RB'/A'\\ -/AB/R\quad /(P-Q)\end{bmatrix}$	$\begin{bmatrix}-/R\ /RB'\\ B/R\ /(zP)\end{bmatrix}$
Scattering matrices			
$N_1\ N_2$ $N_3\ N_4$	$\begin{bmatrix}-/Z_1Z_2 & /Z_1\\ Z_4-Z_3/Z_1Z_2 & Z_3/Z_1\end{bmatrix}$	$\begin{bmatrix}-/RB'/A'(P-Q) & /(B'zPB+R)\\ P-Q & (P-Q)/AB/R\end{bmatrix}$	$\begin{bmatrix}/RB'zP & -/R(R+B'zPB)/R\\ zP & -zPB/R\end{bmatrix}$
Inverse Scattering	$/\underline{N}$	$\begin{bmatrix}-/AB & /(A'zPA)\\ R & B'/A'\end{bmatrix}$	$\begin{bmatrix}-B\ A/(P-Q)A'\\ -R\quad B'\end{bmatrix}$
Inverse System	$/\underline{N}$	$\begin{bmatrix}/A & /AB/RB'/A'\\ -Q & /A'\end{bmatrix}$	$\begin{bmatrix}A & -B/RB'\\ Q & A'\end{bmatrix}$
Inverse Interface	$\underline{I}$	$\begin{bmatrix}-/RB'zP & /R(R-B'zPB)/R\\ zP-Q & zPB/R\end{bmatrix}$	$\begin{bmatrix}/RB'/A'(P-Q) & -/R(R+B'/A'(P-Q)/AB)/R\\ P & -(P-Q)/AB/R\end{bmatrix}$

The table compares the orthogonal network with orthogonal forwards and backwards multistage optimal control. We use z the (monadic) forward shift operator rather than time coefficients. The vectors are augmented and their relationships defined in terms of square and nonsingular connection matrices. Note $A/(P-Q)A' = z/P+B/RB'$. All the standard network equations can be derived. Note that Nicholson[1] defines the control variables (e_c, E_o) and (i^c, I^o) the other way round hence the inside out appearance of his admittance matrix. Nevertheless he effectively arrives at our results using only scattering theory, not Roth's diagram. The internal structure of the orthogonal connection matrices is shown in fig(3.4).

3.8 The scattering structure (inverse forward form)

It is important to note that the matrix derivations above, and for the rest of this chapter are carried out without reference to the variables they transform. Indeed though some of the variables are equal to zero, this does not affect the results. We first wish to investigate the discrepancy between the system matrix and the scattering matrix[1 - 6, 24 - 30]. This is defined by the E matrix in the space filter and represents the interface between successive stages. The interface was defined as unity by Kron ($I^o(t+1)=i^c(t)$ and $E_o(t+1)=e_c(t)$) and ignored by Nicholson who presents an apparently trivial decomposition of the system matrix. We will also use the forward form of the scattering matrix as the decomposition is simpler this case, we use the inverse form as this is the normal system matrix, and gives the same results as Nicholson. It is surprising how apparently dual forms show occasional discrepancies. The derivation of the (top right hand element of

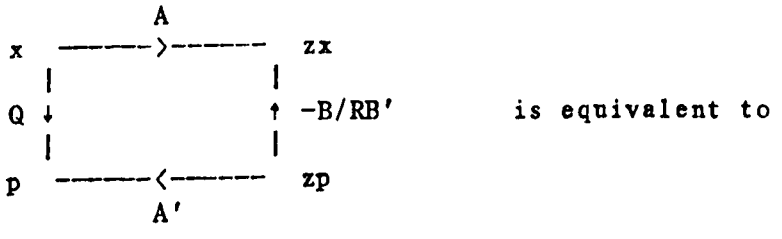
the) interface matrix is difficult so we just present the results. Note the + sign in the top right hand element of the interface matrix. We have always had a - sign in terms of this form before. The left hand side of the interface matrix is the same as Nicholson's matrix, but he uses zeros in the right hand side.

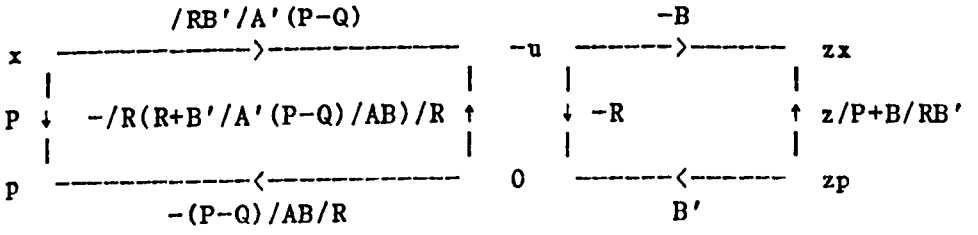
$$\begin{bmatrix} A & -B/RB' \\ Q & A' \end{bmatrix} =$$

$$\begin{bmatrix} /RB'/A'(P-Q) & -/R(R+B'/A'(P-Q)/AB)/R \\ P & -(P-Q)/AB/R \end{bmatrix} * \begin{bmatrix} -B & z/P+B/RB' \\ -R & B' \end{bmatrix}$$

system matrix = interface matrix * scattering matrix

Transformation diagram of decomposition of system matrix





Nicholson's transformation diagrams are in fact essentially signal flow diagrams. The 'branches' do not commute but there is a law to be obeyed at the 'nodes'. The decomposition in scattering form can be written as a commutative diagram with pairs of orthogonal variables at the nodes and scattering matrices as mappings (drawn as double arrows) commuting under the star product. Note that the scattering matrix can be further decomposed into a star product as in Chapter III for electrical networks. So each time step in the dynamic programming process consists of three substages. In the next section we consider the scattering process between the time steps.

6.9 The Chandrasekar equations

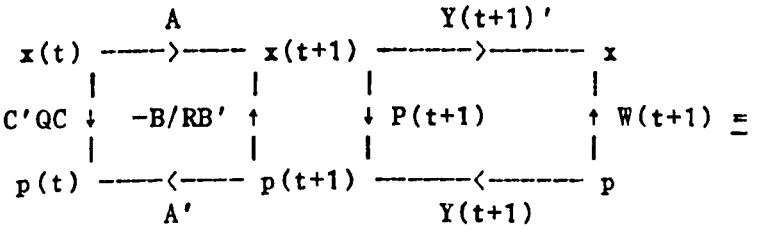
Define the set of equations

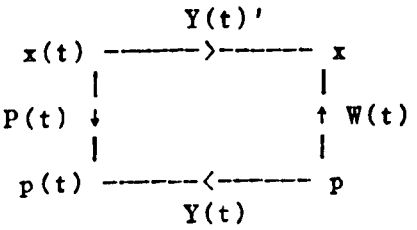
$$\begin{bmatrix} x \\ p(t+1) \end{bmatrix} = \begin{bmatrix} Y(t+1)' & W(t+1) \\ P(t+1) & Y(t+1) \end{bmatrix} \begin{bmatrix} x(t+1) \\ p \end{bmatrix}$$

where p is a vector multiplier, previously assumed to be zero and $x=Z'y(n)$ is a terminal constraint

$$\begin{bmatrix} Y'_o & W_o \\ P_o & Y_o \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \qquad \begin{bmatrix} Y'_n & W_n \\ P_n & Y_n \end{bmatrix} = \begin{bmatrix} I & B/RB' \\ C'QC & I \end{bmatrix}$$

then from the equivalence of the transformation diagrams





it is easy to see that

$$\begin{bmatrix} Y' & W \\ P & Y \end{bmatrix} = \begin{bmatrix} A & -B/RB' \\ C'QC & A' \end{bmatrix} *z \begin{bmatrix} Y' & W \\ P & Y \end{bmatrix}$$

$$= \begin{bmatrix} zY'/(I+B/RB'zP)A & zW-zY'B/RB'/(I+zPB/RB')zY \\ C'QC+A'zP/(I+B/RB'zP)A & A'/(I+zPB/RB')zY \end{bmatrix}$$

The equations for P and W are a dual pair of Riccati equations. The equations for Y and Y' are a symmetrical pair of Chandrasekar equations[2 - 6]. This derivation, based on the ideas of Kailath, et al demonstrates that the Chandrasekar equations hold in the time varying case. In the general case we have $W-zW=-zY'B/RB'/A'Y$. For the time invariant case it is easy to show that $P-zP=zYC'QCY$, from which we may write the coupled difference equations in PB and YC'

PB=zPB+(zYC')Q(CY')B and

YC'=A'/(I+(zPB)/RB')zYC',

which are the Chandrasekar equations of discrete optimal control as given by Kailath et al. Under certain conditions - ie n(n+1)/2>n(m+k) where n,m,k are the number of states, inputs and outputs, respectively - these are more efficient than the Riccati equation.

The following differential scattering equation may also be verified

$$\begin{bmatrix} z/Y' & z(-/Y'W) \\ P/Y' & Y-P/Y'W \end{bmatrix} = \begin{bmatrix} A & -B/RB' \\ C'QC & A' \end{bmatrix} \begin{bmatrix} /Y' & -/Y'W \\ z(P/Y') & z(Y-P/Y'W) \end{bmatrix}$$

6.10 Continuous optimal control

Continuous optimal control is isomorphic to the discrete case: z is replaced by the Laplace operator s where z=exp(s). Consider optimal control of the linear system

dx(t)/dt=A(t)x(t)+B(t)u(t), where y(t)=C(t)x(t),

subject to the terminal constraint x=Z'y(n),

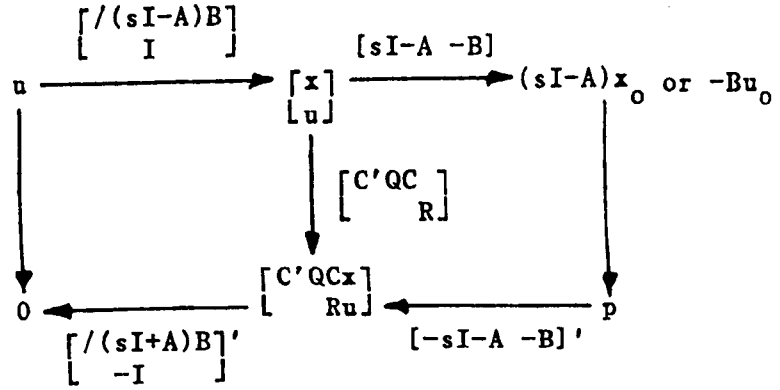
with minimum performance index $\int y(t)'Q(t)Y(t)+u(t)'R(t)u(t)dt$.

Over an infinite time horizon this is equivalent to minimising

$$\oint y^*(s)Qy(s)+u^*(s)Ru(s)ds/2$$

by Parseval's theorem where the integral occurs around the infinite left half s-plane. Writing $G(s)=C/(sI-A)B$ we have $G^*(s)Q(s)G(s)u(s)+R(s)u(s)=0$ is sufficient to minimise the integral giving $u(s)=-/R(s)G^*(s)Q(s)y(s)$.

Roth's diagram for the continuous case looks like this (see Fig(6.5) for the substructure)



The solution is thus represented by the left hand square. The diagram also gives the solution for singular systems [11]. We again stress that the objects between the mappings are modules of polynomials in s, that is sets of time evolutions of the state vectors.

The right hand square represents the Euler-Lagrange equation. Minimising the Hamiltonian

$$\oint y'Qy+u'Ru+2p(sx-Ax-Bu)ds$$

defines the solution of the two point boundary problem

$$\left(\frac{\partial}{\partial[x'\ u']}-s\frac{\partial}{\partial[sx'\ su']}\right)\left(\frac{1}{2}\begin{bmatrix}x\\u\end{bmatrix}'\begin{bmatrix}C'QC & \\ & R\end{bmatrix}\begin{bmatrix}x\\u\end{bmatrix}-\begin{bmatrix}x\\u\end{bmatrix}'\begin{bmatrix}A\\B\end{bmatrix}p+\begin{bmatrix}sx\\su\end{bmatrix}'\begin{bmatrix}I\\0\end{bmatrix}p\right)=0$$

therefore

$$\begin{bmatrix}C'QC & \\ & R\end{bmatrix}\begin{bmatrix}x\\u\end{bmatrix}-\begin{bmatrix}A+sI\\B\end{bmatrix}'p=0$$

$$\begin{bmatrix}sx(t)\\sp(t)\end{bmatrix}=\begin{bmatrix}A & -B/RB'\\-C'QC & -A'\end{bmatrix}\begin{bmatrix}x(t)\\p(t)\end{bmatrix}$$

$$\begin{bmatrix}x(o)\\p(n)\end{bmatrix}=\begin{bmatrix}x\\C'QCx(n)+C'Zp\end{bmatrix}$$

where p is a vector multiplier. Adjoint state variable and terminal constraint relationships are then introduced with

$$\begin{bmatrix}x\\p(t)\end{bmatrix}=\begin{bmatrix}Y(t)' & W(t)\\P(t) & Y(t)\end{bmatrix}\begin{bmatrix}x(t)\\p\end{bmatrix}$$

where $Y'(n)=Z'C$ and $W(n)=0$. Differentiating,

$$\begin{bmatrix}0\\sp(t)\end{bmatrix}=s\begin{bmatrix}Y' & W\\P & Y\end{bmatrix}\begin{bmatrix}x(t)\\p\end{bmatrix}+\begin{bmatrix}Y' & W\\P & Y\end{bmatrix}\begin{bmatrix}sx(t)\\0\end{bmatrix}$$

therefore

$$s\begin{bmatrix}Y' & W\\P & Y\end{bmatrix}\begin{bmatrix}x(t)\\p\end{bmatrix}=\begin{bmatrix}-Y'sx(t)\\sp(t)-Psx(t)\end{bmatrix}$$

$$=\begin{bmatrix}-Y' & 0\\-P & I\end{bmatrix}\begin{bmatrix}sx(t)\\sp(t)\end{bmatrix}$$

$$=\begin{bmatrix}-Y' & 0\\-P & I\end{bmatrix}\begin{bmatrix}A & -B/RB'\\-C'QC & -A'\end{bmatrix}\begin{bmatrix}x(t)\\p(t)\end{bmatrix}$$

$$=\begin{bmatrix}-Y' & 0\\-P & I\end{bmatrix}\begin{bmatrix}A & -B/RB'\\-C'QC & -A'\end{bmatrix}\begin{bmatrix}I & 0\\P & Y\end{bmatrix}\begin{bmatrix}x(t)\\p\end{bmatrix}$$

Therefore

$$s\begin{bmatrix}Y' & W\\P & Y\end{bmatrix}=\begin{bmatrix}Y'(B/RB'-A) & Y'B/RB'Y\\-PA-A'P+PB/RB'P-C'QC & (PB/RB'-A')Y\end{bmatrix}$$

$$\begin{bmatrix}Y_o' & W_o\\P_o & Y_o\end{bmatrix}=\begin{bmatrix}I & 0\\0 & I\end{bmatrix}\text{ and }\begin{bmatrix}(sY')_o & (sW)_o\\(sP)_o & (sY)_o\end{bmatrix}=\begin{bmatrix}A & -B/RB'\\-C'QC & -A'\end{bmatrix}$$

$$\begin{bmatrix}Y_n' & W_n\\P_n & Y_n\end{bmatrix}=\begin{bmatrix}C'Z & C'QC\\0 & Z'C\end{bmatrix}$$

The equations for P and W are a dual pair of Riccati equations. The equations for Y and Y' are a symmetrical pair of Chandrasekar equations. No mention of time invariance has yet been made. It is not difficult to show that in the time invariant case the coupled nonlinear differential equation pair

$$s(YC')=((PB)/RB'-A')(YC')$$

$$s(PB)=(YC')Q(CY')B$$

in YC' and PB may be used instead of integrating the backwards Riccati equation and is more efficient in certain cases (see above).

Further Y, P and W can be shown to obey the matrix differential equations

$$s\begin{bmatrix} /Y' & -/Y'W \\ P/Y' & Y-P/Y'W \end{bmatrix}=\begin{bmatrix} A & -B/RB' \\ -C'QC & -A' \end{bmatrix}\begin{bmatrix} /Y' & -/Y'W \\ P/Y' & Y-P/YW' \end{bmatrix}$$

An unusual approach to the solution of the Riccati equation is to write $K=\sum K_i s^i$, $i=0,1,\ldots$ and equate powers of s. This technique could also be applied to the Chandrasekar equations and would be particularly effective in the z-domain.

6.11 Conclusions

A major result of this chapter is to put optimal control into a multivariable frequency response context. Specifically we have made it unnecessary to use the state space approach at all in either discrete or continuous time by showing the structure of optimal systems as modules of polynomials in z or s (a particularly efficient realisation from the point of view of computer implementation.) This has been touched upon before by Chang[20] and Kucera[19] for special cases in the

s-plane, and in more detail by Kalman[31] in the z-plane though not with respect to optimal systems. Our treatment was however carried out on far more general, though rather nonrigorous grounds. In the discrete time case we justified dropping the summation sign in the minimisation by invoking the Principle of Optimality. This appears to be a valid approach. Kalman, however, defines the product (composition) of two polynomials (polynomial functions) in z^{-1} as the 'ordinary product of polynomials with deletion of all terms corresponding to nonnegative powers of z', apparently a necessary consequence of the fact that the minimisation occurs only over positive time, (and appears to be related to the difference between the 'scientific' and 'control theory' versions of the z-transform.) We appear to have avoided the problem by writing all modules as direct sums with the initial conditions eg. $x(z)\oplus x_o$, the latter defined to be zero after t=0.

Again in the s-plane we invoked Parseval's theorem followed by the residue theorem

$$\min_u \int_{t=0}^{\infty} x(t)Q(t)x(t)+u(t)R(t)u(t)dt=$$

$$\min_u \frac{1}{2\pi j}\oint_{-j\infty}^{j\infty} x^*(s)Q(s)x(s)+u^*(s)R(s)u(s)ds=$$

infinite semicircle in rhp

$$\frac{1}{2\pi j}\min_u \int_{-j\infty}^{j\infty} \text{real poles in the right half plane}=$$

$$\frac{1}{2\pi j}\min_u x^*(s)Q(s)x(s)+u^*(s)R(s)u(s)$$

and drop the summation because the last form obviously includes all the poles in the right half plane. We have assumed (1) R(s)=R*(s), Q(s)=Q*(s) or the $\underline{R}$ and $\underline{Q}$ matrices are symmetrical in the state space case, (2) infinite time horizon, though we know that all the results hold over a finite interval, and particularly that the feedback matrix is constant in the infinite interval case: and yet we allow a time

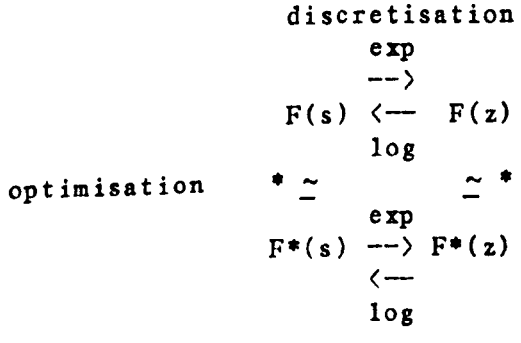
varying system. It is the fact that the minimisation occurs over real poles that causes the problem here; thus modules in stable dual systems take the form

$\mathrm{Hom}(X(s),K[s]^{+})$ and $\mathrm{Hom}(X(z),K[z]^{+})$

where $K[s]^{+}$ are polynomials with real poles,

and $K[z]^{+}$ are polynomials in positive time.

Chang and Kucera take a mechanistic approach to the continuous time problem by introducing the technique of spectral factorisation whereby a rational transfer function may be decomposed into the product of two rational transfer functions having only positive and negative poles respectively and further by partial fraction expansion into sums of such terms. Chang in fact 'cheats' and drops the integral on the grounds that it is sufficient (though not necessary) for each term in the integrand to be minimised, though giving a detailed analysis of the relationship between optimality and stability (system poles in lhp.) Kucera deals with modules of rational transfer functions in discrete time and, like Chang, derives an explicit solution for the optimal controller in the single-input single-output case, in a similar form. The relationship between adjoint pairs of discrete and continuous systems is often shown in the following diagram



We have advocated the use of direct polynomial manipulation rather than use of the state space form in optimal control, along the lines of Kucera[19]. Our theory also applies to matrices of polynomials, rational transfer functions and matrices thereof, etc, though in no case have we rigorously proved anything - the ideas are all results of the Homology Theorem and its associated Hypothesis - however we have not as yet found any counterexamples! For instance we can specialise the Hypothesis for the siso system $y=f(s)u$ where we wish to minimise $\int qy^2+u^2dt$ and speculate that the optimal control law $u^o=-f^*qy=-f(-s)qy$. For example for a second order system $(as^2+bs+c)y=u$ we have $u^o=-(as^2-bs+c)^{-1}qy$ which is easy, if tedious to check by state space methods (note that this is actually a simple rational transfer function example). We have

$$y=\begin{bmatrix}0 & 1\end{bmatrix}\left(s\begin{bmatrix}1 & \\ & 1\end{bmatrix}-\begin{bmatrix}-b/a & -c/a\\ 1 & 0\end{bmatrix}\right)^{-1}\begin{bmatrix}1/a\\ 0\end{bmatrix}u$$

therefore $$f^*=\begin{bmatrix}1/a & 0\end{bmatrix}\left(-s\begin{bmatrix}1 & \\ & 1\end{bmatrix}-\begin{bmatrix}-b/a & 1\\ -c/a & 0\end{bmatrix}\right)^{-1}\begin{bmatrix}0\\ 1\end{bmatrix}$$

$$=1/(as^2-bs+c)\ \text{QED.}$$

It is easy to see that this generalises to higher order cases.

This work is intended to be taken more in the vein of Heaviside (or Kron) than Laplace, ie the methods appear to work. A large body of deep results already exists within General Homology Theory which seems a good candidate to make the theory rigorous - ie, to establish the conditions under which it does work. A number of related fields remain open for investigation eg, the relationship between the Homology Theorem as applied to modules of polynomials in s, the Laplace transform and to modules in z, the shift operator, and further between the Fourier transform and Fourier series representations. It is likely that a major application of our theory will be to time series analysis. In a sense most of the major problems of optimal control, noninteracting control, etc have been solved, whereas though Box and Jenkins have analysed single variable time series in detail,

no general theory of multivariable time series analysis yet exists. The important question here is whether the category of modules of rational polynomials, RATPOL, induces a homology theory. In the stochastic domain Dodson[33,34] and others have constructed homology theories where the base sets are 'fuzzy' or 'hazy', that is over the categories FUZ and HAZ, and describes a remarkably practical application in the paper industry. Again the 'duality' (certainty equivalence) principle as proposed for Kalman filtering is unsatisfactory in the general stochastic case. Khabie-Zeitoune suggests that the theory of stochastic differential equations may be important here. Finally part of Ref[10] is reproduced in Appendix(1) showing the relevance to transmission zero theory.

6.12 References

(1) Nicholson, Structure of Interconnected Systems, IEE Control Engineering Monograph 5, Peter Perigrinus, 1978.

(2) Kailath, Some New Algorithms for Recursive Estimation in Constant Linear Systems, IEEE Trans. Information Theory, Nov 1973.

(3) Morf, Sidhu and Kailath, Some New Algorithms for Recursive Estimation in Constant Linear Discrete-time Systems, IEE Trans. Automatic Control, Aug 1974.

(4) Ljung, Kailath and Friedlander, Scattering Theory and Linear Least Squares Estimation, Part I - Continuous-time Problems, Proc. IEEE, Vol 64, No 1, Jan 1976.

(5) Friedlander, Kailath and Ljung, Scattering Theory and Linear Least Squares Estimation, Part II - Discrete-time Problems, J. Franklin Inst., Vol 301, Jan/Feb 1976.

(6) Verghese, Friedlander and Kailath, Scattering Theory and Linear Least Squares Estimation, Part III - The Estimates, IEEE Trans. Automatic Control Vol AC25, No 4, Aug 1980.

(7) Kron, Diakoptics, MacDonald, London, 1963.

(8) Kron, Basic Concepts of Multidimensional Space Filters, Proc. IEE, Nov 1959.

(9) Wonham, Linear Multivariable Control, A Geometric Approach, Berlin, Springer Verlag, 1974.

(10) Bowden, An Introduction to Homological Systems Theory: Topological Analysis of Invariant Zeros, Matrix and Tensor Quarterly, Vol 13, No 2, Dec 1980.

(11) El-Sayed, Optimal Observations in Decentralised Systems, IEEE Trans AC-26 No 3 June 1981.

(12) Lewis, Riccati Equation Solution to the Minimum Time Output Control Problem, Ibid.

(13) Lewis, Riccati Equation Computation of Supremal (F,G)-Invariant and Reachability Subspaces, Ibid.

(14) Yoshikawa, Condition for Existence of Solutions of Linear Equations with Arbitrary Vectors and its Relation to Controllability, Ibid.

(15) Sima, An Efficient Schur Method to solve the Stabilising Problem, Ibid.

(16) Jonckheere, On the Existence of a Negative Semidefinite Solution to the Discrete Time Algebraic Riccati Equation, Ibid.

(17) Lewis, A Generalised Inverse Solution to the Discrete Time Singular Riccati Equation, IEEE Trans AC-26 No 2 April 1981.

(18) Pratzel-Woltes, Brunovsky Equivalence of System Matrices: The Reachable Case, Ibid.

(19) Kucera, Algebraic Approach to Discrete Linear Control, IEEE Trans Feb 1975.

(20) Chang, Synthesis of Optimum Control Systems, New York, McGraw Hill, 1961.

(21) Hermann and Martin, Applications of Algebraic Geometry to Systems Theory, Part I, IEEE Trans AC-22, No 1, 1977.

(22) Hermann and Martin, A new application of Algebraic Geometry to Systems Theory, IEEE Conf. Decision and Control, Dec 1976.

(23) IEEE CAS-28, June 1981, No 6, Special Issue on Adaptive Signal Processing.

(24) Nicholson, Structure of Kron's Polyhedron Model and the Scattering Problem, Proc IEE, Vol 120, No 12, Dec 1973.

(25) Nicholson, Sequential Least Squares Estimation and the Electrical Network Problem, Proc IEE, Vol 118, No 11, Nov 1971.

(26) Nicholson, Concepts of Scattering in the Linear-Optimal Control Problem, Proc IEE, Vol 118, No 12, Dec 1971.

(27) Nicholson, Properties of the Generalised Inverse Matrix in the Electrical Network Problem, Univ Sheffield, Dept Control Eng, Research Report No 9.

(28) Nicholson, Transformations of Linear Graph Theory in Multivariable System Problems, Ibid, No 10.

(29) Nicholson, The Structure of Kron's Polyhedron Model and the Scattering Problem, Ibid, No 14, June 1973.

(30) Nicholson, System Concepts in Kron's Polyhedron Model and the Scattering Problem, Ibid, No 26, May 1974.

