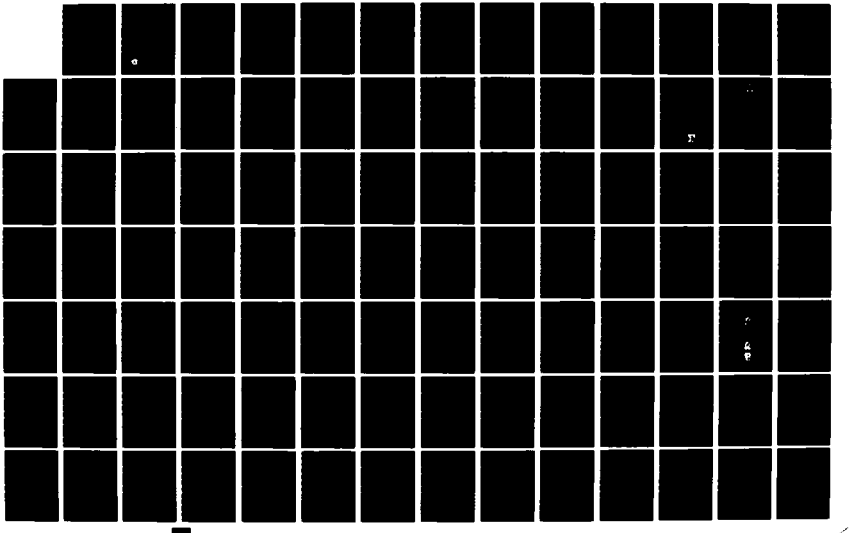
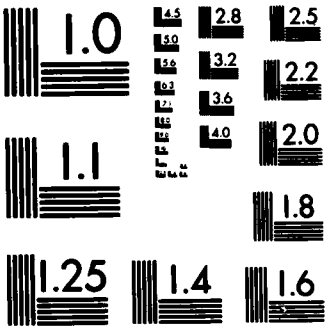
AD-A133 491 ANALOG FAULT DIAGNOSIS OF LARGE-SCALE ELECTRONIC CIRCUITS(U) NOTRE DAME UNIV IN DEPT OF ELECTRICAL ENGINEERING R LIU 01 AUG 83 N00014-78-C-0444 1/2

UNCLASSIFIED F/G 14/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(12)

FINAL REPORT

Analog Fault Diagnosis of Large-Scale Electronic Circuits

by

Ruey-wen Liu
Principal Investigator

Department of
ELECTRICAL ENGINEERING

UNIVERSITY OF NOTRE DAME, NOTRE DAME, INDIAN

FINAL REPORT

Analog Fault Diagnosis of Large-Scale Electronic Circuits

by

Ruey-wen Liu
Principal Investigator

Prepared for

Office of Naval Research
Electronic and Solid State Sciences Program
800 N. Quincy Street
Arlington, Virginia 22217

Contract No. N00014-78-C-0444

1 August 1983

TABLE OF CONTENTS

I. INTRODUCTION

During the period June 1, 1978 to May 13, 1982, the research "Fault Diagnosis of Large Scale Analog Systems" is sponsored by the Office of Naval Research under Contract Number N00014-78-C-04444. This research is undertaken at the University of Notre Dame, Notre Dame, Indiana under the direction of the Principal Investigator, Dr. Ruey-wen Liu. The long-term objective of this research is to develop a practical and reliable Automatic Test Program Generator (ATPG) which will allow us to locate the faulty component(s) of a large analog circuit when it is faulty. The short-term objective is to search for viable and amenable concepts under which long-term objectives can be achieved. During this short period, some significant progress has been made.

To be sure, the fault analysis of analog circuit is an uncharted area of research. There is no precedence with which we can follow. Indeed, this has been an exciting, enjoyable and satisfying research. We continuously discover new problems which, usually dictates the requirements of new methods. As such, new directions of research have been continuously searched and persued. In the end, we believe we have a very practical and reliable method at hand. Some highlights of this development will be presented In Sectin IV. In the meantime, twenty-two publications have been published, one Ph.D dissertation and three Master theses have been completed, and a final report of a Workshop on Analog Automatic Test Program Generation has been completed.

Recently, Dr. Ruey-wen Liu was invited to give an hour-long tutorial address to the 1983 IEEE International Symposium on Circuits and Systems in May, 1983. This was a rare honor for the Principal Investigator. It also shows that the general academic public has begun to show interest in this exciting area of research.

The rest of the Final Report is organized in this way. The background of the analog analysis is presented in Section II. Some important issues are given in Section III. Some highlights of our research activities will be discussed in Section IV. A conclusion is given in Section V. Finally, a record of publications is in Section VI, and selected reprints are attached in the Appendix.

II. THE BACKGROUND

During the past quarter century, the engineering community has witnessed a tremendous strides in the art of electronics design. On the contrary, electronics maintenance has changed little since the day of the vacuum tube. As such, our ability to design a complex electronic circuit is quickly outdistancing our ability to maintain it. In turn, the price reductions which have accompanied modern electronics technology have been paralleled by increasing maintenance and operation costs. Indeed, many industries are finding that the life cycle maintenance costs for their electronic equipment now exceeds their original capitol investment.

Given the above, it is quickly becoming apparent that the electronics maintenance process, like the design process, must be automated. Unfortunately, the 50 years of progress in circuit theory, on which our electronics design automation has been predicated, does not exist in the maintenance area. As such, the past decade has witnessed the inauguration of a basic research program to lay the foundations for a theory of electronics maintenance and a parallel effort to develop operational electronic maintenance codes.

Thus far the greatest success has been achieved in the digital electronics area to the point that commercialized test programs are now readily available. On the other hand, the analog testing is still in its infancy. This is not without reasons.

For one reason, the analog fault diagnosis had a late start. The research and theory development of digital testing started in the mid 1960's when the large-scale computers were readily available. Not until a decade later did a commercialized test program first become available. On the other hand, it was not until the mid 1970's that the test technology community began to face up to the analog test problem. Indeed, <u>even in a predominantly</u>

digital world, analog systems were not disappearing. Analog systems were proving to be among the most unreliable and least readily tested of all electronic systems. Assuming the same speed for the development of digital testing, a commercially available analog testing program would not have been ready until the mid 1980's.

There are two main reasons for this seemingly slow development of an effective method for analog fault diagnosis. One reason is that this is an uncharted area of research. It has no precedence to follow. More time has been spent to find what the real problems are and where the difficulty is. Another reason is that analog fault diagnosis has inherited certain difficult problems which are not shared by digit fault diagnosis. These will be explained later.

III. IMPORTANT ISSUES OF ANALOG FAULT DIAGNOSIS

From many experiences accumulated in the past years, successful ones as well as unsuccessful ones, we can determine two major issues for analog fault diagnosis, not shared by its counter part, the digital fault diagnosis. The two major issues are the tolerance problem and the problem of modeling and simulation of faulty components:

- Tolerance: Possibly the single greatest unknown in the design of an analog testing program is the effect of the tolerances of the "good" component on the performance of a testing program. This tolerance problem has absolutely no counterpart in the digital testing problem. The effect of these tolerances can completely dominate the performance of a testing program. In an analog circuit, unlike digital circuits, the actual values of circuit parameters almost always deviates from the nominal values. Therefore, any analog testing program has to face up to the problem of tolerance problem.
- Modeling and Simulation of Faulty Components: Unlike the digital testing, a complete modeling (and thus simulation) of faulty components is not available for the development of a testing problem. The modes of faulting is too many to encounter. For example, a faulty resistor may have an infinite number of possible resistances (outside of the tolerance). In fact, it can even be nonlinear. A faulty capacitor may have a model of parallel RC. A faulty operational amplifier may have a model of 22 transistors 12 resistors and a capacitor! A good transistor may behave like a faulty one if its bias is switched due to a fault which occurred elsewhere! In fact, in a nonlinear analog environment, we are still in the process of developing viable CAD models for nominal devices, let alone for faulty devices. As such, a thorough test of the performance of a testing

program is impossible. Furthermore, each testing program has to be designed based soley on the nominal values of the circuit.

After discussions with potential users, we also find that there are three important measures for the effectiveness of a testing program.

- Test Points. Due to the practical restriction that there are usually only a few nodes accessible for measurement and testing, the number of required test points has to be as small as possible.
- Post-Fault Computation. Since the post-fault computation is directly related to the per unit cost, it is important to keep the post-fault computation time short and simple.
- Robustness. This issue has been raised many times; only because it is indeed the dominant issue at hand. We want the testing program to be reliable when the fault/tolerance ratio is small.

These two major issues and there effective measures have guided our research in the past years.

IV. HIGHLIGHTS OF RESEARCH EFFORTS

The fault analysis of analog circiut, to be sure, is an uncharted area of research. There is no precedence with which we can follow. Indeed, this has been an exciting, enjoyable and satisfying research. We continuously discover new problems which usually dictates the requirements of new methods. As such, new directions of research have been continuously searched and pursued. In the end, we believe we have a very practical and reliable method at hand. Some highlights of this development will be presented here.

The initial phase of our research was placed on understanding the problem. One of the important problems we have learned in the period was the trade-off problem between the number of text points and the complexity of fault diagnosis computation. It is known that when all the nodes can be used as test points, the fault diagnosis equation to be solved is linear and the computation is relatively simple. However, when the number of test points are reduced then the fault diagnosis equation becomes more and more nonlinear. Solving these simultaneous nonlinear equations involves complex computation and the results are less reliable. Under this constraint, we have attempted to pick the "best" trade-off point. We have found that if the number of test points are reduced at particular locations, the fault diagnosis equations become a set of "nice" nonlinear equations, i.e., the sequentially linear equations. This set of nonlinear equations has a nice property that can be solved iteratively by a linear equation. Therefore, we can reduce the number of test points but not increase the complexity of computation. This part of research was a successful one, see [1] and [2], but the result was not good enough to have practical use. It still requires more than practically allowed test points to achieve sequentially-linear diagnosable circuits. As such, we have to look for new directions.

In the next phase, many ideas have been tried and tested: the modular approach [3], the accessibility approach [5], a comparison study of frequency-domain approach and time-domain approach [8] and the fualt directory approach [12]. We have also investigated the estimation, theory aspect [9,10,11], the computational aspect [4] and the feedback system design aspect [6] of the problem. All these are novel and useful but only for a particular situation. We are still left with the lack of a central thought or framework under which all these approach can be applied. This leads to the third and the final phase of our research.

At this point, we begin to realize that we have been trying to solve an impossible problem, i.e., the isolation of all possible combinations of faults with a limited few test points. With limited test points available, we can isolate all possible faults only in theory, but not in practice. Because it is theoretically possible, it kept us on the wrong track for a long time. The difficulty in the implementation of such a theory is that the theory is based on the ideal case, i.e., no tolerance for non-fault elements. This tolerance effect forces us to look for new directions again.

With a limited few test points, our goal is to isolate only a limited combination of faults. We look for only the cases when numbers of faulty components are limited by a few, say k. This is the k-fault diagnosis problem. By turning our direction in this way, we recognize that we will miss some situations, such as when the number of fault elements are greater than k. However, the k-fault diagnosis problem makes a lot of sense. First, a system is faulty usually because only a few of its components are faulty. A large number of components become faulty at the same time is a rare occurrence. Therefore, we do not miss too much by considering the k-fault diagnosis problem. Secondly, the k-fault diagnosis problem was found to be mathematically

tractable and computationally simple. Finally, by doing so, we do no less than our counterpart, the fault diagnosis of digital systems. In our counterpart, most of their successful ones are only for single-fault diagnosis, i.e., k=1. In the last two years, our attention has soley been devoted to the k-fault diagnosis problem. Fortunately, we have been very successful and had some breakthrough results [14,16,17,18,19,20,21,22].

At this stage, we have only considered the linear circuits. The cases of nonlinear circuits and analog/digital hybrid circuits are left for future research. In the case of linear circuits, the following has been achieved:

1. The k-faulty diagnosability problem is completely solved.
2. A method for the design of the location of the test points so that a circuit becomes k-fault diagnosable is available.
3. A simple and robust computational method is developed to implement the above theories.
4. The number of test points required is small even for large and complex circuits.

As a final test to our porogram, we have tested our method on a circuit jointly supplied by the NAVAIR and the Naval Air Engineering Center. This circuit is a video amplifier which consisted of 21 nodes and 38 components, 8 of which were transistors. Even for digital circuits, 38 components required 7 test points. We needed only 5 test points for our analog circuit. It is well known that the diagnosis of analog circuits is a much harder problem than that of digital circuits. This demonstrates that our method required very few test points.

In the meantime, for the same NAVY circuit, the computation time for each fault-isolation test was less than 10 seconds on the IBM370. This is amazing, for it translates to less than one dollar per unit cost!

Finally, our method has been demonstrated to be reliable. The fault deviation to the tolerance deviation ratio is about 200% to 5%. Although the ratio may be a little high, this is the first method which was shown to have the capability of taking any tolerance at all.

V. CONCLUSIONS

In summary, we started with an uncharted area of research. During the four years, the research experiences have sometimes been painful, but always exciting and rewarding. Many directions have been tried and failed. But each time we have learned something new. Finally, a practical, robust, and yet elegant method is at hand. This method has been tested by a circuit supplied by the NAVAIR and the Naval Air Engineering Center. The result is satisfactory and successful.

The problem of fault diagnosis began to attract academic interest. It has always been that the problem of fault diagnosis is considered as a problem belonging to repairing shops. This point of view has been changed. There have been special sessions on analog fault diagnosis in the IEEE International Symposium on Circuits and Systems in the past two years and definitely will also be in the next two years. The principle investigator has been invited to give talks on analog fault diagnosis problems at the Allerton Conference on Communication, Control and Computing [3], the National Electronic Conference [7], the IEEE Conference on Decision and Control [14], the IEEE International Symposium on Large Scale Systems [16] and many times at the IEEE International Symposium on Circuits and Systems [5,12,18,19]. The last talk [19] was an hour-long tutorial speech. He was also invited to submit a paper by the Journal of Society of Instrument and Control Engineers of JAPAN [15].

Up to this point, we have just completed the first phase of our research, the deterministic model approach. Next, we will take the stochastic model approach, and finally, the artificial intelligence model approach.

The stochastic model will incorporate the uncertainties, such as the tolerance into the model. We have started and nearly completed this part of the research. Though this part of the research was carried out after the termination date of the ONR contract, it may be relevant to report that some breakthrough results have been obtained. Theoretically, we achieved the Cramer-Rao lower bound, and hence, we have obtained the most efficient algorithm. Practically, it recovered all the faults that were missed by using the deterministic model approach before. With these encouraging results, we are very optimistic that the problem of analog fault diagnosis will be completely understood in the very near future.

VI. LIST OF PUBLICATIONS (1978-82)
(Under ONR Contract N00014-78-C-0444)

A) Publications

1. R. Liu and V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part I-The Theory", IEEE Trans. on Circuits and Systems, (A special issue on Automatic Analog Fault Diagnosis), pp. 490-495, July 1979.

2. V. Visvanathan and R. Liu, "Sequentially Linear Fault Diagnosis: Part II-The Design of Diagnosable Systems", (Special Issue) IEEE Trans. on Circuits and Systems, pp. 558-564, July 1979.

3. C.S. Lin and R. Liu, "A Single-Fault Diagnosis Theory", (Invited) Allerton Conf. on Communication, Control and Computing, pp. 172-178, 1979.

4. L.C. Suen and R. Liu, "The Determination of the Rank of a Large Noisy Matrix", (Invited) Proc. Workshop on Large Scale Networks and Systems, Houston, TX, pp. 40-43, 1980.

5. R. Liu, V. Visvanathan and C.S. Lin, "Accessibility of Large-Scale Electronic Circuits", (Invited) Proc. 1980 IEEE International Symposium on Circuits and Systems, Tokyo, Japan, 1980.

6. C.A. Desoer, R. Liu, J. Murray and R. Saeks, "Feedback System Design: The Fractional Representation Approach to Analysis and Systems", IEEE Trans. on Automatic Control, pp. 399-412, 1980.

7. R. Liu, C.S. Lin, A. Deng and V. Raman, "System Diagnosis - A New System Problem", (Invited) Proc. National Electronics Conference, pp. 280-285, 1980.

8. C.S. Lin, "A Comparison of Frequency-Domain Approach and Time-Domain Approach: A Case Study of Fault Analysis of Analog Circuits", Notre Dame Technical Report, TREE-802, May 23, 1980.

9. Y.F. Huang, E. Fogel and J.B. Thomas, "An Alternative Consideration in Singular Linear State Estimation", Proc. 23rd Midwest Symp. on Circuits and Systems, August 1980.

10. E. Fogel and Y.F. Huang, "Reduced-Order Optimal State Estimator for Linear Systems with Partially Noise Corrupted Measurement", IEEE Trans. on Automatic Control, pp. 994-996, 1980.

11. E. Fogel and K. McGill, "Spectral Analysis of Previewing Controllers", IEEE Trans. on Automatic Control, pp. 959-967, 1980.

12. C.S. Lin and R. Liu, "A Fault Directory Approach to Analog Fault Analysis - A Case Study", (Invited) Proc. 1981 IEEE International Symposium on Circuits and Systems, Chicago, IL, 1981.

13. R. Saeks and R. Liu, Proc. of the Workshop on Analog Automatic Test Program Generation, University of Notre Dame, IN, May 26-27, 1981.

14. Z.F. Huang, C. Lin and R. Liu, "Node-Fault Diagnosis and a Design of Testability", (Invited), Proc. 20th IEEE Conf. on Decision and Control, pp. 1037-1042, 1981.

15. R. Saeks and R. Liu, "Fault Diagnosis in Electronic Circuits", (Invited), J. of Society of Instrument and Control Engineers, Vol. 20, pp. 214-216, Tokyo, Japan, 1981.

16. C.S. Lin, Z.F. Huang and R. Liu, "Topological Conditions for Single-Branch Fault", (Invited), Proc. 1982 IEEE Int'l. Large Scale Systems Symp., pp. 302-307, 1982.

17. Z.F. Huang, C. Lin and R. Liu, "Topological Conditions on Multiple-Fault Testability of Analog Circuits", Proc. 1982 Int'l Symp. on Circuits and Systems, Rome Italy, May 1982.

18. C.S. Lin, Z.F. Huang and R. Liu, "Linear Analog Fault Analysis - Theory and Implementation", (Invited) 1983 IEEE International Symp. on Circuits and Systems, pp. 1090-1093, May 1983.

19. R. Liu, C.S. Lin, Z.F. Huang and L.Z. Hu, "Analog Fault Diagnosis - A New Circuit Theory", 1983 IEEE International Symp. on Circuits and Systems, pp. 931-939, 1983.

20. Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and a Design of Testability", IEEE Trans. on Circuits and Systems, pp. 257-265, May 1983.

21. C.S. Lin, Z.F. Huang and R. Liu, "Topological Conditions for Single-Branch Fault", IEEE Trans. on Circuits and Systems, (to appear).

22. C.S. Lin and R. Liu, "Identifiability of an Imbedding Unknown Subnetwork", TR No. EE833, University of Notre Dame, Notre Dame, IN, June 1983.

B) M.S. and Ph.D. Theses

1. V. Visvanathan, "Sequantially-Linea Fault Analysis", M.S. Dissertation, University of Notre Dame, Notre Dame, IN, 1978.

2. Yih-Fang Huang, "System Identification Using Number-Set Description of System Uncertainties and a Reduced-Order Optimal State Estimation", M.S. Dissertation, University of Notre Dame, Notre Dame, IN, 1979.

3. An-Chang Deng, "Rank of a Large Sparse Matrix", M.S. Dissertation, University of Notre Dame, Notre Dame, IN, 1981.

4. Chen-Shang Lin, "Reactability, Excitability and Testability of Analog Networks", Ph.D Dissertation, University of Notre Dame, Notre Dame, IN, 1979.

C) Workshop Report

1. R. Saeks and R. Liu, Workshop on Analog Automatic Test Program Generation, University of Notre Dame, Notre Dame, IN, May 26-27, 1981.

APPENDIX: Reprints of Publications

Sequentially Linear Fault Diagnosis:

Part I - The Theory

by

R. Liu
V. Visvanathan

Sequentially Linear Fault Diagnosis: Part I—Theory

RUEY-WEN LIU, MEMBER, IEEE, AND V. VISVANATHAN

Abstract—A good solution to the tradeoff problem between the cost of computation and the cost of test points in the sequentially linearly diagnosable systems. Conditions under which a system is sequentially linearly diagnosable are developed in Part I. A design procedure for the test points to fulfill these conditions is given in Part II.

I. Introduction

In fault diagnosis of analog systems we encounter two classes of faults—catastrophic faults and soft faults [17]. The former occurs due to an extreme change in the performance of some component(s) (a fuse blows or a circuit shorts out) and results in outright failure of system performance. Diagnosis of such faults [1] is usually based on the acceptable premise that the system failure is due to a catastrophic fault in one (or a few) of its components. On the other hand, when the values of the parameters of the system drift there is a degradation of system performance which is called a soft fault [17]. The drift in parameter values is usually caused by permanent overstress (high temperature, continued overload operation, material stress, etc.) or aging. In such cases one has to make the decision that the system performance has sufficiently degraded for the system to be declared faulty, and the problem of diagnosing the fault is more difficult than in the catastrophic case. This is because the faulty condition of the system may be caused by changes in the values of many parameters. However, if we can determine all the parameter values, not only can we decide if a soft fault has occured, but also locate the faulty parameters. This approach will be used in this paper.

In this paper we consider soft faults in linear large-scale dynamical systems (LSDS), i.e., an interconnected system whose components are linear and time invariant. When we study such systems from the point of view of fault diagnosis, we consider it to have two parts—the system parameters and the system structure. The system parameters are real-valued variables. The system structure consists of all the parts of the system that are known (or assumed) to be invariant. This may include the location of the inputs and outputs, the connections between the components and the dynamical nature of the components.

Let the vector p denote a subset of the system parameters which is to be diagnosed. Its relation to the transfer function is given by

$$H(s) = h(s,p) \tag{1}$$

where the function h is determined by the system structure and the nominal values of all the parameters not included in the subset p, i.e., the parameters which are assumed to be invariant or fault free. Since from q input and m outputs of any linear time-invariant system we can only determine the $m \times q$ transfer function matrix $H(s)$, we have the following definition.

Definition 1

The parameters p of an LSDS are said to be *diagnosable* if p can be determined from $H(s)$, i.e., h is injective.

Note that given an LSDS the function h may vary according to which subset of the parameters of the LSDS are to be diagnosed, and, therefore, one subset of parameters may be diagnosable while some other subset of the same LSDS may not.

Since the function h maps p into the space of symbolic transfer function matrices determining conditions under which it is injective is a difficult problem. We propose instead an approach based on the time-domain representation of the transfer function matrix $H(s)$. This approach is based on the following assertion.

Assertion 1 [2]

Two minimal (controllable and observable) linear time-invariant state equation representations $\mathfrak{R} = [A, B, C, D]$ and $\hat{\mathfrak{R}} = [\hat{A}, \hat{B}, \hat{C}, \hat{D}]$ with state spaces of the same dimension n are realizations of the same transfer function matrix $H(s)$ if and only if

$$D = \hat{D}$$

and

$$CA^iB = \hat{C}\hat{A}^i\hat{B}, \qquad i = 0, 1 \cdots 2n-1.$$

The Markov parameters D and CA^iB, $i = 0, 1 \cdots 2n-1$ are well-defined functions of p. We can therefore write

$$\text{vec}\left[D; CB; \cdots; CA^{2n-1}B\right] = F(p). \tag{2}$$

Clearly we have Assertion 2.

Manuscript received December 4, 1978; revised April 16 1979. This paper was supported in part by the Office of Navel Research under Grant N00014-78-C-0444. This two-part paper is part of a dissertation submitted by V. Visvanathan to the Department of Electrical Engineering, University of Notre Dame, in partial fulfillment of the requirement for the Master of Science degree.

R-w. Liu is with the University of Notre Dame, Notre Dame, IN 46556.

V. Visvanathan is with the University of California, Berkeley, CA 94720.

0098-4094/79/0700-0490$00.75

Assertion 2

The parameters p of the LSDS are diagnosable if and only if F is injective.

Note that F unlike h is a real-valued function of real variables.