(31) Kalman and Falb, Topics in Mathematical Systems Theory, McGraw Hill, 1969.

(32) Reid, Riccati Differential Equations, Academic Press, 1972.

(33) Dodson, Generalised Hypergraphs, To appear in Analele Universitatii din Timisoara, Stiinte Matematice, 1981.

(34) Dodson and Radivoiovici, Homotopy Theory for Generalised Hypergraphs, Preprint of paper for presentation at the 6th Hungarian Colloquium on Combinatorics, organised by the Janos Bolyai Mathematical Society at Eger, Hungary, 6-11 July 1981.

(35) Kucera, A contribution to Matrix Quadratic Equations, IEEE Trans AC-7, 1972.

(36) Luenberger, Canonical Forms for Linear Multivariable Systems, IEEE Trans AC, June 1967.

(37) Noton, Two Level form of the Kalman Filter, IEEE Trans AC-16, No 2, pp 128-39, April 1971.

CHAPTER VII Example: Leontief's Economy

Mathematical economics has progressed a long way since Adam Smith's 'invisible hand' guided the economy along its course. Just over a century ago (Lausanne, 1874) Leon Walras[12] proposed a national economic model involving many self optimising decision units - or entrepreneurs - consumers and producers who act solely on behalf of their own interests, coordination being achieved by indirect, nonlegislative price controls. How this apolitical economy allocates the consumer held factors of production between manufacturers and distributes the resulting national product amongst the consumers has since been the subject of mathematical analysis. Except for the solution of a simple Paselian system by Wald in 1935 the problem remained unsolved until the proof of the existence of a general competitive equilibrium by Arrow and Debreu, McKenzie, Gale and Nikaido[4] in the 1950's (using the concepts of Pareto optimality and 'core' of an economy from game theory). Gale made the conditions for equilibrium weak enough to be acceptable to economists. The problem is closely related to the decomposition and interaction coordination of a multilevel, hierarchical system as investigated by Pearson, Mesarovic, et al. The solution is similar to the proof of Pontryagin's Maximum principle. Both rely on finding a mapping with certain continuity properties from a certain space into itself, and then use the Brouwer or Kakutani fixed point theorems to prove existence of an optimum. Both fixed point theorems are proved using Homology theory. The major practical problem with the work on Walras' model is a preoccupation with existence theorems at the expense of analytical solutions.

Also in the 1930's both Wassily Leontief[5, 6] and John von Neumann formulated and solved simple linear multisector economic models. The former, for which Leontief was awarded the Nobel prize, is a constant, linear commodity flow model: Schwartz has proved that free enterprise is the optimal strategy. The latter which has been relatively little used is an important analysis of capital stock accumulation in an

expanding economy for which von Neumann proved the existence of a minimax moving equilibrium, again in the context of game theory (although appeal to convexity has since given a simpler solution).

More recently the dual simplex linear programming model has been widely applied in economics practice, the Dantzig-Wolfe algorithm being particularly useful for large scale systems. Less restrictive cost functions can be used and generalised linear programming allows arbitrary (convex) constraints. The state space realisation is now coming to the fore in econometric and optimal control applications, though unfortunately this work has become completely detached from that of mathematical economics thus losing the structural insight that has been achieved in this field. Our examination of the dynamic Leontief model (for which useful data is available) in the light of implications of Kron's work on optimal control theory, expands considerably upon the structure of the model.

7.1 Static equilibrium

Consider an n-sector economy producing one commodity per sector where x(i) is the gross amount of the ith commodity produced in a particular time interval and X_{ij} is the amount of the ith commodity needed to produce the jth commodity. Assuming that the X_{ij} are defined so as to remove any ambiguity resulting from joint production (or by-production) then the amount of the total output of the ith sector available for investment, stockbuilding, consumption and export ie final total output is

$$y_i = x_i - \sum_{j=1}^{n} X_{ij}$$

or in matrix notation y=x-Xe where e is the n-unit vector. 'Input-output' tables (X,x,y) of the UK economy are readily available for many years with n from about 10 (highly aggregated) to about 100 (medium aggregation) though it is difficult to obtain accurate data on the intra-industry transfers X_{ii} and the leading diagonal of X is often given as zero with gross output appropriately redefined as $x(i)-X_{ii}$. Regular (yearly) and still larger tables are available for foreign economies (though full use has not been made of them because of the lack of adequate computing techniques).

Assuming that linear technical coefficients of production (ie the ratios of each component needed to make each commodity) $a(i,j)=dX_{ij}/dx(j)=X_{ij}/x(j)$ exist and are approximately constant (over a few time intervals) ie constant returns to scale then

$$A=X/diag(x)$$

where diag(x) is the diagonal matrix of x(i) and

$$y=x-Xe=x-(X/diag(x))(diag(x)e)=x-Ax=Dx$$

therefore x=/(I-A)y. Fig(7) shows the structure of the published input-output tables before and after the above factorisation.

Fig (7) The structure of published input-output tables. Either input (prices) or output (expenditure) must be shown as index numbers in the associated time series, the other being at constant prices to the i/o base year. Hence in the table input=output.

As published $\underline{e}=[1\ 1\ \ldots\ 1]'$	purchases by industry group: total intermediate output	consumers + government expenditure: capital formation + stockbuilding	exports of goods and services	total final output	total output
sales by industry group	$\underline{X}$	$\underline{Xe}$	$\underline{u}$	$\underline{Bsx}$	$\underline{n}$	$\underline{y}$	$\underline{x}=\underline{p}$
total intermediate input	$\underline{e}'\underline{X}$	$\underline{e}'\underline{Xe}=a$	$\underline{e}'\underline{u}$	$\underline{e}'\underline{Bsx}$	$\underline{e}'\underline{n}$	$\underline{e}'\underline{y}$	$\underline{e}'\underline{x}$
taxes less subsidies plus income from employment	$\underline{w}'$	$\underline{w}'\underline{e}$	taxes on expenditure, investment + exports			taxes	T+E
gross profits and trading income	$-s\underline{p}'\underline{B}$	$-s\underline{p}'\underline{Be}$	zero because profits all occur in industry sector				P
imports of goods, services + sales by final buyers	$\underline{m}'$	$\underline{m}'\underline{e}$	imports of consumption, investment and exports			imports	M
total final input	$\underline{v}'$	$\underline{v}'\underline{e}$					Y
total input	$\underline{p}'$	$\underline{p}'\underline{e}$	C+G	I+S	X	Y	Y+a

After factoring $A_{ij}=X_{ij}/x_i$	purchases by industry group: total intermediate output	consumers + government expenditure: capital formation + stockbuilding	exports of goods and services	total final output	total output
sales by industry group	$\underline{A}$	$\underline{Ax}$	$/\underline{Rp}$	$\underline{Bsx}$	$\underline{n}$	$\underline{y}$	$\underline{y}+\underline{Ax}$
total intermediate input	$\underline{p}'\underline{A}$	$\underline{p}'\underline{Ax}$	$\underline{p}'/\underline{Rp}$	$\underline{p}'\underline{Bsx}$	$\underline{p}'\underline{n}$	$\underline{p}'\underline{y}$	$\underline{p}'\underline{x}$
taxes less subsidies plus income from employment	$\underline{x}'\underline{Q}$	$\underline{x}'\underline{Qx}$	taxes on expenditure, investment + exports			taxes	T+E
gross profits and trading income	$-s\underline{p}'\underline{B}$	$-s\underline{p}'\underline{Bx}$	zero because profits all occur in industry sector				P
imports of goods, services + sales by final buyers	$\underline{m}'$	$\underline{m}'\underline{x}$	imports of consumption, investment and exports			imports	M
total final input	$\underline{v}'$	$\underline{v}'\underline{x}$					Y
total input	$\underline{p}'$	$\underline{p}'\underline{x}$	C+G	I+S	X	Y	

A is analogous to the Keynesian propensity to consume (save) relating consumption (saving) to national income. /(I-A) is therefore known as a matrix multiplier and as such is usually given an interesting economic interpretation. The initial component of production x_0 required for final demand is just y. But to produce this y we now need an extra total intermediate output $x_1=Ax_0$ which in turn needs a component $x_2=Ax_1$ etc. Hence the total production required for an output y is

$$x=\sum x_i=(I+A(I+A(...)))y=(I+A+A^2+...)y=/(I-A)y.$$

This series can be used to calculate the inverse and represented by a return difference signal flow graph. The economic output is therefore a control input and is called the demand in static equilibrium or supply in disequilibrium.

We will only deal with A matrices which are 'indecomposable' (Nikaido) or 'connected' (Schwartz) ie are not similar to

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}=/TAT$$

where A_{11} and A_{22} are square: A_{21} and/or A_{12} is a zero matrix and T is a permutation matrix. That is, an autarchy situation with respect to intermediate and/or total outputs must not exist for any subeconomy.

Now defining A=B implies $A_{ij}=B_{ij}$: A>=B implies $A_{ij}>=B_{ij}$: A≥B implies $A_{ij}>=B_{ij}$ and A≠B: A>B implies $A_{ij}>B_{ij}$. If A>=0 it is called nonnegative: A≥0, semipositive and A>0, positive. We have

(A,x,y)≥0.

The necessary and sufficient conditions for a solution were first investigated by Hawkins and Simon (1949) and Georgescu-Roegen[4] and are that all principle leading minors of the D matrix are positive ie that D is positive definite. This condition for 'workability' of the economy was extended by Nikaido who showed that all principal minors of D must also be positive.

7.2 Graph and network theory

Another interesting necessary and sufficient condition has been given by T M Whitin[2] (1954) based on the proof (derived using Grassman algebra) by Bott[1] that the determinant of the D matrix is given by the sum over all possible oriented trees of the carrier of the open oriented graph of the system y=x+Xe. Thus if a tree can be constructed for the Leontief matrix it can be shown that trees can be constructed for all leading diagonal submatrices and if the Leontief determinant is strictly positive it follows that the principle minors are all positive, fulfilling Georgescu-Roegen's 'necessary and sufficient conditions for the existence of static equilibrium'. It is economically unreasonable to suppose that a tree cannot be constructed in a workable economy for this would imply the existence of some industry that does not, directly or indirectly, receive some share of the consumer's dollar. Also Nikaido shows that A is indecomposable if and only if every production sector is directed to every other by a chain of branches ie if an oriented tree spanning only those nodes exists.

A general electrical analogy of the I-A matrix inversion problem was constructed by Franksen[12] to which tearing can be applied directly. He recognised by physical reasoning that the correct contravariant analog of electric current is commodity flow or state and that the analog of electric potential is economic price or costate, for which data is available in the input-output tables.

7.3 The dual system of prices and static optimality

Now if the price of the ith commodity y(i) is p(i) and v(i) is the value added (amount paid to labour) per unit x(i) then, assuming income eqnal to expenditure in each sector (ie no hoarding) we have

$$x_j p_j = x_j \sum_{i=1}^{n} a_{ij} p_i + x_j v_j, \quad i=1,\ldots,n \text{ or}$$

$v=p-A'p=D'p$ ie $p=/(I-A')v$.

The adjoint system has a solution iff D' is positive definite and the Brauer-Solow condition states the equivalence of workability and profitability. The existence of a tree in either the primal or the dual graph is thus a necessary and sufficient condition for the complete solution of the system. We have the dual Leontief pair

$y=x-Ax$ and $v'=p'-p'A$

Therefore $p'y=p'x-p'Ax=p'(I-A)x=v'x$

which is the Walras' law stating the equality of consumer's expenditure to income (analogous to Tellegen's theorem - the law of conservation of power).

Consider the least squares problem of minimising $x'Qx-y'Ry$ where we are weighting the total amount produced with respect to that consumed as a control variable. This represents the consumer's choice of distribution of final demand and right to work. Q and R are positive definite. We have

$(d/dy)(y'/(I-A')Q/(I-A)y-y'Ry)=0$

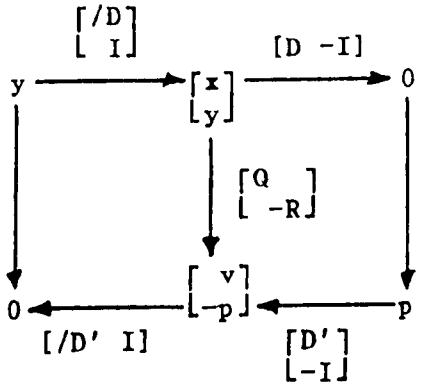
therefore $y=/R/(I-A)'Qx$.

Now from $p=/(I-A)'Qx$ it does not seem unreasonable to associate p with Ry and hence $v=Qx$. R is a matrix representing the utility of consumption of a commodity and Q, the disutility of production, both having positive diagonal elements (except possibly for waste products) and relatively small off diagonal elements. Price is at least monotonically increasing with demand for that good and decreasing with that of others: the value added (amount payed to labour) per unit produced is normally monotonically increasing (returns to scale on labour) with the amount produced of that good and decreasing with that of others. Again we have the further restriction that most variables are strictly positive, hence Franksen's treatment in terms of linear programming. Finally all the above is entirely equivalent to producers maximisation of profits with respect to wages as control (firms' right to choose distribution of prices and wages).

Now taking a Kron type approach we can reformulate as an ordinary least squares problem

$$\min \begin{bmatrix} x \\ y \end{bmatrix}' \begin{bmatrix} Q & \\ & R \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \text{ subject to } [D \ -I] \begin{bmatrix} x \\ y \end{bmatrix} = 0$$

where D=I-A as usual. The homological structure of the economic network is then as follows



and we have $[/D'\ I]\begin{bmatrix} v \\ -p \end{bmatrix}=0$

or $v=(I-A)'p$. Note the appearance of Franksen's negprices $n=-p$, defined purely from physical reasoning by Franksen ie so that commodities will only flow from a high potential to a low one, but appearing quite naturally in Roth's diagram. The connection matrices also appear in Franksen's work. Note also that the minimum $v'x-p'y=0$.

Franksen in fact ends up with equations of the form

$$x=[/D\ I]\begin{bmatrix} u \\ r \end{bmatrix} \text{ and } n=-p=[I\ -/D']\begin{bmatrix} 1 \\ w \end{bmatrix}$$

where r are the 'unemployed remainders' or factors of production going to investment and stockbuilding: 1 is the 'acquisition or opportunity cost' of investing the resources. 'The manner in which these equations have been derived clearly shows that, from the viewpoint of physical theory, they are constraints or auxiliary conditions... In electrical network theory the bilateral equilibrium constraints are but expressions of Kirchoff's two laws.' Elsewhere Franksen compares the conventional approach (which we normally use to solve the equations) with Kron's orthogonal (or complete) network approach which considers transformations of the form

$$\begin{bmatrix} x \\ u \end{bmatrix}=\begin{bmatrix} /D & I \\ I & 0 \end{bmatrix}\begin{bmatrix} u \\ r \end{bmatrix} \text{ and } \begin{bmatrix} w \\ n \end{bmatrix}=\begin{bmatrix} 0 & I \\ I & -/D' \end{bmatrix}\begin{bmatrix} 1 \\ w \end{bmatrix}$$

Further we will see in the next section that we can write

$$\begin{bmatrix} w \\ n \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & -/D \end{bmatrix}' \begin{bmatrix} /D'B'Rs & \\ & Q/BDs^{-1} \end{bmatrix} \begin{bmatrix} 0 & I \\ I & -/D \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}$$

where we have included an impedance type transformation. Franksen, though aware of the orthogonal approach never expands upon it in an economic context.

7.4 A continuous dynamic model

Leontief later extended the input-output model to include a vector of stockbuilding per unit time. Less naively we will assume the dynamics to be mainly due to returns on fixed capital investment. We have

$$\dot{x} = Ax + Bsx + u$$

where x = total output

A_x = intermediate output

u = consumer's expenditure + government expenditure + exports

B_{sx} = gross domestic capital formation and stockbuilding.

Schirnov[7] introduced the dynamic adjoint equation

$p = A'p - B'sp + w$

Where p = total input

$A'p$ = intermediate input

w = income from employment + taxes on expenditure + imports

$B'p$ = gross profits on investment and trading income.

Output and input are terms used in the tables instead of commodity flow and price. This is because both are published in units of dollars. All variables are at factor cost, that is rationalised to remove the ambiguities of economic definition. These are essentially the fundamental equations of national income/expenditure accounting. All the variables can be found in the tables from which the B matrix must be estimated and economically justified. Values for the A matrix are published at regular intervals.

The equations are those of a continuous descriptor system[14], an econometricist would say that they are in structural form. The system of equations is fully justified both on topological and economic grounds. It models transients in the economy due to fluctuations in the distribution of supply or demand. For balanced growth, assuming that money is neither injected into nor taken out of the economy, it will be seen that this is consistent with a state of deflation - the basis of monetary policy. From the first equation and an optimality criterion on the consumers we will now demonstrate that the second equation can be derived. Assume by analogy with the static case that a cost function of the form

$\int x'Qx-u'Rudt$

is minimised where R and Q are symmetric positive definite. That is labour minimises total output with consumption as control. In the same way industry maximises a cost function of the form $\int p'Rp-w'/Qwdt=\int p'u-w'xdt$ that is maximises profits = turnover - wages. This is the dual cost function of control theory. Then from Parseval's theorem, writing

$x=/(I-A-Bs)u=Gu$

We have $u=/RG*Qx=/R/(I-A+Bs)'Qx$

and, assuming linearity we can associate u with /Rp giving w=Qx (elasticity of supply and demand) and

$p=/(I-A+Bs)'w=A'p-B'sp+w.$

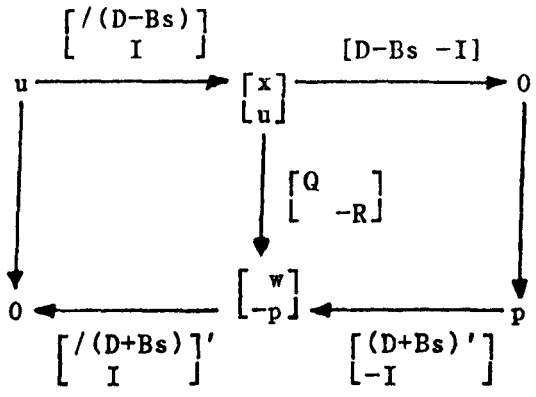
In a similar way it can be shown that a state of optimality exists for the producers. This situation is known in economics as dynamic equilibrium in that there is no incentive for either party to change their policy. p has assumed the role of a Lagrange multiplier and the system minimises the Hamiltonian

$\int x'Qx-u'Ru+p'(x-Ax-Bsx-u)dt.$

Taking a Kron type approach the equivalent ordinary least squares problem is

$$\min \begin{bmatrix} x \\ u \end{bmatrix}' \begin{bmatrix} Q & \\ & -R \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \text{ subject to } \begin{bmatrix} D-Bs & -I \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = 0$$

and Roth's diagram looks like this



From the diagram $\begin{bmatrix} /(D+Bs)' & I \end{bmatrix} \begin{bmatrix} w \\ -p \end{bmatrix} = 0.$

Now the gross domestic product

$$p'x=p'(Ax+Bsx+u)$$

$$=(p'A-sp'B+w')x.$$

Therefore the difference between total income and total expenditure

$$w'x-p'u=sp'Bx+p'Bsx=s(p'Bx)$$

assuming B constant. This is the Walras Law in the dynamic case. It gives the effect of hoarding on returns from investment. It is equivalent to dH/dt in Pontryagin's maximum principle where H is the Hamiltonian.

Further $w'x-p'u=x'Qx-u'Ru=s(p'Bx)$

and $\min\int x'Qx-u'Rudt=p'Bx.$

Defining a Riccati matrix K by

$p=K(t)Bx$ then $B'sp=B'sKBx+B'KBsx$

assuming B is constant and from

$\dot{x}=Ax+Bsx+/Rp$

$\dot{p}=A'p-B'sp+Qx$

it is not difficult to show that

$B'sKBx=B'K/RKBx+(A'-I)KBx+B'K(A-I)x+Qx.$

This must be true for all x giving the dynamic matrix Riccati equation

$B'sKB=B'K/RKB+(A'-I)KB+B'K(A-I)+Q$

showing that K is symmetric as one would expect. Livesey has previously derived a simpler form of this equation. By comparing coefficients we can introduce the Chandrasekar equations.

$$\begin{bmatrix} sY'B & sW \\ B'sKB & B'sY \end{bmatrix}=\begin{bmatrix} Y'(/RKB-A+I) & Y'/RY \\ B'K/RKB+(A'-I)KB+B'K(A-I)+Q & (B'K/R-A+I)Y \end{bmatrix}$$

For an infinite planning horizon sK=0 is well known and

B'K/RKB+(A'-I)KB+B'K(A-I)+Q=0.

Now consider the special case of zero hoarding,

w'x-p'u=x'Qx-u'Ru=s(p'Bx)=0.

Therefore x'B'K/RKBx=x'Qx for all x

therefore B'K/RKB=Q

which involves n(n+1)/2 equations in n(n+1)/2 unknowns which will, assuming R and Q are constant, almost always give time invariance for K, showing that zero hoarding at least involves an infinite planning horizon, further it gives a zero value for the cost function. Substituting in the time invariant Riccati equation

B'K(I-A)+(I-A')KB=2Q.

This can be solved using the tensor algorithm of Chapter VI. Now for Lyapunov stability of the system it is sufficient to find a positive definite function V(t) such that sV(t) is negative definite. Let

$$V=\begin{bmatrix}x\\p\end{bmatrix}'\begin{bmatrix}B\\-B'\end{bmatrix}'\begin{bmatrix}P_a & P_b\\P_c & P_d\end{bmatrix}\begin{bmatrix}B\\-B'\end{bmatrix}\begin{bmatrix}x\\p\end{bmatrix}$$

where $\underline{P}$ is arbitrary. Then

$sV=s(\underline{x}'\underline{B}'\underline{PB}\underline{x})=s\underline{x}'\underline{B}'\underline{PB}\underline{x}+\underline{x}'\underline{B}'\underline{PB}s\underline{x}$

where $\begin{bmatrix}B\\-B'\end{bmatrix}s\begin{bmatrix}x\\p\end{bmatrix}=\begin{bmatrix}I-A & -R\\-Q & I-A'\end{bmatrix}\begin{bmatrix}x\\p\end{bmatrix}$

and writing $sV=\underline{x}'\underline{L}\underline{x}$ we have $\underline{L}=$

$$\begin{bmatrix}I-A' & -Q\\-R & I-A\end{bmatrix}\begin{bmatrix}P_a & P_b\\P_c & P_d\end{bmatrix}\begin{bmatrix}B\\-B'\end{bmatrix}+\begin{bmatrix}B'\\-B\end{bmatrix}\begin{bmatrix}P_a & P_b\\P_c & P_d\end{bmatrix}\begin{bmatrix}I-A & -R\\-Q & I-A'\end{bmatrix}=$$

$$\begin{bmatrix} (I-A')P_a B - QP_c B + B'P_a(I-A) - B'P_b Q & -(I-A')P_b B' + QP_d B' - B'P_a/R + B'P_b(I-A') \\ -/RP_a B + (I-A)P_c B - BP_c(I-A) + BP_d Q & /RP_b B' - (I-A)P_d B' + BP_c/R - BP_d(I-A') \end{bmatrix}$$

and letting $\underline{B}'\underline{P}\underline{B} = \begin{bmatrix} B'KB & 0 \\ 0 & 0 \end{bmatrix}$

we have $\underline{L} = \begin{bmatrix} B'K(I-A) + (I-A')KB = 2Q & -B'K/R \\ -/RKB & 0 \end{bmatrix}$

Then if $\begin{bmatrix} B'KB & 0 \\ 0 & 0 \end{bmatrix} > 0$ and $\begin{bmatrix} 2Q & -B'K/R \\ -/RKB & 0 \end{bmatrix} < 0$ we have stability.

Otherwise, as is necessary for economic growth, we have instability - in a control theory sense.

Similarly letting $\underline{B}'\underline{P}\underline{B} = \begin{bmatrix} P & -B' \\ B & 0 \end{bmatrix}$ gives $L_a = 0$.

Finally it is interesting to note the four conditions

(1) $p'u = w'x$ ... no money withdrawn by hoarding

(2) $s(p'Bx) = 0$ ... a condition on the money injected into the economy or withdrawn by legislation

(3) $u = /Rp$ and $w = Qx$ ... linearity

(4) $\min \int x'Qx - u'Rudt = 0$ ... optimality

7.5 The discrete time model

Leontief's original work was in discrete time

$$x(t)=Ax(t)+B(x(t+1)-x(t))+u(t).$$

We can introduce the discrete analog of the adjoint system

$$p(t+1)=A'p(t+1)+B'(p(t)-p(t+1))+w(t+1)$$

where $u(t)=/Rp(t)$

$$(t+1)=Qx(t+1)$$

and the variables are defined as above. To prove the second equation from the first one take

$$\int_n(x'Qx-u'Ru)$$

$$Ax+B(z-1)x+u=/(I-A-B(z-1))u=G(z)u.$$

For optimality by the discrete Parseval theorem

$$/RG'(z^{-1})Qx$$

≈/R(I-A'-B'(z^{-1}-1))Qx

&iving p=/(I-A'-B'(z^{-1}-1))w

herefore p=A'p+B'(z^{-1}-1)p+w QED.

ow rewriting the equations

I-A+B)x(t)=Bx(t+1)+/Rp(t)

-A'+B')p(t+1)=B'p(t)+Qx(t+1)

d putting p(t)=K(t)x(t)

e matrix manipulation gives the discrete matrix Riccati equation

)=/(B/((I-A+B)'K(t+1)-Q)B'+/R)(I-A+B).

e equation can be made symmetrical by writing either

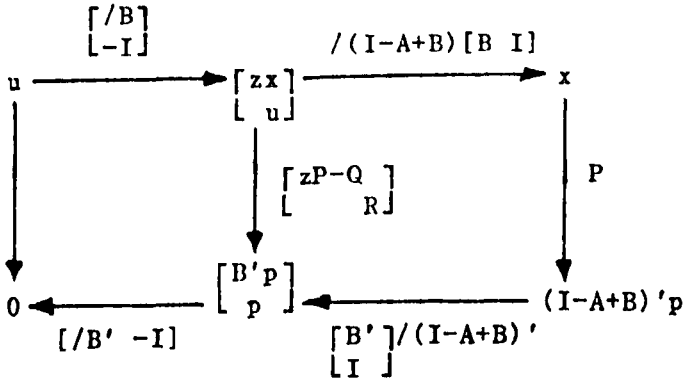
I-A+B)'K or P=K/(I-A+B).

the first case Roth's diagram can be written down from the ordinary least squares

lem

$$\min \sum H(t) \text{ where } \begin{bmatrix} zx \\ u \end{bmatrix}' \begin{bmatrix} zP-Q & \\ & R \end{bmatrix} \begin{bmatrix} zx \\ u \end{bmatrix}$$

subject to $[B\ I]\begin{bmatrix} zx \\ u \end{bmatrix}=(I-A+B)x$ as follows



and $P=(I-A+B)'/([B\ I]/\begin{bmatrix} zP-Q \\ R \end{bmatrix}\begin{bmatrix} B' \\ I \end{bmatrix})(I-A+B)$

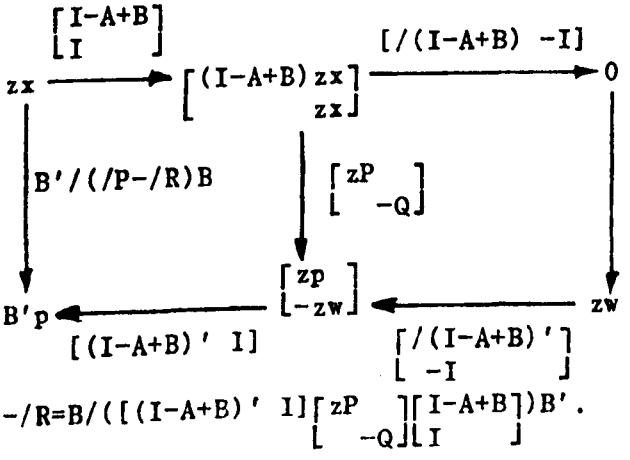
which is the first symmetrical version of the backward Riccati equation. This can be reversed from left to right and treated as a section torn out of a space filter as we have seen in the optimal control case, hence the boundary operator can be found.

For the second case we have

$$H(t)=\begin{bmatrix} zp \\ -zw \end{bmatrix}'/\begin{bmatrix} zP & \\ & -Q \end{bmatrix}\begin{bmatrix} zp \\ -zw \end{bmatrix}$$

subject to $[(I-A+B)'\ I]\begin{bmatrix} zp \\ -zw \end{bmatrix}=B'p$

Roth's diagram is



and $/P-/R=B/([(I-A+B)'\ I]\begin{bmatrix} zP & \\ & -Q \end{bmatrix}\begin{bmatrix} I-A+B \\ I \end{bmatrix})B'$.

It can be seen that in fact the equations of optimal control are a special case of the equations of input-output analysis, in the sense that they have less structure.

7.6 History

Leontief[5, 6] derived the discrete commodity flow equation only and showed how the system of equations could be built up into a partitioned matrix and solved stepwise. He considered that for a closed system it would be necessary to introduce behavioural equations, and appears to make no mention of the pricing mechanism. Mathur (1964) applied this analysis to the economy of India.

Pontryagin[7] (USSR, 1970) in his 'optimal interbranch model of socialist reproduction' starts from the continuous dynamic model and after defining a Hamiltonian function of social utility introduces the dual vector 'as optimal programming of economic resource prices.' From Pontryagin's maximum principle the equations sx=dH/dx and sp=-dH/dx give the adjoint system. This analysis is closest to our own. Brody follows similar lines and also considers time optimal paths.

Livesey[10] (1971) describes the application of control theory to input-output analysis in the Cambridge Growth Model. Starting from continuous dynamic commodity flow he notes that Theil's decision rules define a quadratic utility function in $\int x'Qx+y'Ry dt$ and hence derives the Chandrasekar-Riccati equations

$dY/dt=YA-YB/RB'K$ and $-dK/dt=Q+KA+A'K-KB/RB'K$

using the control theory definitions of A and B. He notes that 'any model of the economic system that involves quantities and prices is nonlinear' and further quotes Morishima: 'Once initial output and final demand are known the entire evolution of the system is defined' - a two point boundary value problem.

Brody[8] (Hungary, 1972) follows Goodwin and 'simply equates price increases with excess demand and production increases with profit.' (He is however aware of the pricing mechanism.) This produces a skew symmetric matrix with pure imaginary eigenvalues except for the equilibrium point. The vectors p and x will rotate on a 2n dimensional hypersphere. Cross multiplying and adding gives p'sp+x'sx=0. Brody states that this sum of squares has no economic meaning and further that the mass of empirical data gives no indication of pure sine oscillations. He then invokes measure invariance and reckons in relative rather than absolute quantities giving the Volterra-Goodwin equations

$$(sp)/p=(-Cx)/x \text{ and } (sx)/x=(C'p)/p$$

where by abuse of notation a/b=[a(i)/b(i)]. Cross multiplying and adding now gives $p'x+p'sx=s(p'x)=0$ ie the price of total output is kept constant. These equations have no explicit (closed, analytic) solution but can readily be simulated on a computer and show many phenomena readily observable in real economic systems. The system was used to simulate the US economy over the period 1958-1968.

Franksen's[12] (1969) important contribution has been dealt with above. Teldahl's[13] (1975) peculiar relativistic distributed economic system appears to be modelled along the same lines but via De Rham's theorem. This was apparantly used in practice on a model of the Swedish economy with 121 sectors.

Finally a note on the estimation of B, Q and R. This chapter is at least a justification for approximating the economy with a second order differential/autoregressive model. The problem is the estimation of the capital coefficient matrix, as figures for intermediate output are not available for non input-output years. John Sutton of the Department of Economic Studies, Sheffield University has suggested that intermediate output should be estimated by subtracting the factored sum of government expenditure, consumers expenditure and capital formation from total output in the Treasury's 'Blue Book'.

7.7 References

(1) Bott and Mayberry, Matrices and Trees, in Economic Activity Analysis, Morgenstern, Wiley, 1954, p391.

(2) Whitin, An economic application of 'Matrices and Trees', ibid, p401.

(3) Debreu, Theory of value: an axiomatic study of economic equilibrium, Wiley/Cowles, 1959.

(4) Nikaido, Convex structures and economic theory, Academic press, 1968.

(5) Leontief, An open dynamic system for long range projection of economic growth, in Economic analysis in an input-output framework, 1964.

(6) Leontief, The dynamic inverse, in Brody and Carter, Input-output techniques, North-Holland, 1972.

(7) Smirnov (USSR Academy of Sciences), Problems of constructing an optimal interbranch model of socialist reproduction, ibid, Part I Dynamic analysis, 1972.

(8) Brody (Hungarian Academy of Sciences), Optimal and time optimal paths of the economy, ibid, p62.

(9) Brody (Univ Zambia), An input-output model of the market, Ibid p574.

(10) Livesey (Univ Cambridge), Control theory and input-output analysis, 1971, Fifth international conference on input-output techniques, Geneva. Also personal communication.

(11) Vogt et al, A dynamic Leontief model for a production system, Proc IEEE, Vol 63, No 3, March 1975.

(12) Franksen, Mathematical programming in economics by physical analogies, Simulation. Part I, The analogy between engineering and economics, June 1969. Part II, The economic network concept, July 1969. Part III, System equilibrium and mathematical programming, August 1969.

(13) Teldahl, Tensors and Sociodynamics, Matrix and Tensor Quarterly. A preamble to Tensors and Sociodynamics, Vol 24, No 4, June 1974, pp 133-4. Part I: The Postulate of Bi-Duality, Vol 25, No 2, Dec 1974, pp 48-58. Part II: A Socio-Dynamic Model, Vol 25, No 4, June 1975, pp 122-35. Part III: An Economic Interpretation, Sept 1976, pp 15-29. Reprinted as Tensors and Sociodynamics, 1975, MTQ Reprint, Sweden. Also a condensed version of Socio-Ekonomisk Futurologi, mimeograph, University of Stockholm,

Institute of National Economy, 734 pp, Stockholm, Sweden, Oct 1973.

(14) Yip and Sincovec, Solvability, Controllability and Observability of Continuous Descriptor Systems, IEEE Trans AC-26 No 3 June 1981.

(15) Franksen, On the Formulation of an Economic Equation of Motion, Matrix and Tensor Quarterly, Dec 1974, pp 39-47.

CHAPTER VIII General Homology Theory: the Structure of Reality?

We give the formal definition of a cochain complex. The Eilenberg-Steenrod axioms for General Homology Theory in an Admissable Category and the de Rham Thercom are stated. The axioms can be related to the definitions of the more specific objects used elsewhere in this work. We start with a discussion attempting to show why homology theory describes the mathematics of so many diverse subjects. We conclude that it is more than the mathematics of physics, it describes the structure of information.

8.1 Homology Theory

Homology theory is the mathematics of reality: that is, at least, much of the reality we experience is described by homology theory. As such it is a superset of the mathematics of physics. Conventional science describes the evolution of an 'objective' universe in R^3. Define a vector $\underline{k}(x,y,z,t)$ representing a point in the space-time continuum, ie the electric, magnetic, gravitational and other fields at that point, its mass density and the direction in which that mass is moving and spinning, its charge, etc, then $\underline{k} \subset K$ represents at least from the point of view of Newtonian physics, the entire evolution of the universe (the 'great machine theory'.) This structure is Misnor and Wheeler's superspace and is described by homology theory.

Twentieth century physics has shown that life is rather more complicated than this. Einstein's theory shows that as our viewpoint gets larger the universe, the macrocosm, starts to distort (badly). Quantum theory says that as our viewpoint gets smaller the universe gets grainy: microcosmic decisions can be predicted only statistically. Further Bell's theorem (1962) proved that either the statistical predictions of quantum theory - or that some of our other ideas - were wrong. Quantum theory won out, the Clauser-Freedman experiment (1972) amongst others, has shown that Einstein's Principle of Separability (action at a distance) one of the basic axioms of Western Science fails. Aspect's experiment (1982) showed that there were no hidden variables involved, leading to the conclusion that we live in one of many parallel universes (Moorcock's multiverse, otherwise known as the many worlds theory.) The multiverse holds at least on the medium scale and is described by homology theory.

Consider the set K of all conceivable evolutions of the universe, ie the set whose elements are evolutions of the universe disregarding the laws of physics, common sense, etc. The universe at any point in space-time bears no special relationship to that immediately ahead or behind, ie K can take any value at that point. We must look at subsets of this superspace. Consider the set of all conceivable states of the universe. A subset of this is the set of all possible (or syntactically correct) states, ie states of the universe that obey the laws of physics etc. Now consider the subset of K representing all physically possible evolutions of the universe Q: further consider the subset R of Q containing all physically possible evolutions of the universe through the point $\underline{k}(x,y,z,t_o)$, the universe as we know it exists now. This can be considered an optimal set in some sense. Homology theory allows us to study the quotient sets K\Q, K\R and Q\R. The structure R becomes more hazy as we look further into the past or future. Classically there is a double cone in space-time (Bergman's paradigm) representing what we see as 'now' in the past (light travelling from a distant star has taken a

finite time to arrive) and the limits to what we can predict as the future. There is a similar cone in R which contains a double decision tree, going both forwards and backwards in time through $\underline{k}(x,y,z,t_0)$. The tree splits in the future every time a quantum decision is made: it appears to be symmetrical in the past due to our inability to trace our own history accurately. Decision trees of this sort can be used to represent subjective realities rather than the objective universe. It can be seen that these structures are supersets of the multistage decision problems of dynamic programming. For example binary replication of DNA involves a decision tree of a similar form to the above. The subsets of all conceivable evolutions of DNA can be defined. In a similar way a sociological space, considering the world as a sphere of interacting intelligences can be defined. In fact any decision tree in a dynamic programming sense is described by homology theory. Kron distinguished between the state (or network) space in which he worked, and real space, ie. the universe. As these are both special cases of superspace R we can see that the two concepts are not so intrinsically different. Thus homology theory describes more than just physics, it can describe subjective problems: problems with deep structure and subjective criteria - in this sense it is the mathematics of reality. This technique of predicting all possible futures is the basis of the geometric method in control theory.

We have defined three special cases of General Homology Theory: The first case is Singular Homology Theory which describes topological networks with impressed functions. DeRham Cohomology Theory describes the isomorphic system of functions impressed on topological spaces. Axiomatic sheaf cohomology theory is the third case: it is used for more abstract problems, eg to prove the de Rham theorem. It is built of chain complexes, an abstraction of the polytopes used in singular homology theory, and thus serves as a generalisation of the structure of abstract optimisation problems. The remarkable thing is that these <u>are</u> all special cases of General Homology Theory - and that General Homology Theory is based on a

(structurally) simple set of axioms. And that these axioms therefore appear to be the basis for a description (of at least a large part) of reality as we know it.

Homology provides an algebraic picture of topological spaces, assigning to each space X a family of Abelian groups $H_0(X),\ldots,H_n(X),\ldots$, and to each continuous map $f:X\to Y$, a family of group homomorphisms $f_n:H_n(X)\to H_n(Y)$. Properties of the space or the map can often be found from properties of the groups H_n or the homomorphisms f_n. A similar process associates homology groups to other mathematical objects, for example, to a group or to an associative algebra.

Complexes provide a means of calculating homology. Each n-dimensional 'singular' simplex T in a topological space X has a boundary consisting of singular simplices of dimension n-1. If K_n is the free abelian group generated by all these n-simplices, the function d assigning to each T the alternating sum dT of its boundary simplices determines a homomorphism $d:K_n\to K_{n-1}$. This yields a 'complex' which consists of abelian groups K_n and boundary homomorphisms d, in the form

$$0\leftarrow K_0\xleftarrow{d}K_1\xleftarrow{d}K_2\xleftarrow{d}K_3\xleftarrow{d}\ldots.$$

Moreover $dd=0$, so the kernel Z_n of $d:K_n\to K_{n-1}$ <u>contains</u> the image dK_{n+1}. The factor group $H_n(K)=Z_n\backslash dK_{n+1}$ is the nth homology group of the complex K or of the underlying space X. Often a smaller or simpler complex will suffice to compute the same homology groups for X. Given a group G there is a corresponding complex whose homology is appropriate to the group, for example the one dimensional homology of G is its factor commutative group $G\backslash[G,G]$.

Homomorphisms of appropriate type are associated with each type of algebraic system. Under composition of homomorphisms the systems and their homomorphisms constitute a category. If A and C are abelian groups, the set Hom(C,A) of all group homomorphisms $f:C\to A$ is also an abelian group. For C fixed, it is a

covariant functor on the category of all abelian groups A, each homomorphism a:A->A' induces the map a*:Hom(C,A)->Hom(C,A') which carries each f into its composite af with f.

$$\begin{matrix} f & a \\ C & \to A \to A' \end{matrix}$$

induces a*:Hom(C,A)->Hom(C,A')

which takes a:f->af

with f<Hom(C,A), af<Hom(C,A'), a<A*.

For A fixed Hom is contravariant. Each c:C'->C induces the map c* in the opposite direction, Hom(C,A)->Hom(C',A), sending f to the composite fc.

$$\begin{matrix} c & f \\ C' & \to C \to A \end{matrix}$$

induces c*:Hom(C,A)->Hom(C',A)

which takes c:f->fc

with f<Hom(C,A), fc<Hom(C',A), c<c*.

Thus Hom(?,A) applied to a complex K=? turns the arrows around to give a complex

$$\mathrm{Hom}(K_0,A) \xrightarrow{d^*} \mathrm{Hom}(K_1,A) \xrightarrow{d^*} \mathrm{Hom}(K_2,A) \to \ldots$$

so if $a(i):K_i \to A$ then $a(i+1) \rangle a(i)d:K_{i+1} \to A$ and the adjoint $d^*:a(i) \to a(i+1)$ where d* means postmultiply by d or in the matrix case where the contravariant variables are treated as row vectors, the conjugate transpose of d. Here the factor group $\ker(d^*) \backslash \mathrm{im}(d^*)$ is the nth cohomology group $H^n(K,A)$ of K with coefficients A. According to the provenance of K, it yields the cohomology of a group G or a space

X.

Definition A cochain complex C^* consists of a sequence of K-modules and homomorphisms

$$\ldots \to C^{q-1} \to C^q \to C^{q+1} \to \ldots$$

defined for all integers q such that at each stage the image of a given homomorphism is contained in the kernel of the next. The homomorphism $C^q \to C^{q+1}$ (which we will refer to as d^q or simply d) is called the qth coboundary operator. $dd=0$ but the sequence is not long exact. The kernel $Z^q(C^*)$ of d^q is the module of qth degree cocycles of the cochain complex C^*, and the image $B^q(C^*)$ of d^{q-1} is the module of qth degree coboundaries. The qth cohomology module $H^q(C^*)$ is defined to be the quotient module

$$H^q(C^*)=Z^q(C^*)\backslash B^q(C^*)=\ker(d^q)\backslash \mathrm{im}(d^{q-1}).$$

This of course only makes sense if the sequence is not long exact.

Let C^* and D^* be cochain complexes. A cochain map $C^* \to D^*$ consists of a collection of homomorphisms $C^q \to D^q$ such that for each q, the diagram

$$\begin{array}{ccc} C^q & \to & D^q \\ \uparrow & & \uparrow \\ C^{q+1} & \to & D^{q+1} \end{array}$$

commutes. It follows that a cochain map sends the module of q-cocycles of C^* into the module of q-cocycles of D^* and maps the module of q-coboundaries of C^* into the module of q-boundaries of D^*, and thus induces a homomorphism of the cohomology modules

$$H^q(C^*) \to H^q(D^*).$$

The composition $C^*->E^*$ of two cochain maps $C^*->D^*$ and $D^*->E^*$ induces on the cohomology modules the homomorphism $H^q(C^*)->H^q(E^*)$, which is the composition of $H^q(C^*)->H^q(D^*)$ and $H^q(D^*)->H^q(E^*)$. A sequence of cochain maps

$$0->C^*->D^*->E^*->0$$

forms a short exact sequence if for each q,

$$0->C^q->D^q->E^q->0$$

is a short exact sequence of K-modules. A homomorphism between short exact sequences $0->C^*->D^*->E^*->0$ and $0->\underline{C}^*->\underline{D}^*->\underline{E}^*->0$ of cochain complexes consists of cochain maps $C^*->\underline{C}^*$, $D^*->\underline{D}^*$ and $E^*->\underline{E}^*$ such that we have a commutative diagram

$$\begin{array}{ccccccccc} 0 & -> & \underline{C}^* & -> & \underline{D}^* & -> & \underline{E}^* & -> & 0 \\ & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & -> & C^* & -> & D^* & -> & E^* & -> & 0 \end{array}$$

Proposition Given the short exact sequence $0->C^*->D^*->E^*->0$ of cochain maps, there are homomorphisms

$$H^q(E^*)\xrightarrow{d}H^{q+1}(C^*)$$

for each q such that the sequence

$$...->H^{q-1}(E^*)\xrightarrow{d}H^q(C^*)->H^q(D^*)->H^q(E^*)\xrightarrow{d}H^{q+1}(C^*)->...$$

is exact, and such that given the above homorphism of short exact sequences of chain complexes, the following diagram is commutative

$$\begin{array}{ccc} H^q(\underline{E}^*) & \xrightarrow{d} & H^{q+1}(\underline{C}^*) \\ \uparrow & & \uparrow \\ H^q(E^*) & \xrightarrow{d} & H^{q+1}(C^*) \end{array}$$

It is shown in Warner[2] that the required homomorphism is $d:Z^q(E^*)\backslash B^q(E^*)\to H^{q+1}(C^*)$. The proof by 'diagram chasing' is easy but tedious.

An extension of a group A by a group C is a group $B\supset A$ with $B\backslash A\sim C$, in diagrammatic language, an extension is just a sequence

$$E:0\to A\to B\to C\to 0$$

of abelian groups and homomorphisms in the sense that the kernel of each homomorphism is exactly the image of the preceding one. B can be considered as a product of C and A. The set $Ext^1(C,A)=\{B\}$ of all extensions of A by C turns out to be an abelian group and a functor of A and C, covariant in A and contravariant in C.

Does the homology of the complex K determine its cohomology? The answer is almost yes, provided each K^n is a free abelian group. In this case $H^n(K,A)$ is determined 'up to a group extension' by $H_n(K)$, $H_{n-1}(K)$ and A. Specifically the Universal Coefficient Theorem (a special case of the Kunneth theorem, which treats tensor and torsion products) gives an exact sequence

$$0\to Ext^1(H_{n-1}(K),A)\to H^n(K,A)\to Hom(H_n(K),A)\to 0$$

or $Hom(H_n(K),A)\sim H^n(K,A)\backslash Ext^1(H_{n-1}(K),A)$

involving the functor Ext^1 just introduced. If the K^n are not free groups there is a more complex answer, involving 'spectral sequences'.

8.2 Admissable categories

Let X be a nonempty set. A class T of subsets of X is a topology on X iff T satisfies the following axioms:

(1) X and {0} belong to T.

(2) The union of any number of sets in T belongs to T.

(3) The intersection of any two sets in T belongs to T.

The members of T are called open sets. (X,T) is a topological space.

By a topological pair[1] (X,A) we mean a topological space X and a subspace A<X. In case A is the empty subspace 0<X we will not distinguish between the topological pair (X,0) and the topological space X although they are logically different. Thus the topological spaces are special cases of topological pairs. By a subpair (X',A')<(X,A) we mean a pair st X'<X and A'<A. The three spaces X,A,0 together with their inclusion relations 0<A<X make up six pairs (0,0), (A,0), (X,0), (A,A), (X,A), (X,X). These are the pairs associated with (X,A). In case A=0 or X=A some of these pairs become equal. By a map f:(X,A)->(Y,B) we mean a (continuous) map f:X->Y satisfying f(A)<B. In case A=0 the condition f(A)<B is always satisfied and hence every map from (X,0) into (Y,B) is just a map f:X->Y. In particular we shall not distinguish between the map f:(X,0)->(Y,0) and the map f:X->Y. For example, let (X',A')<(X,A). The inclusion map i:X'->X obviously satisfies the condition i(A')<A and hence is a map from (X',A') into (X,A). In case (X',A')=(X,A), i is the identity map on (X,A). Transitivity: consider any three pairs (X,A), (Y,B), and (Z,C). If f:(X,A)->(Y,B), g:(Y,B)->(Z,C) are maps then the composition h=g.f

satisfies $h(A)=g(f(A))\subset g(B)\subset C$ and hence is a map $h:(X,A)\to(Z,C)$. Next consider an arbitrary $f:(X,A)\to(Y,B)$. Let $(X',A')\subset(X,A)$ and $(Y',B')\subset(Y,B)$ satisfy $f(X')\subset Y'$ and $f(A')\subset B'$. There exists $g:(X',A')\to(Y',B')$ defined by taking $g(x)=f(x)$, for any $x\subset X'$. In case $(Y',B')=(Y,B)$, g is called the restriction of f to $(X',A')\subset(X,A)$ denoted $g=f|(X',A')$. By the lattice (see Wonham) of an arbitrary (X,A) we mean the six associated pairs together with all the identity maps on the pairs and the inclusion maps in the diagram

$$\begin{array}{ccccc} & & (X,0) & \to & (X,A) & \to & (X,X) \\ & & \uparrow & & \uparrow & & \\ (0,0) & \to & (A,0) & \to & (A,A) & & \end{array}$$

By the cylinder $(X,A)xI$ over (X,A) we mean the pair (XxI,AxI) which consists of the topological product XxI of X with $I=[0,1]$, the closed unit interval and its subspace AxI. The maps $k_0,k_1:(X,A)\to(X,A)xI$ defined by $k_0(x)=(x,0)$ and $k_1(x)=(x,1)$ are canonical imbeddings of (X,A) into $(X,A)xI$.

By an admissable category[1] for a homology theory we mean a category, C whose objects are topological pairs and whose morphisms are maps of topological pairs satisfying

(1) If (X,A) is an object in C, then C contains the lattice of the pair (X,A).

(2) If $f:(X,A)\to(Y,B)$ is a morphism in C then C contains (X,A) and (Y,B) together with all the maps that f defines from members of the lattice of (X,A) into that of (Y,B).

(3) If $f:(X,A)\to(Y,B)$ and $g:(Y,B)\to(Z,C)$ are morphisms in C then C contains $g.f:(X,A)\to(Z,C)$.

(4) If (X,A) is an object in C then C contains the cylinder (X,A)xI and the two canonical imbeddings k_0, k_1 of (X,A) into (X,A)xI.

(5) There is a singleton space, 0 in C. If 0 is any singleton space in C then C contains every f:0->X<C.

For example the category of all cellular pairs and maps of cellular polytopes (CW complexes) such that A is a subpolytope (or subcomplex) of X is an admissable category. The category of all pairs (X,A) and all maps of such pairs is the largest admissable category. Each example contains all singleton spaces. Choose the distinguished singleton space 0 in condition (5) as the space that consists of all real numbers 0 only and denote it by 0 even though this may not always be true. Let C be an arbitrary admissable category. Then (X,A) is admissable iff it is in C. Similarly for f:(X,A)->(Y,B). Consider f,g:(X,A)->(Y,B). Then f and g are homotopic in C iff there exists an admissable h:(X,A)xI->(Y,B) st $f=h.k_0$ and $g=h.k_1$ (continuously deformable). In the examples f and g are homotopic in C iff they are homotopic in the usual sense.

8.3 The Eilenberg-Steenrod axioms

Let C denote an admissable category. By a homology theory[1] on C we mean a collection of three functions

$$\underline{H}=(H,d,*)$$

H assigns to each (X,A) in C and each q in I, the set of integers, an abelian group $H_q(X,A)$ which will be called the q-dimensional homology group of the topological pair (X,A) in the homology theory, (or the q-dimensional (relative) homology group of the topological space X modulo its subspace A.) In case A=0 it is called the

q-dimensional (absolute) homology group of the space X.

* assigns for any $f:(X,A)\to(Y,B)$ in C and for any $q<I$ a homomorphism

$$f_*=f_{*q}:H_q(X,A)\to H_q(Y,B)$$

the homomorphism induced by f in $\underline{H}$.

d assigns for any (X,A) in C and for any $q<I$, a homomorphism

$$d=d(X,A,q):H_q(X,A)\to H_{q-1}(A)$$

the boundary operator on $H_q(X,A)$ in $\underline{H}$.

Further H,* and d are required to satisfy the seven axioms:

(1) Identity. If $i:(X,A)\to(X,A)$ is the id map on (X,A) in C then the induced homomorphism $i^*:H_q(X,A)\to H_q(X,A)$ is the id automorphism of $H_q(X,A)$ for any $q<I$.

(2) Composition. If $f:(X,A)\to(Y,B)$ and $g:(Y,B)\to(Z,C)$ are maps in C then

$$(g.f)_{*q}=g_{*q}.f_{*q}$$

for any $q<I$. Hence for any q, $(H_q,*q)$ constitutes a covariant functor from C to A, the category of all abelian groups and homomorphisms. Writing $H_q(f)=f_{*q}$ for any f in C the functor is denoted by H_q and is called the q-dimensional homology functor in $\underline{H}$.

(3) Commutativity. If $f:(X,A)\to(Y,B)$ in C defined by $g(x)=f(x)$, for any $x<A$ then $d.f^*=g^*.d$ ie

$$\begin{array}{ccc} H_{q-1}(A) & \xrightarrow{g^*} & H_{q-1}(B) \\ d\uparrow & & d\uparrow \\ H_q(X,A) & \xrightarrow{f^*} & H_q(Y,B) \end{array}$$

for any $q<I$.