Assertion 2 requires that $F(\cdot)$ be injective for the parameters p to be diagnosable. This is a very strong requirement. A less strict but equally useful condition is that $F(\cdot)$ be injective in a generic sense.[1]

Definition 2

A function f: $\mathbf{R}^n \to \mathbf{R}^m$ is *injective in a generic sense* if

$$V \triangleq \{x \in \mathbf{R}^n \mid \exists x' \in \mathbf{R}^n,\ x \neq x' \text{ and } f(x) = f(x')\}$$

is a proper variety.

Definition 3

A set of parameters p is *diagnosable in a generic sense* if $F(\cdot)$ is injective in a generic sense.

Since the "probability" that the parameters will take on the exact values of the points on the proper variety equals zero, generic diagnosability is an equally good property as far as fault diagnosis is concerned.

It is not an easy task to determine a necessary and sufficient condition under which this property holds for an arbitrary nonlinear function. At this stage it is pertinent to mention that if the Jacobian of the function $F(\cdot)$ has full column rank in a generic sense it is not necessary that $F(\cdot)$ is injective in a generic sense. We will illustrate this point with an example.

Example 1: Consider the function

$$y(x) = x^2$$

whose domain is the set of real numbers. The Jacobian of the function is

$$J_y(x) = 2x.$$

$J_y(x) = 0$ if and only if $x = 0$. Hence, $J_y(x)$ has full column rank in a generic sense. However, the set V of Definition 2 is, in this case

$$V = \mathbf{R} - \{0\}$$

which is not a proper variety. Hence, $y(x)$ is not injective in a generic sense.

Let us consider a tradeoff problem between the cost of computing and the cost of test points. In general, (1) or (2) are nonlinear and the dimension of p is large. Therefore, the computation cost is high and the accuracy is low. This situation can be alleviated by an increase of the number of test terminals. For example, if every parameter can be measured directly, the function f or F becomes linear and decoupled. On the other hand, every additional test point is associated with an additional cost, and there is a limit of numbers of test points to be added to a circuit. This is a major tradeoff problem for fault diagnosis problems.

[1]This concept was introduced to the problem of fault diagnosis by Sen and Saeks [3]. See [10] or the Appendix for the definition of genericity.

In this paper, we propose a solution to the tradeoff problem. It is found that under certain conditions, the nonlinear equation (2) can be made to be sequentially linear (see (31)), i.e., a set of nonlinear equations which can be solved by solving a set of *linear* equations in a sequential manner. Therefore, an optimal solution to the tradeoff problem becomes the least number of test points which makes (2) sequentially linear. Any additional test points to the optimal solution will not reduce the computation cost appreciably. One less test point will cause (2) to be "genuinely" nonlinear, and, therefore, an increase of computation cost by an appreciable amount. The two-part paper gives a partial solution to the tradeoff problem.

II. The LSDS Model and Partitioning

The LSDS is assumed to be an interconnection of single input-single output (SISO) components. The ith component may be represented as

$$\mathfrak{p}_i g_i(s) \tag{3}$$

where $g_i(s)$ is a *known* transfer function and $\mathfrak{p}_i$ the parameter to be diagnosed, and it has a state equation representation (4), where $\boldsymbol{x}_i$ is the state vector and $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$ are, respectively, the scalar input and output of the component.

$$\begin{aligned} \dot{\mathfrak{x}}_i &= \mathfrak{A}_i \mathfrak{x}_i + \mathfrak{B}_i \mathfrak{p}_i \mathfrak{a}_i \\ \mathfrak{b}_i &= \mathfrak{C}_i \mathfrak{x}_i + \mathfrak{D}_i \mathfrak{p}_i \mathfrak{a}_i. \end{aligned} \tag{4}$$

The LSDS is described by the component-connection model (5) [5], [7], [18].

$$\begin{aligned} \dot{x} &= Ax + BPa \\ b &= Cx + DPa \end{aligned} \tag{5a}$$

$$\begin{bmatrix} a \\ y \end{bmatrix} = \begin{bmatrix} L_{ab} & L_{au} \\ L_{yb} & L_{yu} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} \tag{5b}$$

where $A = \text{diag}\ (\mathfrak{A}_i)$ and B, C, D, and P are similarly defined. The vector $x = \text{vec}\ (\mathfrak{x}_i)$ and a and b are similarly defined. Equation (5b) describes the connection between the components and the input–output terminals u and y of the system. Without loss of generality it may be assumed that both the numerator and denominator of $g_i(s)$ are monic polynomials [14].

It has been shown [4] that without loss of generality the LSDS description (5) with each component being a SISO system applies to LSDS where some of the components may be MIMO with a transfer function as follows:

$$\begin{bmatrix} p_{11} g_{11}(s) & p_{12} g_{12}(s) & \cdots & p_{1j} g_{1j}(s) \\ \vdots & & & \vdots \\ p_{i1} g_{i1}(s) & & \cdots & p_{ij} g_{ij}(s) \end{bmatrix}.$$

By redefining the connections, each one of the above ij scalar transfer functions becomes an SISO component.

We now introduce a partitioning of the components and rewrite (5) based on it. The partitioning leads naturally to the fault diagnosis equations developed in the next

section. Consider the SISO component described by equations (3) or (4). Let

$$m_i \triangleq (\text{degree of the denominator polynomial of } g_i(s)) - (\text{degree of the numerator polynomial of } g_i(s)). \quad (6)$$

Definition 4

The quantity m_i defined in (6) is called the minimum delay order of the ith component and the component is called an m_i-minimum delay component.

Partition the components of the LSDS into the classes $S_0, S_1, \cdots S_j, \cdots S_K$ in the following manner. The class S_j consists of all the j-minimum delay components of the LSDS. The largest minimum delay order of all the components in the LSDS is K. On the basis of this partitioning we can rewrite (5) as in equation (7). The partitioning of the vectors a, b, and x and the matrices A, B, C, D, P, and L is conformable with the partitioning into the classes $S_0 \cdots S_K$. For example, the matrix P_1 is a diagonal matrix whose diagonal entries are the parameters of the components that belong to the class S_1 and L_{32} consists of the gains of the connections directed to the components in class S_3 from those in class S_2.

Let $p \in \mathbf{R}^n$, $p_0 \in \mathbf{R}^{n_0}$, $p_1 \in \mathbf{R}^{n_1} \cdots p_K \in \mathbf{R}^{n_K}$ be, respectively, the diagonal entries of the matrices P, P_0, $P_1 \cdots P_K$. Note that $n = n_0 + n_1 + \cdots n_K$. Since $g_i(s)$ is monic if it is part of an m-minimum delay component, it can be easily shown that for values of s whose magnitude is sufficiently large

$$g_i(s) = \frac{1}{s^m} + \sum_{j=m}^{\infty} \frac{\mathcal{C}_i \mathcal{A}_i^j \mathcal{B}_i}{s^{j+1}}. \quad (8)$$

It then follows that for the LSDS description (7)

$$D_0 = I, D_1 = 0, C_1 B_1 = I$$

and

$$D_i = 0$$
$$C_i A_i^j B_i = 0$$
$$C_i A_i^{i-1} B_i = I, \qquad j = 0, \cdots, i-2, \; i = 2, \cdots, K. \quad (9)$$

Therefore,

$$DP = \operatorname{diag}(D_0 P_0, D_1 P_1 \cdots D_K P_K) = \operatorname{diag}(P_0, 0, \cdots, 0). \quad (10)$$

As a consequence, $\det [I - L_{ab} DP] = \det [I - L_{00} P_0]$ and by Theorem 1 of Singh and Liu [7] the LSDS (7) has a state equation representation in the composite state space of all its components if and only if $\det [I - L_{00} P_0] \neq 0$.

Assertion 3

Det $[I - L_{00} P_0] \neq 0$ in a generic sense.

Proof: Det $[I - L_{00} P_0]$ is a polynomial in the entries of the matrix which are themselves polynomials in the variables p_0. Hence, the equation

$$\det [I - L_{00} P_0] = 0 \quad (11)$$

can be written as

$$\begin{bmatrix} \dot{x}_0 \\ \dot{x}_1 \\ \vdots \\ \dot{x}_K \end{bmatrix} = \begin{bmatrix} A_0 & & & 0 \\ & A_1 & & \\ & & \ddots & \\ 0 & & & A_K \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} B_0 P_0 & & & 0 \\ & B_1 P_1 & & \\ & & \ddots & \\ 0 & & & B_K P_K \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_K \end{bmatrix} = \begin{bmatrix} C_0 & & & 0 \\ & C_1 & & \\ & & \ddots & \\ 0 & & & C_K \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} D_0 P_0 & & & 0 \\ & D_1 P_1 & & \\ & & \ddots & \\ 0 & & & D_K P_K \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_K \end{bmatrix} \quad (7a)$$

$$\left[\begin{array}{c} a_0 \\ a_1 \\ \vdots \\ a_K \\ \hline y \end{array}\right] = \left[\begin{array}{ccc|c} L_{00} & \cdots & L_{0K} & L_{0u} \\ L_{10} & \cdots & L_{1K} & L_{1u} \\ \vdots & & \vdots & \vdots \\ L_{K0} & \cdots & L_{KK} & L_{Ku} \\ \hline L_{y0} & \cdots & L_{yK} & L_{yu} \end{array}\right] \left[\begin{array}{c} b_0 \\ b_1 \\ \vdots \\ b_K \\ \hline u \end{array}\right]. \quad (7b)$$

$$q(p_{01}, p_{02}, \cdots, p_{0n_0}) = 0 \tag{12}$$

where $q(\cdot, \cdot, , \cdot)$ is a polynomial and hence, a variety V, in $\mathbf{R}^{n_0}$. Since

$$\det [I] \neq 0$$

the point $p = 0$ does not lie on V and hence, V is a proper variety. The proof is completed.

Due to Assertion 3 in what follows, we consider only those fault conditions for which $\det [I - L_{00}P_0] \neq 0$. Let,

$$\begin{aligned}
\tilde{A}_0 &\triangleq A_0 + B_0P_0(I - L_{00}P_0)^{-1}L_{00}C_0 \\
\tilde{B}_0 &\triangleq B_0P_0(I - L_{00}P_0)^{-1} \\
\tilde{C}_0 &\triangleq C_0 + P_0(I - L_{00}P_0)^{-1}L_{00}C_0 \\
\tilde{P}_0 &\triangleq P_0(I - L_{00}P_0)^{-1}.
\end{aligned} \tag{13}$$

Due to (10) and (13), (7) is equivalent to equation (14).

$$\begin{aligned}
\dot{x} &= \hat{A}x + \hat{B}a \\
b &= \hat{C}x + \hat{D}a
\end{aligned} \tag{14a}$$

$$\begin{bmatrix} a \\ y \end{bmatrix} = \begin{bmatrix} \hat{L}_{ab} & L_{au} \\ L_{yb} & L_{yu} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} \tag{14b}$$

where

$$\begin{aligned}
\hat{A} &= \operatorname{diag}(\tilde{A}_0, A_1, \cdots, A_K) \\
\hat{B} &= \operatorname{diag}(\tilde{B}_0, B_1P_1, \cdots, B_KP_K) \\
\hat{C} &= \operatorname{diag}(\tilde{C}_0, C_1, \cdots, C_K) \\
\hat{D} &= \operatorname{diag}(\tilde{P}_0, 0, \cdots, 0)
\end{aligned} \tag{15}$$

and

$$\hat{L}_{ab} = \begin{bmatrix} 0 & L_{01} & \cdots & L_{0K} \\ L_{10} & L_{11} & \cdots & L_{1K} \\ \vdots & & & \\ L_{K0} & L_{K1} & \cdots & L_{KK} \end{bmatrix}. \tag{16}$$

Note that

$$\hat{D}\hat{L}_{ab}\hat{D} = 0. \tag{17}$$

III. Conditions for Generic Diagnosability

Let $S(s)$ represent the matrix transfer function of the LSDS described by (14) and $Z(s)$ the matrix transfer function of the components described by (14a).

$$S(s) = L_{yb}Z(s)\left[I - \hat{L}_{ab}Z(s)\right]^{-1}L_{au} + L_{yu}. \tag{18}$$

With a power series expansion of $Z(s)$ and $[I - \hat{L}_{ab}Z(s)]^{-1}$ which are valid for values of s whose magnitude is sufficiently large, we have

$$S(s) = L_{yb}\left[\sum_{i=0}^{\infty}\left(\hat{D} + \sum_{j=0}^{\infty}\frac{\hat{C}\hat{A}^j\hat{B}}{s^{j+1}}\right)\left(\hat{L}_{ab}\left(\hat{D} + \sum_{j=0}^{\infty}\frac{\hat{C}\hat{A}^j\hat{B}}{s^{j+1}}\right)\right)^i\right]L_{au} + L_{yu}. \tag{19}$$

Since $S(s)$ has a state equation representation,

$$S(s) = \sum_{i=0}^{\infty} M_i/s^i \tag{20}$$

where M_i's are the Markov parameters of $S(s)$. We want to equate the coefficients of (19) and (20).

It follows from (15) and (9) that

$$\hat{C}\hat{A}^j\hat{B} = \operatorname{diag}(m_{0j}, m_{1j}, \cdots, m_{jj}, P_{j+1}, 0, \cdots, 0)$$

where

$$\begin{aligned}
m_{0j} &= \tilde{C}_0\tilde{A}_0^j\tilde{B}_0 \\
m_{ij} &= C_iA_i^jB_iP_i, \qquad 1 \leq i \leq j.
\end{aligned}$$

Note that m_{ij} is a function of only P_i. Equating the coefficients of (19) and (20), we arrive at Lemma 1.

Lemma 1

$M_i = \phi_i(P_0, \cdots, P_i)$ where $i = 0, 1, 2, \cdots$.

Imposing (17), the explicit expressions of ϕ_i can be obtained:

$$M_0 = L_{y0}\tilde{P}_0L_{0u} + L_{yu} \tag{21a}$$

$$M_i = \left[L_{yi} + L_{y0}\tilde{P}_0L_{0i}\right]P_i\left[L_{iu} + L_{i0}\tilde{P}_0L_{0u}\right] + f_i(P_0, \cdots, P_{i-1}), \qquad i = 1, \cdots, K. \tag{21b}$$

The above equations indicates the sequential nature of the fault diagnosis equations. Note from (21b) that for each i, ϕ_i is a *linear* function with respect to P_i.

Recall that $p_i \in \mathbf{R}^{n_i}$ are the diagonal entrices of P_i. Taking the vec of equation (21) and the appropriate dot product [8], we have

$$\operatorname{vec}\left[M_0 - L_{yu}\right] = \operatorname{vec}\left[L_{y0}\tilde{P}_0L_{0u}\right] \triangleq \psi_0(p_0) \tag{22a}$$

$$\operatorname{vec}\left[M_i - f_i(P_0, \cdots, P_{i-1})\right] = N_ip_i, \qquad i = 1, 2, \cdots, K. \tag{22b}$$

where

$$N_i = \left[L_{iu} + L_{i0}\tilde{P}_0L_{0u}\right]^T \odot \left[L_{yk} + L_{y0}\tilde{P}_0L_{0k}\right] \tag{23}$$

is a function of P_0 only.

Note that the existence of the solution to (22) is guaranteed. If ψ_0 is injective, and if N_i has full column rank, the solution is unique. Since the column rank of N_i depends on the value of P_0, the post-fault values of system parameters, it does not shed a light in the design of diagnosable systems. As such, we will treat the problem in a generic sense.

Lemma 2

The union of a finite number of proper varieties is a proper variety.

Proof: Consider the $k_1 + k_2 \cdots + k_m$ polynomials in n indeterminates with coefficients in $\mathbf{R}$:

$$\begin{aligned}
&q_{11}, q_{12}, \cdots, q_{1k_1} \\
&q_{21}, q_{22}, \cdots, q_{2k_2} \\
&q_{m1}, q_{m2}, \cdots, q_{mk_m}.
\end{aligned}$$

The Boolean variable b_{ij}, $i=1,\cdots,m$, $j=1,\cdots,k_i$, is defined as follows:

$$b_{ij}=0, \quad \text{if } q_{ij}\neq 0$$
$$b_{ij}=1, \quad \text{if } q_{ij}=0. \tag{24}$$

It follows from this definition that the proper varieties $V_i\subset\mathbb{R}^n$, $i=1,\cdots,m$ described as

$$q_{ij}=0, \quad j=1,\cdots,k_i$$

can be equivalently described by the Boolean equations

$$b_{i1}\cdot b_{i2}\cdot b_{ik_i}=1, \quad i=1,\cdots,m$$

while the Boolean equation

$$b_{ij}+b_{i'j'}=1 \tag{25}$$

is equivalent to the polynomial equation

$$q_{ij}q_{i'j'}=0. \tag{26}$$

The union of the proper varieties V_i, $i=1,\cdots,m$, can now be described as

$$b_{11}\cdot b_{12}\cdot b_{1k_1}+b_{21}\cdots b_{2k_2}+\quad +b_{m1}\cdot b_{m2}\cdots b_{mk_m}=1. \tag{27}$$

We can now express this sum of products as the product of sums,

$$\bar{b}_1\cdot\bar{b}_2\cdot\cdots\cdot\bar{b}_{\bar{m}}=1 \tag{28}$$

where $\bar{b}_i$ is a sum of some of the Boolean variables b_{ij} of (27). Since (25) is equivalent to (26) each $\bar{b}_i$ associates with a polynomial which is a product of some of the polynomials q_{ij}. Therefore, (28) represents a variety $\bar{V}$. Clearly $\bar{V}\neq\mathbb{R}^n$ and therefore it is a proper variety. The proof is completed. Theorem 1 follows directly from Lemma 2 and (22).

Theorem 1

The parameters p of the LSDS are diagnosable in a generic sense if $\psi_0(p_0)$ is injective in a generic sense and the matrices N_i, i.e.,

$$\left[L_{1u}+L_{10}\tilde{P}_0L_{0u}\right]^T\odot\left[L_{yi}+L_{y0}\tilde{P}_0L_{0i}\right], \qquad i=1,2\cdots K$$

have full column rank in a generic sense.

Definition 5

The parameters p of the LSDS are said to be *quasi-sequentially linearly diagnosable in a generic sense* if the conditions of Theorem 1 are satisfied.

Assertion 4

The matrix N_i, i.e.,

$$\left[L_{1u}+L_{10}P_0(I-L_{00}P_0)^{-1}L_{0u}\right]^T \odot\left[L_{yi}+L_{y0}P_0(I-L_{00}P_0)^{-1}L_{0i}\right]$$

has full column rank in a generic sense if and only if it is full column rank for some $p_0=p_0^*$.

Proof:

Necessity: The proof follows from the definition of a generic property.

Sufficiency: By Assertion 3 there exists a proper variety $V_1\subset\mathbb{R}^{n_0}$ such that $\forall p_0\notin V_1$ the matrix

$$\left[L_{1u}+L_{10}P_0(I-L_{00}P_0)^{-1}L_{0u}\right]^T \odot\left[L_{yi}+L_{y0}P_0(I-L_{00}P_0)^{-1}L_{0i}\right]$$

exists. When the matrix does exist it does not have full column rank if and only if all its $n_i\times n_i$ minors (since there exists a p_0^* where the matrix is full column rank, it has at least n_i rows) are zero, i.e., if the parameter values p_0 are common zeros of the polynomials in the numerator of each minor. Let V_2 be the variety so defined. Since there exists a p_0^* where the matrix is full column rank, $V_2\neq\mathbb{R}^{n_0}$, hence it is a proper variety. By Lemma 2,

$$V\triangleq V_1UV_2$$

is a proper variety. Since $\forall p_0\notin V$ the matrix is defined and is full column rank, the proof is completed.

Due to Assertion 4, to check if the parameters p of an LSDS are quasi-sequentially linearly diagnosable in a generic sense we need to check the rank of the matrices at only one set of parameter values (say the nominal values).

Note that if the parameters p of an LSDS is quasi-sequentially linearly diagnosable then all P_i's, $i=1,2,\cdots$, can be solved from the *linear* equations (22b), in a sequential manner. The only exception is p_0, which has to be solved from (22a), generally a nonlinear equation. In the next section we propose a canonical form of the LSDS for which (22a) becomes a linear equation in p_0 and as a consequence the fault diagnosis equations (22) are indeed sequentially linear.

We have so far assumed that all the parameters of the LSDS need to be diagnosed. The extension of Theorem 1 to the case where only a subset of the parameters p are to be diagnosed is straightforward. Let $p'=\text{Col}\,(p'_0,\cdots,p'_K)$ be a vector which consists of the parameters to be diagnosed. Equation (22a) may be rewritten as

$$v=\psi'_0(p'_0) \tag{29}$$

where v is a vector determined by M_0, L_{yu} and the nominal values of the parameters which are in p_0 but not in p'_0. We have Corollary 1 immediately.

Corollary 1: The parameters p' of the LSDS are diagnosable in a generic sense if the function $\psi'_0(\cdot)$ is injective in a generic sense and for $i=1,\cdots,K$ the submatrix of the matrix N_i, i.e.,

$$\left[L_{1u}+L_{10}P_0(I-L_{00}P_0)^{-1}L_{0u}\right]^T \odot\left[L_{yi}+L_{y0}P_0(I-L_{00}P_0)^{-1}L_{0i}\right]$$

consisting only of the columns that correspond to parameters in p'_i is full column rank in a generic sense.

Definition 6

The parameters p' of the LSDS are said to be *quasi-sequentially linearly diagnosable in a generic sense* if the conditions of Corollary 1 are satisfied.

IV. The Canonical LSDS

Definition 7

If in an LSDS (7) the matrix L_{00} is diagonal it is called a *Canonical LSDS*.

Let l_{0i} and p_{0i} represent the (ii)th entry of the matrices L_{00} and P_0, respectively, of a Canonical LSDS. If we now redefine p_{0i} as

$$\hat{p}_{0i} \triangleq \frac{p_{0i}}{1 - l_{0i}p_{0i}} \tag{30}$$

the fault diagnosis equations (22) for the Canonical LSDS reduce to

$$\operatorname{vec}\left[M_0 - L_{y_u}\right] = \left[L_{0u}^T \odot L_{y0}\right] p_0 \tag{31a}$$

$$\operatorname{vec}\left[M_i - f_i(P_0, P_1, \cdots, P_{i-1})\right] = \left(\left[L_{iu} + L_{i0}P_0L_{0u}\right]^T \odot \left[L_{yi} + L_{y0}P_0L_{0i}\right]\right)p_i, \qquad i = 1,2,\cdots,K. \tag{31b}$$

Note that (31a) and (31b) are sequentially linear.

The major problem of diagnosis of an LSDS is with the 0-minimum delay components since it requires that a nonlinear function be injective in a generic sense. On the other hand, for a Canonical LSDS the corresponding condition reduces to that of the matrix $L_{0u}^T \odot L_{y0}$ being full column rank. Therefore, it would be advantageous if an LSDS can be reduced to a Canonical LSDS. As such, an LSDS can be made sequentially linearly diagnosable. A strategy of test-point location whereby this is achieved is given in the second half of this two-part paper.

V. Appendix
Genericity

Let

$$p = (p_1, p_2, \cdots, p_n) \in \mathbf{R}^n$$

and consider polynomials $\psi(\lambda_1, \cdots, \lambda_n)$ with coefficients in $\mathbf{R}$. A *variety* $V \subset \mathbf{R}^n$ is defined to be the set of common zeros of a finite number of polynomials $\psi_1, \cdots, \psi_k$:

$$V = \{p: \psi_i(p_1, \cdots, p_n) = 0, i = 1,2,\cdots,k\}.$$

V is *proper* if $V \neq \mathbf{R}^n$ and *nontrivial* if $V \neq \emptyset$. In this setup, a *property* π is merely a function π: $\mathbf{R}^n \to \{0,1\}$, where $\pi(p) = 1$ (or 0) if π holds (or fails) at p. Let V be a proper variety. We say that π is generic relative to V provided $\pi(p) = 0$ only if $p \in V$; and that π is generic provided such a V exists.

The usefulness of this concept is based on the following facts. Let $p_0 \in V$, where V is nontrivial and proper. If π is generic relative to V and if π fails at p_0, π can be made to hold if p_0 is shifted by a suitable perturbation, arbitrarily small. It can also be shown that such a V has zero "probability" (Lebesgue measure). Therefore, π holds almost everywhere in $\mathbf{R}^n$.

References

[1] P. K. Srimani, B. P. Sinha, and A. K. Choudhury, "Minimal test point location in a system graph," *Proc. 20th Midwest Symp. Circuits Syst.*, (Lubbock) pp. 567–570, 1977.

[2] B. L. Ho and R. E. Kalman, "Effective construction of linear state variable models from input-output data", *Proc. 3rd Ann. Allerton Conf. Circuit and System Theory*, pp 449–459, 1965.