(1) to (3) are the algebraic axioms.

(4) Exactness. If (X,A) in C and $i:A\to X$ and $j:X\to(X,A)$ denote inclusion maps then the infinite sequence

$$\ldots\to H_q(A)\xrightarrow{i^*}H_q(X)\xrightarrow{j^*}H_q(X,A)\xrightarrow{d}H_{q-1}(A)\to\ldots$$

of groups and homomorphisms, the homology sequence of (X,A) is exact.

(5) Homotopy. If $f,g:(X,A)\to(Y,B)$ in C are homotopic in C then $f_{*q}=g_{*q}$, for any $q<I$.

(6) Excision. If $U<X$ is open and the closure $Cl(U)<Int(A)$ the interior of $A<X$ and if the inclusion, $e:(X-U,A-U)\to(X,A)$ is in C then the induced homomorphism $e_{*q}:H_q(X-U,A-U)\to H_q(X,A)$ is an isomorphism for any $q<I$. e is the excision and e_{*q} the q-dim excision homomorphism.

(7) Dimension. $H_q(0)=0$, $(q\#0)$ for any $q<I$.

The first six axioms constitute a generalised homology theory. $G=H_o(0)$ is the coefficient group of $\underline{H}$.

In a similar way we can define generalised (contravariant) cohomology theories with coboundary operators using superscripts. For example taking $C=\{(Z,B)\}$ where $Z=\{Z^q(C^*)\}$, $B=\{B^q(C^*)\}$ and

$$H^q(Z,B)=Z^q(C^*)\backslash B^q(C^*)$$

$$d:H^q(C^*)\to B^{q+1}(C^*)$$

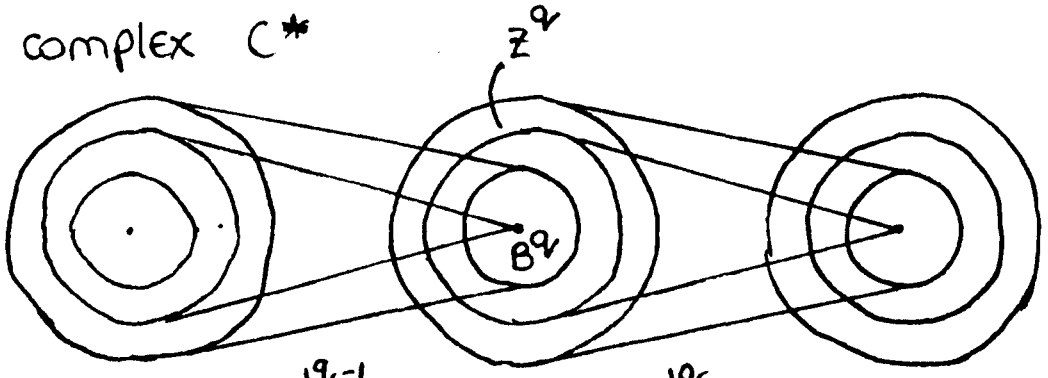
we have constructed the singular cohomology theory, see Fig(8.1). Construction of homology for discrete optimal control is shown in Fig(8.2). Fig(8.3) shows the twisted isomorphisms between homology and cohomology sequences and Fig(8.4) gives the complete space filter including homology and scattering sequences. We finish this discussion in Chapter X and show how the axioms may be related to definitions in optimal control theory.

8.4 The De Rham Theorem

An n-dimensional manifold is a space which is not necessarily a Euclidean space nor is it a domain in Euclidean space, but which from the viewpoint of the short sighted observer living in the space, looks just like a domain in Euclidean space. For example the 2-sphere cannot be considered part of the Euclidean plane, however our observer on the sphere sees that he can describe his immediate vicinity by two coordinates and so fails to distinguish between this and a small domain in the plane. Let M be a paracompact differentiable manifold. Its differentiable structure will not be invoked. Let K be a fixed principle ideal domain. The most important cases are (1) K is the ring of integers when a K-module is an abelian group, (2) K is the field of real numbers when a K-module is a real vector space.

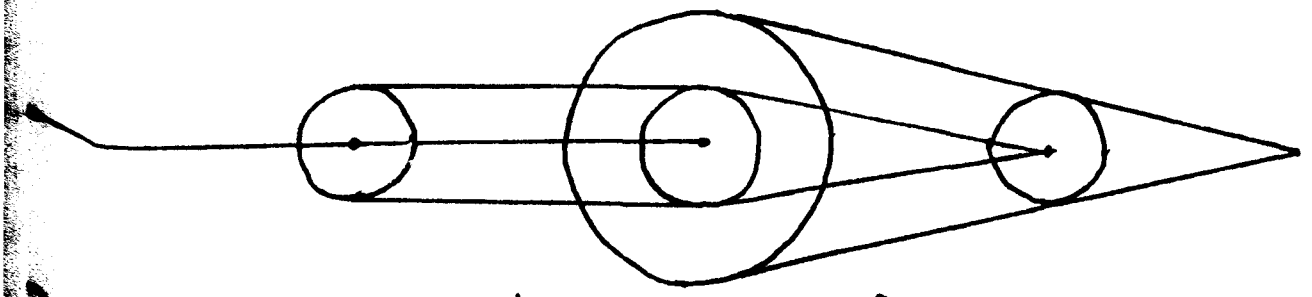
Fig(8.1) Construction of exact sequences from complexes

chain complex C^*



$$\cdots \longrightarrow C^{q-1} \xrightarrow{d^{q-1}} C^q \xrightarrow{d^q} C^{q+1} \longrightarrow \cdots$$

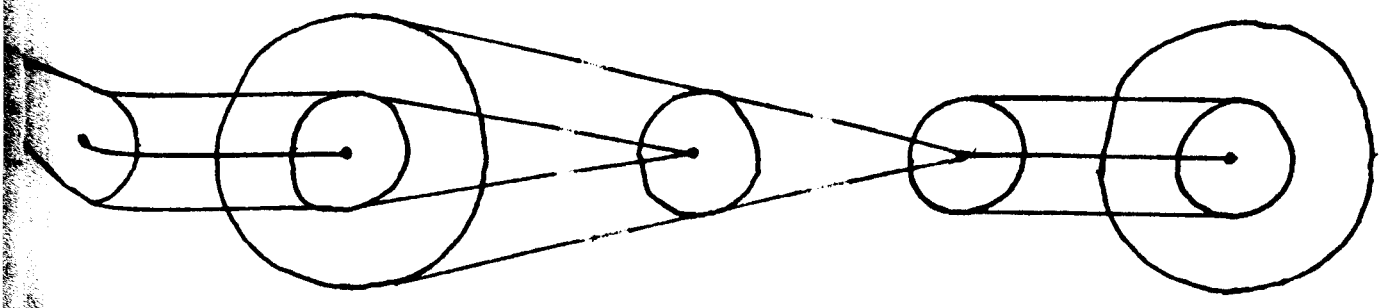
duces a series of short exact sequences $q = 0, 1, 2 \ldots$



$$\longrightarrow \underset{=\mathrm{im}(d^{q-1})}{B^q(C^*)} \xrightarrow{i} \underset{=\ker(d^q)}{Z^q(C^*)} \xrightarrow{p} \underset{=\frac{Z^q(C^*)}{B^q(C^*)}}{H^q(C^*)} \longrightarrow 0$$

hich are related by the boundary operators since if

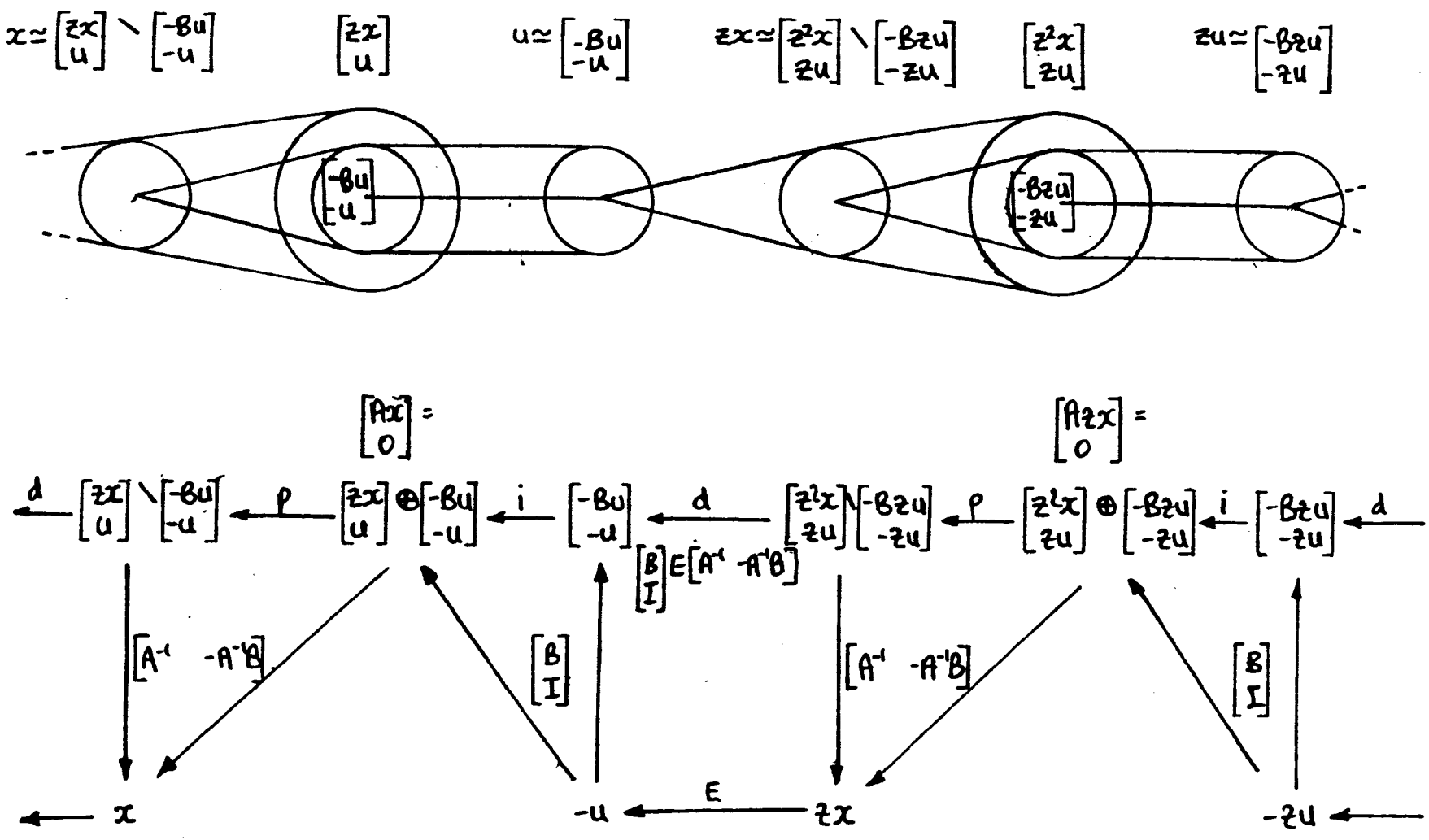
$d^q : Z^q \to 0$ then $d^q : H^q \to 0$



$$B^q \xrightarrow{i} Z^q \xrightarrow{p} H^q \xrightarrow{d} B^{q+1} \xrightarrow{i} Z^{q+1} \xrightarrow{p} \cdots$$

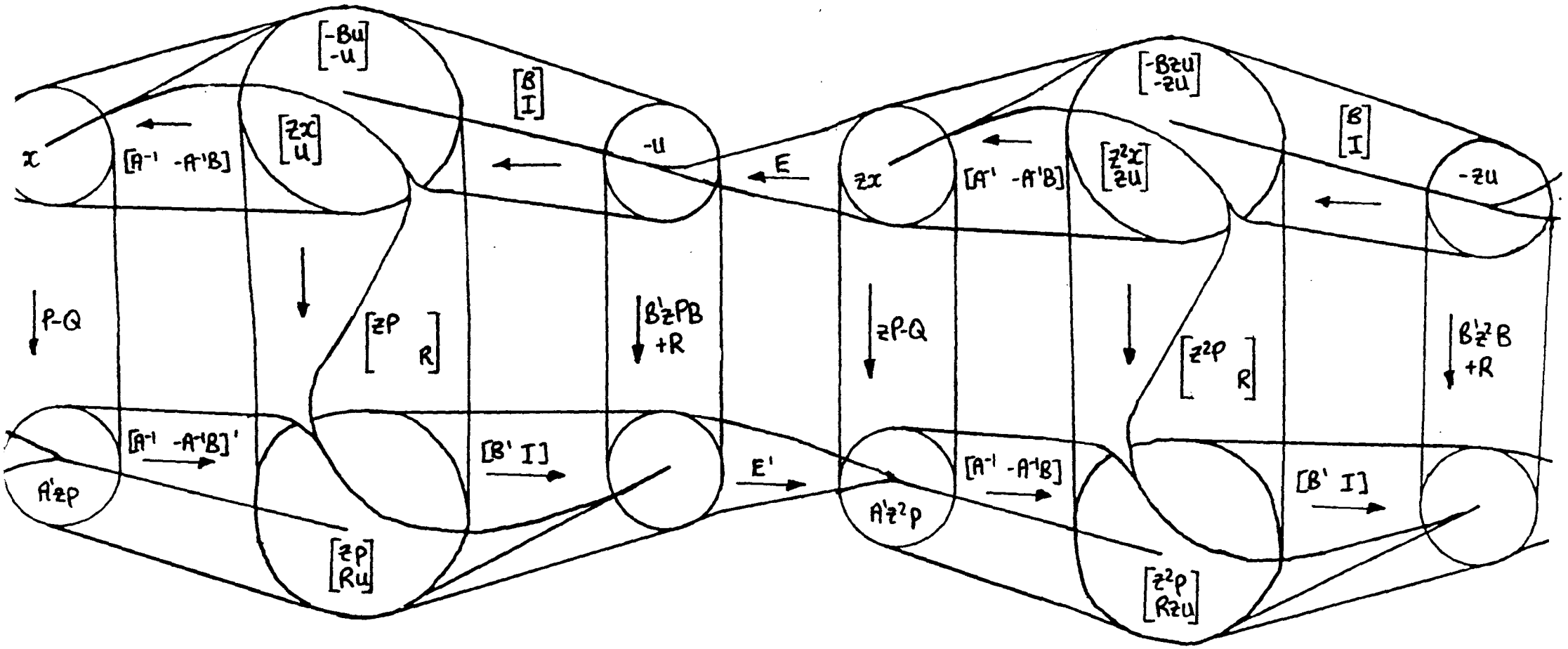
ving a long exact sequence which is associated with

$$H^q(B) \longrightarrow H^q(Z) \longrightarrow H^q(Z,B) \xrightarrow{d} H^{q+1}(B) \longrightarrow H^{q+1}(Z) \longrightarrow \cdots$$



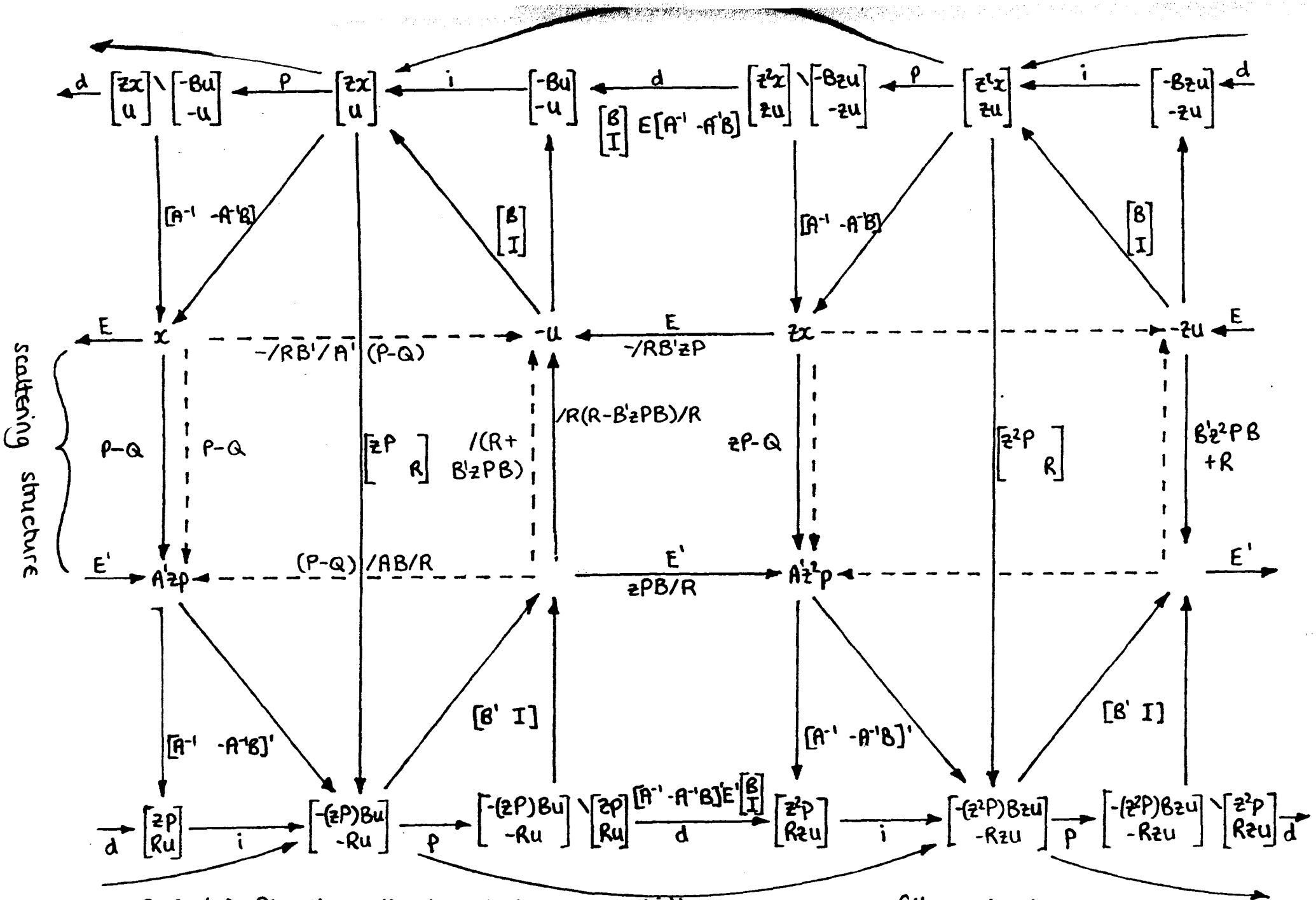
$$x = \begin{bmatrix} A^{-1} & -A^{-1}B \end{bmatrix} \left(\begin{bmatrix} zx \\ u \end{bmatrix} \setminus \begin{bmatrix} -Bu \\ -u \end{bmatrix} \right) = \begin{bmatrix} A^{-1} & -A^{-1}B \end{bmatrix} \left\{ \begin{bmatrix} zx \\ u \end{bmatrix} + \left\{ \begin{bmatrix} B \\ I \end{bmatrix} u \right\} \right\} = \begin{bmatrix} A^{-1} & -A^{-1}B \end{bmatrix} \left\{ \begin{bmatrix} zx \\ u_1 \end{bmatrix} + \begin{bmatrix} B \\ I \end{bmatrix} u_2 \right\}$$ which is independant of u_2

fig(8.2) Substructure of discrete optimal control as a long exact sequence



Fig(8.3) Internal substructure of discrete optimal control as a multidimensional space filter

What happens at the ends?



Fig(8-4) Discrete optimal control as a multidimensional space filter showing homology sequences and scattering structure

A p-form a on a differentiable manifold M is called closed if da=0, where d is the differential operator. It is called exact if there is a (p-1) form b such that a=db. Since dd=0, every exact form is closed. The quotient space of the real vector space of closed p-forms modulo the subspace of exact p-forms is called the pth de Rham cohomology group of M.

$$H^p_{deR}(M)=(\text{closed p-forms})\backslash(\text{exact p-forms})$$

For each integer $p\geq 0$ we let ${}_{inf}S_p$ denote the real vector space generated by the differentiable singular p-simplices in M. Hence the elements of ${}_{inf}S_p(M,R)$ are precisely the differentiable singular p-chains in M with real coefficients. For $p<0$ we let ${}_{inf}S_p(M,R)$ be the zero vector space. The boundary operator d induces linear transformations

$$d_p : {}_{inf}S_p(M,R) \to {}_{inf}S_{p-1}(M,R)$$

for each integer p, which for $p\leq 0$ are simply the zero transformation. Now dd=0, and the pth differential singular homology group of M with real coefficients is defined by

$${}_{inf}H_p(M,R)=\ker(d_p)\backslash\mathrm{im}(d_{p+1}),$$

and is moreover a real vector space.

We define a linear mapping of the de Rham cohomology into the dual space of the real differential singular homology

$$H^p_{deR}(M) \to {}_{inf}H_p(M,R)^*$$

The de Rham theorem states that this is an isomorphism[2]. It is the isomorphism (information preserving map) between distributed or continuous systems and their appropriate discrete representations. De Rham's theorem shows that any such system, is isomorphic to any homological torn form of itself.

8.5 References

(1) Hu, Homology Theory, Holden Day, 1966.

(2) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresman, 1971.

(3) Vick, Homology Theory, Academic Press, 1973.

(4) Flanders, Differential Forms with Applications to the Physical Sciences, Academic Press, 1963.

(5) Bourgin, Modern Algebraic Topology, Macmillan, 1963.

(6) Hilton and Wylie, Homology Theory, Cambridge University Press, 1960.

CHAPTER IX Computer Programming: Orthogonal Languages

A little accepted fact in the computer field is that programming languages have developed almost as rapidly over the last thirty years as has hardware. Unfortunately Basic and Fortran (and Cobol in the commercial world) are still the most widely used languages, though with the advent of microprocessors Pascal is becoming more popular. Note that Pascal does not stand for 'Programming and Scientific Computational Algorithmic Language' but was in fact named after Blaise Pascal!

Early computers were programmed in machine language. This rapidly became too laborious for writing long programs and the first high level (or source) language, Fortran, was written and compilers developed which could translate it into machine (or object) code. Basic was the first interactive language, that is instead of compiling the source into object code and then running it, the program is treated as a sequence of commands to be obeyed by a program known as an 'interpreter'. Compiled code is in general much faster - though compilation itself may be slow, it only has to be done once. Good Basic systems provide both an interpreter and a compiler - debugging can be carried out interactively with the interpreter eg single lines of program may be executed either directly or indirectly, and then the program compiled for fast execution. Lisp and Pop-2 interpreters even allow source to be read into the user program and executed, allowing program changes at run time.

Algol60 was the first of the structured, modular languages - as opposed to the sequential nature of Fortran - and was biased more towards ease of use than efficiency. By structured we mean having a general, English-like, nested syntax. By modular we mean a procedure orientated language. Compared to Fortran or Basic they are easier to read, easier to program, easier to maintain, easier to translate to other language, easier to debug, there is less chance of hidden bugs, degenerate

cases tend to take care of themselves (you don't find yourself having to think 'what if....') and it discourages 'spaghetti' coding. A lot is talked in the literature on programming techniques such as stepwise refinement, though flow charts are not approved of - developing a feel for structured programming is most important. The emphasis is on short procedure definitions rather than long sequential programs.

The art then developed in two directions. Pascal was designed to be the most efficient structured language possible, while Algol 68 became the first ORTHOGONAL language and will probably remain the most general programming language ever written for the near future - it allows definition of the most abstract quantities and operators possible - and although compilers remain large and slow, the code produced for production runs can be as fast as that from Fortran. Except for minor differences in syntax Pascal, like Algol 60 and to a certain extent BCPL and the real time language Coral66, is a subset of Algol 68. ADA[14] is becoming the most popular structured language in the USA, it is based on Algol68 but with a more Pascal like syntax, generic typing and real time facilities thrown in.