[3] N. Sen and R. Saeks, "A measure of testability and its application to test point selection—Theory," *Proc. 20th Midwest Symp. Circuits and Systems* (Lubbock), pp. 576–583, 1977.

[4] M. N. Ransom and R. Saeks, "Fault isolation via term expansion," *Proc. 3rd Pittsburgh Symp. Modeling and Simulation*, (Univ. of Pittsburgh), vol. 4, pp. 224–228, 1973.

[5] R. Saeks, S. P. Singh, and R. W. Liu, "Fault isolation via components simulation," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 634–640, Nov. 1972.

[6] T. N. Trick and C. J. Aljajian, "Fault diagnosis of analog circuits" *Proc. 20th Midwest Symp. Circuits Syst.* (Lubbock), pp. 576–583, 1977.

[7] S. P. Singh and R. W. Liu, "Existence of state equation representation of linear large-scale dynamical systems," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 239–246, May 1973.

[8] C. G. Khatri and C. R. Rao, "Solutions to some Functional equations and their applications to characterization of probability distributions," *Sankhya*, Ser. A, vol. 30, pp. 167–180, 1968.

[9] S. Seshu, "The future of diagnosis," *Proc. Seminar on Automatic Checkout Techniques*, Battelle Memorial Institute, Columbus, OH, 1962.

[10] W. M. Wonham, *Linear Multivariable Control*. Berlin: Springer-Verlag, 1974.

[11] H. Neudecker, "Some theorems on matrix differentiation with Special reference to kronecker matrix products," *Amer. Stat. Assoc. J.*, pp. 953–963, Sept. 1969.

[12] F. Harary, *Graph Theory*. Reading, MA: Addison-Wesley, 1969.

[13] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*. Englewood Cliffs, N.J: Prentice-Hall, 1974.

[14] V. Visvanathan, "Sequentially-linear fault diagnosis," M.S.E.E. thesis Univ. of Notre Dame, Notre Dame, IN., Aug. 1978.

[15] R. Liu and V. Visvanathan, "Diagnosability of large-scale dynamical systems," *Proc. 20th Midwest Symposium on Circuits and Systems*, pp. 585–589, 1977.

[16] ——, "Sequentially Linear Fault Diagnosis: Part II—The design of diagnosable systems," *IEEE Trans. Circuits Syst.*, this issue, pp. 558–564.

[17] S. R. Liberty, L. Tung, and R. Saeks, "Fault prediction—Towards a mathematical theory," in *Rational Fault Analysis*, Ed. R. Saeks and S. R. Liberty, pp. 135–142, 1977.

[18] R. Saeks and R. A. DeCarlo, *Interconnected Dynamical Systems*. New York: Marcel Dekker, to be published.

Ruey-wen Liu (S'53–M'59) was born in Kiangsu, China, in 1930. He received the B.S., M.S., and Ph.D. degrees in 1954, 1955, and 1960, respectively, from the Department of Electrical Engineering at the University of Illinois at Urbana.

He was a Research Assistant Professor at the Department of Theoretical and Applied Mechanics at the University of Illinois at Urbana from February to June 1960, and received a National Science Foundation Fellowship at the Stanford University in the summer of 1962. He was a Visiting Associate Professor from 1965 to 1966 (and a Visiting Professor from 1977 to 1978) at the University of California, Berkeley. He also held visiting professorships at the National Taiwan

University, Republic of China in the spring of 1969, the Universidad de Chile, Santiago de Chile in the summer of 1970, and the Institute of Mathematics, Academia Sinica, Republic of China in the summers of 1976 and 1978. His main interest has been in the area of nonlinear circuits and systems. Currently, he is the Chairman of the Technical Committee on Nonlinear Circuits and Systems. His other interests include policy analysis of socio-economic systems, fault diagnosis of electronic circuits, and applications of system identification to medical diagnosis problems. Since September 1960, he has been with the University of Notre Dame, Notre Dame, IN, and is currently a Professor at the Department of Electrical Engineering.

Dr. Liu is a member of Sigma Xi and the New York Academy of Sciences.

V. Visvanathan was born in New Delhi, India, on March 2, 1955. He received the B. Tech. degree in electrical engineering from the Indian Institute of Technology, Delhi, India, in 1976, and the M.S.E.E. degree from the University of Notre Dame, Notre Dame, IN, in 1978.

He is currently with the Department of Electrical Engineering and Computer Sciences and the Electronics Research Laboratory at the University of California, Berkeley, where he is a student in the Ph.D. program and is a Research Assistant. His research interests are in the areas of fault analysis and computer-aided design of electronic circuits.

Sequentially Linear Fault Diagnosis: Part II - The Design of Diagnosable Systems

by

R. Liu
V. Visvanathan

Sequentially Linear Fault Diagnosis: Part II—The Design of Diagnosable Systems

V. VISVANATHAN AND RUEY-WEN LIU, MEMBER, IEEE

Abstract—Based on the results developed in Part I a strategy of test point location by which the parameters of an LSDS are made diagnosable in a generic sense, is developed. By appropriate test point placement, an LSDS is reduced to a canonical LSDS. Next, an algorithm for the synthesis of the test points required to make the parameters of the canonical LSDS sequentially linearly diagnosable in a generic sense, is given.

I. Introduction

SYSTEM designers have so far concentrated on the performance specifications of the systems they build and have given little thought to the diagnosability of the end product. At a seminar nearly two decades ago, the late Prof. Seshu in a paper about the future of diagnosis [9] stated that there would be increased concern for diagnosability in circuit design to the extent that "the circuit designer is going to be required to supply the diagnostic tests for the circuit he designs." Recently, studies [3]–[6] have been made on the tests of diagnosability from a given set of test points. In this paper, a design procedure for the test points is given so that a given circuit is linearly sequentially diagnosable.

We first present a strategy of test point location which reduces an LSDS to a canonical LSDS. We then discuss an algorithm to determine the test point locations required to make the parameters of the canonical LSDS sequentially linearly diagnosable in a generic sense. We approach the problem from a graph-theoretic point of view. We view the present work as a preliminary result in a field that is largely unexplored.

II. Preliminaries

We begin this section with the definition of a test point.

Definition 1

An input or output terminal introduced in a system for the sole purpose of diagnosis is called a *test point*.

An important property that must be satisfied by all test points is that during normal operation we should be able to put them in a state in which they do not affect the system. The test points then come into the picture only when fault diagnosis needs to be done.

Manuscript received December 4, 1978; revised April 16, 1979. This paper was supported in part by the Office of Navel Research under Grant N0014-78-C-0444. This two-part paper is part of a dissertation by V. Visvanathan to the Department of Electrical Engineering, University of Notre Dame, in partial fulfillment of the requirement for the Master of Science degree.

V. Visvanathan is with the University of California, Berkeley, CA 94720.

R-w. Liu is with the University of Notre Dame, Notre Dame, IN 46556.

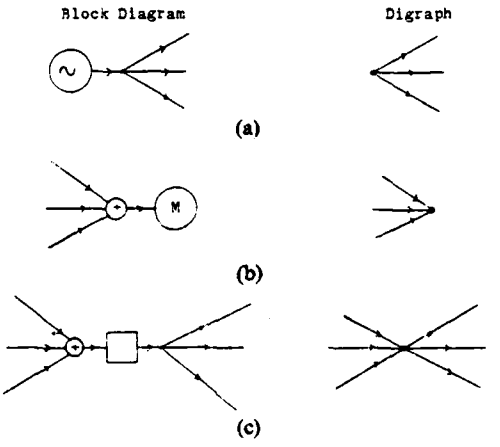


Fig. 1. Equivalent representations.

The component-connection model of the LSDS has a digraph representation, with the components, input signal generators and output detectors and meters being nodes and the connections being directed edges. The equivalence between the block-diagram representation of the LSDS [16, eq. (5)] and the digraph representation is illustrated in Fig. 1. In the rest of this paper the terms system node, input node, output node, and edge will be used interchangeably with component, input terminal, output terminal, and connection, respectively.

III. A Decomposition Strategy

A major problem of diagnosis is with the 0-minimum delay components since it requires that a nonlinear equation be injective in a generic sense. As we mentioned in [16] it is not an easy task to determine the necessary and sufficient conditions under which this property exists. A simple sufficient condition can be derived by considering [16, eq. (21a)]. It follows from this equation that

$$\text{vec}\,(M_0 - L_{yu}) = \left[L_{0u}^T \otimes L_{y0}\right] \text{vec}\left[\tilde{P}_0\right] \qquad (1)$$

where $\otimes$ denotes the Kronecker cross product [11]. From the definition of $\tilde{P}_0$ [16, eq. (13)] it follows that

$$P_0 = (I + \tilde{P}_0 L_{00})^{-1} \tilde{P}_0. \qquad (2)$$

If the matrix $[L_{0u}^T \otimes L_{y0}]$ has full column rank then equations (1) and (2) can be solved for a unique P_0. However, for the matrix $[L_{0u}^T \otimes L_{y0}]$ to be full column rank it is

0098-4094/79/0700-0558$00.75

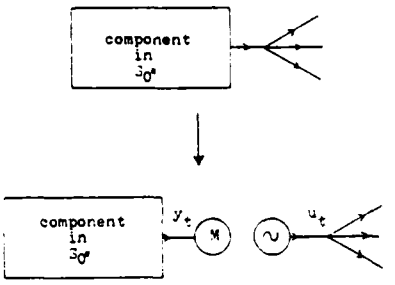


Fig. 2. Decomposition.

necessary that the number of both the input terminals and output terminals be greater than the number of 0-minimum delay components. We propose instead a strategy of decomposition which in most cases will require a fewer number of test points.

Consider the subgraph of the LSDS which consists of the 0-minimum delay nodes and all the edges between them except the self-loops and find its minimum node cover [12]. Partition the set S_0 [16, section II] into $S_{0'}$ and $S_{0''}$ where $S_{0'}$ consists of the nodes that are not part of the node cover and $S_{0''}$ consists of those that are. Break the output edges of the nodes in $S_{0''}$ and introduce an input and output node in each edge as shown in Fig. 2. The signals measured at these output nodes will be described by the vector y_t and those introduced at the input nodes by the vector u_t. Under normal operation the corresponding nodes are connected, i.e., $u_t = y_t$. When fault diagnosis needs to be done the connection is broken.

The vector u which represents the inputs of the LSDS other than those introduced by tearing is partitioned as

$$u = \begin{bmatrix} u' \\ \hline u'' \end{bmatrix} \tag{3}$$

where u'' consists of the inputs that reach only the components in $S_{0''}$.

Based on this partitioning of u, the partitioning of S_0 and the introduction of u_t and y_t, the LSDS description [16, eqs. (7) and (10)] changes to that in (4) where the rationale behind the partitioning of some of the vectors and matrices is made self-evident by the corresponding subscripts.

$$\begin{bmatrix} \dot{x}_{0'} \\ \dot{x}_{0''} \\ \dot{x}_1 \\ \vdots \\ \dot{x}_K \end{bmatrix} = \begin{bmatrix} A_{0'} & & & & 0 \\ & A_{0''} & & & \\ & & A_1 & & \\ & & & \ddots & \\ 0 & & & & A_K \end{bmatrix} \begin{bmatrix} x_{0'} \\ x_{0''} \\ x_1 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} B_{0'}P_{0'} & & & & \\ & B_{0''}P_{0''} & & & 0 \\ & & B_1P_1 & & \\ & & & \ddots & \\ 0 & & & & B_KP_K \end{bmatrix} \begin{bmatrix} a_{0'} \\ a_{0''} \\ a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$\begin{bmatrix} b_{0'} \\ b_{0''} \\ b_1 \\ \vdots \\ b_K \end{bmatrix} = \begin{bmatrix} C_{0'} & & & & \\ & C_{0''} & & & 0 \\ & & C_1 & & \\ & & & \ddots & \\ 0 & & & & C_K \end{bmatrix} \begin{bmatrix} x_{0'} \\ x_{0''} \\ x_1 \\ \vdots \\ x_K \end{bmatrix} + \begin{bmatrix} P_{0'} & & 0 & \cdots & 0 \\ 0 & P_{0''} & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & & \cdots & & 0 \end{bmatrix} \begin{bmatrix} a_{0'} \\ a_{0''} \\ a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$\begin{bmatrix} a_{0'} \\ a_{0''} \\ a_1 \\ \vdots \\ a_K \\ \hline y \\ y_t \end{bmatrix} = \left[\begin{array}{ccccc|ccc} L_{0'0'} & 0 & L_{0'1} & \cdots & L_{0'K} & L_{0'u'} & 0 & L_{0'0} \\ L_{0''0'} & 0 & L_{0''1} & \cdots & L_{0''K} & L_{0''u'} & L_{0''u''} & L_{0''0''} \\ L_{10'} & 0 & L_{11} & \cdots & L_{1K} & L_{1u'} & 0 & L_{10''} \\ & & & & & & & \\ L_{K0'} & 0 & L_{K1} & \cdots & L_{KK} & L_{Ku'} & 0 & L_{K0''} \\ \hline L_{y0'} & 0 & L_{y1} & \cdots & L_{yK} & L_{yu'} & 0 & L_{y0''} \\ 0^{0'} & I & 0^{1} & \cdots & 0^{K} & 0^{u'} & 0 & 0^{0''} \end{array}\right] \begin{bmatrix} b_{0'} \\ b_{0''} \\ b_1 \\ \vdots \\ b_K \\ \hline u' \\ u'' \\ u_t \end{bmatrix} \tag{4}$$

Note that the L matrices associated with $b_{0''}$ are now associated with u_t and that $y_t = b_{0''}$. Also, $L_{0'0'}$ is a diagonal matrix. We now decompose the LSDS (4) into two subsystems which are defined as follows.

Subsystem 1 (SS1) is defined by the component-connection model:

$$\begin{bmatrix} \dot{x}_{0'} \\ \dot{x}_1 \\ \vdots \\ \dot{x}_K \end{bmatrix} = \begin{bmatrix} A_{0'} & & & 0 \\ & A_1 & & \\ & & \ddots & \\ 0 & & & A_K \end{bmatrix} \begin{bmatrix} x_{0'} \\ x_1 \\ \vdots \\ x_K \end{bmatrix}$$

$$+ \begin{bmatrix} B_{0'}P_{0'} & & & 0 \\ & B_1P_1 & & \\ & & \ddots & \\ 0 & & & B_KP_K \end{bmatrix} \begin{bmatrix} a_{0'} \\ a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$\begin{bmatrix} b_{0'} \\ b_1 \\ \vdots \\ b_K \end{bmatrix} = \begin{bmatrix} C_{0'} & & & 0 \\ & C_1 & & \\ & & \ddots & \\ 0 & & & C_K \end{bmatrix} \begin{bmatrix} x_{0'} \\ x_1 \\ \vdots \\ x_K \end{bmatrix}$$

$$+ \begin{bmatrix} P_{0'} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{0'} \\ a_1 \\ \vdots \\ a_K \end{bmatrix}$$

$$\begin{bmatrix} a_{0'} \\ a_1 \\ \vdots \\ a_K \\ \hline d_{0'} \\ d_1 \\ \vdots \\ d_K \\ d' \end{bmatrix} = \left[\begin{array}{cccc:cc} L_{0'0'} & L_{0'1} & \cdots & L_{0'K} & L_{0'u'} & L_{0'0''} \\ L_{10'} & L_{11} & & L_{1K} & L_{1u'} & L_{10''} \\ & & & & & \\ L_{K0'} & & & L_{KK} & L_{Ku'} & L_{K0''} \\ \hline I & 0 & \cdots & 0 & 0 & 0 \\ 0 & I & & & 0 & 0 \\ & & & & & \\ 0 & 0 & & I & 0 & 0 \\ L_{y0'} & L_{y1} & \cdots & L_{yK} & L_{yu'} & L_{y0''} \end{array}\right] \cdot \begin{bmatrix} b_{0'} \\ b_1 \\ \\ b_K \\ \hline c_1 \\ c_2 \\ \\ \\ \\ \end{bmatrix}.$$

Subsystem 2 (SS2):

$$\dot{x}_{0''} = A_{0''}x_{0''} + B_{0''}P_{0''}a_{0''}$$
$$b_{0''} = C_{0''}x_{0''} + P_{0''}a_{0''}. \tag{6}$$

The connections between SS1, SS2, and the inputs and outputs of the LSDS is given by

$$\begin{bmatrix} c_1 \\ c_2 \\ \hline a_{0''} \\ \hline y \\ y_t \end{bmatrix} = \left[\begin{array}{cccccc:ccc} 0 & 0 & \cdots & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & I \\ \hline L_{0''0'} & L_{0''1} & \cdots & L_{0''K} & 0 & 0 & L_{0''u'} & L_{0''u''} & L_{0''0''} \\ \hline 0 & 0 & \cdots & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & I & 0 & 0 & 0 \end{array}\right] \cdot \begin{bmatrix} d_{0'} \\ d_1 \\ \cdot \\ \cdot \\ d_K \\ d' \\ \hline b_{0''} \\ \hline u' \\ u'' \\ u_t \end{bmatrix}. \tag{7}$$

It is easy to verify that (5)–(7) are equivalent to the LSDS description in (4). As a result of the decomposition components in the set $S_{0''}$ become a subsystem separate from the rest. Note that SS1 is a canonical LSDS. The vectors $d_0, d_1, \cdots, d_K$ are considered outputs of SS1 for the consistency of the model (5)–(7) with (4), and it is only outputs d' that are connected to output terminals of the LSDS. The parameters of SS1 are therefore sequentially linearly diagnosable in a generic sense, by [14, theorem 1], if

$$[L_{0'u'}; L_{0'0''}]^T \odot L_{y0'}$$

and

$$[(L_{iu'}; L_{i0''}) + L_{i0'}P_{0'}(L_{0'u'}; L_{0'0''})]^T \odot [L_{yi} + L_{y0'}P_{0'}L_{0'i}], \qquad i = 1, \cdots, K$$

are full column rank in a generic sense. In other words, under this condition the parameters of SS1 can be sequentially linearly diagnosed in the generic sense from the LSDS input–output measurements, without the information of SS2. It remains to show that the parameters of SS2 can be linearly diagnosed from the LSDS input–output measurements if the parameters of SS1 are known. This is done in two-steps, both with an application of a result in [15].

Next, we want to show that SS2 is diagnosable from the component input–output measurements, i.e., the parameter $P_{0''}$ can be determined from $(a_{0''}, b_{0''})$. Note that (6), with the subscript 0″ omitted for simplicity, is equivalent

to

$$\dot{x} = Ax + Ba$$
$$b = PCx + Pa. \tag{8}$$

This is because of the commutative property of single-input single-output components [16, eq. (3)] between p_i and $g_i(s)$. Equation (8) can be decomposed into a component-connection model:

$$\dot{x} = Ax + Bc$$
$$d = Cx + c \tag{9}$$
$$s = Pr \tag{10}$$
$$\begin{bmatrix} c \\ r \\ b \end{bmatrix} = \begin{bmatrix} 0 & 0 & I \\ I & 0 & 0 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} d \\ s \\ a \end{bmatrix}. \tag{11}$$

The parameters of component (9) are known, while the parameters of component (10) are to be determined. Equation (11) is the connection equation. Formulating the matrix (12) of [15],

$$Q = \begin{bmatrix} -I & 0 \\ 0 & I \end{bmatrix}$$

which has full column rank. According to Corollary 1 of [15], (s,r) can be determined from (a,b), without the information of P. Since P is a diagonal matrix, it can be determined from (s,r) by (10). Therefore, P can be determined from (a,b), the input–output measurement of SS2. In fact, P can be determined by solving a set of linear equations because (10) is linear with respect to p_i, and (s,r) can be determined from (a,b), independently from P.

Finally, from the component-connection model (5)–(7), we want to show that the input–output measurements (a_{0^-}, b_{0^-}) of SS2 can be determined from (u,y) with the assumption that the parameters of SS1 are known but the parameters of SS2 are not. Again, formulating the matrix (12) of [15],

$$Q = \begin{bmatrix} -I & 0 & 0 \\ 0 & -I & 0 \\ Q_1 & Q_2 & 0 \\ 0 & 0 & I \end{bmatrix} \tag{12}$$

where

$$Q_1 = L_{yu'} + L_{y0'} P_{0'} L_{0'u'}$$
$$Q_2 = L_{y0'} + L_{y0'} P_{0'} L_{0'0'} \tag{13}$$

which has full column rank. From Corollary 1 of [15], (a_{0^-}, b_{0^-}) can be determined from (u,y) without knowing the parameters of SS2.

Definition 2

For a canonical LSDS [16, definition 7], the matrices

$$L_{0u}^T \odot L_{y0} \tag{14}$$

and

$$[L_{iu} + L_{i0} P_0 L_{0u}]^T \odot [L_{yi} + L_{y0} P_0 L_{0i}], \qquad i = 1, \cdots, K \tag{15}$$

are called *diagnosis matrices*.

The discussion above can be summarized in the next theorem.

Theorem 1

If the diagnosis matrices of SS1 are full column rank in a generic sense, then the LSDS (4) is sequentially linearly diagnosable in a generic sense.

Note that the diagnosis matrices of SS1 concerns only with $P_{0'}$ which are those memoryless components which are not part of the node cover.

Before we discuss the design of the diagnosis matrices in Section V, in the next section we will discuss their graphical structure since it gives an insight into their design.

IV. The Structure of the Diagnosis Matrices

Consider the $n_i \times q$ matrix $[L_{iu} + L_{i0} P_0 L_{0u}]$ and the $m \times n_i$ matrix $[L_{yi} + L_{y0} P_0 L_{0i}]$. The rows of the matrix $[L_{iu} + L_{i0} P_0 L_{0u}]^T ([L_{yi} + L_{y0} P_0 L_{0i}])$ correspond to input (output) nodes and the columns correspond to i-minimum delay system nodes. The matrix $[L_{iu} + L_{i0} P_0 L_{0u}]^T (L_{yi} + L_{y0} P_0 L_{0i})$ has a nonzero entry if the input (output) node and the system node that correspond to that position in the matrix have an edge or a directed path through a 0-minimum delay node between them. The value of the entry is the total gain along the above mentioned directed paths between the two nodes computed by multiplying the gains of edges in series and adding the gains of edges in parallel. Thus due to these two matrices we have certain directed paths from input to output nodes through the i-minimum delay system nodes.

The matrix $[L_{iu} + L_{i0} P_0 L_{0u}]^T \odot [L_{yi} + L_{y0} P_0 L_{0i}]$ is a $qm \times n_i$ matrix. Each of its columns corresponds to an i-minimum delay system node while each of its rows corresponds to an ordered pair of input and output nodes. There is a nonzero entry in any position of this matrix if there exists a directed path between the input and output nodes of the ordered pair corresponding to the row through the i-minimum delay node corresponding to the column and possibly some 0-minimum delay nodes. Such a directed path is called an *i/o path*. An *i/o* path is identified by the ordered pair of its end nodes. Note that not all directed paths from an input node to an output node through an i-minimum delay node is in *i/o* path through that i-minimum delay node. Only those directed paths that contribute a nonzero entry to the diagnosis matrix are considered as *i/o* paths. Fig. 3 considers various directed paths between an input and an output node as candidates for *i/o* paths through the i-minimum delay node A. All of them except for Figs. 3(e) and 3(f) are valid *i/o* paths through the node A. In the case of the 0-minimum delay components the diagnosis matrix reduces to $L_{0u}^T \cdot L_{y0}$ and an *i/o* path through a 0-minimum delay node goes through no other system node. Note that in Fig. 3(d) we have the same *i/o* path through the 0-minimum delay node B_4 and the 1-minimum delay node A.

If we construct a bipartite graph [13] with X and Y, the two disjoint subsets of nodes, corresponding to the rows

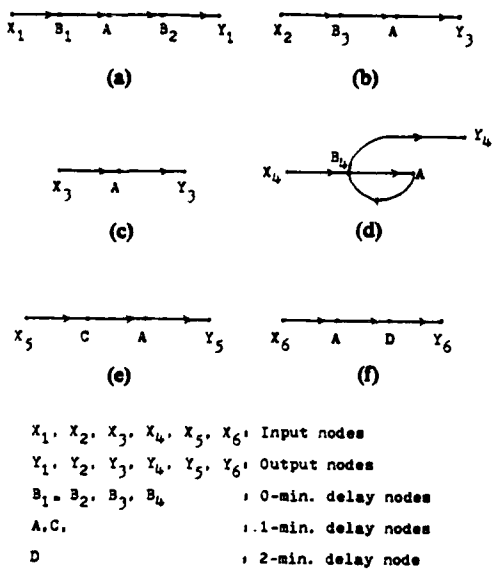


Fig. 3. Candidates for i/o paths.

and columns of the i-minimum delay diagnosis matrix, and draw an edge between two nodes if the corresponding entry in the matrix is nonzero, then a complete matching of Y into X [13], is a necessary and sufficient condition for the diagnosis matrix to be full column rank in a generic sense in the space of nonzero entries of the matrices L_{iu}, L_{i0}, L_{0u}, L_{yi}, L_{y0}, L_{0i}, and P_0. However, since we require that the diagnosis matrices be full column rank in a generic sense in the space $\mathbf{R}^{n0}$ complete matching is only a sufficient condition for sequentially linear diagnosis in a generic sense.

V. Design of the Diagnosis Matrices

In this section we develop an algorithm for the design of diagnosis matrices that have full column rank in a generic sense. Therefore, we assume that all the parameters of the canonical LSDS need to be diagnosed. The modification of the algorithm to the case where only some of the parameters of the canonical LSDS need to be diagnosed is straightforward.

If the canonical LSDS has no existing input and output terminals then the strategy of adding a minimum number of test points to make the parameters sequentially linearly diagnosable in a generic sense is trivial. We first choose the number of input terminals q and the number of output terminals m which minimize the quantity $(q+m)$ subject to the constraint

$$qm > \max_{i=0,1,\cdots,K} n_i. \tag{16}$$

The matrices L_{iu} and L_{yi}, $i=0,1,\cdots,K$ are then chosen so that

$$L_{iu}^T \odot L_{yi} = I \quad \text{or} \quad \left[\begin{array}{c} I \\ \hline 0 \end{array}\right]. \tag{17}$$

Then by [16, association 4 and theorem 1] the parameters of the canonical LSDS are diagnosable in a generic sense.

The procedure for constructing the matrices L_{iu} and L_{yi} so that $L_{iu}^T \odot L_{yi}$ is as given in (17) is best understood by using the bipartite graph discussed in the previous section. We start with the two disjointed subsets of nodes corresponding to the qm possible i/o paths and the n_i i-minimum delay system nodes. Each i-minimum delay system node is matched to a different i/o path. In the digraph of the canonical LSDS we then create the i/o path corresponding to the system node by introducing a directed edge of gain 1 to the system node from the corresponding input node and from the system node to the corresponding output node. This procedure is carried out for $i=0,1,\cdots,K$.

However, since all systems have input and output terminals which already exist, and have not been introduced for the purpose of diagnosis, we should make the best possible use of the i/o paths created by these inputs (inputs created in the canonical LSDS due to the decomposition of the LSDS can also be considered as existing inputs) and outputs. Stated from a matrix point of view, some of the rows of L_{iu}^T and L_{yi}, $i=0,1,\cdots,K$, are already fixed and what we need to do is add a minimum number of extra input and/or output nodes (under normal operation the input signals will equal zero and the output signals need not be measured) and appropriate rows in the above mentioned matrices to make the parameters of the canonical LSDS sequentially linearly diagnosable in a generic sense. Though this is a conceptually simple problem it is not an easy task to create an efficient algorithm which will solve this problem for any given canonical LSDS.

The algorithm that we present makes the parameters of the canonical LSDS sequentially linearly diagnosable in a generic sense but it has not been proved that it will use the minimum possible number of test points required. Starting with the first matrix $(L_{0u}^T \odot L_{y0})$ the algorithm sequentially makes each one of them of full column rank in a generic sense. In the algorithm we start with the minimum number of test points required, i.e., the minimum number required to satisfy the inequality (16). Then, when we are working with the j-minimum delay diagnosis matrix we first create the possible i/o paths through j-minimum delay system nodes, which go between the existing i/o nodes and the test points and the test points themselves and help increase the rank of the diagnosis matrix. This operation is done in an efficient manner by considering the system nodes in the increasing order of the number of input or output nodes to which they are connected. Note that we are allowed to create edges only between test points and system nodes and not between existing i/o nodes and system nodes since such an edge would change the normal operation of the system. At this stage extra test points are added only if they are needed to make the diagnosis matrix of full column rank. As a result of the algorithm each diagnosis matrix has the structure shown in Fig. 4. Clearly the matrix in Fig. 4 is of full column rank. Notation for the algorithm is as follows.

$$\begin{bmatrix} A_1 & A_3 & A_4 \\ 0 & A_2 & 0 \\ 0 & 0 & I \\ A_5 & A_6 & A_7 \end{bmatrix}$$

← i/o paths between existing i/o nodes
← i/o paths between test points and existing i/o nodes
← i/o paths between test points

A_1 : full column rank

A_2 : column echelon form (full column rank)

Fig. 4. Structure of a diagnosis matrix.

The input and output nodes that are part of the canonical LSDS before any test points are added are called existing nodes while the test points (which are added on by the algorithm) are called test nodes:

- a number of existing input nodes,
- b number of existing output nodes,
- u signal at the a existing input nodes,
- y signal at the b existing output nodes,
- n_i number of i-min delay system nodes,
- $\bar{P}_0$ nominal value of 0-minimum delay parameters,
- K maximum delay order of all the system nodes in the canonical LSDS.

Algorithm

1) Input g, the subgraph of the canonical LSDS which consists of all the nodes, and the edges between the i-minimum delay nodes, $i=0,1\cdots K$, and the input, output, and 0-minimum delay nodes.

2) Add to g, r input test nodes and s output test nodes where r and s are chosen such that

$$r+s \quad \text{is a minimum}$$

subject to

$$(a+r)(b+s) \geq \max_{i=0,1,\cdots,K} n_i.$$

3) Let $i \leftarrow 0$.

4) If $i=0$, $L = L_{0u}^T \odot L_{y0}$; and else

$$L = \left[L_{iu} + L_{i0}\bar{P}_0 L_{0u} \right]^T \odot \left[L_{yi} + L_{y0}\bar{P}_0 L_{0i} \right].$$

5) From the columns of L select a basis for its column space.

6) To each one of the i-minimum delay nodes that do not correspond to the basis selected in step 5 (these are called the remaining nodes) assign two numbers called the U-number and the Y-number, where the U-number is the number of existing input nodes that are connected to this system node either via an edge or a directed path through a 0-minimum delay node, and the Y-number is the number of existing output nodes to which this node is connected either via an edge or a directed path through a 0-minimum delay node.

7), 16) Form a stack $R_x(R_y)$ of the remaining nodes (nodes in S_x) such that nodes with a smaller U-number (Y-number) are closer to the top of the stack. Arbitrarily order nodes with the same U-number (Y-number).

8), 17) Arbitrary assign an ordering 1 through $s(r)$ to the test output (input) nodes. Let the set $S_x(S_y) \leftarrow \emptyset$.

9), 18) If $R_x(R_y)$ is empty go to step 16(25).

10), 19) Pop the stack $R_x(R_y)$. Let $j \leftarrow 1$.

11), 20) Create a directed edge of gain 1 in g from this popped node (test input node j) to test output node j (this popped node).

12), 21) If a new i/o path through i-minimum delay nodes has been created go to step 9(18).

13), 22) Delete the directed edge created in step 11(20).

14), 23) Let $j \leftarrow j+1$. If $j \leq s(r)$ go to step 11(20).

15), 24) Assign the system node to the set $S_x(S_y)$ and go to step 9(18).

25) Let, $\bar{n}_i \leftarrow$ the number of system nodes in the set S_y and, $\hat{n}_i \leftarrow$ the number of i/o paths in g through i-minimum delay nodes which are between test input nodes and test output nodes.

26) Add to g $\bar{r}$ test input nodes and $\bar{s}$ test output nodes where $\bar{r}$ and $\bar{s}$ are chosen such that

$$\bar{r}+\bar{s} \quad \text{is a minimum}$$

subject to

$$(r+\bar{r})(s+\bar{s}) \geq \bar{n}_i + \hat{n}_i.$$

27) Let $r \leftarrow r+\bar{r}$ and $s \leftarrow s+\bar{s}$.

28) Create $\bar{n}_i$ new i/o paths through the i-minimum delay system nodes by adding to g appropriately directed edges of gain 1 between the system nodes in S_y and the test input and output nodes so that one and only one of these i/o paths goes through each system node in S_y.

29) Let $i \leftarrow i+1$. If $i < K$ go to step 4.

30) Output g.

31) Stop.

Note that in steps 7–15, we try to create new i/o paths by creating edges from a system node to a test output node; and in steps 16–24 we try to create more i/o paths by creating edges from test input nodes to system nodes.

The algorithm that we have presented has some shortcomings. In order to minimize the total number of test points we have introduced both input and output test points. However, an input test point requires the introduction of a signal generator and could therefore be called an *active* test point while an output test point requires only a measurement and therefore is a *passive* test point. In a practical situation one might wish to use more passive than active test points. Such a requirement can be incorporated into the algorithm by defining a cost function which assigns a greater cost to an active test point than to a passive test point and then minimizing this cost function rather than the total number of test points. Another drawback is that we have assumed that an input can be connected to any component and that the outputs of any set of components can be added to create a system output. In other words, we have assumed that there are no constraints on the connections. This may not always be true. One type of constraint that is imposed on the connections is that the inputs or outputs of two components may not be the same physical quantity. For example if the outputs of two components are voltage and current they cannot be added to create a system output. Another kind of con-

straint imposed on the connections is that they may have to satisfy certain physical laws. In electrical circuits for example the connections have to satisfy KVL and KCL. Physical systems will in some cases have certain constraints on the possible connections and, therefore, appropriate constraints have to be introduced into the algorithm for the design of the diagnosis matrices.

In spite of these weaknesses the algorithm is useful since it provides the basis for a systematic design of diagnosable systems. In contrast, existing tests of diagnosability can be applied only in a limited way to the design of diagnosable systems. To amplify the point, these techniques work as follows. The designer makes an "intelligent" choice of test points and checks the condition for diagnosability. If the condition is not satisfied, he chooses another set of test points and continues the process until the condition is satisfied. Such a technique might be effective for a small system but is inadequate for a system that a truly large scale.

VI. Summary and Conclusions

It is long been recognized that the problem of fault diagnosis of LSDS is one of exploiting the structure of the LSDS in a suitable manner. Saeks *et al.* [5] introduced the component-connection model of LSDS to the fault diagnosis problem since it gave a better insight into the structure of LSDS. In this two-part paper we have continued in the same spirit. By writing the fault diagnosis equations in terms of the Markov parameters of the LSDS rather than the transfer function evaluated at different frequencies as in [3]–[5], we are able to use to our advantage the property of the minimum delay order of a component to write, in Part I, the sequentially linear fault diagnosis equations. The condition for diagnosability as a rank test on a set of matrices follows from these equations.

In Part II we explore the "automatic" design of diagnosable systems. We first present a strategy of test point location, which reduces the problem of sequentially linear diagnosis of an LSDS to that of a canonical LSDS. We then discuss an algorithm to determine the test point locations required to make the parameters of the canonical LSDS sequentially linearly diagnosable in a generic sense. The algorithm is based on the graphical structure of the diagnosis matrices which is also discussed.

Acknowledgment

The authors wish to thank R. Saeks for many useful suggestions.

References

[1] P. K. Srimani, B. P. Sinha, and A. K. Choudhury, "Minimal test point location in a system graph," *Proc. 20th Midwest Symp. Circuits Syst.*, (Lubbock), pp. 567–570, 1977.
[2] B. L. Ho and R. E. Kalman, "Effective construction of linear state variable models from input–output data," *Proc. 3rd Ann. Allerton Conf. Circuit and System Theory*, pp. 449–459, 1965.
[3] N. Sen and R. Saeks, "A measure of testability and its application to test point selection—Theory," *Proc. 20th Midwest Symp. Circuits and Systems*, (Lubbock), pp. 576–583, 1977.
[4] M. N. Ransom and R. Saeks, "Fault isolation via term expansion," *Proc. 3rd Pittsburgh Symp. Modeling and Simulation* (Univ. of Pittsburgh), vol. 4, pp. 224–228, 1973.
[5] R. Saeks, S. P. Singh, and R. W. Liu, "Fault isolation via components simulation," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 634–640, Nov. 1972.
[6] T. N. Trick and C. J. Aljajian, "Fault diagnosis of analog circuits," *Proc. 20th Midwest Symp. Circuits Syst.* (Lubbock), pp. 576–583, 1977.
[7] S. P. Singh and R. W. Liu, "Existence of state equation representation of linear large-scale dynamical systems," *IEEE Trans. Circuit Theory*, vol. CT-20, pp. 239–246, May 1973.
[8] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhya*, Ser. A, vol. 30, pp. 167–180, 1968.
[9] S. Seshu, "The future of diagnosis," presented at *Proc. Seminar Automatic Checkout Techniques*, Battelle Memorial Institute, Columbus, OH 1962.
[10] W. M. Wonham, *Linear Multivariable Control*. Berlin: Springer-Verlag, 1974.
[11] H. Neudecker, "Some theorems on matrix differentiation with special reference to kronecker matrix products," *Amer. Stat. Assoc. J.*, pp. 953–963, Sept. 1969.
[12] F. Harary, *Graph Theory*. Reading, MA: Addison-Wesley, 1969.
[13] N. Deo, *Graph Theory with Applications to Engineering and Computer Science*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
[14] V. Visvanathan, "Sequentially-Linear Fault Diagnosis," M.S.E.E. thesis, Univ. of Notre Dame, Notre Dame, IN, Aug. 1978.
[15] R. Liu and V. Visvanathan, "Diagnosability of large-scale dynamical systems," *Proc. 20th Midwest Symposium on Circuits and Systems*, pp. 585–589, 1977.
[16] ——, "Sequentially Linear Fault Diagnosis: Part I—Theory," *IEEE Trans. Circuits Syst.*, this issue, pp. 490–496.
[17] S. R. Liberty, L. Tung, and R. Saeks, "Fault prediction—Towards a mathematical theory," in *Rational Fault Analysis*, Ed. R. Saeks and S. R. Liberty. pp. 135–142, 1977.
[18] R. Saeks and R. A. DeCarlo, *Interconnected Dynamical Systems*, New York: Marcel Dekker, to be published.

V. Visvanathan, for a photograph and biography please see page 496 of this issue.

Ruey-wen Liu (S'53–M'69), for a photograph and biography please see page 495 of this issue.

A Single-Fault Diagnosis Theory

by

R. Liu
C. Lin

A SINGLE-FAULT DIAGNOSIS THEORY

CHEN-SHANG LIN
R. LIU
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

A theory for the study of analog circuit single-fault-diagnosis problem is developed in this paper. First, the concepts of fuzzy, precise, and source components along with canonical circuits are introduced. A sufficient condition and an algorithm are then given to determine the diagnosibility of a canonical circuit and the identifiability of the faulty fuzzy-component in the circuit. Finally, an example is included to illustrate the application of the theory.

I. INTRODUCTION

Frequently, a faulty circuit results from very few faulty elements in the circuit. This observation has been used in several papers concerning fault-diagnosis problem of analog circuit. Most of them [1,2,3] employ the "simulation approach". In this approach, the various element values are changed so as to simulate element failures and all the test-point voltages are computed. These results are used either to prepare a "fault dictionary" which would be supplied to diagnosis technicians, or to compare directly with the measured data to determine the faulty elements under some criterion. The computation time and memory size are greatly reduced by the use of very-few-fault observation.

In this paper, the single-fault case is studied. By taking advantage of this information, a sufficient condition based on graph theory is obtained. The process involved a decomposition of a circuit into subcircuits to satisfy certain properties.

A novelty of the paper is the way it is treated for those elements whose nominal characteristics are fuzzily specified, or specified with a large tolerance.

II. PRELIMINARY DEFINITIONS AND ASSUMPTIONS

A circuit component may be a circuit element, a subcircuit or a functional block. Each component is characterized by its characteristic, such as the resistance of a resistor, the gain of an operational amplifier,

or the characteristic curves of a transistor.

Definitions.

1. A nominal characteristic is the designed characteristic of the component.

2. An actual characteristic is the characteristic of the component at the time of testing.

3. A precise component, or p-comp, is one whose nominal characteristic is precisely specified. It is fault-free if its actual characteristic is the same as its nominal one.

4. A fuzzy component, or f-comp, is one whose nominal characteristic is not precisely specified, but within a specified range. It is fault-free if its actual characteristic is within this range.

5. A source component, or s-comp, is one whose terminal-voltages and terminal-currents can be measured at the time of testing.

Examples of fuzzy components are transistors, operational amplifiers and electrolytic capacitors.

Remarks.

1. For fault diagnosis problems, the nominal characteristics are usually given and the actual characteristics are unknown.

2. In a realistic case, the actual characteristic of a precise component may deviate from its nominal one under fault-free case, but the deviation is small, say less than five percent.

Assumptions.

1. The actual characteristic of a component can be uniquely determined from its terminal-voltages and terminal-currents.

2. The nominal characteristic of a p-comp is voltage-controlled, i.e., its terminal-currents is uniquely determined by its terminal-voltages.

Definitions.

6. A circuit is diagnosable w.r.t. a set of measurements if all faulty components can be located from the measurements.

7. A component of a circuit is identifiable w.r.t. a set of measurements if its actual characteristic can be determined from the measurements.

III. THE MAIN THEOREM

In this section, a restricted class of circuits will be considered. It is a class of circuits crucial to one-component-fault problems. The extention to the general circuit will be considered in the later section.

Definition 6. A circuit is canonical if it satisfies the following:

(C1) All components are two-terminal.

(C2) It contains only p-comp and f-comp.

(C3) Its graph is non-separable.

(C4) There is neither f-comp only loops, nor p-comp only cutsets.

(C5) There is at most one component connecting any two nodes.

The canonical circuit may be excited by any number of s-comps, connecting across any two nodes, with any waveforms, as long as Assumption 1 is satisfied. With the above condition, the following theorem can be stated.

Theorem 1. A canonical circuit is diagnosable and its faulty f-comp is identifiable, if

(A1) at most one component is at fault.

(A2) all node voltages are measurable.

Proof: The proof of Theorem 1 is based on the following algorithm.

Step 1. Calculate the terminal currents for all p-comps, from their nominal characteristics and measured node voltages. These currents are called the estimated currents.

Step 2. Calculate the terminal currents for all f-comps, from the estimated currents obtained in Step 1. This is done by the fundamental cutset equations [4,5] associated with the tree containing all and only f-comps. The existence of such a tree is guarenteed by Condition (C4). These obtained currents again are called the estimated currents.

Step 3. Determine the characteristic of f-comps from their estimated currents and measured node voltages, called the estimated characteristic.

Step 4. Label each f-comp with the letter "T" or "F", according to the following rule. If the estimated characteristic is within its specified range, label it "T", otherwise, "F". There may be zero, one, or more than one f-comps labeled F. In the last case all f-comps labeled F have to be in the same fundamental loop.

Step 5. Locate the fault component according to the following rule.

1. If there is no F-labeled f-comp, then there is no faulty component in the circuit.

2. If there is one F-labeled f-comp, then this f-comp is at fault.

3. If there is more than one F-labeled f-comps, then the p-comp which defined the fundamental loop containing these f-comps is at fault.

It remains to be shown that the decision made in Step 5 is a correct one. Since there are at most one faulty components, there are only three

possibilities.

1) There is no faulty components. Therefore, the estimated currents of all p-comps are actual currents. As a consequence, all estimated currents and estimated characteristics of f-comps are actual ones. Therefore, there will be no F-labeled f-components.

2) There is one f-comp at fault. In this case, all p-comps are fault-free. By the same reason as above, all estimated currents and estimated characteristics of f-comps are actual ones. Therefore, there will be one F-labeled f-comp.

3) There is one p-comp at fault. In this case the estimated current of this, and only this, p-comp is at error. As a consequence, all the estimated currents are at error for those, and only those, f-comps in the fundamental loop defined by this p-comp. Therefore, their estimated characteristics will also be at error, causing a switch in their labelling from T to F. Since, in view of Condition (C5), there are at least two f-comps in every fundamental loop, there will be two or more f-comps labeled by F. This completes the proof that the circuit is diagnosable.

If one f-comp is at fault, from the above discussion, all estimated characteristics of f-comps are the actual ones. In particular, the actual characteristic of the faulty f-comp can be obtained. Therefore, it is identifiable. The theorem is proved.

The definition of canonical circuit and the conditions of Theorem 1 seem to be very restrictive. Actually this is not the case. The only crucial condition is that there exists no f-comps only loops and that there is at most one component at fault. The others can be alleviated by a proper grouping of circuit elements into components. This is illustrated by the next example.

IV. AN EXAMPLE

An example will be given which apparently does not satisfy Conditions (C1), (C4) and (A2).

Example 1. Consider a voltage regulator as shown in Figure 1. The measurable node voltages are at nodes (1), (2), (3) and (4), and the s-comps are at the input and output terminals. From these measurements, we want to locate the one-element-fault.

Note first that node voltages at nodes (A) and (B) are not measurable. This can be alleviated by grouping elements (Q1,Q2,D2,R5) into a 4-terminal f-comp and (D3,D4) into a two-terminal f-comp. The other f-comps are

D1 and C. The rest of them are p-comps. Next, the above 4-terminal f-comp is further represented by three coupled two-terminal f-comps. The resulting circuit is shown in Figure 2. Note that all node voltages are now measurable.

In order to check if the circuit is canonical or not, the s-comps are first removed. It is seen that there is a cutset of P1 and P3. This situation can be alleviated in the following way. Calculate the estimated currents for P1 and P3, and see if the cutset equation of P1, P3 and S1 is satisfied or not. If it is not satisfied, then the fault is clearly at either P1 or P3. No further calculations need to be done. If the cut set equation is satisfied, then both P1 and P3 are fault-free. In addition, the estimated currents are the actual currents. Therefore, P1 and P3 become s-comps. The remaining circuit no longer has any P-comp, only cutsets.

Since P3 is now considered as a s-comp, when removed the component F2 is separable from the circuit. A close check, the actual current of F2 can also be found. Therefore, it can also be considered as a s-comp. The resulting circuit is shown in Figure 3, which is canonical except that components F2, F3 and F4 are coupled.

By the same reasoning a partial answer can be obtained and it is summarized in Table 1. Note that only the partial labels are used for the decision and the coupling relation between (F2, F3, F4) is not used.

Observed Labels			Elements at Fault
F1	F5	F6	
T	T	T	(F2, F3, or F4), or no fault
T	T	F	F6 or P4
T	F	T	F5
T	F	F	P5
F	T	T	F1
F	T	F	P2
F	F	T	F5
F	F	F	More than one fault

Table 1. Diagnosis Diagram for Example 1.