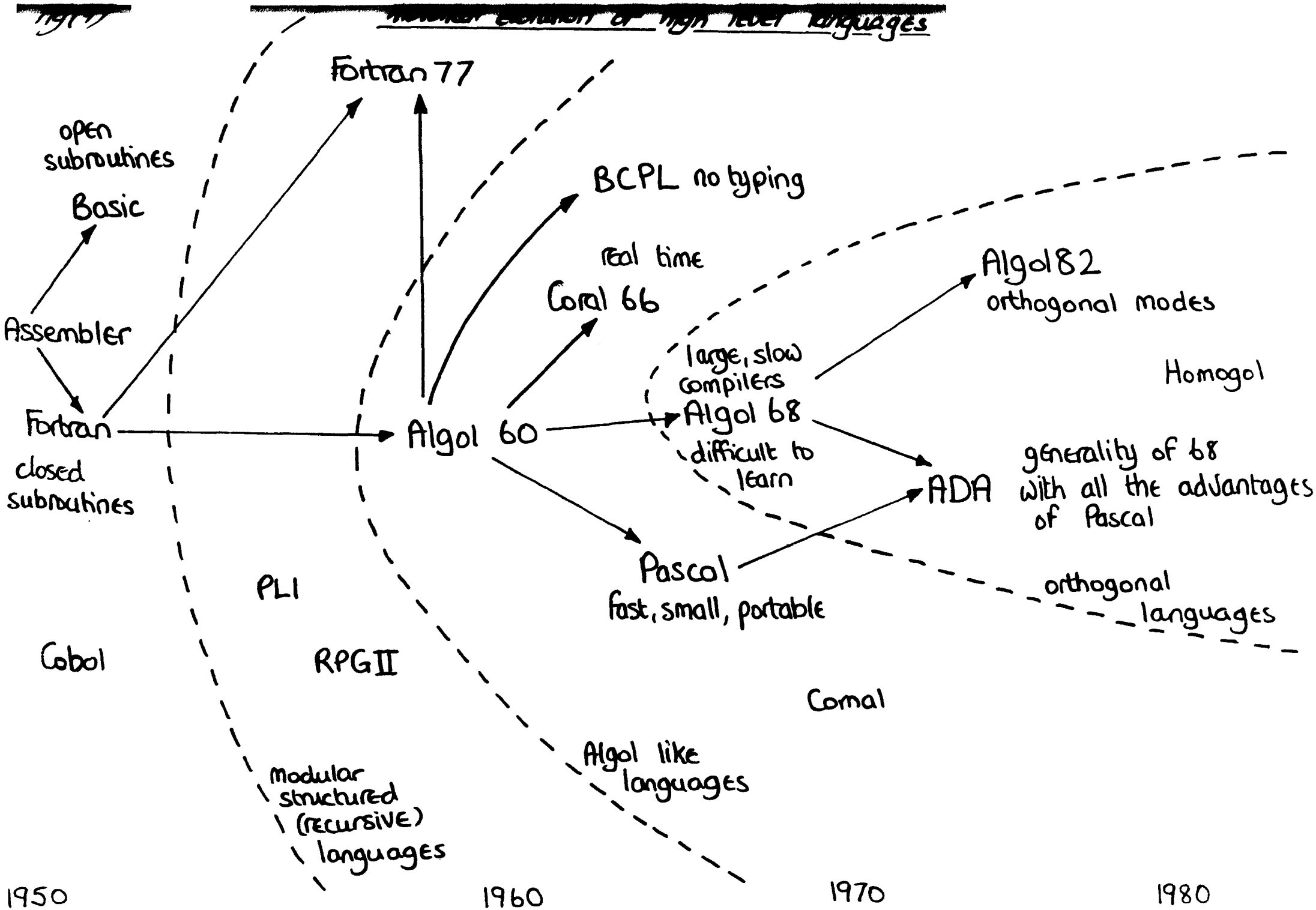
9.1 Comparison of languages

Almost all theory of the type discussed in this thesis is conventionally programmed as matrix algorithms in Fortran. Dynamic arrays (whose size may be specified at run time) and flexible arrays (whose size may be changed at run time) are not allowed. Matrices are added, multiplied, inverted etc. by calling library subroutines, usually with one operation per line. Hence much computer memory is wasted as matrices must be dimensioned to their largest likely size, and programs tend to be long and unreadable, one matrix equation taking up many lines with explicit work space being defined for the result of each stage. Algol 60 allows

dynamic arrays but not flexible arrays, and further allows recursive procedure calls rationalising many function definitions. Pascal does not have dynamic arrays but does have an explicit pointer system whereby arrays of fixed sizes may be specified as types and created at run time on the 'heap' (similar to a stack) and destroyed when no longer needed. If procedures whose results are pointers to arrays are defined recursively this process can be made semiautomatic. Algol 68 removes all these difficulties. Dynamic and flexible arrays of any rank are allowed. An automatic garbage collector destroys these when they are no longer needed and the space on the heap is wanted for something else. Slices of arrays and arrays of arrays can be manipulated and arbitrary (monadic or infix diadic) operators may be defined allowing matrix equations to be programmed exactly how they are written optimising writing, reading, correcting and maintainance of programs. The same operator may even be defined between objects of different type, the results being understood from the context. Undefined operations may be used, the definition being understood by a set of rules known as coercion. Objects may even simultaneously have more than one type. Sylvester's expansion theorem gives arbitrary functions of a matrix. Further Algol 68 objects may be defined in any category necessary - sets, groups, rings, modules and their morphisms may be manipulated. Covariant and contravariant types may be defined. It should further be possible to define objects representing arbitrary homology theories or having the structure represented by Kron's algebraic diagram for the space filter or polyhedron model or the scattering structure (Nicholson suggested implementing this as hardware) or perhaps a procedure called ROTH which would perform the manipulations achieved by reading around the squares in Roth's diagram. This would represent an extremely general scientific program, perhaps a software library or computer language, which would be able to deal with any topological or optimisation problem that was a special case of the theory.

Pascal is more usually available as a P-code interpreter than a compiler, with non of the usual advantages. Pascal is first translated to P-code and then interpreted. This is slow and non-interactive, and does not allow mixed language programming. Their existence is partly historical - Pascal was designed for a hypothetical stack orientated machine which executed P-code - and partly due to ease of implementation and transportability. Pascal is rapidly becoming available on most machines that offer Fortran, and many that don't - ie microprocessors. It is rapidly and satisfactorily replacing Fortran for both applications and systems work - which does not mean to say of course that one stops using programs that are already written in Fortran! In fact mixed language programming enables one to call Fortran libraries from Algol or Pascal. These are a highly liked teaching languages, much preferred to Fortran, and give a strong feel for other structured languages. Unlike Fortran, and particularly, as microcomputer users will know, Basic, Pascal has a highly standardised syntax - it does not vary between machines. UCSD Pascal is often available as a complete system including compiler library, dedicated editor, debugger, assembler, compiler or interpreter, linker, filing system and graphics (historically usually turtle (r,theta) graphics). Pascal compilers are usually one pass compilers with look ahead and are hence very fast, and produce very efficient code. Also because unlike Algol 60, the language was designed to achieve this, they can be quite small. Interpreters tend to be at least faster than Basic. An interesting point with high level languages is that the errors tend mainly to come out in the compilation stage. Execution errors are rare with Algol 68. Lastly Pascal has a heap and some degree of garbage collection is specified in the Report, though this is rarely implemented. Algol68 garbage collection is automatic. The main disadvantages of Pascal are that it does not allow dynamic (and hence flexible) arrays, you cannot define your own operators and the assignment operator does not work on structured types.

The two main computing environments that have developed then are commercial data processing and scientific programming. A third field that is rapidly progressing is that of Artificial Intelligence. Here there are two main schools of thought. One is that expert systems which can carry out specific tasks eg. in robotics should be developed. The other is that it should be possible to develop general intelligent programs using the theory of natural language originating from image processing and language translation - the study of syntax and semantics. Computer scientists have developed a number of modular, structured, highly recursive, interpretive programming languages for AI work, Spitbol (Snobol IV) is a pattern matching language, Lisp is a theoretically based language having as its core (the minimum number of basic operators) only five symbols, and Pop-2 combines the facilities of a Lisp type language with some of those of Algol 68. An interpreter for Lisp can be written in Lisp in a remarkably short number of lines. The ideal aim of the natural language community is to develop a (minimal) program that can learn, (in a general sense) - the core of an intelligent program. That is that for example, rather than programming the computer to understand English, a program could be written that could be taught to understand it. The Lisp interpreter itself was a major breakthrough in this sense though it cannot be taught in English, only in Lisp. The mathematics of natural language is moving more towards the structures used in homology theory - Kron predicted this when talking about the space filter as a set of self organising polyhedra, a model for an artificial brain.


Fortran 77
open subroutines
Basic
BCPL no typing
real time
Coral 66
Assembler
Algol 82
orthogonal modes
large, slow compilers
Homogol
Fortran
Algol 60
Algol 68
difficult to learn
generality of 68 with all the advantages of Pascal
ADA
closed subroutines
Pascal
fast, small, portable
orthogonal languages
PL1
Cobol
RPG II
Comal
Algol like languages
modular structured (recursive) languages
1950
1960
1970
1980

9.2 Automatic matrix manipulation

There are roughly two kinds of people who deal with matrices on computers. The first bracket is the requirement for processing large amounts of numerical data. The most common examples are in statistics, for which large packages such as SPSS or GENSTAT are available, Linear Programming, for which large packages such as XDLA (ICL) and smaller programs are available, and transformations such as Eigenvalue Analysis and Matrix Inversion. The other bracket is matrix manipulation, usually the requirement for writing algorithms, in fields such as Time Series Analysis, Network Analysis, the Finite Element Method, Decomposition Theory and Modern Control Theory. We stress that for large amounts of manipulation on relatively small matrices, using relatively small amounts of machine time, for non production run jobs and testing algorithms, Algol 68, Pascal or even Basic are much better languages than Fortran. Consider the following example. A common simple problem in estimation is, given a rectangular matrix, $\underline{A}$ and a vector $\underline{y}$, to find the best estimate $\underline{x}^0$ of $\underline{x}$ such that $\underline{y} = \underline{A}\underline{x}$. The least squares solution is

$$\underline{x}^0=(\underline{A}'\underline{A})^{-1}\underline{A}'\underline{y} \quad \ldots \quad (1).$$

9.2.1 Basic

This is certainly the simplest language to use from the point of view of a beginner. Its disadvantages are that it is slow, inaccurate, will only take one matrix operation per line and requires to be given workspace. The program can be designed using an interpreter (Prime Basic or ICL SOBS) and compiled under VBASIC or JBAS for production runs. The example program would look something like this.

```
 10  input n,m
 20  dim a(m,n), x(n),y(m),ws(n,n), at(n,m)
 30  mat  input a
 40  mat  input y
 50  mat at=trn (a)
 60  mat ws=at*a
 70  mat ws=inv(ws)
 80  mat x=at*y
 90  mat x=ws*x
100  mat print x
```

9.2.2 Fortran

A typical Fortran[1, 3, 6 - 10] semicompiled subroutine library for matrix manipulation would include the following routines: set a matrix to zero or the unit matrix, add, subtract, multiply two matrices (of appropriate size), scalar multiplication, inverse, determinant, transpose, read or write a matrix, set one matrix equal to another, negate a matrix, eigenvalue analysis, and perhaps exponentiation or redimension. These subroutines must be supplied with the input and output matrices, their real and virtual sizes (if different and redimension not available) and sufficient workspace. Typical examples of Fortran matrix subroutine libraries are: ICL Scientific Subroutines, National Algorithms Group (the NAG library - also available for Algol60 and 68), the

Sheffield University Control Engineering library (ICL only) and Melsa and Jones' Linear Control Theory package. For further information see the appropriate reference. The subroutines can be included in a Fortran program by including the appropriate library statement before the program description segment and compiling using FORTRANL or FORLOAD under Maximop. The Sheffield Control Engineering package is the most general as regards manipulating arrays and provides facilities for dynamically redimensioning Fortran matrices inside the calling segment[3] to minimise storage of arbitrarily sized arrays without recompiling the program. The NAG library includes a large set of accurate and efficient routines for inversion, determinants and eigenvalue analysis particularly for complex and special sparse matrices. Matrix manipulation in Fortran on a Prime machine is horrific, particularly with arrays over 64K, nevertheless it is widely used.

Using the Sheffield library matrix manipulation is made almost as simple as using Basic and it is fast. An optimising compiler is available for still greater speed. Matrices can be redimensioned at run time (under TRACE 2) using

```
call redimension(ind,nar,a,b,...,nds,n1,n2,...)
```

where
- ind — is a constant initially set to zero.
- nar — is the number of arrays, with the same number of dimensions, and size, to be redimensioned.
- a,b,.. — are the arrays to be altered.
- nds — is the number of dimensions and
- n1,n2,.. — is the size of each dimension.

The example program would look something like this:

```
      master
      dimension a(1,1),x(1,1),y(1,1),ws(1,1),at(1,1)
      read(1,10)n,m
 10   format(2i0)
      ind = 0
      call  redimension (ind,1,a,2,m,n)
      call  redimension (ind,1,x,2,n,1)
      call  redimension (ind,1,y,2,m,1)
      call  redimension (ind,1,ws,2,n,n)
      call  redimension (ind,1,at,2,n,m)
      call  mread (a,m,n)
      call  mread (y,m,1)
      call  mtrans (at,a,m,n)
      call  mmult (ws,at,a,n,m,n)
      call  minv (ws,ws,n,n)
      call  mmult (x,at,y,n,m,1)
      call  mmult (x,ws,x,n,n,1)
      call  mwrite (x,n,1)
      stop
      end
      finish
```

Note that each matrix must have <u>two</u> dimensions.

9.2.3 <u>Algol 68</u>

Even using the Sheffield library the example is hardly an elegant way of implementing equation (1). Algol 68[2, 10, 13] provides this elegance. It is by far the most appropriate language for large scale matrix manipulation (large sets of small matrices). Its disadvantage is that compilation involves running a large slow program. The automatic garbage collector for flexible (dynamically redimensionable) arrays is quite an overhead on execution.

To understand how the garbage collector works one must have some idea of the memory structure involved. Typically, above the program area is a static data space consisting of variables, arrays etc. defined at compile time. Above this

is an area called the heap which is initially empty. Every dynamic data allocation at run time will create a new area of workspace on the heap of appropriate type. Often, eg. when redimensioning a flexible array or reassigning a pointer away from the heap, areas will become unused. Unless these are cleared up the heap will eventually collide with the stack (used for return addresses etc.) which is simultaneously growing down from the top of memory. A routine called the 'garbage collector' is therefore initiated at regular intervals (particularly when space is at a premium) to 'scavenge' redundant areas on the heap either by reordering the pointers or shifting the data around - depending on the algorithm used - thus delaying or preventing the crash.

The least squares problem would be programmed as follows:

```
'with'matlib'from'alib'
'begin'
  'int'm,n; read ((m,n));
  [1:m,1:n]'real'a;[1:n]'real'x;[1:m]'real'y;
  in(a);in(y);
  x:=/('trans'a * a)*'trans'a*y;
out(x)
'end'
'finish'
```

The simplification is yet more noticeable with larger problems. The algorithm of Ref(4) has been implemented, as has a general linear time series model estimation program, with ease. The library MATLIB is available from the author[2].

Notes

(1) Duplication of data can be avoided by using references to arrays. However real arrays must be used as procedure parameters when used as workspace.

(2) Flexible arrays may be created and destroyed eg. by x:=((0)) the overhead is the garbage collector. Experiments carried out at Norwich indicate that compiled Algol68 is as fast as Fortran: However as soon as a reference to the garbage collector is made run time increases by a factor of around 40%. This though does not get much worse with further calls until a situation is achieved where the heap is so small that scavenging occurs on virtually every operation.

(3) When a vector of matrices is required it must be defined [1:n]'ref'[,]'real'x; and each element filled up like so: x[i]:='loc'[1:m,1:m]'real'; assignment must then be made to the x[i] by a forced coercion eg. ('ref'[,]'real''val'x[i]):='unit'm;

(4) The 'proc' sylvester = ('proc'('real')'real',[,]'real)[,]'real: gives any scalar function of a matrix.

(5) By defining addition, inner and outer multiplication etc for 'mode''tensor'=[,,,]'real', much more general manipulations can be carried out.

(6) Using 'mode''object' = 'union'('real',[]'real'[,]'real',[,,]'real',[,,,]'real') and defining addition, multiplication etc. between objects of 'mode''object' the size of the necessary library is considerably lessened.

(7) Despite all this it is important to avoid duplication of number crunching - for instance performing inv(a) more than once in a line for the sake of clarity leads to much more computer time being used. That is the only thing Algol68 will not do is optimise your algorithm for you.

9.2.4 Pascal

Pascal[3, 10, 11, 12] is not such a natural language as Algol68 for matrix manipulation, nevertheless it has advantages over Fortran, and it is very generally available nowadays. A Pascal version of MATLIB is available at NELP. The following notes cover most salient points.

(1) Matrix operators cannot be defined. Procedures do not return anything. Functions may not return arrays but may return pointers to arrays. Therefore all manipulations must be done in Forward Polish Notation (easier than it sounds).

(2) Flexible arrays are not allowed. Even dynamic arrays are not allowed. This does not make programs too inflexible as the dimensions may be specified initially as constants.

(3) Function and procedure parameters must be type specified only in advance. They therefore may take arrays of one size only. Occasionally this is quite satisfactory and results in efficient use of store. This restriction does not apply in the Paisley College implementation of the Amsterdam Prime V-mode Pascal compiler.

(4) If matrices of more than one size are to be dealt with - or if flexible matrices are required - the following type may be used

```
type matrix = ↑ record
     x:array[1:M,1:N]of real;
     m,n:integer end;
```

where M and N are constants equal to the largest dimensions used (this is the old Fortran trick but it now has to produce an array larger than any matrix used) and m and n are variables representing the current desired size of the matrix. This can be very wasteful on space. The array part of a variable var a:matrix may now be referred to as a↑.x, its elements as a↑.x[i,j], and its bounds as a↑.m and a↑.n. The with statement can be useful in this context.

(5) Functions which return pointers to objects must create their own workspace if these objects are to be referenced later. This workspace cannot have the same name as the function has, as is allowed in Algol 68. eg.

```
function trn(x:matrix):matrix;
    var z:matrix;i,j:integer;
    begin
        new(z);
        for i:=1 to x↑.m do for j:=1 to x↑.n do
        z↑[j,i]:=x↑[i,j]; (* data moves*)
        z↑.m:=x↑.n; z↑.n:=x↑.m;
        trn:=z  (*pointer move*)
    end;
```

The procedure new does two things. It creates workspace for the pointer z to point to, and it makes z point to this space. Garbage collection is very different to Algol68. Two other procedures, mark and release (or sometimes dispose) are supplied which allow the user to store the current size of the heap and to release space no longer in use. One has to be careful not to create redundant areas on the heap with no associated pointer as these cannot be destroyed. If the heap and the stack collide an execution error will occur.

(6) Given the appropriate matrix library the program may now look like this.

```
program pseud(input,output);
   const size = 10 (*say*);
   type matrix = ↑record
        x:array[1:size,1:size] of real;
        m,n :integer end;
   var a,y,x:matrix;
   function inv(x:matrix):matrix;extern;
   function trn(x:matrix):matrix;extern;
   function mm(x,y:matrix):matrix;extern;
   procedure inn(var x:matrix);extern;
   procedure out(x:matrix);extern;
   begin new(a);new(y);new(x);
       inn(a);inn(y);
       x:=mm(inv(mm(trn(a),a)),mm(tm(a),y));
       out(x)
   end.
```

(7) Note that during the evaluation of the nested functions the values of each bracketed subexpression must still exist. The garbage collector may be called at the end of each expression to reclaim this workspace. It is then important to make sure that the contents of any pointers to be later referenced are not destroyed i.e. before calling the garbage collector we must not assign

new(x); x:= expression; but x↑:= expression↑.

This last statement is invalid in Pascal for structured objects like matrix: we must therefore define an explicit copy function which puts the results of 'expression' into the workspace which x points to. Finally a distinction must be made between objects of mode matrix created for convenience and presumably not to be destroyed by the garbage collector and workspace created by calling functions of mode matrix which should be destroyed when no longer wanted.

9.3 The Homology Theorem

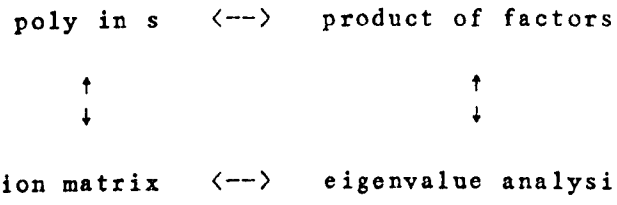
The problem with implementing the homology theorem is that there appear to be a number of alternative modes of attack. Sets, groups, rings and modules may be implemented as Algol68 structures, and operators between objects in these categories defined. Further the type 'mode''category' may be defined and coerced to a simpler object when necessary. An abstract system (in the sense of Kalman) may also be defined as an Algol68 structure. The adjoint system is given by the homology theorem. The isomorphism between the two is given by the cost function. The object is to make the implementation as non-mechanistic as possible. It is possible to set up a library of structures, operators and procedures, equivalent to the axioms of the categories desired, then by defining the known part of the (arbitrary) system structure it will be possible to solve for any well-defined part of the system remaining simply by calling the relevant procedure, from a program. We are thus working with three levels of data: the specified category (mode or type), the values of the homomorphisms (eg matrices) in that category, and the values of the objects in the category (eg initial conditions of vectors).

Homomorphisms are procedures. A distinction is necessary between eg, premultiplication of a vector by a matrix using an infix * operator defined for arrays, and calling a procedure that has the same effect, upon the vector. It is possible to define an infix operator for composition of homomorphisms - and hence of procedures!

```
'op' . =('proc'('category')'category'f,g)'proc'('category')'category':
```

The unary adjoint operator, * may be defined for particular categories, (obviously matrix transpose etc) or better, for general categories ... presumably in terms of a procedure Hom(x,y) the group of homomorphisms from X to Y. The ring of polynomials is a K[z] module. It is quite easy to define addition, multiplication (time convolution for polynomials in z), inversion of objects of type polynomial along the lines suggested by Kalman's treatment of algebraic systems theory. So we have (1) procedure/structure libraries for matrices, polynomials, rational transfer functions etc and their manipulation, (2) an abstract algebraic procedure/operator library dealing with homomorphisms within arbitrary categories, (3) a program calling these libraries and containing our system definition and procedure calls for extracting the solution. The solution may take the form of (1) an algebraic formula with no values given eg $u(s)^{o}=(as^2+bs+c)x(s)$, (no need to use a symbol manipulation language), (2) a numerical formula, eg as in (1) but with a, b and c replaced by numbers, (3) a direct numerical answer, eg the optimal control at specified points in time. Note a fourth form of solution is to actually derive the optimal control in terms of the system transfer function itself.

Any transfer function or time evolution (the two things are isomorphic, one is the impulse response of the other) may be represented in a number of possible ways: (1) as a time series or a function of time, (2) as a polynomial in z or s, (3) as a factorised polynomial in $(s+a_i)$ or $(z^{-1}+b_i)$, (4) as a continuous or discrete system (A,B,C), (5) as a system matrix $G(s)=C/(sI-A)B$ or $G(z^{-1})$, (6) A is a companion matrix from which the system eigenvalues may be obtained. Particularly it is necessary to consider that any object may take any of these forms, whichever is most efficient to manipulate under particular circumstances. The three approaches, algebraic, frequency response and state space can all be considered simultaneously. The diagram shows relationships between a dynamic system and its roots



Now consider implementation of these objects. A polynomial is a vector of coefficients, a factorised polynomial is a set of factors. This can be seen when considering group multiplication. The product of two polynomials is well known. The product of their factors is the union of the two (unordered) sets. Whereas group addition provides far more problems in factored form. Again the elements of the set consist in general of reals or complex conjugate pairs. The latter have two degrees of freedom per pair and thus may be stored as one complex number, its conjugate being assumed, or as a quadratic. The adjoint of a polynomial in s means negate alternating coefficients, the adjoint of a factored polynomial means negate all eigenvalues.

In Algol68 it is possible to implement recursive modes as will be seen in the next section, thus matrices of polynomials etc. can be defined. It is further possible to define recursive operators to handle these modes so that if + is defined for both matrices and polynomials it is automatically available for matrices of polynomials. In attempting to program self generating operators we come up against the limitations of Algol68 showing that it is not truly orthogonal. A number of improvements could be made to the language. (There is a proposal to implement generic types in the next release of Algol68C, along the lines of ADA, but this involves recompiling the code for a particular operator every time it is used with a different mode of parameter.) In Algol68 'op' is not an object but is part of the syntax of the language, as a result operators cannot be passed as parameters. The same goes for MODE and modes. Logically one should not need to use 'conformity clauses' to find the mode of an operand as an operator can be defined for different modes. They do however appear to be necessary to distinguish between operands of

recursively defined modes. In Algol68C diadic operators cannot be used as monadic operators which is a nuisance as * cannot be used as the adjoint. Operators do not distinguish between contexts that require different result types, hence the action must depend entirely on the operands. Further it is by no means obvious how coercions on user defined modes work.

9.4 Algol82

Algol82 (Ada82) is a superset of Algol68 (Ada) designed by the author for use in control engineering, systems theory, electronic and electrical, civil and mechanical engineering, time series analysis, mathematics, electromagnetic theory and other field problems, physics, econometrics and computer graphics. Its use can be learned quickly and easily as it allows mathematical equations, using sophisticated mathematical structures to be programmed exactly as they are written in derivations, as opposed to designing an algorithm, as is necessary to write scientific programs in Fortran. This results in much shorter programs, which are likely to have far fewer bugs, are self documenting, easy to modify, and often faster and more accurate than conventional methods, (consider integration of ODEs using expA compared to Runge-Kutta.) Algol68 itself, though an orthogonal language like Ada and hence effectively a superset of most other programming languages, is difficult to learn (Ada is easier) and necessitates a large slow compiler. Because Algol82 programs are so short they take a comparitively short time to compile (depending on the implementation). Algol68 runs about the same speed as Fortran unless the heap is used in which case it runs about 40% slower. Algol82 binary packages can of course be run on any appropriate machine without an Algol82 compiler available.

It is hoped eventually to be able to supply a compiler which would allow the user to write his program in three separate modules - the first supplying the system equations, the second the problem to be solved and the third the algorithm to be used to solve it. The latter may be chosen automatically or perhaps for instance defined by the form or category in which the problem is specified.

Algol82 itself appears to have a certain amount of inbuilt intelligence in that it contains recently discovered properties of the inherent structure of mathematics itself - ie the morphology of physics - in a similar way to that in which Lisp programmers include properties of natural language and learning into their programs. The author in fact believes that the 'expert system' approach to AI problems is wrongheaded and that it is time to reattack problems of basic structure. Optimisation is automatically available from a Roth type structure embedded in the language which enables automatic solution of electric circuits, economic models, control problems etc. To implement this structure it is further necessary to distinguish between covariant and contravariant types within the language. The recursive Roth type structure describes a twisted isomorphism between covariant and contravariant exact sequences of operators and objects, eg current and voltage, states and costates, prices and stocks etc.

The initial version of the language will allow programmers to define objects in the following modes (types): sets, matrices, vectors, polynomials in z (the shift operator) or s (the Laplace transform), factored polynomials (sets of roots), rationals (quotients of eg integers or polynomials), all with integer, real or complex coefficients. Further all normal operators are available to manipulate these types, eg +, -, *, /, transpose etc, conjugate, scalar functions of matrices, tensor products, zero, unit, sigma, bigpi and transput. Further the types available are split into two sets: basic and concatenable. Basic types are eg, reals, integers, chars, bits and booleans. Concatenable types are eg, set, matrix,

tensor, vector, polynomial, factored polynomial, rational (pair), complex (pair) etc and these types may be concatenated indefinitely to produce new modes like eg, polynomial matrices, matrix polynomials (this distinction holds in that the operations would be carried out the other way round), sets of polynomials of complex matrices, vectors of arrays of rational factored polynomials. A basic type always appears at the right hand side of the mode definition. Further all normal operators and functions are automatically available for use with these deep structures. (This facility is available in no other language to the authors knowledge.) Further the user can insert new modes and new operators into the base set. The only time conventional numerical techniques are used is if an operator recurses down to for instance

```
'op'*=([,]'real'a,b)[,]'real':
```

in which case conventional matrix multiplication is used. For arrays of other objects Gaussian elimination is used eg, for the inverse of arrays of rational polynomials, arrays of rational integers (rationals) etc, in which case an exact answer is always obtained. Later versions of Algol82 may contain mode definitions and efficient operators for sparse matrices along the lines of TORRIX: and for scattering matrices. The mode definitions for the current Algol82 bootstrap are as follows

```
'mode''array'='ref'[,]'type',
      'poly'='ref'[]'type',
      'rational'='struct'('notarraymode'num,den),
      'complex'='struct'('amode're,im),
      'type'='struct'('amode'mat),
      'scalar'='union'('real','int','char','bit','bool','cmplx'),
      'notarraymode'='union'('rational','scalar','poly','complex'),
      'amode'='union'('notarraymode','array');
```

A system may be represented as a polynomial or matrix in s or z and the corresponding object in the users program may thus be considered to be in the appropriate category at the appropriate time. By changing the definition of the category under which the program is to run the user may force a different algorithm to be used (polynomial multiplication is much more efficient than a product of companion matrices.) Solutions may be printed as discretised time evolutions, functions of frequency, functions of the Laplace transform, or the z transform. In this regard many ideas were obtained from the UMIST implementation of Professor Rosenbrock's work on polynomial matrices. Indeed it should be easy and efficient to rewrite their package in Algol82. Algol82 should thus reduce programming of packages in engineering and time series analysis etc to an almost trivial level.