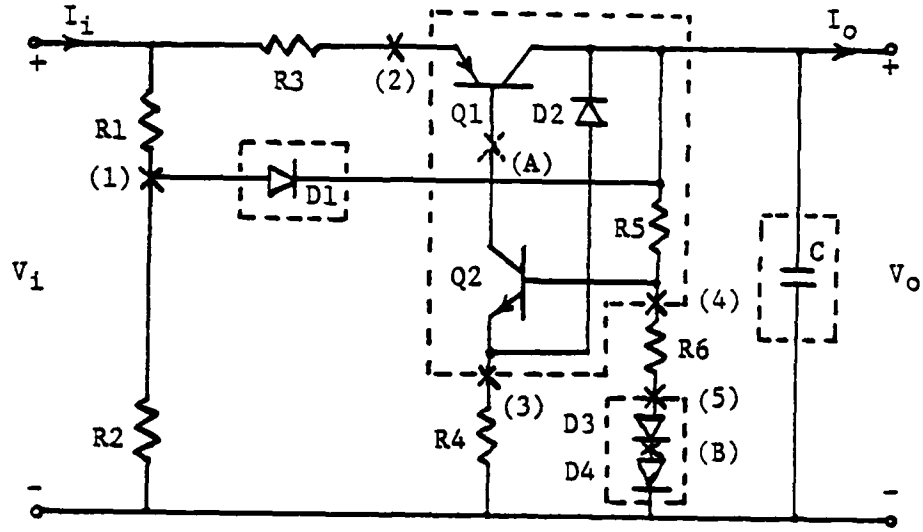


Figure 1

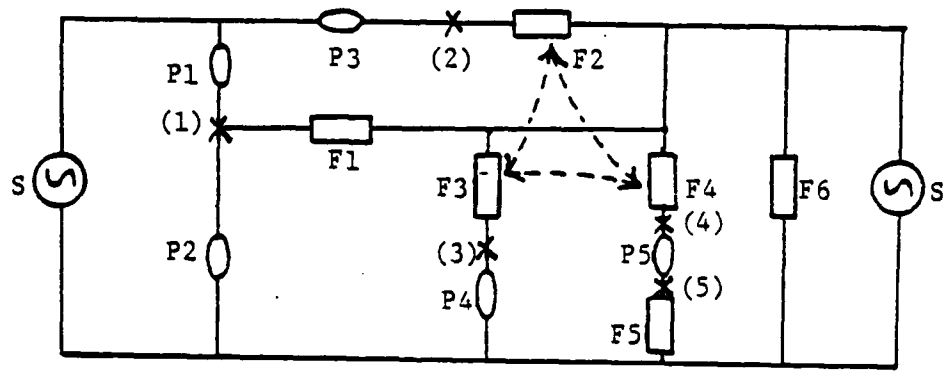


Figure 2

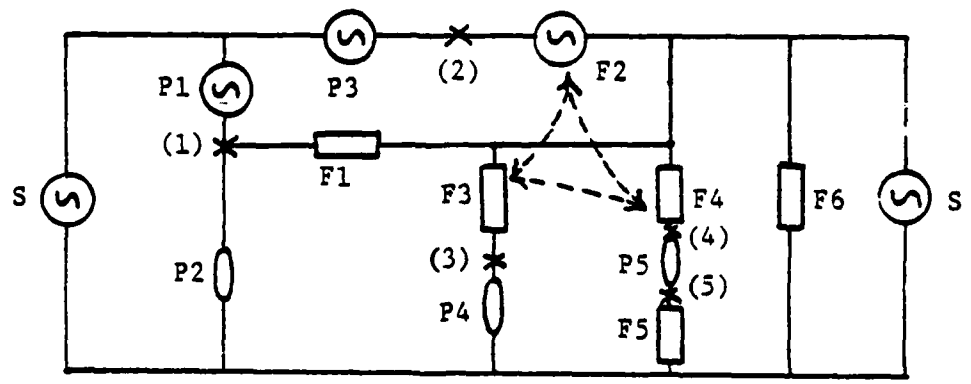


Figure 3

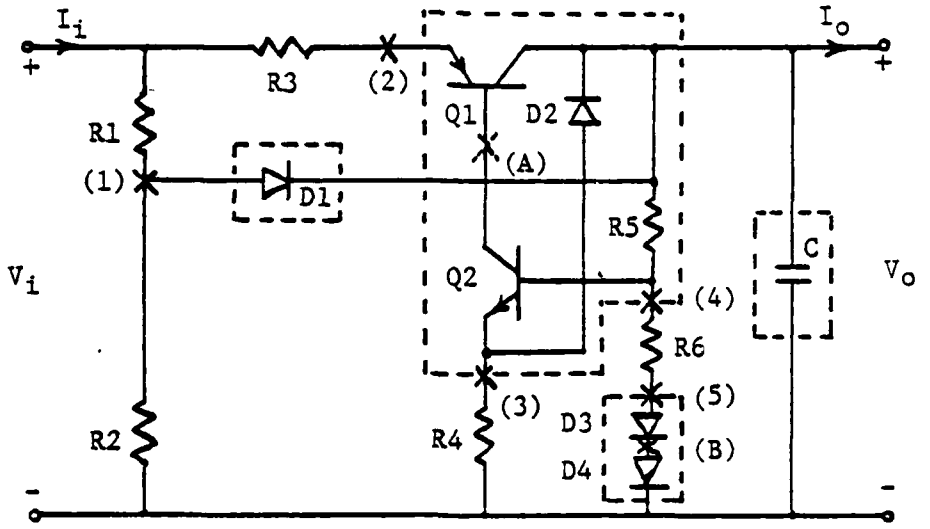


Figure 1

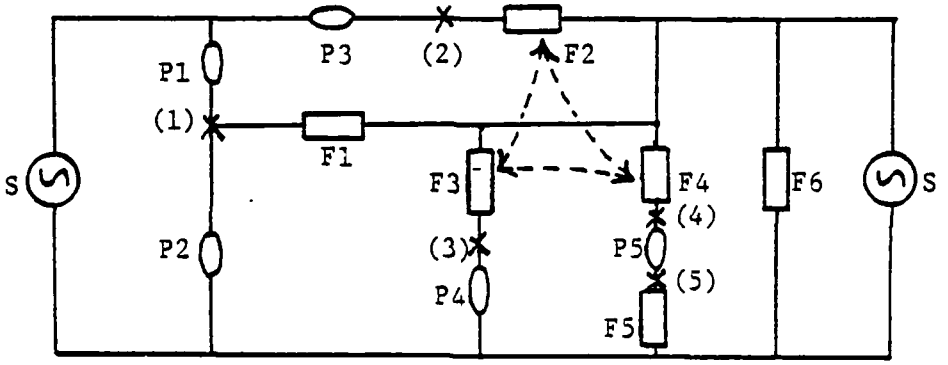


Figure 2

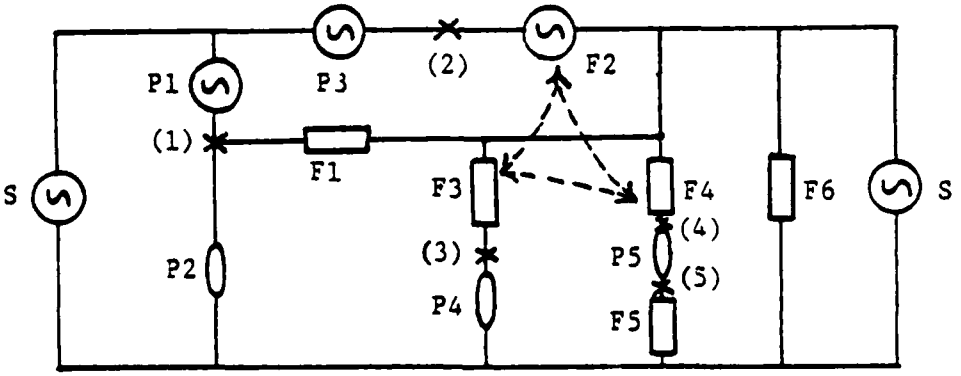


Figure 3

REFERENCES

[1] J. Lee and S.D. Bedrosian, "Fault Isolation Algorithm for Analog Electronic Systems Using the Fuzzy Concept", IEEE Trans. Circuits and Systems, Vol. CAS-26, No. 7, pp. 518-522, July 1979.

[2] R.F. Garzia, "Fault Isolation Computer Methods", NASA Contract Rep., CR-1758, Computer Sciences Corp., Feb. 1971.

[3] E.C. Neu, "Combinatorial Analysis for Check-out Techniques", Joint National Meeting of the American Astronautical Society, pp. 17-20, June 1969.

[4] S. Seshu and M.B. Reed, Linear Graphs and Electrical Networks, Addison-Wesley 1961.

[5] W. Chen, Applied Graph Theory, North-Holland, 1971.

ACKNOWLEDGEMENT

This paper was supported in part by the Office of Naval Research under Grant N0014-74-C-0444.

The Determination of the Rank of a Large Noisy Matrix

by

R. Liu
C. Sung

THE DETERMINATION OF THE RANK OF A LARGE NOISY MATRIX*

L. C. Suen
Bell Laboratories
Naperville, IL 60540

and

R. Liu
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

The conventional methods used for the determination of the rank of a noisy matrix are based on singular values. It is shown that the singular-value approach will create an inconsistancy especially when the matrix is large. The residual-number approach is presented for this case.

I. INTRODUCTION

Let the matrices A and $\tilde{A}$ be related by

$$\tilde{A} = A + E$$

where the matrix E is the noise (error) matrix whose elements are assumed to be small. A rank degeneracy problem [1] is the determination of the rank of A from the matrix $\tilde{A}$. We are especially interested in the case when the rank of $\tilde{A}$ is full rank but the rank of A is not. Under this case, a square matrix $\tilde{A}$ is usually said to be a ill-conditioned or near singular [2], [3].

One may conjecture that a nearly singular matrix must have a near-zero determinant, or a near-zero eigenvalues. This conjecture turns out to be false. There exists a matrix which is near singular but whose determinant and eigenvalues are equal to unity [4]. The current and conventional methods for the determination of the rank and the ill-conditioness of a matrix are based on its singular values [3], [5].

Let $A \in R^{mxn}$, with $m \geq n$. There exist orthogonal matrices U and V such that

$$V^T A U = \begin{bmatrix} \Sigma \\ 0 \end{bmatrix}$$

where $\Sigma = \text{diag.}\ (s_1, s_2, \ldots, s_n)$ and $s_1 \geq s_2 \geq \ldots \geq s_n \geq 0$. The numbers $s_1, s_2, \ldots, s_n$ are unique and are called the singular values. The rank of A is r if and only if $s_r \neq 0$ and $s_{r+1} = 0$. When a matrix is corrupted by a noise, a characterization of the rank of a matrix is given by Golub et al [1]. A matrix A is said to have numerical rank (δ,ϵ,r) with respect to the norm $\|\cdot\|$ if

$$r = \inf\{\text{rank } B \mid \|A - B\| \leq \epsilon\}$$

and

$$\epsilon < \delta \leq \text{Sup}\ \{\eta \mid \|A - B\| \leq \eta \Rightarrow \text{rank } B \geq r\}$$

The numerical rank is related with the singular values as follows.

Theorem 1 [1]. A matrix A has numerical rank (δ, ϵ, r) with respect to $\|\cdot\|_2$, if and only if

$$s_r \geq \delta > \epsilon \geq s_{r+1}.$$

Consequently, the numerical rank of a matrix is completely determined by its singular values. However, the singular-value approach will result as an inconsistancy. If the rank of an nxn matrix is $(n - 1)$, then by deleting an appropriate column of the matrix, the rank of the resulting matrix should remain to be $(n - 1)$. This is not true as shown in the following example if the singular-value approach is used.

Example 1 [1]. Consider the matrix

$$A(n) = I_n - \frac{e(n)e^T(n)}{n}$$

where I_n is the nxn identity matrix and $e^T(n) = [1, 1, \ldots, 1]$. It is easy to show that the singular values of this matrix are $(1, 1, \ldots, 1, 0)$. When any i columns are deleted, the remaining $nx(n - i)$ matrix has singular values $(1, 1, \ldots, 1, \sqrt{\frac{1}{n}})$. Clearly, when n is large, there is always one and only one singular value near-zero. In other words, for any (ϵ,δ) such that $1 \geq \delta > \epsilon > 0$, the numerical rank of the $nx(n - i)$ submatrix is always one less than its full rank, i.e., $(n - i - 1)$, for any $i < n$. This clearly is an inconsistency because if the rank of an nxn matrix is $r < n$, then there exists a nxr submatrix whose rank is full rank r.

In this paper, we propose to use the residual numbers to determine the rank of a matrix. Certain nice properties will be presented.

II. RESIDUAL NUMBERS

We consider $A \in R^{mxn}$. Let the rank of A denoted by $\rho(A)$ and the column-space of A by $R(A)$. Let a_i denote the i-th column-vector

*This research is supported in part by the ONR Grant N00014-78-C-0444.

of A, and A_i the remainder of A with a_i deleted. Without loss of generality, assume that $m \geq n$.

Definition 1. The n residual numbers r_i of A is defined by

$$r_i(A) \triangleq a_i^T(I - A_iA_i^+)a_i \tag{1}$$

for $i = 1, 2, \ldots, n$, where A_i^+ is the pseudo-inverse of A_i [6] and I the identity matrix.

It is well known that

$$r_i = \min\{||a_i - x||^2 \mid x \in R(A_i)\}$$

Consequently, a_i is a linear combination of the column-vectors of A_i if and only if $r_i = 0$. Therefore, A has full rank if and only if $r_i \neq 0$ for all i. These observations lead to an alternate definition of the rank of a matrix A. It also provides a residual-number test for the rank of A, when not corrupted by noise.

Definition 2. An $m \times n$ matrix A has rank r if,

(i) there exists an $m \times r$ submatrix B such that

$$r_i(B) > 0 \quad , \quad i = 1, 2, \ldots, r$$

(ii) there is no $m \times (r+1)$ submatrix B such that

$$r_i(B) > 0 \quad , \quad i = 1, 2, \ldots, (r+1).$$

An important relation between a matrix A and its residual numbers is given by Theorem 2.

Theorem 2. When $(A^TA)^{-1}$ exists,

$$r_i = d_i^{-1} \tag{2}$$

where d_i is the i-th diagonal element of $(A^TA)^{-1}$; otherwise,

$$r_i = \lim_{\epsilon \to 0^+} d_i^{-1}(\epsilon) \tag{3}$$

where $d_i(\epsilon)$ is the i-th diagonal element of $(A^TA + \epsilon I)^{-1}$ for $\epsilon > 0$.

As a consequence, a relation between the residual numbers and the singular values of A can be established. Let $s_1, s_2, \ldots, s_n$ be the singular values of A.

Corollary 2.1. If A has full column-rank i.e., $\rho(A) = n$, then

$$\sum_{i=1}^{n} \frac{1}{r_i} = \sum_{i=1}^{n} \frac{1}{s_i^2} \tag{4}$$

In view of Theorem 2, it is reasonable to conjecture that when $\rho(A) < n$,

$$\sum_i \frac{1}{r_i} = \sum_i \frac{1}{s_i^2}$$

where the summation is over non-zero r_i's and s_i's. It turns out that this is false, as shown by the following counter example.

Example 2. Consider

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

It is easy to show that $r_1 = 0$, $r_2 = 0$, $r_3 = 1$ while $s_1 = \sqrt{2}$, $s_2 = 1$, $s_3 = 0$.

Clearly, the above conjecture is not satisfied.

The case of noisy matrices will be considered next.

III. THE MAIN THEOREM

Let us now consider the case when a matrix $A \in R^{mxn}$, with $m \geq n$, corrupted by a noise matrix V, i.e.,

$$\tilde{A} = A + V \tag{5}$$

The problem is to determine the rank of A from the measured $\tilde{A}$. Assume that

1) $E[v_{ij}] = 0$

2) $E[v_{ij}v_{kl}] = \sigma^2$ if $i = k$, $j = l$
$= 0$ if otherwise

where $E[x]$ is the expectation of x. Then,

1a) $E[V] = 0$ (6)

2a) $E[V^TV] = m\sigma^2 I$ (7)

where m is the number of rows of V. From (5),

$$\tilde{A}^T\tilde{A} = A^TA + A^TV + V^TA + V^TV \tag{8}$$

Let the elements of A be bounded, i.e., $|a_{ij}| \leq K$ for some K. It can be shown that when $m \to \infty$,

$$\frac{1}{m} A^TV \to 0$$

$$\frac{1}{m} V^TA \to 0$$

$$\frac{1}{m} V^TV \to \sigma^2 I$$

Therefore, (8) reduces to

$$\tilde{A}^T\tilde{A} = A^TA + m\sigma^2 I \tag{9}$$

Consequently, the singular values $\tilde{s}$ of $\tilde{A}$ is related to the singular values s of A by

$$\tilde{s}_i^2 = s_i^2 + m\sigma^2 \tag{10}$$

for $i = 1, 2, \ldots, n$. This shows that if $s_r = 0$

then $\tilde{s}_r$ is near-zero if the noise is small. We want to derive a similar relation for the residual numbers.

Theorem 3.

$$r_i(\tilde{A}) = r_i(A) + m\sigma^2 + \sum_{j=1}^{\infty} (-1)^{j+1}(m\sigma^2)^j a_i^T (A_i^{+T} A_i^+)^j a_i \tag{11}$$

In addition, the third term can be estimated by

$$m\sigma^2 ||A_i^+ a_i||^2 \geq \sum_{j=1}^{\infty} (-1)^{j+1}(m\sigma^2)^j a_i^T (A_i^{+T} A_i^+)^j a_i \geq \frac{m\sigma^2}{1+\alpha_r} ||A_i^+ a_i||^2$$

where $\alpha_r = \frac{m\sigma^2}{s_r^2}$, and s_r is the smallest non-zero singular value of A.

As a consequence, we have Corrollary 3.1.

Corollary 3.1.

$$r_i(A) + m\sigma^2 + m\sigma^2 ||A_i^+ a_i||^2 \geq r_i(\tilde{A}) \geq r_i(A) + m\sigma^2 + \frac{m\sigma^2}{1+\alpha_r} ||A_i^+ a_i||^2$$

for $i = 1, 2, \ldots, n$.

Note that when $\alpha_r \to 0$, both sides of the bounds are the same, and we have an equality. Next we need an estimation of $||A_i^+ a_i||^2$. This is given by the next two theorems.

Theorem 4.

a) $||A^+ a_i||^2 = 1$ if $a_i \notin R(A_i)$

$$= \frac{||A_i^+ a_i||^2}{1 + ||A_i^+ a_i||^2} \quad \text{if } a_i \in R(A_i)$$

b) $$\sum_{i=1}^{n} ||A^+ a_i||^2 = r$$

where r is the rank of A.

Theorem 5. If $\rho(A) = r < n$, then there exists $a_i \in R(A_i)$ such that

$$||A_i^+ a_i||^2 \leq \frac{r}{n-r}$$

Combining Corollary 3.1 and Theorem 5, we obtain the main theorem.

Theorem 6. If

$$r_i(\tilde{A}) > mn\sigma^2$$

for $i = 1, 2, \ldots, n$, then $\rho(A) = n$.

Note that $mn\sigma^2$ may not be large even if m is large. For example, if we are interested in the effect on the rank of A due to the round-off error from a computer operation. Typically, $\sigma = 10^{-6}$. Therefore, even if $m = n = 100$, $mn\sigma^2 = 10^{-8}$, which is a small number.

Based on Theorem 6, we will give a definition of the numerical rank of A and an algorithm to compute it.

Definition 3. An mxn matrix A has a numerical rank (σ, r) if

(i) there exists an mxr submatrix B such that

$$r_i(B) > mr\sigma^2 \quad , \quad i = 1, 2, \ldots, r.$$

(ii) there is no mx(r+1) submatrix B such that

$$r_i(B) > m(r+1)\sigma^2 \quad , \quad i = 1, 2, \ldots, (r+1).$$

The above definition is an extension of Definition 2 from the noise-free case to the noisy case. An algorithm to obtain the numerical rank of a given matrix will now be presented.

Algorithm.

Step 0. Let $B \leftarrow A$, $r \leftarrow n$.

Step 1. Compute $r_i(B)$ for $i = 1, 2, \ldots, r$.

Step 2. Choose the smallest of $r_i(B)$, say $r_s(B)$.

Step 3. If $r_s(B) \leq mn\sigma^2$ and $r \geq 1$, then $B \leftarrow B_s$, $r \leftarrow r-1$, go to Step 1.
Otherwise, numerical rank = r, stop.

It can be shown that the above algorithm yields the numerical rank (σ, r) of Definition 3.

Let us now go back to Example 1. It can be shown that the residual numbers of $A(n)$ are $(0, 0, \ldots, 0)$. When any column is deleted, the residual numbers of the remaining $nx(n-1)$ matrix are $(\frac{1}{2}, \frac{1}{2}, \ldots, \frac{1}{2})$, which are independent of n. Since

$$\frac{1}{2} > n(n-1)\sigma^2$$

if σ is small enough, it can be concluded from Definition 3 that the numerical rank of $A(n)$ is $(\sigma, n-1)$. It can be further shown that when i columns are deleted from $A(n)$, the residual numbers of the remaining $nx(n-i)$ submatrix are the same and equal to $\frac{i}{i+1} \geq \frac{1}{2}$. Therefore, all such submatrices are of full rank as they should be.

The differences between the residual numbers and the singular values can be seen from (4).

$$\sum_{i=1}^{n} \frac{1}{r_i} = \sum_{i=1}^{n} \frac{1}{s_i^2} \qquad (4)$$

Usually, the residual numbers do not differ very much from each other, i.e.,

$$r_i \approx r_j$$

On the other hand, when the numerical rank is $n - 1$, it is usually that $s_{n-1} >> s_n$.

Therefore, Eq. (4) becomes

$$\frac{n}{r_i} \approx \frac{1}{s_n^2}$$

or

$$s_n^2 \approx \frac{1}{n} r_i$$

i.e., the smallest singular value is much smaller than the residual numbers when n is large. Put it into another way, the smallness of the residual numbers are distributed equally among themselves, while the smallness of the singular values are concentrated on the smallest one. It can be unusually small when n is large, and therefore, it may not be accurately served as an indicator for the rank of a matrix. In other words, when a singular value is near-zero, there are at least two possibilities. The matrix may be near singular, or the matrix may be too large.

In conclusion, the singular-value approach may work well when the matrix is small, but fail when the matrix is large, in which case, the residual-number approach is better.

IV. REFERENCES

[1] G. H. Golub, V. Klema and G. W. Stewart, "Rank Degeneracy and Least Squares Problems", Technical Report TR-456, Computer Science Department, Stanford University, June 1976.

[2] D. K. Faddeev and V. N. Faddeeva, Computational Methods of Linear Algebra, W. H. Freeman and Company, 1963.

[3] G. W. Steward, Introduction Matrix Computation, Academic Press, 1973.

[4] B. Noble, Applied Linear Algebra, Prentice-Hall, Inc., 1969.

[5] G. H. Golub, "Least Squares, Singular Values, and Matrix Approximation", Aplikace Mathematiky, Vol. 13, pp. 44-51, 1968.

[6] C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and Its Applications, John Wiley and Sons, Inc., 1971.

Accessibility of Large-Scale Electronic Circuits

by

R. Liu
C. Lin
V. Visvanathan

ACCESSIBILITY OF LARGE-SCALE ELECTRONIC CIRCUITS

R. Liu*, V. Visvanathan**, C. Lin*

University of Notre Dame
Notre Dame, IN 46556

University of California
Berkeley, CA 94720

ABSTRACT

The problem of fault diagnosis of large-scale analog circuit is studied. Any fault diagnosis procedure is limited by the number of circuit parameters to be diagnosed. When such limit is exceeded by large-scale circuits, some kind of tearing process has to be implemented before a fault diagnosis procedure can be applied. In this paper, a tearing process via accessibility of subnetworks is presented. The necessary and sufficient condition for accessibility is obtained. The implementation of this tearing process is discussed. The tearing process can be applied to nonlinear circuits.

I. INTRODUCTION

In the study of Large-Scale Dynamical Systems (LSDS), in order to simplify a problem we often reduce it from the level of the overall system to that of its components or subsystems. Tearing or Diakoptics [1] is such an approach for the analysis of large-scale networks. For the fault diagnosis of LSDS there is to technique which is equivalent to tearing. Existing methods of fault diagnosis (for example [3-6]) attack the problem at the LSDS level.

The easiest way to transfer the problem of fault diagnosis from the level of the overall system to that of the subsystems is to have direct access to the inputs and outputs of each subsystem. However, such direct access may not be available to us. In such a case if we can determine the inputs and outputs of the components of interest from the LSDS inputs and outputs, we have effectively accessed them. Intuitively, we can say that this would be possible if a mapping existed from the space of input-output waveforms of the LSDS to the space of input-output waveforms of the components. Such a map would be the basis of our tearing approach. In this paper, we explore these concepts and determine the necessary and sufficient conditions for the existence of such a map which takes as much advantage of the known information as possible. We then lay the intuitive basis for a strategy of tearing which simplifies the problem of fault diagnosis. The results presented are a generalization of an earlier work by Saeks, Singh and Liu [2] and Liu and Visvanathan [7].

[1]This research is supported in part by ONR Grant No. N00014-78-C-0444.