As far as solution of say ordinary differential equations is concerned the transformation

$\dot{x}=Ax \rightarrow z\underline{x}=\exp A\underline{x}$

always gives exact solutions (on a machine with an infinite word length) and is faster than say Runge-Kutta, etc, which always gives a truncation error from the Taylor series expansion. The Runge-Kutta method and its like were invented in the last century and are no longer appropriate for linear problems. Unfortunately they are still widely taught and used in University and industry. Writing a program to call a Runge-Kutta routine in Fortran (several hundred lines of very obscure programming) requires considerably more effort than writing a program to solve ODEs by $\exp A$ in Algol82 ... with no truncation error, no convergence problems, no need to resolve for new boundary conditions, total readability and maintainability and the Algol82 program would solve any set of ODEs without recompiling.

Partial differential equations are catered for by the inclusion of operators such as curl, div and grad. Automatic discretisation of analytic equations for numerical solution using two dimensional z and s transforms giving finite difference type solutions is available. Higher dimensional pde's may be treated by the use of mode tensor (or even differential forms) and the corresponding use of covariant and contravariant etc types. The Kronecker and other delta functions, inner, outer, exterior, cross, dot and other products and operators eg contraction would be supplied. The Hodge star operator gives the adjoint system. The general monadic star operator, * is in fact a basic concept in Algol82: it transposes matrices, takes s to -s, z to z^{-1}, grad to div and div to curl etc.

The current implementations of Algol82 on the ICL1904S and the Prime 550 at NELP use the Algol68R and Algol68C compilers, respectively and an album of Algol82 objects as a bootstrap. That is we are extending the language in the same way as a Lisp programmer does by typing definitions of new objects into the Lisp interpreter. Lisp is one of the few languages other than Algol68 which is truly ortnogonal. It is more like using a subroutine library in Fortran than performing a conventional Pascal or BCPL bootstrap, but because Algol68 is an orthogonal language it allows definition of new syntax and redefinition of old syntax under new circumstances, which can be held in an 'album'. Indeed one can redefine the syntax of Algol68 so that it looks pretty well like any language one cares to consider. Hence Algol82 is a new programming language.

There are two reasons why Algol82 can be bootstrapped in the above way. One is because Algol68 is orthogonal ie, consistent in that arbitrary modes can be defined, operators with different meanings depending on the modes of the operands can be defined. Ideally all Algol68 syntax should consist of Algol68 objects. Secondly the objects in Algol82 eg vectors, sets, matrices, polynomial matrices, rational polynomials, etc, all obey certain laws such as commutativit-

associativity, distributivity, existence of +, −, *, /, 1 and 0 etc, and thus are objects in certain mathematical categories such as groups, rings, modules, semigroups, abelian rings etc. The final version of Algol82 is intended to allow objects in these more sophisticated modes to be defined. This version will require some mathematical knowledge to use but this is vastly worthwhile because of the resulting reduction in overhead necessary. Objects will be available in all the above categories with full concatenation of objects and homomorphisms and automatic availability of operators and functors for these deeper structures. A large body of results from general homology theory is thus available for use in our language.

9.5 References

(1) Bowden, Comparison of Fortran Matrix Schemes on the 1904S at NELP, Computer Centre Document No 445.

(2) Bowden, Use of Algol 68 Maximop macros with subfile albums, NELP Computer Centre Document No. 436.

(3) Bowden, Applications Programming news: Dynamic Array Dimensioning in Fortran, SCAN - NELP Computer Centre Magazine Issue No. 15, May 1978.

(4) Khabie-Zeitoune, Closed form likelihood in multivariate weakly stationary time series, Mathematics Department, North East London Polytechnic.

(5) Bowden, Pascal at NELP, NELP Computer Centre Document No 449.

(6) Numerical Algorithms Group, Fortran Library Manual, Mark 5, Vol II.

(7) ICL Scientific Subroutines, 1900 series (Technical Publications 4096).

(8) Bowden, University of Sheffield Control Engineering Library, NELP Computer Centre Document No 179.

(9) Bowden, A Linear Control Theory Package, NELP Computer Centre Document No 188.

(10) Bowden, Automatic matrix manipulation in Fortran, Basic, Pascal and Algol68, NELP Computer Centre Document No 448.

(11) Jensen and Wirth, Pascal User Manual and Report, 2nd Printing, Springer-Verlag, 1978.

(12) Cornelius and Thomas, The Hull V-mode Pascal Compiler, Dept of Computer Studies, Univ of Hull, 1981.

(13) Woodward and Bond, Algol68R Users Guide, Division of Computing and Software Research, Royal Radar Establishment, MOD Malvern, 2nd Ed 1974.

(14) Goodwin, Why Programming Environments need Dynamic Data Types, IEEE Trans SE-7, No 5, Sept 81.

(15) Allison and McBride, Lisp Users Guide, Dept Computer Science, Queen's University, Belfast, 1974.

(16) Bourne et al, Algol68C Reference Manual, University of Cambridge Computing Laboratory, 1975.

CHAPTER X Conclusions

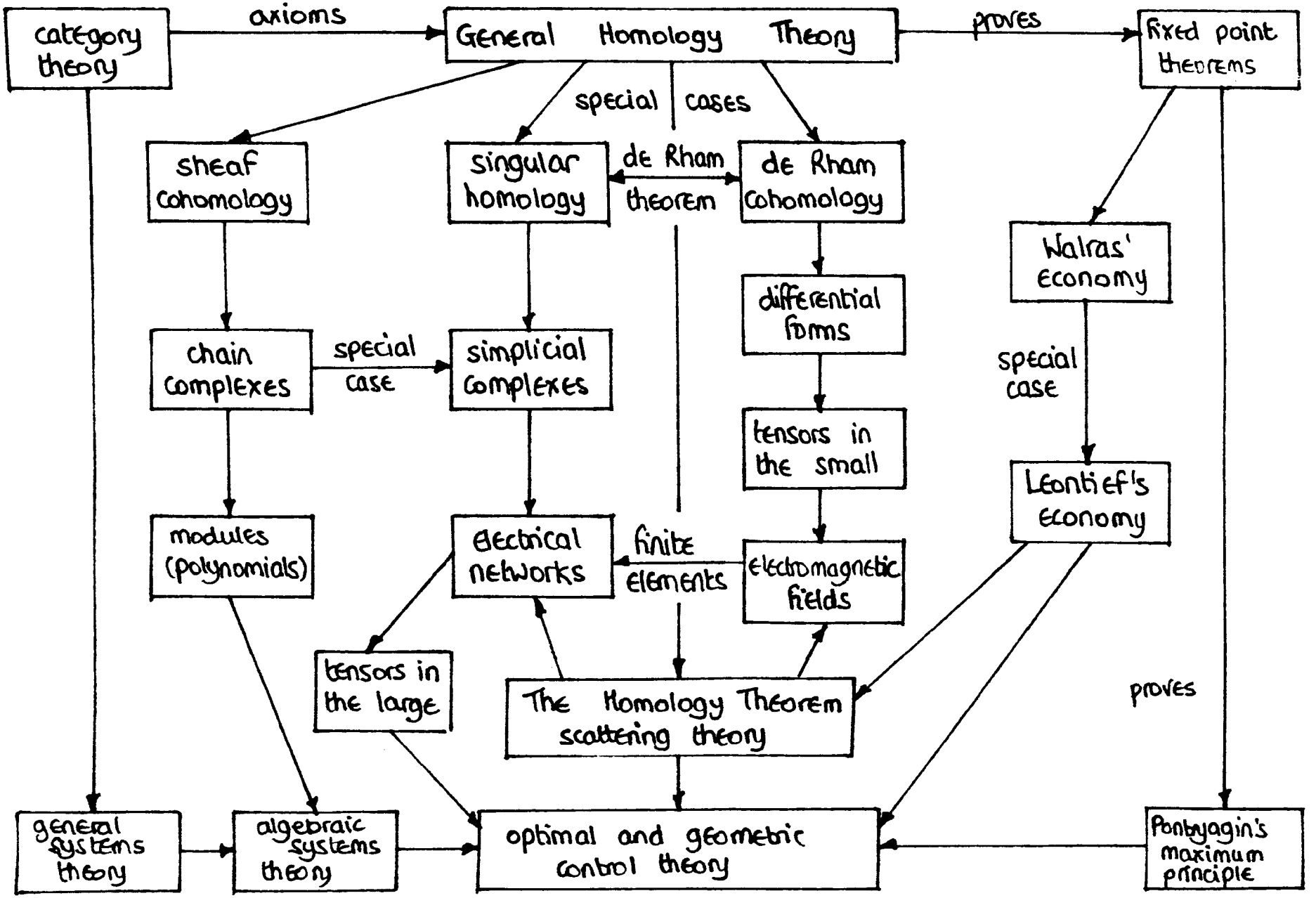
It is hoped that the appearance of yet another unifying work in the field of engineering has been justified. The review in the Introduction shows just how much fragmentation still exists in this area. This thesis collects together most of the relevant work, (much of which is quite difficult to obtain elsewhere) into one volume, in what the author believes is a less obscure form than has existed before. The theory is presented as a coherent whole, many previous gaps in knowledge, especially from an interdisciplinary point of view, are filled in, new algorithms are derived, and homology theory is introduced as the rigorous background necessary, in much the same way as Nicholson used scattering theory. A substantial body of deep results in homology theory already exists: a plethora of new avenues is thus opened for further research. The relationship between topology and optimisation is emphasised in a similar way to the use of direct and energy methods in electromagnetic theory. Structural or systemic rather than brute force methods, global rather than local methods are emphasised. Kron's First Generalisation Postulate can be applied directly to optimal control systems. The advantages of the Kron type approach are: a consistent approach to specifying a problem in a rational way leading to a natural form of solution directly programmable in a high level computer language.

Roth's diagram exhibits many of these advantages and indeed should perhaps be used for educational purposes providing a unifying method of teaching general optimisation and topological problems, particularly in the substructure form introduced in this thesis. The remarkable way in which the Riccati equation can be displayed on Roth's diagram in an easily visualisable form and hence appears almost directly from the original optimisation problem could prove an invaluable aid in teaching this otherwise rather obscure derivation - in a similar way to that in which the Chandrasekar equations can be so simply derived from scattering theory. Further Roth's diagram leads directly to

the set of equations which we have called Kron's lemma. This most important yet almost completely unknown formula actually goes a long way to completing the set needed for naive matrix manipulation. Returning to Fig(10), the usual way of solving general optimal control problems is via Pontryagin's Maximum Principle. The latter is proved by the use of Fixed Point Theorems such as Brouwer or Kakutani ... which in their turn are proved using Homology Theory. Control Engineers are using a black box method whose origin they do not in general understand. Our work suggests an alternative approach via the Homology Theorem which would involve a much more geometric approach to optimisation.

10.1 Generalised networks

Kron, Branin, Franksen and others devoted much effort into trying to define a 'generalised network' that would include as special cases both conventional electric circuits, more general networks which could be represented as graphs, and still more general systems to which tearing, or general diakoptical philosophy could be applied. Branin generalised the conventional electric circuit to higher dimensional networks. Kron applied tearing to arbitrary systems represented as graphs, whilst elsewhere lamenting the 'non-physical' nature of this type of approach. Kalman and others represented arbitrary dynamical systems as electric circuits. Franksen tried to define economic network elements. We now propose that the answer to this problem has been known almost as long as the problem. The generalised network necessary to apply the diakoptical philosophy is not a graphical concept but an abstract one: that of general homology theory - specifically in the form of Kron's algebraic diagram of the multidimensional space filter. At the specific level many systems can be modelled as chain complexes. We have seen this in the case of electrical neworks and also of multistage optimal



Fig(10) Conceptual diagram

control. At the general level - singular homology theory, de Rham cohomology theory - the study of simplicial complexes and of differential forms or generalised fields respectively - and further the theory of K[z]-modules used in the study of general systems - are all special cases of general homology theory. A conceptual diagram of these ideas is given in Fig(10). Homology theory includes all the usual diakoptical concepts. Covariant and contravariant variables and the scattering structure are defined within the space filter along with pairs of adjoint boundary operators, Roth's isomorphisms give impedance type network elements. Always there is an inherent topology and an equivalent optimisation problem - minimisation of energy or of distance or of a cost or disutility function. In the distributed case the network can be defined by a, perhaps infinitesimal, discretisation of the field as in the finite element method. Homology theory describes the topology of abstract structures. A generalised network is an abstract structure with a topology.

10.2 General systems

This work has dealt mainly with optimal systems, though reference has been made to non-interacting control, pole placement, etc. The relationships between homological systems theory and current algebraic and general systems theories remain to be fully investigated.

The most general definition of a system, according to Mesarovic is a subset of the Cartesian product of the input and state spaces $S<XxU$ with $S=\{(x,u):g(x,u)=0\}$. We can further define a subset of S, S^0 such that for all (x^0,u^0) in S^0, (x^0,u^0) are optimal according to some performance criteria $f^0(x^0,u^0)=\min_u f(x,u)$. Now various structures may be put upon S and it seems likely that the homology theorem may hold

in many cases. Our hypothesis is that S must be of a sufficient category for the axioms of homology theory to hold. Mesarovic has defined categories of systems. Kalman takes a more narrow view, considering a dynamical system S to consist of a tuple containing a time set, the input, state and output spaces, and a state transition map (isomorphism), input map (onto) and output map (1:1). He concentrates on controllability, realisability and the like. His major result is that the natural state space of a constant, linear, discrete system admits the structure of a finite free K[z] module with convolution as multiplication, a result essential to our treatment. This work originated in the treatment of finite automata as finite semigroups by Arbib. Kalman has further, in a short paper, treated dual systems with adjoint maps and defined the costate as $X^*=Hom(X,K)$, noting the duality between epimorphisms and monomorphisms and the properties of the star operator in a discrete linear system context.

The geometric control theory of Wonham et al is concerned with the concept of subspaces of the state space which are invariant under the state transition map, specifically the (A+BF)-invariant subspace of a system $sx=Ax+Bu$, $y=Cx$ with control law $u=Fx+Bu$. This is a generalisation of the familiar idea in linear algebra of an eigenspace, ie a subspace $\underline{V}$ spanned by some or all of the eigenvectors of a matrix A such that $A\underline{V}<\underline{V}$ or $AV=VE$ where V is the matrix of eigenvectors of A and E is, in the distinct eigenvalue case, the diagonal matrix of corresponding eigenvalues of A, and $\underline{V}=spanV$. A subspace $\underline{V}<\underline{X}$ is called an (A+BF)-invariant subspace if there exists F such that $(A+BF)\underline{V}<\underline{V}$. An immediate result is that $\underline{V}$ is an (A+BF)-invariant subspace iff $A\underline{V}<\underline{V}+\underline{B}$. It can be seen that F is irrelevant hence the term (A,B)-invariant subspace is often used! Wonham goes on to define controllability subspaces and derives important applications in non-interacting control, pole-placement, etc. $\underline{V}$ turns out to be the maximal invariant subspace of (A,B) contained in kerC. Determining such a subspace and calculating the spectrum of $A^2=(A+BF)|\underline{V}$ where $(A+BF)V=VA^2$, yields the transmission zeros of the system, a

result examined in Appendix I. The latest theory of general systems is the algebraic-geometric approach of Hermann and Martin, described in the Introduction. The ideas of topological duality and the geometric theory are combined. (Typically $f^*(p)(x)=p(f(x))$ where f is a polynomial map $C^n \to C^m$, $p \in PF(C^n)$ the associated ring of polynomials and $f^*:PF(C^m) \to PF(C^n)$.) Hermann's work is the closest to our standpoint and the two complement each other.

10.3 Natural language

Chapter IX describes approaches to programming our theory. However systems theory is rapidly becoming applicable to computer programming. Mesarovic has applied his 'fundamental theorem' to a formal or 'symbol-manipulating' system represented as an ordered sextuple

K=(E,S,T,R,P,f) where

E is a denumerable set and represents expressions,
$S \subset E$ represents sentences,
$T \subset S$ represents theorems of S,
$R \subset S$ represents refutable sentences,
$P \subset E$ are unary predicates,
N denotes the set of integers.

Two mappings are given by $g:E \to N$ and $f:E \times N \to E$ such that g is an injection and $f(e,n) \in S$ whenever e is a predicate, $e \in P$. Then for any $e \in E$, g(e) is the Goedel number of E. Mesarovic constructs a general system for K by establishing the following correspondences:

Predicates P are inputs of the system.
Expressions E are the states.
Sentences S are outputs.

The state representation r:ExP->S is r(e,p)=f(p,g(e)). There may yet be a further application of homological systems theory. Artificial intelligence is based to a large extent on the concept of natural language, the basic tools of which are syntactic and semantic grammars. Tang and Huang[2] define a semantic grammar as follows:

A word w=[a,v] consists of a semantic component v, the feature vector of the word and a syntactic component a, which can be expressed by a label which corresponds to a terminal or nonterminal of a context free grammar. We can write v=m(a). A syntactic category is the set of all words with the same syntactic component. A sentence Z is a sequence of words $[a_1,v_1],\ldots,[a_n,v_n]$. The syntactic rule Z' of Z is $a_1,\ldots,a_n$. Let F denote a (finite computational) algorithm and $A\to A_1\ldots A_n$ denote a production rule of a (Chomsky) context free grammar G=(N,T,P,S). Then $m(A)=F(m(A_1)\ldots m(A_1))$ means F takes $m(A_1)\ldots m(A_n)$ as its inputs and assigns the output vector of F to A. When $A_1\ldots A_n$ is null F still generates an output vector. The notation for context free grammars is included in the following definition. A semantic grammar G' is a 5-tuple (N',T',P',S',f) where

(1) S' is a set of terminal words,

(2) T' is a set of nonterminal words,

(3) P' is a finite set of ordered pairs (P,F). P is an element $Nx(NuT)^+$ where $N=\{A\mid[A,v]\in N'\}$, the set of all nonterminal semantic categories, $A^+=A_1uA_2u\ldots$, N is finite. $T=\{a\mid[a,v]\in T'\}$, the set of all terminal semantic categories, T is finite. Each element in P' is called a production. Let $(P,F)\in P'$ and P=(A,r) then we denote the production by a->r:F. F is an algorithm such that if P is $A_1\ldots A_n$ then $m(A)=F(m(A_1),\ldots,m(A_n))$, where $A_1,\ldots,A_n\in(NuT)^+$ and F is called the feature transfer function.

(4) N n T=0.

(5) $S'=\{[S,v]\mid[S,v]\in N'\}$ where S is the distinguished nonterminal syntactic category (start symbol) in N.

(6) G=(N,T,P,S) is the underlying (Chomsky) context-free grammar, or base, of G'.

(7) f is a Boolean function defined over $\{v\mid[S,v]\in S'\}$.

A sentence Z is well formed if Z' is acceptable under G. It seems likely that we can formulate G' as a general system using similar techniques to those Mesarovic has used for treating symbol manipulating systems. Homological analysis of semantic nets could also prove productive.

10.4 The future

The set of basic tools used in this thesis is introduced in the Glossary and extended in Chapter II. The single most important and probably confusing concept is the construction of a short exact sequence from a (short) chain complex and the fact that the former is isomorphic to a short exact sequence in standard form involving a quotient space. Once this is properly understood it can be seen that the latter is related to the adjoint sequence by Roth's twisted isomorphism, and we can drop from abstract categories to functional analysis, matrix algebra or whatever at leisure. This essentially is the Homology Theorem for Optimal Systems. We make no more than an hypothesis on the categories or types of systems for which the theorem holds. The author believes however that it must be possible and would be useful to rewrite the axioms of General Systems Theory in terms of those of General Homology Theory. In fact the two are probably isomorphic.

Chapters III and IV cover most of the relevant work that has been published on network tearing and field theory. The former appears in general to be well understood and has if anything been rather overworked by a generation of authors fascinated by the mystique and beauty of Kron's work, the only remaining confusion being caused by the incredible variety of notation being used by authors from varying backgrounds.

Many questions however remain open with regard to distributed systems. Most of this work has either been carried out by engineers interested in practical solutions to real problems or by mathematicians investigating the structure of de Rham's cohomology theory. Flanders' work on differential forms certainly helps to bridge the gap between these two extremes but Branin's somewhat naive approach to the homological structure of Maxwell's equations leaves a lot unanswered, particularly regarding the existence of boundary conditions in Roth's diagram. If this problem could be satisfactorily concluded then the whole question of efficient computer solution of partial differential equations could be reopened in the light of network tearing.

In Chapter V we come to the conclusion that tensor algebra as a manipulative tool probably has rather limited application, though the concepts arising from the theory are basic to our entire philosophy and constitute the basis for identifying the homological structure of real systems. (Of course it depends how one looks at it - a partitioned matrix is a fourth rank tensor.) Scattering theory is established along with homology theory and Kron's algebraic diagrams as the underlying justification for analogies in science and engineering, the physical structure of general systems.

These ideas culminate in the analysis of optimal control systems in Chapter VI where we find that the three parallel themes can be represented simultaneously on the same diagram showing a system torn in time and reconnected stage by stage, with the scattering structure along the middle and the homological structure along the top and bottom of Kron's algebraic diagram for the multidimensional space filter. Analogies abound. The partitioned system, the torn system, the reconnected system and the continuous system are all isomorphic in a similar way to that in which Kron's intersection network is called a miniature copy of the original. The interface between the stages may seem slightly artificial but this appears to be

caused by the fact that we are tearing in time rather than between spatial dimensions. This needs investigating more thoroughly as do the problems incurred by integrating only over real poles or positive time. (This asymmetry of time is currently one of the basic problems of modern physics.) The analysis should apply equally to the study of systems described by rational polynomial matrices. It is expected that a major application of this theory will be to multivariable Time Series Analysis (an area wide open for further research). It is also closely related to the theory of noninteracting control.

In Chapter VII all the concepts thus far developed are used to identify and analyse the homological structure of Leontief's macroeconomic model. The analysis appears to be complete and describes the core of the Cambridge Growth Model. We come across the classical use of fixed point theorems to prove the existence of a general equilibrium in the Walras model. A further example considered by the author is that of dynamically optimising timeslicing on a multiaccess computer according to the priority of the users. If $\underline{p}_c$ is a vector giving priority numbers for different classes of user then the priority of a particular user is given by $\underline{p}_u=\underline{C}\underline{p}_c$ where $\underline{C}$ is a matrix of 0's and 1's picking out the class of a particular user. Now if $\underline{y}_u$ is the actual timeslice taken by each user then it is easy to see that the total timeslice of each class in one operating system cycle is given by $\underline{y}_c=\underline{C}'\underline{y}_u$. If the dynamics of $\underline{y}_u$ are identified a quadratic cost function on $(\underline{y}_c,\underline{p}_c)$ may be minimised and the optimal dynamics of $\underline{p}_u$ (which may for instance be the maximum time slice allowed to each user) found from the adjoint system.

The Eilenberg-Steenrod axioms of General Homology Theory are stated in Chapter VIII. Some of the relationships with General Systems Theory can be seen. The definition of an admissable category depends upon the same lattice diagram as Wonham's geometric analysis. The * operator can be used to define the adjoint system and relate the boundary and coboundary operators. The algebraic axioms

(identity, composition and commutativity) appear to be related to Mesarovic's functors between categories of systems. The exactness axiom implies optimality. The homotopy axiom implies continuity. The excision axiom describes the necessary degree of connectedness of the underlying spaces. All these concepts arise in Optimal Control Theory. They must in fact form a core or kernel for the theory. What is needed now is cooperation between mathematicians and systems engineers to discover the systemic interpretations of the many deep results of homology theory in this context. The Universal Coefficient Theorem is said to determine cohomology from homology. That is it gives the relationship between a system and its adjoint. However it is very difficult for a control engineer to visualise the meaning of this without help from a mathematician. And the theorem is only a special case of the Kunneth theorem! We need to understand the Meyer-Veitoris sequence, the applications of fibre bundle and sheaf theory and more. And one day we may be able to present the mathematicians with a computer language capable of manipulating objects in these categories. But before this can happen Control Engineers must start implementing their packages in Algol68 or Ada. There are some hopeful signs. UMIST are considering rewriting their CAD package in ADA.

The concept of an orthogonal language (Chapter IX) is probably the most underestimated idea ever to come out of computer science. Unfortunately no existant computer language (with the possible exception of Lisp which is too basic for our purposes) is truly orthogonal so that the attempt to bootstrap Algol68 or Ada up to the levels mentioned above is fraught with difficulty. Nevertheless an initial attempt to introduce recursive operators on mutually recursive modes has been successful. (I am indebted to Chris Cheyney of the Computer Laboratory, Cambridge, Ted Elsworth of the Computer Science Department, Aston University and Trevor Elliot of NELP for their invaluable help with this.) The continual problem is that there are so many possible ways of achieving the desired result - and that so few of them can actually be implemented in practice. As it is there are at

least two isomorphic ways of defining recursive modes in Algol68, and at least two ways of defining recursive operators, giving four or more combinations. The best approach will probably only become obvious when the initial bootstrap is completed. A further implication of Algol82 is that it should be possible to define recursive categories in mathematics. NB. Authors Note. Algol82 may finally be released using the name Homogol.

Summarising then, once again, the results and implications of this work. (1) In the understanding and teaching of optimisation in any discipline the substructure version of Roth's diagram may provide a unifying, easily understood medium for presentation. As things stand each generation of students is presented with an exponentially increasing amount of theory to absorb. It is necessary to reassess the means of presentation of the material at regular intervals, in order to offset this growth. (2) Derivation of new algorithms. The application of homological systems theory to existing optimisation theories should force the appearance of all algorithms of the LQG form. In fact few new algorithms have been found which only goes to show what a time worn field this is (compared to say time series analysis.) (3) More importantly we now have something approaching a complete theory of structural optimisation, though there are still many gaps that need filling. Philosophically homology theory describes all continuous theory between the scales of quantum theory and relativity (see below). (4) The author believes that if this line of investigation is continued, engineering computer languages which contain a degree of intelligence (sic) will result. Consider the operator + operating upon two arrays of polynomials. The existence of + for pairs of polynomials and pairs of arrays <u>implies</u> the operator upon the concatenated type. This level of inference exists only in artificial intelligence languages as yet. Further extensions of Algol82 using category and homology theory will enhance this inference mechanism still further. Consider the implementation of an English <u>compiler</u>, (a very unusual thing to do, all AI languages are interpreted, or threaded like Forth). A verb is

an operator that takes objects of mode noun. A noun is an operator that takes objects of mode adjective. A conjunction is an operator that takes objects of mode clause. The semantics of the English language relies on inference. This has proved to be the most frustrating property to implement in a computer language. We are suggesting a new mechanism which may well be applicable to AI. It is surprising how the remarkable similarity between the syntax - and semantics - of English and the orthogonal languages has been so little utilised.