II. ACCESSIBILITY FROM INPUT/OUTPUT TERMINALS

The LSDS model consists of three parts, the masked subsystem, the unmasked subsystem and the connection-box as shown in Figure 1. The vectors u and y denote the input and output vectors of the LSDS, c and d of the unmasked subsystem and r and s of the masked subsystem. The connection-box consists of the connections between the above variables. The equations for the three parts are:

a) Unmasked Subsystems

We assume that the unmasked subsystem is a linear dynamical system described by:

$$\begin{aligned} \dot{x}(t) &= A\,x(t) + B\,c(t) \\ d(t) &= C\,x(t) + D\,c(t) \end{aligned} \tag{1}$$

where A,B,C, and D are constant matrices and x is the state vector for the unmasked subsystem.

b) Connection Box

The connection box is described by [10]:

$$\begin{bmatrix} c \\ r \\ y \end{bmatrix} = \begin{bmatrix} L_{cd} & L_{cs} & L_{cu} \\ L_{rd} & L_{rs} & L_{ru} \\ L_{yd} & L_{ys} & L_{yu} \end{bmatrix} \begin{bmatrix} d \\ s \\ u \end{bmatrix} + \begin{bmatrix} R_{cc} & R_{cr} & R_{cy} \\ R_{rc} & R_{rr} & R_{ry} \\ R_{yc} & R_{yr} & R_{yy} \end{bmatrix} \begin{bmatrix} c \\ r \\ y \end{bmatrix} \tag{2}$$

The L's and R's are constant matrices. Note that d,s and u are inputs to the connection-box and c,r and y are the outputs.

c) Masked Subsystem

The inputs and outputs of the masked subsystem are related by some functional form

$$s = f \cdot r \tag{3}$$

which is assumed unknown. For example, the relation (3) could be a state equation or a zero-memory nonlinear function.

Equations (1), (2) and (3) completely describe the LSDS. The unmasked subsystem has been included in the LSDS model to provide us with a greater flexibility. Components that are known to be fault-free or have been independently diagnosied can be

included as part of the unmasked subsystem. We further assume that the LSDS is well-posed, i.e., the initial value solution (x(t), c(t), d(t), r(t), s(t)) exists and is unique for all admissible inputs u(·).

Definition 1.

The masked subsystem of the LSDS is said to be accessible from input/output terminals, or simply accessible, if $\forall t \in [0,\infty)$, (r(t),s(t)) can be uniquely determined from (u(τ), y(τ)) for $\tau \in [0,t]$ by use of equations (1) and (2) but not (3), with the initial state x(0) = 0 of the unmasked subsystem. It is said to be anticipatively accessible if (r(t), s(t)) can be uniquely determined from (u(τ),y(τ)) for $\tau \in [0,t+\delta]$ for some $\delta > 0$ but not for $\delta = 0$.

Theorem 1 [11]

The masked subsystem of the LSDS is accessible if and only if the matrix J

$$J = \begin{bmatrix} R_{cc} + L_{cd}D - I & R_{cr} & L_{cs} \\ R_{rc} + L_{rd}D & R_{rr} - I & L_{rs} \\ R_{yc} + L_{yd}D & R_{yr} & L_{ys} \end{bmatrix}$$

has full column-rank.

Note that the accessibility only depends on the memoryless part of LSDS. The application of Theorem 1 to large-scale networks is considered next.

III. ACCESSIBILITY OF SUBNETWORKS

The above result can be applied to the diagnosis of subnetworks. A given network can be decomposed into three parts as shown in Figure 2. Those R-elements, which are known to be reliable, are placed in the resistive-network box. Those L,C-elements and/or (nonlinear) devices are placed in the masked box. The last box contains all the elements to be diagnosed.

Note that (u,y) and (r,s) are the port-voltages and the port-currents of the overall network and of the subnetworks in the masked box respectively.

The unmasked part has a state-equation representation (1),

$$\dot{x} = Ax + Bc$$
$$d = Cx$$

where A = 0, C = I, and B = diag $(\frac{1}{C_i}, \frac{1}{L_j})$, $C = (i_{C_i}, v_{L_j})$, $d = (v_{C_i}, i_{L_j})$.

Associate each network in Figure 2, construct a LC-reduced network, or simply reduced network, by replacing each inductor by an open circuit and each capacitor by a short circuit as shown in Figure 3. Note that the reduced network is a resistive network.

Theorem 2. [11]

The masked subnetwork in Figure 2 is accessible if and only if (i_C, v_L, r, s) can be uniquely determined by (u,y) from its reduced resistive network.

IV. ACCESSIBILITY OF FURTHER REDUCED SUBNETWORK

For the purpose of testing the accessibility, it is possible to further simplify the network. In order to do this, more notations are needed. In this section, m-terminal masked box is decomposed into m-1 2-terminal X-devices and the set of X-devices-branches is denoted by G_X. Note that the knowledge of all X-devices completely describes the behavior at the terminals of the original masked box. Similarly, the set of all inductor-, capacitor-branches are denoted by G_L, G_C respectively. As for the resistor-branches, since some of them may be in a particular tree we choose in the network, we denote the set of them in the tree by G_R, and the rest in the corresponding cotree by G_G.

Now consider a network N satisfying the following assumptions.

(1) The network N is a connected graph.

(2) The sources, including voltmeters and ammeters are considered as a set of branches with measurable voltages and currents. The symbol G_S is used to denote the set.

(3) $G_X \cup G_C$ contains no loop, otherwise ammeters are inserted to break the loops.

(4) $G_S \cup G_L$ contains no cut set.

(5) All resistors have positive resistance.

These five conditions are assumed to be satisfied the networks considered in this section.

Under above assumptions, there exists a tree (t) which contains (G_X, G_C, G_R) and does not contain (G_S, G_L, G_G). Let the set of all such t be denoted by T. Then for each t in T, we have the following equations:

KCL:

$$\begin{bmatrix} I_X \\ I_C \\ I_R \end{bmatrix} = - \begin{bmatrix} F_{SX} & F_{SC} & F_{SR} \\ F_{LX} & F_{LC} & F_{LR} \\ F_{GX} & F_{GC} & F_{GR} \end{bmatrix}^T \begin{bmatrix} I_S \\ I_L \\ I_G \end{bmatrix} \quad (4a)$$

KVL:

$$\begin{bmatrix} V_S \\ V_L \\ V_G \end{bmatrix} = \begin{bmatrix} F_{SX} & F_{SC} & F_{SR} \\ F_{LX} & F_{LC} & F_{LR} \\ F_{GX} & F_{GC} & F_{GR} \end{bmatrix} \begin{bmatrix} V_X \\ V_C \\ V_R \end{bmatrix} \quad (4b)$$

Ohm's Law:

$$V_R = R\, I_R$$
$$I_G = G\, V_G \quad (4c)$$

where R and G are diagonal matrices with positive

diagonal elements.

Define

$$W = \left[\begin{array}{c|c} F_{GR}^{T}\, G\, F_{GX} & R^{-1} + F_{GR}^{T}\, G\, F_{GR} \\ \hline F_{SX} & F_{SR} \end{array}\right]$$

Then from Equation (4) and Theorem 2, the following lemma can be proved.

Lemma 1.

The X-devices in network N are accessible if and only if the associated matrix W has full column-rank.

Although the size of matrix W is smaller than the original matrices, that is only superficial because some of the submatrices cannot be determined without the knowledge of the original matrices. The following lemmas will provide a better solution.

Lemma 2.

In a network N, the associated matrix W has full column-rank in the generic sense [12] if the matrix F_{SX} has full column-rank.

Lemma 3.

The matrix F_{SX} of a network N has full column-rank if (1) there exists a tree tεT such that $G_R U G_S$ contains no loop in LC-reduced network of N and (2) the matrix W associated with N has full column-rank.

The second condition in Lemma 3 will be referred as Assumption (5) in the sequent paragraphs.

Note that not only the size of the matrix F_{SX} is smaller than W but also the matrix itself can be determined by the subgraph N" constructed by shorting all resistor-branches in the particular tree and opening all resistor-branches in the corresponding cotree in the LC-reduced network N'. We call the subgraph N" RLC-reduced network. Then Theorem 3 follows.

Theorem 3.

Let N be a network satisfying Assumptions (1) to (5), then X-devices in N are generically accessible if and only if the matrix F_{SX} in RLC-reduced network associated with t has full column-rank.

Remarks

1. Theorem 3 enables us to determine the accessibility of X-devices in a network by dealing with a considerably smaller subnetwork N" with only sources and X-devices in this subnetwork.
2. From graph theory, F_{SX} has full column-rank if and only if G_S contains a tree in RLC-reduced network. Therefore, to achieve the accessibility of X-devices, we only need to add some voltmeter-branches into G_S to make it contain a tree in RLC-reduced network.
3. In this theory, in order to access all the X-devices in a network, the number of sources must be at least equal to the number of X-devices as can be seen from the required full column-rank of F_{SX}.
4. The choice of tree in a network is crucial in determining the minimal set of test points to obtain accessibility. Altough the algorithm of finding the tree is still under development, it is quite possible to pick the tree in a network of reasonable scale by inspection as shown in Example 1.

Example 1

Consider the accessibility of the two transistors in the two-stage amplifier circuit as shown in Fig. 4a. It is clear to see that (C_1, T_1, C_2, T_2, C_3) form a loop. From Assumption (2), the loop can be broken by inserting an ammeter A in series with C_2. Then, by choosing the tree consisting (X_1, X_2, X_3, X_4, R_5) in Fig. 4b, the associated RLC-reduced network is shown in Fig. 4c. Apparently, the branches (E_1, E_2, E_3, A) contains a tree in Fig. 4c. Thus, (X_1, X_2, X_3, X_4), so that (T_1, T_2), are accessible after the insertion of the ammeter.

V. APPLICATION TO TEARING PROCESS

The purpose of diakoptic or tearing process [1] is to find a way of partitioning a large-scale network into smaller subnetworks so that the solution of the large-scale network can be obtained by solving the (decoupled) subnetworks. Clearly, this represents a reduction in computation time.

If fault diagnosis is of our interest instead of the network solutions, new tearing process should be developed so that it is compatible to fault diagnosis problem. The accessibility can fulfill such purpose. Let us consider the network in Figure 2. If each masked subnetwork is diagnosable from its input/output pair (r,s) and is accessible, then the entire network is diagnosable from its input/output pair (u,y). This is because (r,s) can be obtained from (u,y) independently from the characteristics of masked subnetworks.

In summary, if accessibility is achieved, one can diagnose the entire network by diagnosing each of the (smaller and decoupled) masked subnetworks individually.

REFERENCES

(1) L.O. Chua and L.K. Chen, "Diakoptic and Generalized Hybrid Analysis", IEEE Trans. Circuits and Systems, vol. CAS-23, No. 12, pp. 694-705, December 1976.

(2) R. Saeks, S.P. Singh and R. Liu, "Fault Isolation via Components Simulation", IEEE Trans. Circuit Theory, vol. CT-19, No. 6, pp. 634-640, November 1972.

(3) V. Visvanathan and R. Liu, "Sequentially-Linear Fault Diagnosis-Part I: Theory", to appear.

(4) M.N. Ransom and R. Saeks, "Fault Isolation via Term Expansion", Proc. 3rd Pittsburgh Symposium on Modeling and Simulation, Univ. of Pittsburgh, 1973, vol. 4, pp. 224-228.

(5) T.N. Trick and C.J. Alajajian, "Fault Diagnosis of Analog Circuits", Proc. 20th Midwest Symposium on Circuits and Systems, Lubbock, Texas, 1977, pp. 211-215.

(6) N. Sen and R. Saeks, "A Measure of Testability and its Application to Test Point Selection -

Theory", Proc. 20th Midwest Symposium on Circuits and Systems, Lubbock, 1977, pp. 576-583.
(7) R. Liu and V. Visvanathan, "Diagnosability of Large-Scale Dynamical Systems", Proc. 20th Midwest Symposium on Circuits and Systems, Lubbock, 1977, pp. 585-589.
(8) V. Visvanathan, "Sequentially-Linear Fault Diagnosis", M.S.E.E. Thesis, Univ. of Notre Dame, Notre Dame, IN., August 1978.
(9) L.M. Silverman and H.J. Payne, "Input-Output Structure of Linear Systems with Application to the Decoupling Problem", SIAM J. Control, vol. 9, No. 2, pp. 199-233, May 1971.
(10) S.P. Singh and R. Liu, "Existence of State Equation Representation of Linear Large-Scale Dynamical Systems", IEEE Trans. on Circuit Theory, vol. 20, pp. 239-246, 1973.
(11) R. Liu, V. Visvanathan and C. Lin, "Tearing in Fault Diagnosis", 1979 IEEE ISCAS.
(12) W.M. Wonham, Linear Multivariable Control, Berlin: Springer-Verlag, 1974.

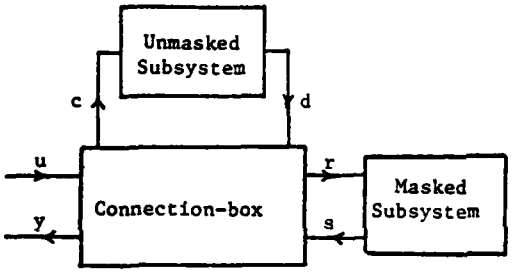


Figure 1

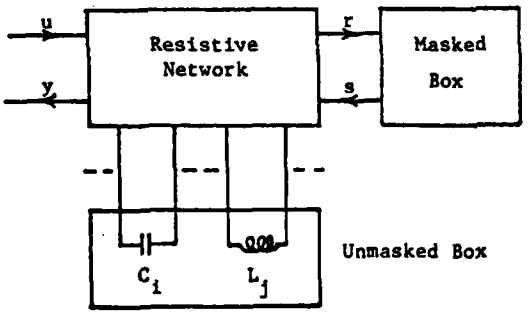


Figure 2

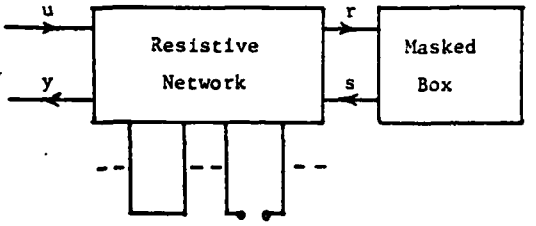


Figure 3

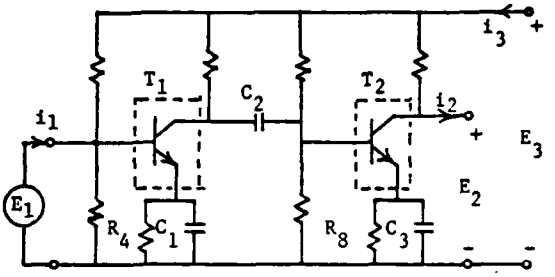


Figure 4a

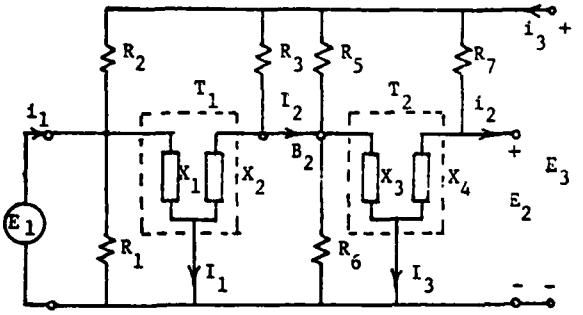


Figure 4b

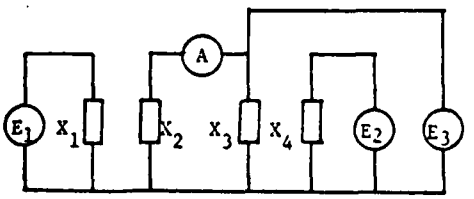


Figure 4c

Feedback System Design: The Fractional Representation Approach to Analysis and Systems

by

R. Liu
C. Desoer
J. Murray
R. Saeks

Feedback System Design: The Fractional Representation Approach to Analysis and Synthesis

C. A. DESOER, FELLOW, IEEE, RUEY-WEN LIU, JOHN MURRAY, AND RICHARD SAEKS, FELLOW, IEEE

Abstract—The problem of designing a feedback system with prescribed properties is attacked via a fractional representation approach to feedback system analysis and synthesis. To this end we let H denote a ring of operators with the prescribed properties and model a given plant as the ratio of two operators in H. This, in turn, leads to a simplified test to determine whether or not a feedback system in which that plant is embedded has the prescribed properties and a complete characterization of those compensators which will "place" the feedback system in H. The theory is formulated axiomatically to permit its application in a wide variety of system design problems and is extremely elementary in nature requiring no more than addition, multiplication, subtraction, and inversion for its derivation even in the most general settings.

I. INTRODUCTION

INTUITIVELY, the linear feedback system design process may be broken down into three steps: modeling, analysis, and synthesis; each of which may be carried out via a multiplicity of time and frequency domain techniques. In engineering practice, however, the three steps are loosely matched to one another. The purpose of the present paper is to use fractional representation models to the analysis and synthesis of feedback systems. Here, if one desires to design a system with prescribed properties the given plant is initially modeled as a quotient of two operators, each of which has the desired properties. Once such a model has been specified a similar model may be formulated for the feedback system constructed from that plant which, in turn, may be used to determine whether or not the feedback system has the desired properties. Moreover, the set of compensators which will cause the feedback system to have the prescribed properties may be completely characterized in terms of such a model. As such, by choosing a model for the plant which is matched to the design criteria the analysis and synthesis processes for a feedback system may be greatly simplified.

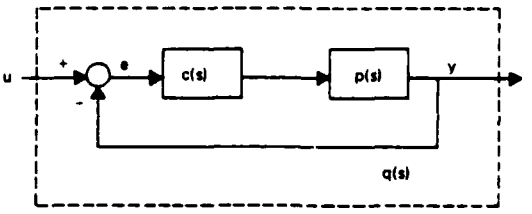


Fig. 1. Single-variate control system.

These ideas are illustrated by the following derivation of the set of stabilizing compensators for the single variate control system of Fig. 1.

We say that a transfer function $p(s)$ is *exponentially stable* (exp. stable) if $p(s)$ is a *proper rational* function with poles having *negative* real parts. Although the plant may naturally be modeled as a quotient of coprime polynomials [16],[19] $p(s) = a(s)/b(s)$ since our ultimate goal is a

Manuscript received April 2, 1979; revised October 15, 1979. Paper recommended by B. Francis, Chairman of the Linear Systems Committee. The work of C. A. Desoer was supported in part by the Joint Services Electronics Program at the University of California, Berkeley, under AFOSR Contract 76-C-0100. The work of R.-W. Liu was supported in part by ONR Contract 78-C-0444. The work of J. Murray and R. Saeks was supported in part by the Joint Services Electronics Program, Texas Tech University, under ONR Contract 76-C-1 3[illegible].

C. A. Desoer is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, CA 94720.

R.-W. Liu is with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

J. Murray and R. Saeks are with the Department of Electrical Engineering, Texas Tech University, Lubbock, TX 79409.

0018-9286/80/0600-0399$00.75

stable system we prefer to model $p(s)$ as a quotient of exp. stable rational functions

$$p(s)=n(s)/d(s)=[a(s)/m(s)][b(s)/m(s)]^{-1} \quad (1.1)$$

where $m(s)$ is strictly Hurwitz polynomial of degree equal to the degree of $b(s)$. Moreover, since $a(s)$ and $b(s)$ are coprime, the rational functions $n(s)$ and $d(s)$ are coprime in the sense that there exist exp. stable rational functions $u(s)$ and $v(s)$ such that

$$u(s)n(s)+v(s)d(s)=1. \quad (1.2)$$

Similarly, we assume that our compensator is modeled as a quotient of exp. stable rational functions, $c(s)=x(s)/y(s)$, which are coprime in the above sense. Now, a little algebra will reveal that the closed-loop system transfer function from input u to output y is given by a ratio of exp. stable rational functions in the form

$$h_{yu}(s)=n(s)[y(s)d(s)+x(s)n(s)]^{-1}x(s). \quad (1.3)$$

Moreover, it can be shown[1] that $h_{yu}(s)$ will be stable if and only if

$$[y(s)d(s)+x(s)n(s)]=k(s) \quad (1.4)$$

has an exp. stable inverse. Since $k(s)$ is, itself, exp. stable this implies that the feedback system will be exp. stable if and only if $k(s)$ is nonzero for all $\mathrm{Re}\, s\geq 0$, including ∞. An exp. stable function with these properties is called miniphase. As such, the problem of synthesizing an exp. stable feedback system reduces to the solution of (1.4) for exp. stable rational functions $x(s)$ and $y(s)$ given exp. stable functions $n(s)$ and $d(s)$ and a miniphase function $k(s)$.

By direct substitution one may verify that

$$y^h(s)=r(s)n(s) \quad \text{and} \quad x^h(s)=-r(s)d(s) \quad (1.5)$$

satisfy the homogeneous equation

$$y^h(s)d(s)+x^h(s)n(s)=0 \quad (1.6)$$

for all exp. stable rational functions $r(s)$. Moreover, since $n(s)$ and $d(s)$ are coprime it follows that all exp. stable rational solutions of (1.6) are of this form [15],[18]. On the other hand, a particular solution of (1.4) may be obtained by multiplying (1.2) by $k(s)$, which yields

$$y^p(s)=k(s)v(s) \quad \text{and} \quad x^p(s)=k(s)u(s). \quad (1.7)$$

As such, if we let $r(s)$ vary over the set of exp. stable rational functions and $k(s)$ vary over the set of miniphase functions we obtain a complete parameterization of the stabilizing compensators for our feedback system in the form

$$c(s)=\frac{x(s)}{y(s)}=\frac{[k(s)u(s)-r(s)d(s)]}{[k(s)v(s)+r(s)n(s)]}=\frac{[u(s)-w(s)d(s)]}{[v(s)+w(s)n(s)]} \quad (1.8)$$

where $w(s)=r(s)/k(s)$ ranges over the exp. stable rational functions.

A comparison of (1.8) with the class of stabilizing compensators derived by Youla, Bongiorno, and Jabr [24],[25],[29] will reveal that the two results differ only in that our $u(s)$, $v(s)$, $n(s)$, and $d(s)$ are exp. stable rational functions while theirs are polynomials.[2] Unlike their analytic derivation, however, the above result was obtained via elementary algebraic operations. Indeed, the only properties of the exp. stable rational functions employed are their closure under addition and multiplication together with the fact that the identity is an exp. stable rational function, i.e., the exp. stable rational functions form a ring with identity. As such, if the exp. stable rational functions of the above derivation were to be replaced by any prescribed ring of single-input single-output systems, (1.8) would yield a complete characterization of the compensators which would "place" the feedback system in that ring. If one works with a ring of rational functions with poles in a prescribed region a solution of the pole placement problem is obtained [18], whereas, if one chooses to work with stable transcendental functions a solution to the stabilization problem for distributed systems is obtained [7],[8] etc. Indeed, with minor modifications the derivation can be extended to noncommutative rings thereby including multivariate and time-varying systems. In each case, a simple solution to a fundamental problem of feedback system design is obtained by virtue of choosing a model for the given plant which is matched to the ultimate goal of the design problem. In particular, if we desire to design a feedback system which lies in a prescribed ring of operators we model the plant as a quotient of operators from that ring.

Consistent with the above philosophy the following section of the paper is devoted to the formulation of an axiomatic theory of fractional system representation. Here, a given system is modeled as a quotient of two operators lying in a prescribed ring H. The corresponding feedback system analysis and synthesis problems are then studied in the succeeding sections. In particular, Section III is devoted to the problem of determining whether or not a feedback system lies in H given that its plant is represented as a quotient of systems from H while Section IV is devoted to the problem of characterizing those compensators which will "place" the feedback system in H. The resultant axiomatic theory of feedback system design is applicable to multivariate, time-varying, distrib-

[1] See the axiomatic derivation of Section III for the details.

[2] From a computational point of view, it is more convenient to represent rational functions as ratios of polynomials, as per Youla *et al.*

TABLE I
EXAMPLES OF THE AXIOMATIC SYSTEM $\{G,H,I,J\}$

G	R(s)	$R_p(s)$	$R(s)^{n\times n}$	$R_p(s)^{n\times n}$	$\hat{B}(\sigma_0)$	$\hat{B}(\sigma)^{n\times n}$	$L_\infty(R)$	B(H)
H	R[s]	$R(\sigma_0)$	$R[s]^{n\times n}$	$R(\sigma_0)^{n\times n}$	$\hat{A}_-(\sigma_0)$	$\hat{A}_-(\sigma_0)^{n\times n}$	$H_\infty(R)$	C(H)
I	$R[s]\neq 0$	$R^{-1}(\sigma_0)$	$M\in R[s]^{n\times n}$ s.t. $\lvert M(s)\rvert \neq 0$	$M\in R(\sigma_0)^{n\times n}$ s.t. $\lvert M(s)\rvert \in R^{-1}(\sigma_0)$	$\hat{A}_-^{-1}(\sigma_0)$	$M\in \hat{A}_-(\sigma_0)^{n\times n}$ s.t. $\lvert M(s)\rvert \in \hat{A}_-^{-1}(\sigma_0)$	$m\in H_\infty(R)$ s.t. $\inf_s \lvert\lvert m(j\omega)\rvert\rvert >0$	$C_0(H)$
J	$m\in R[s]$ s.t. $m(s)=c\neq 0$	$m\in R^{-1}(\sigma_0)$ s.t. $m(s)\neq 0$ for $s\in C_{\sigma_0}$	$M\in R[s]^{n\times n}$ s.t. $\lvert M(s)\rvert \neq 0$ for $s\in C_{\sigma_0}$	$M\in R(\sigma_0)^{n\times n}$ s.t. $\lvert M(s)\rvert \in R^{-1}(\sigma_0)$ & $\lvert M(s)\rvert \neq 0$ for $s\in C_{\sigma_0}$	$m\in \hat{A}^{-1}(\sigma_0)$ s.t. $m(s)\neq 0$ for $s\in C_{\sigma_0}$	$M\in \hat{A}_-(\sigma_0)^{n\times n}$ s.t. $\lvert M(s)\rvert \in \hat{A}_-^{-1}(\sigma_0)$ $\lvert M(s)\rvert \neq 0$ for $s\in C_{\sigma_0}$	$m\in H_\infty(R)$ s.t. $\inf_s \lvert\lvert m(j\omega)\rvert\rvert >0$ & m is outer	CC(H)
Ref.	19	31	19	15,18	4, 31	6,7,8	12	11,15

R(s) = rational functions with real coefficients
$R_p(s)$ = proper rational functions with real coefficients
$X^{n\times n}$ = n by n matrices of elements in X.
A = distributions of the form $g(t)+\sum_{i=0}^{\infty} g_i\delta(t-t_i)$ where $g(t)$ is an integrable function s.t. $g(t)=0$ for $t<0$; g_i is a summable sequence and $0=t_0<t_1<t_2<\cdots$.
$\hat{A}_-(\sigma_0)$ = Laplace transforms of distributions g such that $g(t)e^{-\sigma_1 t}$ is in A for some $\sigma_1<\sigma_0$
$\hat{A}_-^{-1}(\sigma_0)$ = multiplicative subset of $\hat{A}_-(\sigma_0)$ consisting of elements bounded away from zero at ∞.
$\hat{B}(\sigma_0)$ = quotients of elements of the form m/n where $m\in \hat{A}_-(\sigma_0)$ and $n\in \hat{A}_-^{-1}(\sigma_0)$.
R[s] = polynomials with real coefficients
C_{σ_0} = complex numbers with real part greater than or equal to σ_0
$R(\sigma_0)$ = proper rational functions with real coficients which are analytic in C_{σ_0}
$R^{-1}(\sigma_0)$ = proper rational functions with real coefficients which are analytic in C_{σ_0} and nonzero at ∞
B(H) = bounded linear operators on a Hilbert Space H.
C(H) = causal bounded linear operators on a Hilbert space H.
$C_0(H)$ = causal bounded linear operators with a bounded inverse on a Hilbert space H.
CC(H) = causal bounded linear operators with a causal bounded inverse on a Hilbert space H.
$L_\infty(R)$ = essentially bounded Lebesque measurable functions defined on R.
$H_\infty(R)$ = the Hardy space of essentially bounded Lebesque measurable functions defined on R which have an analytic extension into C_{0_+}

uted, and some multidimensional systems and includes the stabilization, pole placement, and feedforward design problems. Several of these applications are illustrated by the examples of Section V. In the final section of the paper a partial generalization of the theory to nonlinear systems is described. This follows the algebraic pattern established in the linear case but is formulated in terms of a left-distributive ring to model the properties of a nonlinear system [23].

II. AXIOMATIC THEORY

Table I displays several examples of the axiomatic system developed below. Reference to it will help in visualizing the breadth and significance of the theory. Additional examples also appear in Section V.

Let G be a (not necessarily commutative) ring with identity and let H be a subring of G which includes the identity. The feedback system and its subsystems will be represented by operators which are elements of G. The compensator will be chosen so that the overall system will be represented by an operator in the subring H.

We define two multiplicative subsets [2],[27] of H,

$$I=\{h\in H \mid h^{-1}\in G\}, \tag{2.1}$$

i.e., I is the set of elements of H which have an inverse in G;

$$J=\{h\in H \mid h^{-1}\in H\}, \tag{2.2}$$

i.e., J is the *subgroup* of H consisting of all invertible elements of H. Note that

$$J\subset I\subset H\subset G. \tag{2.3}$$

Given the above structure we say that a system $g\in G$ has a *right fractional representation* in $\{G,H,I,J\}$ if there exist $n_r\in H$ and $d_r\in I$ such that $g=n_r d_r^{-1}$. Furthermore, we say that the pair $(n_r,d_r)\in H\times H$ is *right coprime* if there exist u_r and v_r in H such that

$$u_r n_r+v_r d_r=1. \tag{2.4}$$

The right fractional representation $n_r d_r^{-1}$ in $\{G,H,I,J\}$ is said to be *right coprime* if the pair (n_r,d_r) is right coprime.

The relationship between our concept of coprimeness and the usual common factor criterion for coprimeness [28] is given by the following properties.

Property 1: Let the pair $(n_r,d_r)\in H\times H$ be right coprime. Let n_r and d_r have a common right factor $r\in H$, i.e., $n_r=x_r r$, $d_r=y_r r$ for some $x_r\in H$ and $y_r\in H$. Then r has a *left inverse* in H.

Proof: Substitute the assumed factorizations of n_r and d_r into (2.4) and obtain

$$u_r n_r+v_r d_r=(u_r x_r+v_r y_r)r=1. \tag{2.5}$$

Since H is a ring, $u_r x_r+v_r y_r\in H$. From (2.5) it follows that $r^{-L}=u_r x_r+v_r y_r$ is a left-inverse of r. ■

Property 2: Let $g=n_r d_r^{-1}$ be a right coprime fractional representation of g in $\{G,H,I,J\}$. Let $g=x_r y_r^{-1}$ be a second (not necessarily coprime) right fractional representation of g in $\{G,H,I,J\}$. Then there exists an r in H such that

$$x_r=n_r r \quad \text{and} \quad y_r=d_r r. \tag{2.6}$$

Proof: Given the two factorizations of g, let $r = d_r^{-1}y_r$; hence $r \in G$. Then

$$y_r = d_r r \tag{2.7}$$

and, performing calculations in the ring G, we obtain

$$x_r = gy_r = (n_r d_r^{-1})y_r = n_r(d_r^{-1}y_r) = n_r r. \tag{2.8}$$

From (2.7) and (2.8), r is a common right factor of x_r and y_r. To show that $r \in H$, consider

$$r = d_r^{-1}y_r = (u_r n_r + v_r d_r)d_r^{-1}y_r = u_r n_r d_r^{-1}y_r + v_r y_r$$

$$= u_r g y_r + v_r y_r = u_r x_r + v_r y_r \in H \tag{2.9}$$

where we used the equality $g = x_r y_r^{-1} = n_r d_r^{-1}$ to derive (2.9). ■

Although G is, in general, a noncommutative ring, the entire theory developed above for right fractional representations can be replicated for left fractional representations. In particular, we say that $g \in G$ has a *left fractional representation* in $\{G,H,I,J\}$ if there exist $n_l \in H$ and $d_l \in I$ such that $g = d_l^{-1}n_l$. Furthermore we say that the pair $(n_l, d_l) \in H \times H$ is *left coprime* if there exist u_l and v_l in H such that

$$n_l u_l + d_l v_l = 1. \tag{2.10}$$

The left fractional representation $d_l^{-1}n_l$ is said to be *left coprime* if the pair (n_l, d_l) is left coprime. With these definitions the existence of a *common left factor* for a left fractional representations of g is characterized by the following properties.

Property 1': Let the pair (n_l, d_l) be left coprime. Let n_l and d_l have a common left factor l in H, i.e., $n_l = lx_l$, $d_l = ly_l$ for some $x_l \in H$ and $y_l \in H$. Then l has a *right inverse* $\in H$.

Property 2': Let $g = d_l^{-1}n_l$ be a left coprime fractional representation of g in $\{G,H,I,J\}$. Let $g = y_l^{-1}x_l$ be a second (not necessarily coprime) left fractional representation of g in $\{G,H,I,J\}$. Then there exists an l in H such that

$$x_l = ln_l \quad \text{and} \quad y_l = ld_l. \tag{2.11}$$

The above properties of a coprime fractional representation have all been derived under the assumption that such a representation exists. Of course, if G denotes the rational matrices and H denotes the polynomial matrices the existence of a coprime representation is implied by classical analysis [16],[19]. Indeed, the classical analysis readily extends to the case where H is taken to be the exp. stable rational matrices or the ring of proper rational matrices with poles in a prescribed region [18]. On the other hand for multidimensional [26], distributed [4],[8], and time-varying systems [11],[15] there is no assurance that an arbitrary $g \in G$ will admit a fractional representation nor even that the set of $g \in G$ which admit such a representation will be a linear space. Moreover, all g's which admit a fractional representation may not admit a coprime fractional representation [26]. In general, the set of $g \in G$ which admit a fractional representation in $\{G,H,I,J\}$ will form a subring of G if and only if the Ore condition[3] is satisfied while criteria for coprimeness have been formulated in various special cases though no general theory exists [1],[4],[26]. The standard condition for the existence of fractional representations which are coprime in the sense of (2.4) is that H be a right principal ideal domain.

Reference to Table I shows that in applications it is important to have conditions under which g will be in H and these conditions should be expressed in terms of its fractional representation.

Property 3: Let $g = n_r d_r^{-1}$ with $n_r \in H$ and $d_r \in I$.

a) If $d_r \in J$, then $g \in H$.

b) If $g = n_r d_r^{-1}$ is a *right coprime fractional representation* of g in $\{G,H,I,J\}$, then $g \in H$ implies that $d_r \in J$.

Proof:

a) We have $d_r \in J$; hence by (2.2), $d_r^{-1} \in H$ and thus $n_r d_r^{-1} = g \in H$.

b) We have $g \in H$. Furthermore, $n_r = gd_r$, $d_r = 1d_r$ implies that d_r is a right common factor of n_r and d_r; hence by Property 1, d_r has a left inverse in H. But $d_r \in I$ by assumption, so d_r^{-1} exists and is an element of G; thus $d_r^{-1} = d_r^{-L} \in H$; hence, by (2.2), $d_r \in J$. ■

Property 3': Let $g = d_l^{-1}n_l$ with $n_l \in H$ and $d_l \in I$.

a) If $d_l \in J$, then $g \in H$.

b) If $g = d_l^{-1}n_l$ is a *left coprime fractional representation* of g in $\{G,H,I,J\}$, then $g \in H$ implies that $d_l \in J$.

Property 4: Let $g = n_r d^{-1}n_l$ where n_r, $n_l \in H$, and $d \in I$.

a) If $d \in J$, then $g \in H$.

b) Let, in addition, $n_r d^{-1}$ be a right coprime fractional representation in $\{G,H,I,J\}$ and $d^{-1}n_l$ be a left coprime fractional representation in $\{G,H,I,J\}$; then $g \in H$ implies that $d \in J$.

Proof:

a) By assumption, $d \in J$; hence $d^{-1} \in H$. So $g = n_r d^{-1}n_l \in H$.

b) Since $d^{-1}n_l$ is a left coprime fractional representation there exist $u_l, v_l \in H$ such that

$$n_l u_l + dv_l = 1, \tag{2.12}$$

thus,

$$n_r d^{-1} = n_r d^{-1}(n_l u_l + dv_l) = n_r d^{-1}n_l u_l + n_r v_l = gu_l + n_r v_l. \tag{2.13}$$

Now $g \in H$ hence (2.13) gives $n_r d^{-1} \in H$. By Property 3, $n_r d^{-1} \in H$ together with the fact that the pair (n_r, d) is right coprime implies $d \in J$. ■

III. Analysis

To start with consider the feedback system Σ_p of Fig. 2. Suppose that the plant is described by a right coprime fractional representation $p = n_r d_r^{-1}$ in $\{G,H,I,J\}$. The

[3] $\{G,H,I,J\}$ satisfies the Ore condition for right fractional representations if, whenever $g \in G$ admits a left fractional representation it also admits a right fractional representation and vice versa.[2]

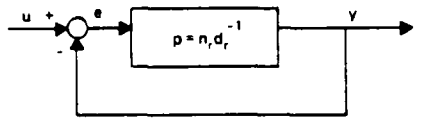


Fig. 2. Unity gain negative feedback system.

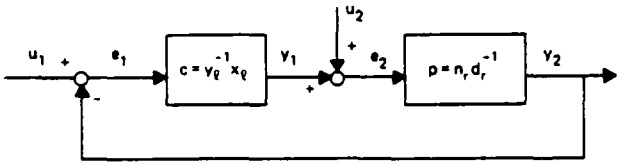


Fig. 3. Feedback system with plant and compensator.

closed-loop dynamics of Σ_p are described by the maps

$$h_{eu}: u \mapsto e\ ; \qquad h_{eu} = (1+p)^{-1} = d_r(d_r+n_r)^{-1} \tag{3.1}$$

$$h_{yu}: u \mapsto y\ ; \qquad h_{yu} = p(1+p)^{-1} = n_r(d_r+n_r)^{-1}. \tag{3.2}$$

Note that

$$h_{eu} + h_{yu} = 1. \tag{3.3}$$

We say that Σ_p is *well defined in* G, (H, respectively), if $h_{eu} \in G$, (H, respectively).

Note that the pairs (n_r, d_r+n_r) and (d_r, d_r+n_r) are right coprime; indeed, the right coprimeness of (n_r, d_r) implies (2.4), hence

$$(u_r - v_r)n_r + v_r(d_r+n_r) = 1 \tag{3.4}$$

while

$$(v_r - u_r)d_r + u_r(d_r+n_r) = 1. \tag{3.5}$$

Theorem 1: Consider the feedback system Σ_p of Fig. 2.

a) Let $p = n_r d_r^{-1}$ be a fractional representation in $\{G,H,I,J\}$ of the element $p \in G$; then Σ_p *is well defined in* G if and only if $d_r + n_r \in I$.

b) Let $p = n_r d_r^{-1}$ be a *right coprime* fractional representation in $\{G,H,I,J\}$ of the element $p \in G$; then Σ_p *is well defined in* H if and only if $d_r + n_r \in J$.

Proof: a) $\Rightarrow$. $h_{eu} \in G$ and $d_r \in I$ imply

$$d_r^{-1} h_{eu} = d_r^{-1}(1+p)^{-1} = d_r^{-1} d_r (d_r+n_r)^{-1} = (d_r+n_r)^{-1} \in G. \tag{3.6}$$

Now $d_r \in I \subset H$ and $n_r \in H$, so $d_r + n_r \in H$. This together with (3.6) implies $d_r + n_r \in I$.

a) $\Leftarrow$. $d_r + n_r \in I$ implies $(d_r+n_r)^{-1} \in G$; hence $h_{eu} = d_r(d_r+n_r)^{-1} \in G$.

b) Follows from Property 3, together with (3.4) and (3.5). ■

Of course, a similar theorem holds for left factorizations.

We now consider the feedback system Σ of Fig. 3 where the plant p is preceded by a compensator c; p and c belong to G and are specified by their coprime fractional representation in $\{G,H,I,J\}$ $n_r d_r^{-1}$ and $y_l^{-1}x_l$, respectively.

To describe the feedback system Σ we consider the map $h_{eu}: (u_1,u_2) \mapsto (e_1,e_2)$. Simple calculations give

$$h_{eu} = \begin{bmatrix} h_{e_1u_1} & h_{e_1u_2} \\ h_{e_2u_1} & h_{e_2u_2} \end{bmatrix} = \begin{bmatrix} (1+pc)^{-1} & -p(1+cp)^{-1} \\ c(1+pc)^{-1} & (1+cp)^{-1} \end{bmatrix}. \tag{3.7}$$

Now let $h_{yu}: (u_1,u_2) \mapsto (y_1,y_2)$. Using the summing node equations it is easy to see that

$$h_{yu} = K(h_{eu} - 1) \quad \text{and} \quad h_{eu} = 1 - Kh_{yu} \tag{3.8}$$

where K is the symplectic matrix

$$K = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{3.9}$$

It is well known that in the case of multivariable rational matrices, one has to consider the four submatrices of h_{eu} in (3.8) because examples show that any one of the submatrices may be unstable while the remaining ones are stable. (For detailed examples, see [30].) Let us calculate

$$\begin{aligned} h_{e_1u_1} &= (1+pc)^{-1} = 1 - pc(1+pc)^{-1} \\ &= 1 - p(1+cp)^{-1}c \\ &= 1 - p\left[y_l^{-1}(y_l d_r + x_l n_r) d_r^{-1}\right]^{-1} c \\ &= 1 - n_r(y_l d_r + x_l n_r)^{-1} x_l \end{aligned} \tag{3.10}$$

$$\begin{aligned} h_{e_2u_1} &= c(1+pc)^{-1} = (1+cp)^{-1}c \\ &= d_r(y_l d_r + x_l n_r)^{-1} x_l \end{aligned} \tag{3.11}$$

$$\begin{aligned} h_{e_2u_2} &= (1+cp)^{-1} = \left(1 + y_l^{-1}x_l n_r d_r^{-1}\right)^{-1} \\ &= \left[y_l^{-1}(y_l d_r + x_l n_r) d_r^{-1}\right]^{-1} \\ &= d_r(y_l d_r + x_l n_r)^{-1} y_l \end{aligned} \tag{3.12}$$

$$h_{e_1u_2} = -p(1+cp)^{-1} = -n_r(y_l d_r + x_l n_r)^{-1} y_l. \tag{3.13}$$

We say that Σ is *well defined in* G, (H, respectively) if and only if each entry of h_{eu} defined in (3.8) belongs to G, (H, respectively).

Theorem 2: Consider the feedback system Σ of Fig. 3. Let $n_r d_r^{-1}$ and $y_l^{-1}x_l$ be a right and left fractional representations of p and c in $\{G,H,I,J\}$.

a) If $y_l d_r + x_l n_r \in I$, then Σ is well defined in G.

b) If $y_l d_r + x_l n_r \in J$, then Σ is will defined in H.

c) If $h_{e_2u_2} \in G$, then $y_l d_r + x_l n_r \in I$ hence if Σ is well defined in G, then $y_l d_r + x_l n_r \in I$.

d) Assume, in addition, that $n_r(y_l d_r)^{-1}$ and $(y_l d_r)^{-1}x_l$ are right coprime and left coprime fractional representation, respectively; then $h_{e_1u_1} \in H$ implies that $y_l d_r + x_l n_r \in J$, and hence, if Σ is well defined in H, then $y_l d_r + x_l n_r \in J$.

Proof: a) and b). If $y_l d_r + x_l n_r \in I$, (J, respectively), then by the definition (2.1) of I, [(2.2) of J, respectively], the formulas (3.10)–(3.13), and the closure of the ring G, (H, respectively), the conclusion follows.

c) If $h_{e_2u_2} \in G$, then so is $d_r^{-1} h_{e_2u_2} y_l^{-1}$ since $d_r \in I$ and $y_l \in I$. Now,

$$\begin{aligned} d_r^{-1} h_{e_2u_2} y_l^{-1} &= d_r^{-1}(1+cp)^{-1} y_l^{-1} \\ &= d_r^{-1}\left(1 + y_l^{-1} x_l n_r d_r^{-1}\right)^{-1} y_l^{-1} \\ &= d_r^{-1}\left[y_l^{-1}(y_l d_r + x_l n_r) d_r^{-1} \right] y_l^{-1} = (y_l d_r + x_l n_r)^{-1} \end{aligned} \tag{3.14}$$

hence the fact that $h_{e_2u_2} \in G$ implies that $(y_l d_r + x_l n_r)^{-1} \in G$ and thus $(y_l d_r + x_l n_r) \in I$.

d) First we prove that the pair $(n_r, y_l d_r + x_l n_r)$ is right coprime. Since $(n_r, y_l d_r)$ is right coprime, there exists $\tilde{u}_r$ and $\tilde{v}_r \in H$ such that

$$\tilde{u}_r n_r + \tilde{v}_r y_l d_r = 1; \tag{3.15}$$

hence

$$(\tilde{u}_r - \tilde{v}_r x_l) n_r + \tilde{v}_r (y_l d_r + x_l n_r) = 1 \tag{3.16}$$

and the claim is established. Similarly, we show that $(y_l d_r + x_l n_r, x_l)$ is left coprime. Now consider

$$h_{e_1u_1} = 1 - n_r (y_l d_r + x_l n_r)^{-1} x_l. \tag{3.17}$$

By assumption, $h_{e_1u_1} \in H$; then the special assumption of d) and Property 4 imply that $y_l d_r + x_l n_r \in J$. This completes the proof. ■

Note, the special assumptions used in d) to the effect that $n_r (y_l d_r)^{-1}$ is right coprime and $(y_l d_r)^{-1} x_l$ is left coprime, imply, in some sense, that p and c have no common factors. More precisely, since J serves as the group of units in our theory these conditions imply that any common factors of p and c must lie in J.

IV. Design

Consistent with our approach of matching the plant model to the goal of the given feedback system design problem the present section is devoted to the problem of characterizing the set of compensators which will "place" a feedback system in a prescribed ring H given that both the plant and compensator are modeled by fractional representations in $\{G, H, I, J,\}$.

Theorem 3: For the feedback system Σ of Fig. 3, let the plant p have a right coprime and a left coprime fractional representation $p = n_r d_r^{-1} = d_l^{-1} n_l$ in $\{G, H, I, J\}$. Let u_r and v_r both in H be such that (2.4) holds. Then for any $w \in H$ such that $w n_l + v_r \in I$, the compensator

$$c = (w n_l + v_r)^{-1} (-w d_l + u_r) \in G \tag{4.1}$$

results in a feedback system Σ well defined in H. For such a compensator, $h_{e_1u_1} \in H^{2\times 2}$ and

$$h_{e_1u_1} = \begin{bmatrix} 1 - n_r(-w d_l + u_r) & -n_r(w n_l + v_r) \\ d_r(-w d_l + u_r) & d_r(w n_l + v_r) \end{bmatrix}. \tag{4.2}$$

Conversely, if Σ is well defined in H and if the compensator $c = y_l^{-1} x_l$ is such that $(n_r, y_l d_r)$ and $(y_l d_r, x_l)$ are right coprime and left coprime respectively, then c is given by expression (4.1).

Proof:

Step 1: Choose any $k \in J$, (hence $k^{-1} \in H$), and solve for y_l and $x_l \in H$ the equation

$$y_l d_r + x_l n_r = k. \tag{4.3}$$

Observe that if (y_l, x_l) is *any solution in H* of (4.3), then

$$k^{-1}(y_l d_r) + k^{-1}(x_l n_r) = 1 \tag{4.4}$$

and

$$(y_l d_r) k^{-1} + (x_l n_r) k^{-1} = 1, \tag{4.5}$$

hence, $(n_r, y_l d_r)$ is right coprime and $(y_l d_r, x_l)$ is left coprime. Thus, the assumptions of Theorem 2, part d) holds for *any* solution of (4.3).

Step 2: Obtain all solutions of the homogeneous equation

$$y_l^h d_r + x_l^h n_r = 0. \tag{4.6}$$

Since $p = n_r d_r^{-1} = d_l^{-1} n_l$, direct calculation shows that for any $r \in H$,

$$y_l^h = r n_l \qquad x_l^h = -r d_l \tag{4.7}$$

are solutions of (4.6).

It remains to show that all solutions of (4.6) are of the form (4.7); so we *assume* that y_l^h and $x_l^h \in H$ and satisfy (4.6). Let $r = -x_l^h d_l^{-1}$; hence

$$x_l^h = r d_l. \tag{4.8}$$

Now using (4.6)

$$\begin{aligned} y_l^h &= y_l^h d_r d_r^{-1} = -x_l^h n_r d_r^{-1} = -x_l^h p \\ &= -x_l^h d_l^{-1} n_l = r n_l. \end{aligned} \tag{4.9}$$

Equations (4.8) and (4.9) show that any solution of (4.6) has the form of (4.7); it remains, however, to show that $r \in H$,

$$\begin{aligned} r &= -x_l^h d_l^{-1} = -x_l^h d_l^{-1}(d_l v_l + n_l u_l) \\ &= -x_l^h v_l - x_l^h d_l^{-1} n_l u_l = -x_l^h v_l + y_l^h u_l \in H. \end{aligned} \tag{4.10}$$

Step 3: Obtain a particular solution of (4.3). From the right coprimeness condition for (n_r, d_r),

$$k v_r d_r + k u_r n_r = k \tag{4.11}$$

hence

$$y_l^p = k v_r, \quad x_l^p = k u_r. \tag{4.12}$$

Hence any solution of (4.3) is of the form

$$\begin{aligned} y_l &= r n_l + k v_r \\ x_l &= -r d_l + k u_r \qquad \text{for some } r \in H \end{aligned} \tag{4.13}$$

and for any such solution $(n_r, y_l d_r)$ is right coprime and $(y_l d_r, x_l)$ is left coprime.