This work coincides with increasing doubt in the minds of many physicists as to the very nature of reality, arising partly from the increasing influence of Eastern mysticism on Western philosophy, and partly due to a number of recent experiments refuting the 'Principle of Separability of Space', one of the basic axioms on which all Western science is based (an axiom which was questioned by quantum physicists, though strongly supported by Einstein). Two possible consequences of this, currently being investigated, are the existence of parallel universes and faster than light transmission of information (rumour has it that Aspect, in France, has achieved this). There is a growing literal belief in Heisenberg's uncertainty principle, that the observer affects what he observes, and a growing conviction that there really is no such thing as mass-energy as substance, normally held to create a disturbance in space-time, the idea is that a particle <u>is</u> a disturbance in space-time. Consequently many scientists suspect that rather than discovering new phenomena researchers may actually be creating these phenomena[3], ie if something is believed strongly enough it may become true - the observer effects what he observes. It is interesting to note here that Western religions, but not Eastern philosophy, are based on the idea of unquestioning faith. These ideas are the basis of the currently popular 'Hadron Bootstrap Theory', the idea that the only thing required for existence is consistency. The mathematical basis of this theory is the scattering matrix.

10.5 Acknowledgements

Finally the author is indebted to the late Gabriel Kron without whom this thesis could have never existed, to Michael Moorcock for his ideas on parallel universes and to the many who have helped with the work, including my supervisor, Professor Nicholson, for being so patient, Peter Marley, my Head of Department, for allowing me to continue with my research at the Computer Centre, North East London Polytechnic, and Valerie for encouraging me to finish it. The thesis was typeset by Runoff 18.2 on the Prime 550 at NELP. It was printed on a NEC Spinwriter with a Times Roman/Scientific printwheel using the pre/postprocessor SRUNOFF written at NELP by Keith Bowden and Stuart Walker.

10.6 References

(1) Kalman, Introduction to the theory of linear dynamical systems, (unknown).

(2) Tang and Huang, A syntactic-semantic approach to Image Understanding and Creation, IEEE Trans PAMI, Vol 1, No 2, April 1979.

(3) Zukav, The Dancing Wu Li Masters, Rider/Hutchinson, 1979.

(4) Aspect et al, Experimental Tests of Local Theories via Bell's Theorem, Physical Review Letters, Vol 47, No 7, pp 460-3, 1981.

(5) Denham, The Geometric Approach, Dept Computing and Control, Imperial College, 1976.

Appendix I: Topological Aspects of Invariant System Zeros

This Appendix is an abbreviated version of Ref[4], and was written in 1975.

Recently Kouvaritakis and Macfarlane: Sinswat, Patel and Fallside: Owens and Davison have produced a geometric analysis of transmission and decoupling zeros linking the frequency response and state space control theories. These are the fixed modes which are invariant under feedback and duality: in the right half of the s-plane they induce a non-minimum phase system which is difficult to control. Along with Morse and Wonham's pole allocation this completes the problem of system synthesis.

Motivation for the topological approach is supplied by the application of homology theory by Roth, Branin, Amari, Kondo and Iri to Kron's analysis and tearing of electrical networks and physical systems: it is known that inherent (energy) minimisation and invariance under coordinate transformation always occurs in physical problems and it is proposed that the optimisation and similarity invariance found in control systems induces an inherent topology mapping the error to zero. The application of a Roth type diagram to dynamical systems is investigated and found to lead to a new definition of transmission zeros. An orthogonalisation procedure will be given, and K-partitioning used for calculating the invariant zeros.

11.1 Transmission zeros in continuous optimal control

Quadratic optimal control of the system

$sx=Ax+Bu$ with $y=Cx$ and cost function $\min\int y'Qy+u'Rudt$

can be shown to lead to a pair of coupled differential equations of the form

$$s\begin{bmatrix}x\\p\end{bmatrix}=\begin{bmatrix}A & B/RB'\\ C'QC & -A'\end{bmatrix}\begin{bmatrix}x\\p\end{bmatrix}$$

where $p=-Kx$ leads to the backwards Riccati equation for K. This system can be displayed on a Roth type commutative diagram as follows

$$\begin{array}{ccccccccc}
 & & & & /(sI-A)B & & C & & \\
0 & \to & u & \to & x & \to & y & \to & 0 \\
 & & R\downarrow & & -K\downarrow & & Q\downarrow & & \\
0 & \leftarrow & . & \leftarrow & p & \leftarrow & . & \leftarrow & 0 \\
 & & & & B' & & /(sI+A')C' & &
\end{array}$$

We wish to investigate how the horizontal sequences in this diagram deviate from being short exact ie

$im/(sI-A)B=\ker C$ or $im/(sI-A')C'=\ker B$.

Now the necessary and sufficent condition for s to be a transmission zero is

$s\in Z(A,B,C)$ itf $x(s)=im(/(sI-A)B) \cap (\ker C\#\{0\})$ where $\cap$ = intersection.

$x(s)$ is the null transmission subspace. That is

$s \in Z(A,B,C)$ iff there exists $x(B) \subset \text{im}/(sI-A)B$ and $x(C) \subset \ker C$ st $x(B)=x(C)$.

$x(s)$ can be shown to have the properties

$x(s) \cap \text{im}(B)=\{0\}$ and therefore $Ax(s) \subset x(s) \oplus \text{im}(B)$.

Also as $\{x: x \subset \ker(C) \cap C^n, Ax \subset x \oplus \text{im}(B)\}$ where C^n is complex n-space is semiordered by inclusion there exists by Zorn's lemma a maximal element x^0 such that

$x^0 \cap (\text{im}(B))=\{0\}$ and therefore $Ax^0 \subset x^0 \oplus \text{im}(B)$,

which is Owens' canonical definition of the state space. The total number of open loop transmission zeros is dim(x) = rank defect of

$$\begin{bmatrix} sI-A & -B \\ C & 0 \end{bmatrix}$$

Owens has illustrated the physical source of the zeros as due to inherent dynamic state feedback within the system structure, by means of a similarity transform taking into account the rank defect.

11.2 Computation of Invariant Zeros: Square systems

Consider the completely controllable and observable proper system

$$\begin{bmatrix} sI-A & -B \\ C & 0 \end{bmatrix}\begin{bmatrix} x \\ u \end{bmatrix}=\begin{bmatrix} 0 \\ y \end{bmatrix}$$

or $\underline{P}(s)\underline{u}=\underline{y}$. The system invariant zeros are those values of s for which $\underline{P}(s)$ is not a monomorphism ie $\ker\underline{P}(s)\neq 0$ or using Schur's formula (K-partitioning)

$$\det\underline{P}(s)=|sI-A||C/(sI-A)B|=|sI-A||G(s)|=0$$

the system transfer function zeros which are not poles of the state space.

We now choose pseudoinverse left and right annihilators of the input and output matrices respectively according to the commutative diagram in Fig (11.1).

$C=[C_1\ C_2]$, $B=\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$. We have $CM=0=NB$ eg

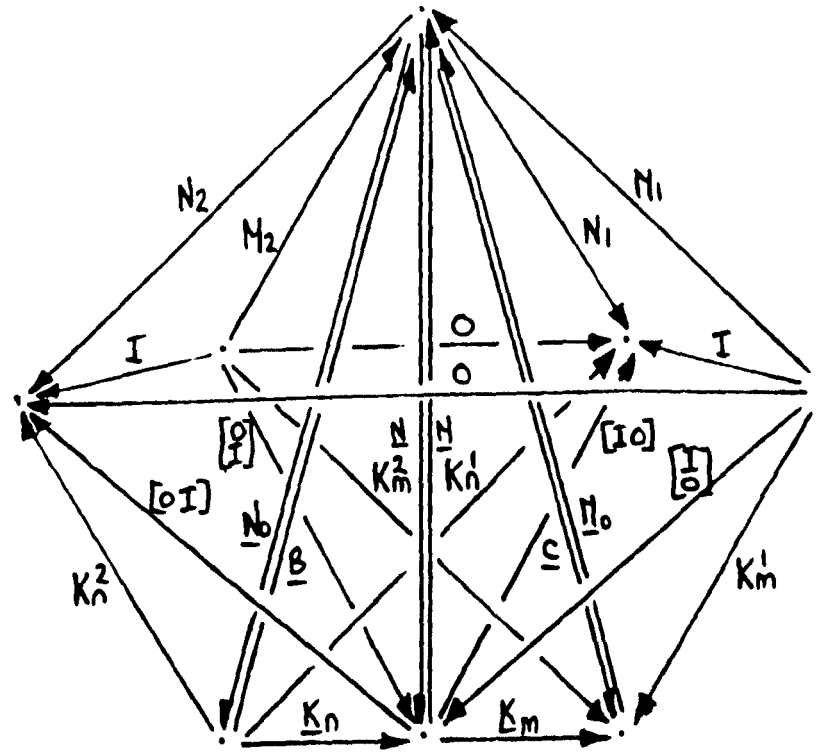
$$M=M_oK_m=\begin{bmatrix} M_o' \\ I \end{bmatrix}K_m=\begin{bmatrix} -/C_1C_2 \\ I \end{bmatrix}K_m,\quad N=K_nN_o=K_n[N_o'\ I]=K_n[-B_2/B_1\ I]$$

$$0 \to \cdot \xrightarrow{M} \cdot \xrightarrow{BC} \cdot \xrightarrow{N} \cdot \to 0$$

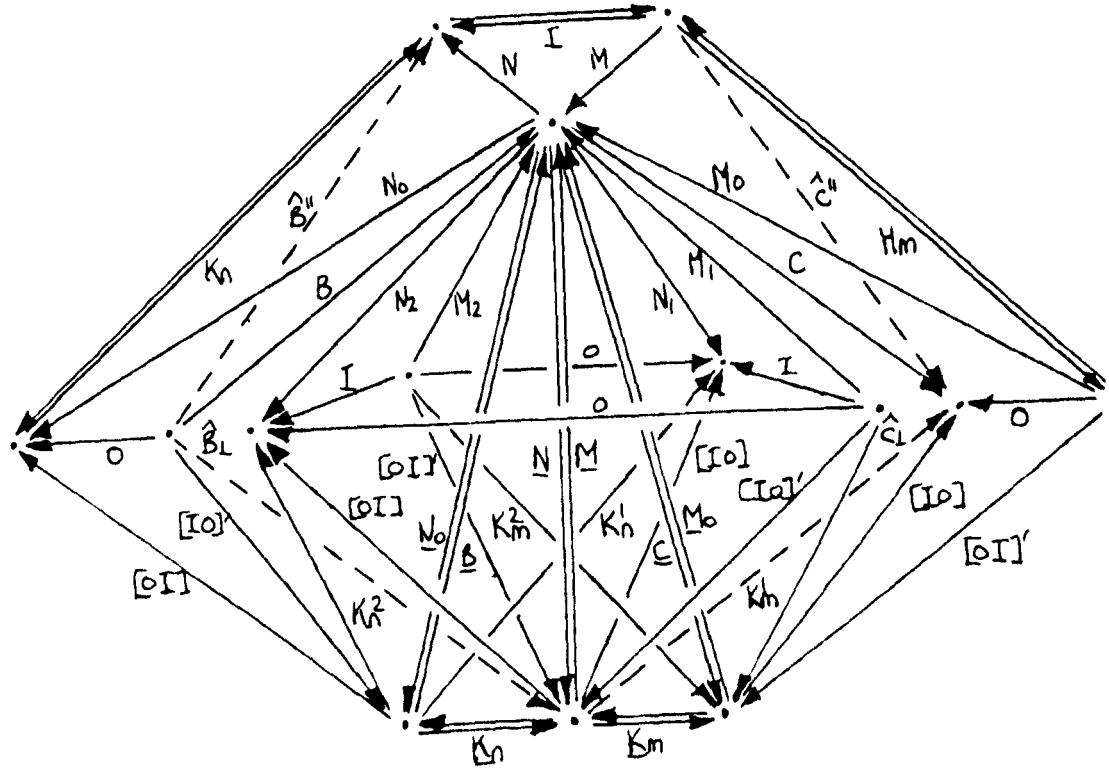
and the above sequence is long exact. K_n and K_m are arbitrary up to $K_nN_oM_oK_m=NM=I$ eg we can choose $K_m=I$. Consider

$$\begin{array}{ccccccccc} & & & B & /(sI-A) & C & & & \\ 0 & \to & u & \to & \cdot & \to & \cdot & \to y \to & 0 \\ & & & N\downarrow & & M\uparrow & & & \\ 0 & \to & u & \to & \cdot & \to & \cdot & \to y \to & 0 \\ & & & B^2 & /(sI-NAM) & C^2 & & & \end{array}$$

Fig(11·1) Internal structure of orthogonalisation procedure



Fig(11·2) Internal structure of extended orthogonalisation



which commutes and by diagram chasing the triple formed by the lower sequence is exact iff that formed from the upper sequence is also. By contradiction it can be shown that the system transmission zeros are the eigenvalues of the matrix NAM. This is the 'NAM algorithm' (Kouvaritakis) where NAM is of lower order than A equal to the number of zeros of the system.

11.3 Orthogonalisation of Rectangular Systems

We extend the method of Sinswat, Patel and Fallside which is more general, and seems more natural than that in the sequel to Kouvaritakis and Macfarlane's paper. A Kron-type orthogonalisation is performed by an extension of the NAM algorithm followed by K-partitioning and the eigenvalue form of Schur's lemma. Consider the commutative diagram in Fig (11.2), where either $\underline{K}_n$ or $\underline{K}_m$ is again arbitrary. Choosing $\underline{K}_m=I$ and using the coordinate transformation defined by $x=>\underline{N}x$ where

$\underline{N}=\begin{bmatrix} C_1 & C_2 \\ 0 & I \end{bmatrix}$ and $/\underline{N}=\underline{M}=\begin{bmatrix} /C_1 & -/C_1C_2 \\ 0 & I \end{bmatrix}$ the system becomes $sx^2=A^2x^2+B^2u$, $y^2=C^2x^2$

where $A^2=\underline{NAM}$, $B^2=\underline{NB}$, $C^2=\underline{CM}=[I\ 0]$.

$$
\begin{array}{ccccccccc}
 & & B & /(sI-A) & C & & & & \\
0 & \rightarrow & u & \rightarrow & . & \rightarrow & x & \rightarrow y \rightarrow & 0 \\
 & & = & N \downarrow\uparrow M & N \downarrow\uparrow M & = & & & \\
0 & \leftarrow & u & \leftarrow & . & \leftarrow & x^2 & \leftarrow y \leftarrow & 0 \\
 & & & B^2 & /(sI-A^2) & C^2 & & &
\end{array}
$$

The transfer function matrix G(s) is now given by $C^2/(sI-A^2)B^2$. Partitioning A and A^2 correspondingly and following the procedure described in Patel G(s) can be factorised as

$$C^2/(sI-A^2)B^2=[I\ 0]/\begin{bmatrix} sI-A^2_{11} & A^2_{12} \\ A^2_{12} & sI-A^2_{22} \end{bmatrix}\begin{bmatrix} B^2_1 \\ B^2_2 \end{bmatrix}$$

$$=/(sI-A^2_{11}-A^2_{12}/(sI-A^2_{22})A^2_{21})(B^2_1+A^2_{12}/(sI-A^2_{22})B^2_2)$$

$$\begin{bmatrix} A^2_{11} & A^2_{12} \\ A^2_{21} & A^2_{22} \end{bmatrix}=\begin{bmatrix} (C_1A_{11}+C_2A_{21})/C_1 & -(C_1A_{11}+C_2A_{21})/C_1C_2+C_1A_{12}+C_2A_{22} \\ A_{21}/C_1 & -A_{21}/C_1C_2 \end{bmatrix}$$

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix}=\begin{bmatrix} C_1B_1+C_2B_2 \\ B_2 \end{bmatrix}$$

It is shown in Patel that the invariant zeros of $B^2_1+A^2_{12}/(sI-A^2_{22})B^2_2$ are identical to those of $C/(sI-A)B$, but as the numerator is of lower order than $G(s)$ the problem of determining invariant zeros is simplified.

The internal structure of the orthogonalisation is given in the commutative diagram Fig (11.1). The restrictions on the arbitrary components are given by $\underline{K}_n\underline{N}_o\underline{M}_o\underline{K}_m=\underline{N}\underline{M}=\underline{I}$ once again, also $K_n[0\ I]/\underline{K}_n/\underline{K}_m[0\ I]'K_m=I$ from the combined diagram Fig (11.2) in which all invertible relations and the final nonorthogonal ($||$) and ortnogonal ($\underline{|}$) transformations are shown. Note that commutation does not occur rouna the perimeter. The new definitions are $[M_1'\ M_2']'=\underline{M}=\underline{M}_o\underline{K}_m=\underline{M}_o[K_{m1}'\ K_{m2}']'$ and $[N_1\ N_2]=\underline{N}=\underline{K}_n\underline{N}_o=[K_{n1}\ K_{n2}]\underline{N}_o$. Transformations can now be made between the ortnogonalised system and its adjoint.

Finally the invariant zeros of the system can be found from the common <u>poles</u> of all minimal order exact pseudoinverses of the system or its adjoint. For a general system $sx=Ax+Bu$, $y=Cx+Du$ with the same number of inputs as outputs, they are the eigenvalues, $P(A-B/DC)$, of $A-B/DC$, the transmission zeros are $P(A-B/DC)-P(A)$ and the decoupling zeros are $P(A-B/DC)\ n\ P(A)$. Similar relations hold for rectangular systems involving the pseudoinverses.

11.4 References

(1) Kouvaritakis and Macfarlane, Geometric approach to analysis and synthesis of system zeros, International Journal of Control, Vol 23, No 2, Feb 1976 (Pts I and II).

(2) Sinswat, Patel and Fallside, A method of computing invariant zeros and transmission zeros of invertible systems, Ibid.

(3) Owens, System transmission zeros: a geometric analysis, Research report No 35, Sheffield University Dept Control Engineering, Sept 1975.

(4) Bowden, An Introduction to Homological Systems Theory: Topological Aspects of Invariant System Zeros, Matrix and Tensor Quarterly, Vol 31, No 2, Dec 1980.

Appendix II: The De Rham Theorem

A sheaf S of k-modules over M consists of a topological space S together with a map p:S->M satisfying

(1) p is a local homeomorphism of S onto P

(2) $p^{-1}(m)$ is a K-module for any m<M

(3) The composition laws are continuous in the topology on S.

This is similar to the definition of a fibre bundle. The map p is called the projection, and the K-module $S_m=p^{-1}(m)$ is called the stalk over m<M. Let U<M be open. A continuous map f:U->S such that p.f=id is called a section of S over U. The 0-section is the section which associates with m<U the zero element of S_m. We let L(S,U) denote the set of sections of S over U. Let f and g belong to L(S,U) and k<K. Define the sections (f+g)(m)=f(m)+g(m) and (kf)(m)=k(f(m)) with m<U. With these operations L(S,U) becomes a K-module. The module of (global) sections of S over M will be denoted L(S).

A continuous map m:S->S^1 such that $p^1.m=p$ which is a homomorphism (of K-modules) on each stalk is a sheaf homomorphism. A sheaf isomorphism is a sheaf homomorphism with an inverse which is also a sheaf homomorphism. Other definitions appear in the usual way. A sheaf S over M is said to be fine if for each locally finite cover of M by open sets there exists an endomorphism such that the set of endomorphisms is a partition of unity. A sheaf of K-modules is said to be torsionless if each stalk is a torsionless K-module X, ie there is no non-zero element x<X for which there exists a non-zero element k<K such that kx=0.

Lemma Let $0\to A^1\to A\to A^2\to 0$ be an exact sequence of K-modules, and let B be a K-module. Then the induced sequence $A^1\otimes B\to A\otimes B\to A^2\otimes B\to 0$ (whose homomorphisms are tensored with the identity homomorphism of B) is exact, but $A^1\otimes B\to A\otimes B$ is not necessarily injective. If however, either A^2 or B is torsionless then the full sequence $0\to A^1\otimes B\to A\otimes B\to A^2\otimes B\to 0$ is exact.

Theorem Let $0\to S^1\to S\to S^2\to 0$ be an exact sequence of sheaves over M, and let T be also a sheaf over M. Then if either T or S^2 is torsionless, then the sequence $0\to S^1\otimes T\to S\otimes T\to S^2\otimes T\to 0$ is exact. If in addition, either T or S^1 is a fine sheaf, then the sequence $0\to L(S^1\otimes T)\to L(S\otimes T)\to L(S^2\otimes T)\to 0$ involving the modules of global sections, is exact.

A sheaf cohomology theory H for M with coefficients in sheaves of K-modules over M consists of

(1) a K-module $H^q(M,S)$ for each sheaf S and each integer q,

(2) a homomorphism $H^q(M,S)\to H^q(M,S^1)$ for each homomorphism $S\to S^1$ and each integer q and

(3) a homomorphism $H^q(M,S^2)\to H^q(M,S^1)$ for each short exact sequence $0\to S^1\to S\to S^2\to 0$ and for each integer q

such that the following properties hold

(1) $H^q(M,S)=0$ for $q<0$, and there is a isomorphism $H^0(M,S)\tilde{}L(S)$ such that for each homomorphism $S\to S^1$ the following diagram commutes

$$\begin{array}{ccc} H^0(M,S^1) & \tilde{} & L(S^1) \\ \uparrow & & \uparrow \\ H^0(M,S) & \tilde{} & L(S) \end{array}$$

(2) $H^q(M,S)=0$ for all $q>0$ if S is a fine sheaf.

(3) If $0->S^1->S->S^2->0$ is exact then

$$...->H^q(M,S^1)->H^q(M,S)->H^q(M,S^2)->H^{q+1}(M,S^1)->...$$

is exact.

(4) The identity homomorphism $id:S->S$ induces the identity homomorphism $id:H^q(M,S)->H^q(M,S)$.

(5) If the homomorphisms $S->S^1$ and $S^1->S^2$ are transitive for every q then so are the homomorphisms $H^q(M,S)->H^q(M,S^1)$ and $H^q(M,S^1)->H^q(M,S^2)$.

(6) For each homomorphism of exact sequences of sheaves

$$\begin{array}{ccccccc} 0 & -> & T^1 & -> & T & -> & T^2 & -> & 0 \\ & & \uparrow & & \uparrow & & \uparrow & & \\ 0 & -> & S^1 & -> & S & -> & S^2 & -> & 0 \end{array}$$

the following diagram commutes

$$\begin{array}{ccc} H^q(M,T^2) & -> & H^{q+1}(M,T^1) \\ \uparrow & & \uparrow \\ H^q(M,S^2) & -> & H^{q+1}(M,S^1) \end{array}$$

The module $H^q(M,S)$ is called the qth cohomology module of M with coefficients in the sheaf S relative to the cohomology theory $\underline{H}$.

An exact sequence $0->A->C^0->C^1->C^2->...$ is called a resolution of the sheaf A. The resolution is called fine (respectively torsionless) if each of the sheaves C^i is fine (respectively torsionless). We shall now show that each fine torsionless resolution of the constant sheaf MxK, canonically determines a cohomology theory for M with coefficients in sheaves of K-modules over M. We obtain a cohomology theory as follows

(1) With each sheaf S and each integer q we associate the qth cohomology module of the cochain complex $L(C^*oS)$, that is we set $H^q(M,S)=H^q(L(C^*oS))$, where o is the tensor product.

(2) With each homomorphism $S->S^1$ and each integer q we associate the homomorphism $H^q(M,S)->H^q(M,S^1)$, induced by the cochain map $L(C^*oS)->L(C^*oS^1)$.

(3) Each short exact sequence $0->S^1->S->S^2->0$ induces a short exact sequence of cochain maps $0->L(C^*oS^1)->L(C^*oS)->L(C^*oS^2)->0$ with which there is an associated homomorphism $H^q(M,S^2)->H^{q+1}(M,S^1)$ that we associate with the short exact sequence $0->S^1->S->S^2->0$ and the integer q.

It can now be shown that the axioms for a cohomology theory are satisfied.

Definition Let H and $\underline{H}$ be two sheaf cohomology theories on M with coefficients in sheaves of K-modules over M. A homomorphism of the cohomology theory H to the theory $\underline{H}$ consists of a homomorphism $H^q(M,S)->\underline{H}^q(M,S)$ for each sheaf S and each integer q, such that tne following conditions hold

(1) For q=0, the following diagram commutes

$$\begin{array}{ccc} \underline{H}^0(M,S) & \sim & L(S) \\ \uparrow & & \uparrow \\ H^0(M,S) & \sim & L(S) \end{array}$$

(2) For each homomorphism S->T and each integer q the following diagram commutes

$$\begin{array}{ccc} \underline{H}^q(M,S) & -> & \underline{H}^q(M,T) \\ \uparrow & & \uparrow \\ H^q(M,S) & -> & H^q(M,T) \end{array}$$

(3) For each short exact sequence of sheaves $0->R->S->T->0$ the following diagram commutes for each integer q

$$\begin{array}{ccc} \underline{H}^q(M,T) & \xrightarrow{d} & \underline{H}^{q+1}(M,R) \\ \uparrow & & \uparrow \\ H^q(M,T) & \xrightarrow[d]{} & H^{q+1}(M,R) \end{array}$$

An <u>isomorphism</u> $H->\underline{H}$ is a homomorphism in which each of the homomorphisms $H^q(M,S)->\underline{H}^q(M,S)$ are isomorphisms.

<u>Theorem</u> Let H and $\underline{H}$ be cohomology theories on M with coefficients in sheaves of K-modules over M. Then there exists a unique homomorphism of H to $\underline{H}$.

<u>Corollary</u> A homomorphism of the cohomology theory H to the theory $\underline{H}$ must necessarily be an isomorphism. Consequently, any two cohomology theories on M with coefficients in sheaves of K-modules over M are uniquely isomorphic.

<u>Theorem</u> Assume that H is a cohomology theory for M with coefficients in sheaves of K-modules over M. Let $0->S->C^0->C^1->C^2->\ldots$ be a fine resolution of the sheaf S. Then there are canonical isomorphisms $H^q(M,S)\underline{\sim}H^q(L(C^*))$ for all q.

Finally it is shown that $0->MxR->E_0(M)->E_1(M)->\ldots$ is a fine torsionless resolution of the constant sheaf MxR where $E_p(M)$ is the sheaf of germs of differential p-forms and that $0->MxK->S_0(M,K)->S_1(M,K)->S_2(M,K)->\ldots$ is a fine torsionless resolution of the constant sheaf MxK where $S_p(M,K)$ is the sheaf of germs of functions on M with values in K.

Given any fine torsionless resolution $0\to MxK\to C_o\to C_1\to C_2\to\ldots$ we have $H_p(M,S)=H_p(L(C^*oS))$. We define homomorphisms $k_p:E_p(M)\to_{inf}S_p(M,R)$ for each integer $p\geq 0$ setting $k_p(w)(s)$ equal to the integral of w over s for each differentiable singular p-simplex s in M. It is an immediate consequence of Stokes theorem that the homomorphisms k_p induce a cochain map $k:E^*(M)\to_{inf}S^*(M,R)$. Let $k_p^*:_{deR}H_p(M)\to_{inf}H_p(M,R)$ denote the induced homomorphism of the cohomology modules, (real vector space). k_p^* is called the de Rham homomorphism.

The de Rham theorem The de Rham homomorphism k_p^* is the canonical homomorphism

$$_{deR}H_p(M)\tilde{=}H_p(M,R)\tilde{=}_{inf}H_p(M,R)$$

for each integer p.

Proof The induced homomorphisms of the associated sheaves form a commutative diagram

$$\begin{array}{ccccccccc} 0\to R\to & S_o(M,R) & \to & S_1(M,R) & \to & S_2(M,R) & \to\ldots \\ \quad id\uparrow & \uparrow k_o & & \uparrow k_1 & & \uparrow k_2 & \\ 0\to R\to & E_o(M) & \to & E_1(M) & \to & E_2(M) & \to\ldots \end{array}$$

Consider the following commutative diagram of cochain complexes in which the rows are exact

$$
\begin{array}{ccccccc}
 & & & & 3 & & \\
0 & \to & S_o^*(M,R) & \to & S^*(M,R) & \to & L(S^*(M,R)) \to 0 \\
 & & \uparrow k & & \uparrow 2 & & \\
 & & 0 & \to & E^*(M) & \to & L(E^*(M)) \to 0 \\
 & & & & 1 & &
\end{array}
$$

1 induces the isomorphism ${}_{de}R^{H_p}(M) \tilde{-} H_p(M,R)$. 2 induces the isomorphism ${}_{inf}H_p(M,R) \tilde{-} H_p(M,R)$. Thus 3 induces de Rham's theorem.

12.1 References

(1) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresmann, 1971.

Appendix III: Amari's Generalised Diakoptics

Kondo, Iri[1] and Amari[2] investigated the topological foundations of tearing with diakoptics and codiakoptics as special cases. Tearing can be said to be a method of solving networks by means of network dissection. Each subcomplex thus dissected also constitutes a complex with the relative topology of the entire complex. We shall deal with the mutual relations between these complexes, so that the mutual relations between the solution of the entire network and those of the subnetworks will be clarified.

Dissection of a complex[1] We use the symbols d for boundary operator and b for coboundary operator. Let X be a connected 2-dimensional complex which is acyclic and acocyclic in the dimension 1, ${}^1X$ a connected closed subcomplex of X and ${}^2X=X-{}^1X$ where ${}^2X$ is an open subcomplex of X. A subcomplex ${}^1X$ of X is called closed when $cl({}^1X)={}^1X$ in X, where $cl({}^1X)$ means the closure of ${}^1X$, namely the set of all elements of ${}^1X$ and all their faces. $st({}^2X)={}^2X$ where $st({}^2X)$ means the star of ${}^2X$, namely the set of all elements of ${}^2X$ and those which have an element of ${}^2X$ as a face. Then every chain of X is split into two parts:

$$C={}^2C+{}^1C \quad {}^2C<{}^2X, {}^1C<{}^1X$$

This kind of splitting of a complex is called dissection. Since ${}^1X$ is closed, a chain of ${}^1X$ has its boundary also in ${}^1X$, ie a cycle of ${}^1X$ is also a cycle of X, and if it bounds in ${}^1X$ it also bounds in X. Therefore we use the same symbol d for the boundary operators in ${}^1X$ as in X. However, ${}^2X$ being an open complement of a closed subcomplex the boundary of a chain of ${}^2X$, when regarded as a chain of X, may lie partly in ${}^2X$ and partly in ${}^1X$. But when a p-chain ${}^2C_p$ of ${}^2X$ is regarded as a chain of the open subcomplex ${}^2X$ itself, the boundary of ${}^2C_p$ in ${}^2X$, which we shall denote by ${}^2d{}^2C$, is to

defined by the intersection

$$d^2C_p = d^2C_p \; n \; {}^jX,$$

ere d^2C_p means the boundary in X of ^2C_p. Therefore a cycle in ^2X is not always a cle in X, but merely a chain of ^2X whose boundary lies in ^1X, and similarly, when a cle in ^2X, together with some chain of ^1X, bounds in X it is defined to bound in ^2X. us defined, the cycles in ^2X are called relative cycles (mod ^1X) and those of X elf are sometimes called absolute cycles.

a dual manner, since ^2X is open, we can define cocycles in ^2X by means of the same boundary operator b as in X, and relative cocycles in ^1X by means of the coboundary erator ^1b:

$$^1C_p = b\,{}^1C_p \; n \; {}^jX.$$

ace, homology and (relative) cohomology are defined in subcomplex ^1X by means of d d ^1b, and cohomology and (relative) homology in ^2X by means of b and ^2d. Moreover ^1X (^2X) is acyclic and acocyclic in the dimension 1 we can define trees and cotrees ^1X (^2X).

clic and acocyclic properties of subcomplexes Throughout this section we assume t X is an n-dimensional connected complex which is acyclic and acocyclic in all the ensions between n-1 and 1, and ^1X is its connected closed subcomplex with ^2X as its n subcomplex such that $X = {}^1X + {}^2X$. Then

If every p-cycle ($0 \leq p \leq n-2$) of ${}^1X$ which bounds in X bounds also in ${}^1X$, then ${}^2X$ is
lic and acocyclic in all dimensions between n-1 and 1. This is a special case of
heorem that the sequence of homomorphisms

$$H^n({}^1X) \to H^n(X) \to H^n({}^2X) \to \ldots \to H^0({}^1X) \to H^0(X) \to H^0({}^2X) \to 0$$

act, where $H^p({}^iX)$ denotes the pth homology group of ${}^iX$. In our case, $H^p(X)$ and
$({}^1X)$ p=2,...,n-1 all vanish, and hence $H^p({}^2X)$ p=2,...,n-1 vanish. Moreover since
$X) \sim H^0(X)$ because of the connectedness of X and ${}^1X$, $H^1({}^2X)=0$.

If ${}^2X$ is acyclic in the dimensions between n-1 and 1, then every p-cycle ($0 \leq p \leq n-2$)
which bounds in X bounds in ${}^1X$.

Under the same assumption as in (1) it is possible to make ${}^1X$ acyclic and
yclic in the dimensions between n-1 and 1.

heorems dual to the above also hold.

ive trees and cotrees We have assumed that the network under consideration is a
cted 2-dimensional complex, acyclic and acocyclic in the dimensions 1. Therefore
irtue of the theorems in the preceding section, every closed subcomplex ${}^1X$ as well
s complement ${}^2X$, an open subcomplex, becomes - after adding 2-cells or meshes to
need be - acyclic and acocyclic in the dimension 1. Hence we can define trees
otrees on ${}^1X$ according to the homology (cohomology) defined by means of $d({}^1b)$,
rees and cotrees on ${}^2X$ according to the homology (cohomology) defined by means of
$b)$ and prove that the definitions based on homology and those based on cohomology
quivalent to each other. We shall therefore call them relative trees and cotrees
mod ${}^2X$ and relative trees and cotrees on ${}^2X$ mod ${}^1X$ respectively, or simply trees
otrees on ${}^1X$ or ${}^2X$. The significance of these concepts - dissection of complex,

and trees and cotrees relative to it - will be clear from the following theorems.

(4) Every union of a tree on ${}^1X$ and a tree on ${}^2X$ is a tree on X.

(5) If the ${}^iX$ part of a tree on X is a tree on ${}^iX$, then the ${}^jX$ part $(i,j=\{1,2\},\ i\neq j)$ is a tree on ${}^jX$.

(4*) Every union of a cotree on ${}^1X$ and a cotree on ${}^2X$ is a cotree on X.

(5*) If the ${}^iX$ part of a cotree on X is a cotree on ${}^iX$, then the ${}^jX$ part $(i,j=\{1,2\},\ i\neq j)$ is a cotree on ${}^jX$.

(6) The branches of a tree on ${}^2X$, together with all the nodes of X, form an a-tree (a wood consisting of a subtrees), where a-1 is the number of independent nodes of ${}^1X$.

(6*) The branches of a cotree on ${}^1X$, together with all the meshes of X, form a b-cotree where b-1 is the number of independent meshes of ${}^2X$.

Projections to and injections from subcomplexes[2] Our aim is to solve a network problem on the original complex X with the help of subcomplexes ${}^1X$ and ${}^2X$, for it enables us to use the topological information more completely than by any other method. Hence our present purpose is to find the relations between the groups of chains, cycles and boundaries of ${}^iX$ over a suitable coefficient domain and those of X itself. To investigate these relations, four fundamental chain transformations ${}^1i$, ${}^2i$, ${}^1p$ and ${}^2p$ will be defined, and they will play important roles as follows.

The projection ${}^{i}p:C\to{}^{i}C$ is naturally defined regarding the ${}^{i}X$ part of a chain C as a chain ${}^{i}C$ of ${}^{i}X$, where C is the group of chains of X and ${}^{i}C$ is the group of chains of ${}^{i}X$ (i=1,2). We shall write ${}^{1}p:C\to{}^{1}C$, ${}^{2}p:C\to{}^{2}C$. This operation ${}^{i}p$ is called the projection of X onto ${}^{i}X$. Physically speaking, to operate ${}^{i}p$ on a chain C means to regard physical quantities (such as currents) represented by C as those of ${}^{i}X$.

Since the elements of ${}^{i}X$ are among those of X, we have an injection ${}^{i}i:{}^{i}C\to C$ by regarding a chain of ${}^{i}X$ as a chain of X itself. We call the operation ${}^{i}i$ the injection of ${}^{i}X$ into X. This means physically that we regard the physical quantities of ${}^{i}X$ as those of X.

The following two relations are easily proved from the facts that $X={}^{1}X+{}^{2}X$ and each ${}^{i}X$ is the complement of the other.

$${}^{1}i\,{}^{1}p+{}^{2}i\,{}^{2}p=1 \quad \text{(identity operator)}$$

$${}^{i}p\,{}^{j}p=d^{ij} \quad \text{(Kronecker's delta)} \quad i,j=1,2$$ This means that ${}^{1}X$ and ${}^{2}X$ have no common element.

Theorem No information is lost by operating ${}^{1}i\,{}^{1}p+{}^{2}i\,{}^{2}p$ on a chain of X, that is, by projecting a chain of X into ${}^{1}X$ and ${}^{2}X$ respectively and then gathering them by the injections from both ${}^{1}X$ and ${}^{2}X$. This theorem shows the validity of using the dissection processes or diakoptics as a method to solve network problems.

Boundary operators in ${}^{i}X$ In subcomplexes ${}^{1}X$ and ${}^{2}X$, the incidence relations are the same as in X. Therefore the boundary operators ${}^{1}d$ in ${}^{1}X$ and ${}^{2}d$ in ${}^{2}X$ are so defined that ${}^{i}d\,{}^{i}X$ is the ${}^{i}X$ part of dC, where C is the injection of ${}^{i}C$ into X. Thus we have

$^1d=^1pd^1i$, $^2d=^2pd^2i$

Since X is closed $d(^1i^1C)$ is always on ^1X only, so that we have

$^2pd^1i=0$

But ^2X is not closed and $d(^2i^2C)$ may have an ^1X part. Therefore we have a non zero operator

$^1pd^2i=^{12}d$

^{12}d is an operator which transforms $^2C^r$ into $^1C^{r-1}$, consisting in taking the ^1X part of the boundary $d(^2i^2C)$ of ^2C in X. Hence ^{12}d represents the connection relations of ^1X and ^2X, and the influences of ^2C are carried over to ^1C as we shall see later.

The coboundary operators ^2b in ^2X and ^1b in ^1X are defined by duality.

$^2b=^2pb^2i$, $^1b=^1pb^1i$

By expanding $d=(^1i^1p+^2i^2p)d(^1i^1p+^2i^2p)$ and using $dd=0$ we obtain $^id^id=0$ (and $^ib^ib=0$).

We can also show that the operators ^1i and ^1p commute with the boundary and coboundary operators, and that ^1i and ^1p are dual mappings, $^1i^*=^1p$. Therefore $(^1i,^1p)$ is a pair of dual chain mappings. (A chain mapping is a chain transformation that commutes with the boundary operator.) In the same way $(^2p,^2i)$ is another pair of dual chain mappings.

Network analysis by dissection For the rest of this chapter we will represent chains by I,i and cochains by E,e. Assume that no mutual couplings exist between ${}^1X$ and ${}^2X$. Consider the admittance matrix to be a mapping $y:E_1 \to I^1$ we may write ${}^1py^2i=0$, ${}^2py^1i=0$, similarly for $z=y^{-1}$. We define the admittance and impedance matrices of ${}^iX$ as ${}^iy={}^ipy^ii$, ${}^iz={}^ipz^ii$ and we see ${}^iy^iz={}^iz^iy=1$ the unit matrix (for ${}^iy^iz={}^ipy^ii^ipz^ii={}^ipyz^ii-{}^ipy^ji^jpz^ii={}^ip^ii=1$, $i\neq j$) and we obtain

$y=({}^1i^1p+{}^2i^2p)y({}^1i^1p+{}^2i^2p)={}^1i^1y^1p+{}^2i^2y^2p$ and $z={}^1i^1z^1p+{}^2i^2z^2p$.

In ordinary cases, mutual couplings even if they do exist concentrate only locally, so that this assumption will be satisfied in most cases, if we dissect (tear) the network appropriately. The more general case in which this assumption is not satisfied can be dealt with by extending the method.

Fundamental equation of diakoptics and codiakoptics Under the above assumptions we introduce the fundamental equation of Amari's method as follows. For ${}^1X$ taking account of Kirchoff's 2nd law $b(E_1-e_1)=0$, we can put $E_1-e_1=bE_o$. Operating on this by ${}^1p$ we have ${}^1E_1={}^1pbE_o={}^1b^1E_o$ for ${}^1pe_1=0$ and ${}^1pb={}^1b^1p$. This is Kirchhoff's 2nd law in ${}^1X$.

Next from ${}^1pd(i^1-I^1)=0$ we have ${}^1d({}^1i^1-{}^1I^1)=-{}^{12}d({}^2i^1)$ using ${}^1pd={}^1d^1p+{}^{12}d^2p$ and ${}^2pI^1=0$. This is Kirchoff's 1st law in ${}^1X$. Since $y={}^1i^1y^1p+{}^2i^2y^2p$, Ohm's law in ${}^1X$ takes the form ${}^1i^1={}^1y^1E_1$ giving ${}^1d^1y^1b^1E_o={}^1d^1I^1-{}^{12}d^2i^1$ or for ${}^2X$, ${}^2b^2z^2d^2i^2={}^2b^2e_1-{}^{21}b^1E_1$

${}^1d^1y^1b^1E_o+{}^{12}d^2d^2i^2={}^1d^1I^1$

$^{21}{}_b{}^1{}_b{}^1E_o+{}^2{}_b{}^2{}_z{}^2{}_d{}^2i^2={}^2{}_b{}^2e_1$

These are Amari's fundamental equations of tearing. The following table gives the matrix coefficients corresponding to each (co)boundary operator, operating on a p-(co)chain:

2-chain		1-chain	0-cochain		1-cochain
d	Cbm	Anb	b	Anb	Cbm
$^1{}_d$	$C^1{}_b{}^1{}_m$	$A^1{}_n{}^1{}_b$	$^1{}_b$	$A^1{}_n{}^1{}_b$	$C^1{}_b{}^1{}_m$
$^2{}_d$	$C^2{}_b{}^2{}_m$	$A^2{}_n{}^2{}_b$	$^2{}_b$	$A^2{}_n{}^2{}_b$	$C^2{}_b{}^2{}_m$
$^{12}{}_d$	$C^1{}_b{}^2{}_m$	$A^1{}_n{}^2{}_b$	$^{12}{}_b$	$A^1{}_n{}^2{}_b$	$C^1{}_b{}^2{}_m$

Hence Amari's fundamental equations in matrix form are:

$$A^1{}_n{}^1{}_bY^1{}_b{}^1{}_bA^1{}_b{}^1{}_nE_o{}^1{}_n+A^1{}_n{}^2{}_bC^2{}_b{}^2{}_mi^{22}{}_m=A^1{}_n{}^1{}_bI^{11}{}_b$$

$$C^2{}_m{}^1{}_bA^1{}_b{}^1{}_nE_o{}^1{}_n+C^2{}_m{}^2{}_bZ^2{}_b{}^2{}_bC^2{}_b{}^2{}_mi^{22}{}_m=C^2{}_m{}^2{}_be_1{}^2{}_b$$

From dd=0 we obtain $^1{}_d{}^{12}{}_d+{}^{12}{}_d{}^2{}_d=0$. Defining $-\underline{C}=C^1{}_n{}^2{}_m=A^1{}_n{}^2{}_bC^2{}_b{}^2{}_m=-A^1{}_n{}^1{}_bC^1{}_b{}^2{}_m$, we interpret $C^1{}_n{}^2{}_m$ as the incidence number between node $^1{}_n$ of ^1X and mesh $^2{}_m$ of ^2X, the mutual influence between $E_o{}^1{}_n$ and $i^{22}{}_m$. Of course the elements of $\underline{C}$ are simple integers 0,1 and -1. Putting

$$^1\underline{Y}=A^1{}_n{}^1{}_bY^1{}_b{}^1{}_bA^1{}_b{}^1{}_n,$$

$$^2\underline{Z}=C^2{}_m{}^2{}_bZ^2{}_b{}^2{}_bC^2{}_b{}^2{}_m,$$

$\underline{e}=E_0\,{}^1n,\ {}^2\underline{i}=i^{22}m,$

$\underline{I}=I^{11}b,\ {}^2\underline{E}=e_1\,{}^2b,$

e fundamental equations can be rewritten in their original form

$$\begin{bmatrix} {}^1\underline{Y} & -\underline{C} \\ \underline{C}^1 & {}^2\underline{Z} \end{bmatrix}\begin{bmatrix} {}^1\underline{e} \\ {}^2\underline{i} \end{bmatrix}=\begin{bmatrix} {}^1\underline{I} \\ {}^2\underline{E} \end{bmatrix}$$

e and ${}^2\underline{i}$ form the diakoptical coordinates of the network. Usually we dissect X in ch a way that either ${}^1X$ or ${}^2X$ is composed of several disconnected subdivisions and mutual couplings exist between them. Either ${}^1\underline{Y}$ or ${}^2\underline{Z}$ will be a block diagonal trix. It is therefore efficient to solve by partitioning because the topological formation in the network has been utilised. The special cases diakoptics and diakoptics are clearly duals.

e procedure of diakoptics In this section we will show the practical procedures of ari's method. We solve the fundamental equation by partitioning the coefficient trix.

lving the first equation we get

$\underline{e}=/{}^1\underline{Y}({}^1\underline{I}+\underline{C}^2\underline{E})$

ere calculating ${}^1\underline{Z}=/{}^1\underline{Y}$, as well as $/\underline{Z}^1\underline{I}$, which is the node voltage vector of the rn subnetworks, corresponds to solving the subnetworks. Substituting in the second uation

$\underline{i}=/(\underline{C}^{11}\underline{ZC}+^{2}\underline{Z})\,(^{2}\underline{E}+\underline{C}^{1}\underline{ZI})$

lves the intersection network, where the denominator is the mesh impedance matrix of e intersection network. The solution $E^{1}b$ of node ^{1}b in ^{1}X is given by

$=^{1}\underline{Z}(\underline{I}+\underline{C}^{2}\underline{i})$.

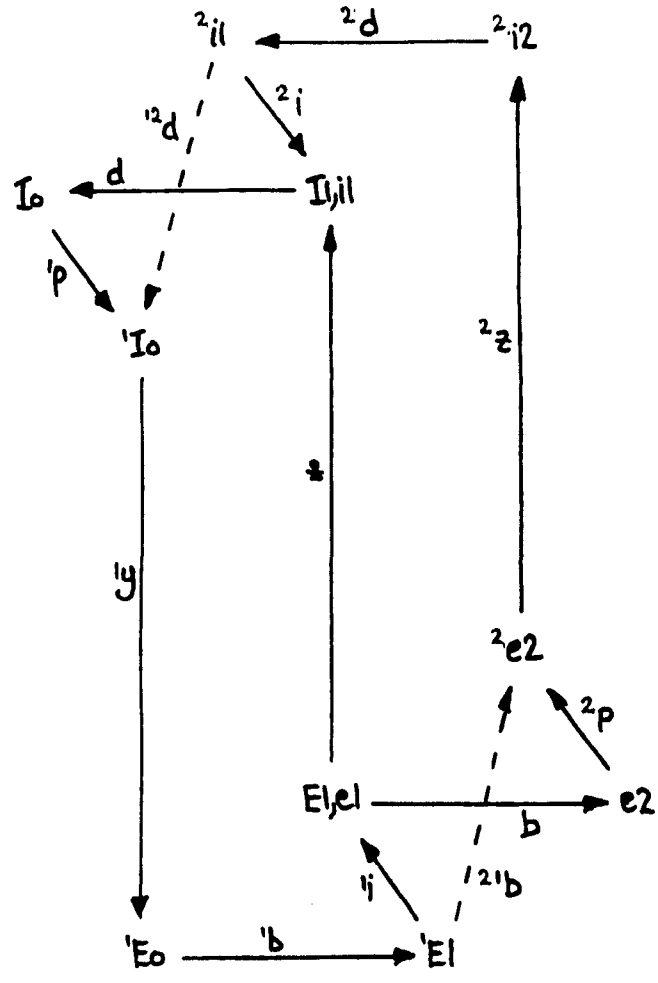
e branch currents and voltages are easily obtained

$b=C^{2}b^{2}mi^{2}m,\quad E^{2}b=Z^{2}b^{2}bi^{2}b,$

$b=Z^{1}b^{1}bE^{1}b,\quad E^{1}b=A^{1}b^{1}nE^{1}n,$

thus see that the diakoptical procedure coincides with our method of partitioning e fundamental equation. These steps are summarised in Fig(13). When the impressed antities $I^{2}n$ and $e^{2}m$ are not yet explicitly given we have only to calculate $^{1}Z^{1}n^{1}n$ d $^{2}Y^{2}m^{2}m$ (the factorised inverse matrices) thereafter if these impressed quantities e given we can immediately calculate the response quantities using $^{1}Z^{1}n^{1}n$ and $^{2}m^{2}m$.

codiakoptical analysis dual to the above results in the same diagram but with the art point in the bottom centre instead of the top centre.



Fig(13) Amari's algebraic diagram for diakoptics

13.1 References

(1) Kondo and Iri, On the theory of Trees, Cotrees, Multi-trees, and Multi-cotrees, RAAG Memoirs, Vol 2, 1958.

(2) Amari, Topological Foundations of Kron's Tearing of Electric Networks, RAAG Memoirs, Vol 3, 1962.

Bibliography

(1) Athens et al, Systems, Networks and Computation: Multivariable Methods, McGraw Hill, 1974.

(2) Athans and Falb, Optimal Control, McGraw Hill, 1966.

(3) Bertalanffy, General Systems Theory, A. Lane, 1972.

(4) Box and Jenkins, Time Series Analysis, Holden Day, 1976.

(5) Brammeller, Practical Diakoptics for Electrical Engineers, Chapman and Hall, 1969.

(6) Brooklyn Polytechnic, Proceedings of a Symposium on Generalised Networks, 1966, Vol XVI.

(7) Van Dixhorn, Physical Structure in Systems Theory: Network Approaches to Engineering and Economics, Academic Press, 1975.

(8) Hammer, Advances in Mathematical Systems Theory, Pennsylvania Univ Press, 1969.

(9) Happ, Diakoptics and Networks, Academic Press, 1971.

(10) Hawkins and Ellis, Large Scale Structure of Space Time, Cambridge Univ. Press, 1973.

(11) Himmelblau, Proceedings of a Conference on Decomposition of Large Scale Systems, Cambridge, 1972, North-Holland/ American Elsevier 1973.

(12) Journal of the Franklin Institute, Special Issue: Modern Aspects of Large Scale Systems Science, Vol 286, No 6, Dec 1968.

(13) Kalman, A new approach to Linear Filtering and Prediction Problems, Journal of Basic Engineering, p35, March 1960.

(14) Kalman and Bucy, New Results in Linear Filtering and Prediction Theory, Ibid, 95, March, 1961.

(15) Kalman and Falb, Topics in Mathematical Systems Theory, McGraw Hill, 1969.

(16) Klir, Approach to General Systems Theory, Van Nostrand Reinhold, 1971.

(17) Klir, Trends in General Systems Theory, Interscience 1972.

(18) Kron, Diakoptics - the Piecewise Solution of Large Scale Systems, London, McDonald, 1963.

(19) Lasdon, Optimisation Theory for Large Systems, Collier MacMillan, 1970.

(20) Luenberger, Optimisation by Vector Space Methods, Wiley, 1968.

(21) MacLane and Birkhoff, Algebra, 2nd Edition, Collier MacMillan, 1979.

(22) Mesarovic and Takahara, General Systems Theory: Mathematical Foundations, Academic Press, 1975.

(23) Mesarovic et al, Theory of Hierarchical, Multilevel Systems, Academic Press, 1970.

(24) Nikaido, Convex Structures and Economic Theory, Academic Press, 1968.

(25) Newman, Readings in Mathematical Economics, I and II, John Hopkins 1968.

(26) Oppenheimer, Lectures on Electrodynamics, Gordon and B, 1970.

(27) RAAG Memoirs, The Unifying Study of Basic Problems in Engineering by means of Geometry, Vols I, II, III and IV.

(28) Reid, Riccati Differential Equations, Academic Press, 1972.

(29) Sage, Optimum Systems Control, Prentice Hall, 1968.

30) Sage, Methodology for Large Scale Systems, McGraw Hill, 1977.

31) Sage and Melsa, Estimation Theory with Applications to Communication and Control, McGraw-Hill, 1971.

32) Steenrod, The Topology of Fiber Bundles, Princeton Univ Press, 1951.

33) Stone, Mathematical Models of the Economy and other essays, Chapman Hall, 1970.

34) Wise, The Hierarchical Structure of Large Scale Systems, Thesis for PhD, Dept of Control Eng, Univ of Sheffield, 1975.

35) Wismer, Optimisation Methods for Large Scale Systems... with Applications, McGraw-Hill, 1971.

36) Warner, Foundations of Differential Manifolds and Lie Groups, Scott Foresman, 1971.

37) Wonham, Linear Multivariable Control: a Geometric Approach, Springer Verlag, 1974.

38) Zadeh, Fuzzy Sets, Information and Control, Vol 8, pp 338-53, 1965.