Step 4: Consider the condition

$$r \in H \text{ and } k \in J \qquad \text{such that } rn_l + kv_r \in I \quad (4.14)$$

or equivalently, if we set $w = k^{-1}r \in H$,

$$w \in H \qquad \text{such that } wn_l + v_r \in I. \quad (4.15)$$

If (4.15) holds,

$$c = (wn_l + v_r)^{-1}(-wd_l + u_r) \in G \quad (4.16)$$

is a compensator in G which can also be written as [see (4.13)]

$$c = (rn_l + kv_r)^{-1}(-rd_l + ku_r). \quad (4.17)$$

If we let $y_l = rn_l + kv_r$ and $x_l = -rd_l + ku_r$, then, by (4.17), $c = y_l^{-1}x$ and, by calculation, we verify that (4.3) holds. Thus for any such compensator, by Theorem 2, the feedback system Σ is well defined in H.

Step 5: Conversely consider a feedback system well defined in H with a compensator $c = y_l^{-1}x_l$ such that $(n_r, y_l d_r)$ and $(y_l d_r, x_l)$ are right coprime and left coprime, respectively. By Theorem 2, (4.3) holds for some $k \in J$, hence by the analysis above, c is also given by (4.1) for some $w \in H$ such that $wn_l + v_r \in I$. The proof is thus complete. ■

The theorem yields a complete parameterization of all possible controllers which will place a plant in H given the existence of:

1) right and left coprime fractional representations of p and

2) a w in H for which $(wn_l + v_r)$ is in I.

In the multivariable case where p is a square matrix whose elements are proper rational functions it is well known that p has left and right coprime fractional representations [19]. In order to obtain a *proper* controller one has to choose w in (4.1) so that $\det[w(s)n_l(s) + v_r(s)] \neq 0$ at infinity. Methods for obtaining such a proper stabilizing controller have been reported in [32] and [33]. Alternatively, one can verify the existence of such a w in our algebraic setting by invoking the fact that n_l and d_r are right coprime and applying linear algebraic arguments thereto. Of course, these arguments apply to distributed systems as well as lumped systems using the formulation of [7] and [8].

In the most general ring theoretic setting neither right nor left coprime fractional representations of p, nor a w such that $(wn + v_r)$ is in I, are assured to exist. At present, the only known counterexample to the latter is, however, in the ring of integers which is of no system theoretic interest.

Conditions 1) and 2) have been conjectured to be both necessary and sufficient conditions for the existence of a compensator, c, which places the feedback system in H [3]. In fact, if c places the feedback system in H, then from (3.7) we obtain left and right fractional representations

$$p = (-h_{e_1u_2})(h_{e_2u_2})^{-1} = (h_{e_1u_1})^{-1}(-h_{e_1u_2}). \quad (4.18)$$

Note that there is no guarantee that these fractional representations are coprime. These representations are, however, coprime when the compensator is in H. Indeed, in that case they satisfy a stronger condition which completely characterizes those plants which can be placed in H by a compensator in H. For an early analogous result, see [10, pp. 85–87].

Corollary 1: For the feedback system Σ of Fig. 3 there exists a c in H which places the feedback system in H if and only if p admits left and right fractional representations $p = d_l^{-1}n_l = n_r d_r^{-1}$ such that n_r is a right factor of $1 - d_r$ and n_l is a left factor of $1 - d_l$.

Proof: If the feedback system is placed in H by a c in H it admits the fractional representations of (4.18). By calculation [see (3.7)]

$$h_{e_2u_2} - ch_{e_1u_2} = 1 \quad (4.19)$$

and

$$h_{e_1u_1} - h_{e_1u_2}c = 1 \quad (4.20)$$

which verifies their coprimeness since c is in H. Moreover, upon rearranging the terms in (4.19) and (4.20) the conditions of the corollary follow. Conversely, if fractional representations exist which satisfy the conditions of the corollary there exists u_r in H such that

$$u_r n_r = d_r = 1 \quad (4.21)$$

(equivalently $p = n_r d_r^{-1}$ is a right coprime fractional representation with $v_r = 1$). Now, by using this right fractional representation in (4.1) (with any left coprime fractional representation) and $w = 0$ we obtain a compensator $c = u_r$ in H, which places the feedback system in H. ■

V. Examples

Example 1: A Single Variate Servomechanism Problem[4]

Here G is the ring of proper rational functions and H is subring of functions analytic in $\mathrm{Re}\, s \geq -1$. Consider the problem of designing a compensator for the unstable plant $p(s) = (s+1)/(s^2-4)$ which will simultaneously place the poles of the feedback system in the region, $\mathrm{Re}(s) < -1$, and cause the system to asymptotically track a step input. Since our transfer functions are commutative we may adopt common right and left fractional representation for $p(s)$. In particular,

$$p(s) = \frac{(s+1)}{(s^2-4)} = \left[\frac{(s+1)}{(s+2)^2}\right]\left[\frac{(s-2)}{(s+2)}\right]^{-1} = n(s)d(s)^{-1} \quad (5.1)$$

while

$$\left[\frac{16}{3}\right]\left[\frac{(s+1)}{(s+2)^2}\right] + \left[\frac{(s+2/3)}{(s+2)}\right]\left[\frac{(s-2)}{(s+2)}\right] = u(s)n(s) + v(s)d(s) = 1. \quad (5.2)$$

[4]The purpose of this example is merely to give a simple illustration of the theory. In this situation, a much more highly developed theory is available in [29].

Here, each of the four rational functions, $n(s)$, $d(s)$, $u(s)$, and $v(s)$, lie in the ring of operators with poles in the region $\mathrm{Re}(s) < -1$ and hence the set of all compensators which will place the feedback system in this ring is given by Theorem 3 with $w(s)$ also in the ring. Moreover, for an arbitrary $w(s)$ the input–output mapping for the resultant feedback system will take the form

$$h_{y_2u_1}(s) = -\left[\frac{(s+1)(s-2)}{(s+2)^3}\right]w(s) + \left[\frac{16(s+1)}{3(s+2)^2}\right]$$
$$= -n(s)d(s)w(s) + n(s)u(s). \tag{5.3}$$

By the final value theorem the feedback system will asymptotically track a step input if and only if $h_{y_2u_1}(0) = 1$ (equivalently $c(s)$ has a pole at zero). As such, to simultaneously place the poles of the feedback system in the region, $\mathrm{Re}(s) < -1$, and cause the feedback system to asymptotically track a step input we must find a $w(s)$ with poles in this region such that $h_{y_2u_1}(0) = 1$. Evaluating (5.3) at $s = 0$ and setting it equal to one yields

$$h_{y_2u_1}(0) = \frac{1}{4}w(0) + \frac{4}{3} = 1, \tag{5.4}$$

implying that $w(0) = -4/3$. As such, the simplest $w(s)$ which will achieve our simultaneous goals is the constant $w(s) = -4/3$ whose poles are trivially in the prescribed region. Adopting this $w(s)$, a little algebra with the expressions of Theorem 3 will reveal that the required compensator takes the form

$$c(s) = \frac{(20s+24)(s+2)}{(3s+4)s} \tag{5.5}$$

while the input–output mapping for the feedback system takes the form

$$h_{y_2u_1} = \frac{(s+1)(20s+24)}{3(s+2)^3}. \tag{5.6}$$

Clearly, $c(s)$ has the required pole at zero (for $h_{y_2u_1}(0) = 1$), although it is by no means obvious that this quasi-stable compensator will transfer the unstable poles of $p(s)$ to the prescribed region. Indeed, this illustrates the underlying power of the proposed design technique in that when one designs the system in terms of $w(s)$ rather than $c(s)$ the pole placement or stabilization process is automatically resolved by working with a $w(s)$ whose poles lie in the prescribed region while the remainder of the design process is simplified by the affine relationship between $w(s)$ and the matrices h_{ew} and h_{yu}. Finally, we note that $c(s)$ has a zero at $s = -2$ which may cancel with the pole of $p(s)$ at $s = -2$. This, however, does not contradict the coprimeness assumptions of Theorem 3 since the common factors involved lie in J which serves as the group of units in our theory. Fortunately, such common factors can never lead to an erroneous design since by assumption the poles and zeros of the rational functions in J lie in the prescribed region. As such, any cancellations which may take place are benign.

Since the previous compensator design was achieved with an especially simple $w(s)$ let us add an additional constraint to the problem by requiring that $h_{y_2u_1}(s)$ have zeros at $\pm j$ (so that the system will be insensitive to a noise source at that frequency). Now, from (5.4) it follows that the above design is the only compensator which will make $h_{y_2u_1}(0) = 1$ with a constant $w(s)$; hence to satisfy this additional design constraint we will work with the first order $w(s)$ in the form

$$w(s) = \frac{as-4}{bs+3}. \tag{5.7}$$

Here, by specifying the zeroth-order coefficients of $w(s)$ we assure that $w(0) = -4/3$ while we are left with the parameters a and b to create the required zeros. Of course, to achieve our stability condition we must have $-3/b < -1$. Substituting the $w(s)$ of (5.7) into (5.3) yields

$$h_{y_2u_1}(s) = \frac{(s+1)\left[(16b-3a)s^2 + (60+6a+32b)s + 72\right]}{3(s+2)^3(bs+3)}. \tag{5.8}$$

To obtain the desired zeros at $s = \pm j$ the equation

$$\left[(16b-3a)s^2 + (60+6a+32b)s + 72\right] = k\left[s^2+1\right] \tag{5.9}$$

must be satisfied. Now, this represents three linear equations in three unknowns and has the unique solution

$$a = -17, \quad b = \frac{21}{16}, \text{ and } k = 72. \tag{5.10}$$

Moreover, $-3/b = -16/7 < -1$; hence this choice of $w(s)$ will also assure the prescribed degree of stabilization. As such, we take

$$w(s) = \frac{-(17s+4)}{(21s/16+3)} = \frac{-(272s+64)}{(21s+48)} \tag{5.11}$$

which yields

$$c(s) = \frac{128(s+2)(s^2+1)}{(7s^2-56s-60)s} \tag{5.12}$$

and

$$h_{y_2u_1}(s) = \frac{384(s+1)(s^2+1)}{(s+2)^3(21s+48)} \tag{5.13}$$

satisfying all of our design criteria.

Example 2: A Multivariate Lumped-Distributed Decoupling Problem

Consider the multivariate, lumped-distributed plant

$$p(s) = \begin{bmatrix} \dfrac{e^{-1/s}}{(s+1)} & \dfrac{(s-1)}{(s+1)} \\ 0 & \dfrac{1}{(s-1)} \end{bmatrix} \tag{5.14}$$

which we desire to stabilize and simultaneously decouple by feedback. For most lumped-distributed systems one can take H to be a ring of matrices whose elements lie in the algebra $\hat{\mathfrak{A}}_-(\sigma_0)$ of stable transfer functions generated by lumped elements and delays while G is a ring of matrices whose elements lie in $\hat{\mathfrak{B}}(\sigma_0)$, the algebra of quotients of elements in $\hat{\mathfrak{A}}_-(\sigma_0)$, as per Table I. In our case, however, although $e^{-1/s}$ is L_2-stable (since it is analytic on the right half-plane and bounded on the imaginary axis [10]) it has a "nasty" singularity at $s=0$ and hence does not lie in $\hat{\mathfrak{A}}(\sigma_0)$ for any $\sigma_0 \leq 0$. As such, we take H to be a ring of 2×2 matrices whose elements are transfer functions lying in the Hardy space $H_\infty(R)$ of functions which are (essentially) bounded on the $j\omega$ axis and admit an analytic extension into the right half-plane (thereby making them L_2-stable) [12]. Similarly, we let G be a ring of 2×2 matrices whose entries are transfer functions lying in the Lebesgue space $L_\infty(R)$ [12]. With this setup I becomes the set of H_∞ functions which are uniformly bounded below on the $j\omega$ axis while J is the set of H_∞ functions whose analytic extension is uniformly bounded below in the right half-plane [12]. Equivalently, J is the set of invertible outer functions in $H_\infty(R)$ [12].

Using these spaces a little algebra will reveal that $p(s)$ has the right and left coprime fractional representations in $\{G,H,I,J\}$ shown below:

$$p(s)=\begin{bmatrix} \frac{e^{-1/s}}{(s+1)} & \frac{(s-1)^2}{(s+1)^2} \\ 0 & \frac{1}{(s+1)} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{(s-1)}{(s+1)} \end{bmatrix}^{-1} = n_r(s)d_r(s)^{-1} \tag{5.15}$$

$$p(s)=\begin{bmatrix} 1 & 0 \\ 0 & \frac{(s-1)}{(s+1)} \end{bmatrix}^{-1}\begin{bmatrix} \frac{e^{-1/s}}{(s+1)} & \frac{(s-1)}{(s+1)} \\ 0 & \frac{1}{(s+1)} \end{bmatrix} = d_l(s)^{-1}n_l(s) \tag{5.16}$$

where

$$\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} \frac{e^{-1/s}}{(s+1)} & \frac{(s-1)^2}{(s+1)^2} \\ 0 & \frac{1}{(s+1)} \end{bmatrix}+\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & \frac{(s-1)}{(s+1)} \end{bmatrix} = u_r(s)n_r(s)+v_r(s)d_r(s)=1 \tag{5.17}$$

and

$$\begin{bmatrix} \frac{e^{-1/s}}{(s+1)} & \frac{(s-1)}{(s+1)} \\ 0 & \frac{1}{(s+1)} \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}+\begin{bmatrix} 1 & 0 \\ 0 & \frac{(s-1)}{(s+1)} \end{bmatrix}\begin{bmatrix} 1 & \frac{-2(s-1)}{(s+1)} \\ 0 & 1 \end{bmatrix} = n_l(s)u_l(s)+d_l(s)v_l(s)=1. \tag{5.18}$$

Upon substitution of these matrices into the expression for $h_{y_2u_1}(s)$ from Theorem 3 one obtains

$$h_{y_2u_1}(s)=-\begin{bmatrix} \frac{e^{-1/s}}{(s+1)} & \frac{(s-1)^2}{(s+1)^2} \\ 0 & \frac{1}{(s+1)} \end{bmatrix}\begin{bmatrix} w_{11}(s) & w_{12}(s) \\ w_{21}(s) & w_{22}(s) \end{bmatrix}\cdot\begin{bmatrix} 1 & 0 \\ 0 & \frac{(s-1)}{(s+1)} \end{bmatrix}+\begin{bmatrix} 0 & \frac{2(s-1)^2}{(s+1)^2} \\ 0 & \frac{2}{(s+1)} \end{bmatrix} \tag{5.19}$$

which will be stable if and only if the $w_{ij}(s)$ are stable.

Now, to decouple the system we require that

$$h^{12}_{y_2u_1}(s)=\frac{(s-1)e^{-1/s}}{(s+1)^2}w_{12}(s)+\frac{(s-1)^3}{(s+1)^3}w_{22}(s)+\frac{2(s-1)^2}{(s+1)^2}=0 \tag{5.20}$$

and

$$h^{21}_{y_2u_1}(s)=\frac{1}{(s+1)}w_{21}(s)=0. \tag{5.21}$$

Clearly, $w_{21}(s)=0$ solves (5.21). On the other hand (5.20) has numerous solutions none of which are, however, stable. As such, the system cannot be decoupled and stabilized simultaneously. Note, since our theory guarantees that all stable feedback systems with plant $p(s)$ take the form of (5.19) if we cannot find stable w's which decouple (5.19) we are assured that it is *impossible* to simultaneously stabilize and decouple $p(s)$ by feedback (using a compensator as specified in Theorem 3) and we need not consider other formulations.

Since we cannot simultaneously stablize and decouple $p(s)$ by feedback the best we can do is to try to stabilize $p(s)$ while preserving its triangularity (which will allow us to sequentially adjust its various outputs). Formally, this can be achieved by taking $w(s)=0$ which yields the input–output mapping

$$h_{y_2u_1}(s)=\begin{bmatrix} 0 & \frac{2(s-1)^2}{(s+1)^2} \\ 0 & \frac{2}{(s+1)} \end{bmatrix}. \tag{5.22}$$

Unfortunately, the first input has been rendered useless by this compensator and hence the goal of being able to sequentially tune the outputs is not achieved. On the other

hand, if we take

$$w(s)=\begin{bmatrix}1 & 0\\ 0 & 0\end{bmatrix}, \tag{5.23}$$

then

$$h_{y_2u_1}(s)=\begin{bmatrix}\dfrac{e^{-1/s}}{(s+1)} & \dfrac{2(s-1)^2}{(s+1)^2}\\ 0 & \dfrac{2}{(s+1)}\end{bmatrix} \tag{5.24}$$

which has the desired property is obtained. In particular, one can tune the second input to control the second output and then adjust the first input to simultaneously cancel out the effects of the second input on the first output and control the first output. Of course, since $w(s)$ is stable so is $h_{y_2u_1}(s)$.

Finally, we note that as we have formulated our theory one can deal only with square matrices (since rectangular matrices are not closed under multiplication). The extension to rectangular matrices is, however, straightforward [19] and yields an identical theory the details of which are left to the reader.

Example 3: A Multidimensional Image Restoration Problem

Let

$$p(z_1,z_2)=\frac{z_1+z_2}{z_1^2+z_1z_2+3} \tag{5.25}$$

denote the discrete two-dimensional transfer function for a device in a digital image processing system. Since this represents an IIR (infinite impulse response) transfer function the image processing device will tend to "smear" the image with the data observed at any one pixel distorting all other pixels at the output of the device. In an effort to reduce this "smearing" effect we would like to place the device in a feedback system whose input–output transfer function minimizes the "smearing" effect. In particular, that means that the input–output mapping for the feedback system should have an FIR (finite impulse response) transfer function with its "point-spread function" concentrated about a single point as closely as possible.

Since the FIR transfer functions are just the polynomials we let H be the ring of polynomials in two variables and G be the ring of rational functions in two variables [16]. Once again employing only a single fractional representation since these rings are commutative we obtain the coprime fractional representation

$$\begin{aligned}p(z_1,z_2)&=[z_1+z_2][z_1^2+z_1z_2+3]^{-1}\\ &=n(z_1,z_2)d(z_1,z_2)^{-1}\end{aligned} \tag{5.26}$$

where

$$\begin{aligned}&\left[-\frac{1}{3}z_1\right][z_1+z_2]+\left[\frac{1}{3}\right][z_1^2+z_1z_2+3]\\ &\quad=u(z_1,z_2)n(z_1,z_2)+v(z_1z_2)d(z_1,z_2)=1.\end{aligned} \tag{5.27}$$

As such, the set of all possible FIR transfer functions which can be obtained from $p(z_1,z_2)$ by feedback takes the form

$$\begin{aligned}h_{y_2u_1}(z_1,z_2)=&-[z_1^3+2z_1^2z_2+z_1z_2^2+3z_1+3z_2]\\ &\cdot w(z_1,z_2)-\frac{1}{3}[z_1^2+z_1z_2]\end{aligned} \tag{5.28}$$

where $w(z_1,z_2)$ is an arbitrary polynomial in two variables. Clearly, $w(z_1,z_2)$ should be low order to keep the "point-spread function" of $h_{y_2u_1}(z_1,z_2)$ as concentrated as possible. Indeed, if we take $w(z_1,z_2)=0$ we obtain

$$h_{y_2u_1}(z_1,z_2)=-\frac{1}{3}[z_1^2+z_1z_2] \tag{5.29}$$

in which the response from a given pixel effects only two adjacent pixels. Note that the fact that these pixels are not centered around the input point does not cause any difficulty since one can always shift the origin of the raster to compensate. Taking this $w(z_1,z_2)$ we obtain the simple compensator $c(z_1,z_2)=-z_1$ which represents a one directional shift and a 180° phase shift.

An alternative design which also yields a "point-spread function" which affects only two pixels, although it is shifted further from the origin, is obtained with $w(z_1,z_2)=-(1/9)z_1$. This yields

$$h_{y_2u_1}(z_1,z_2)=\frac{1}{9}[z_1^4+2z_1^3z_2] \tag{5.30}$$

and

$$c(z_1,z_2)=\frac{z_1^2(z_1+z_2)}{z_1^2+z_1z_2+3}. \tag{5.31}$$

Since two-thirds of the output energy in this design is concentrated at a single point whereas the energy is equally divided in the previous design it may be argued that this represents a superior design. On the other hand, the shift from the origin is greater and the compensator more complex in this case. Finally, since all FIR transfer functions are stable (in an appropriate sense) the feedback systems obtained via either choice of $w(z_1,z_2)$ are stable. Moreover, both compensators are, themselves, stable as is $p(z_1,z_2)$ [6].

Example 4: A Time-Varying Differential-Delay Stochastic Optimal Control Problem

Consider the feedback system of Fig. 4 where the plant represents a cascade of a time-varying function f with an ideal predictor e^s. The system is driven by a stochastic process a, which is derived from white noise by passing it through a miniphase filter with transfer function $(s+2)/(s+1)$. We desire to choose a compensator which will stabilize the system and minimize the performance measure

$$J=E\|b\|^2+E\|d\|^2 \tag{5.32}$$

under the constraint of stability. Here, d is the stochastic process observed at the output of the system, b is the

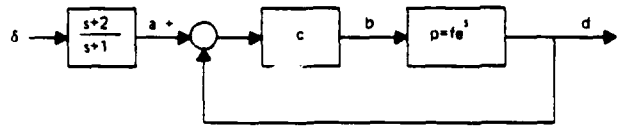


Fig. 4. Stochastic control system.

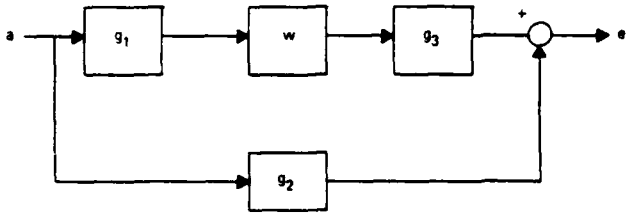


Fig. 5. Open-loop optimization problem.

stochastic process observed at the plant input, and E is the expected value operator.

Since we have a time-varying component, a rational component, and a delay component we formulate our theory in an abstract operator theoretic setting [20] with G taken to be the bounded operators on the Hilbert space $L_2(R)$ and H taken to be the causal bounded operators (which correspond to the stable systems in such a setting) [20],[23]. Note, in this setting we will denote the time-invariant operators by their transfer function and the time-varying multiplication operators by their characteristic function. Of course, one must be careful with such notation since the operational calculus associated with the time-invariant components is only partially valid in such a setting.

Since the inverse of a predictor is the ideal delay which is causal one immediately obtains the right and left coprime fractional representations for p in the form

$$p=[f][e^{-s}]^{-1}=[e^{-s}f^{-1}]^{-1}[1]=n_r d_r^{-1}=d_l^{-1}n_l \tag{5.33}$$

where

$$[f^{-1}][f]+[0][e^{-s}]=u_r n_r+v_r d_r=1 \tag{5.34}$$

and

$$[1][1]+[e^{-s}f^{-1}][0]=n_l u_l+d_l v_l=1. \tag{5.35}$$

Here, we have assumed that f^{-1} exists and is bounded (i.e., f is bounded away from zero) while f and f^{-1} are both causal since multiplication by a function of time is a memoryless operation [20]. From Theorem 3 it now follows that the input–output and input-plant input mappings for our feedback system with compensator defined by a causal operator w will take the form

$$h_{y_2u_1}=-[f]w[e^{-s}][f^{-1}]+1 \tag{5.36}$$

and

$$h_{e_2u_1}=-[e^{-s}]w[e^{-s}][f^{-1}]+[e^{-s}][f^{-1}]. \tag{5.37}$$

As such, our optimization problem reduces to choosing the causal w which minimizes the performance measure of (5.32) where $d=h_{y_2u_1}a$ and $b=h_{e_2u_1}a$.

It is significant to note that even though we are interested in designing an optimal closed-loop system by minimizing over the operator w rather than the compensator we have transformed the problem into the open-loop optimization problem of Fig. 5.

Here we desire to minimize $J=E\|e\|^2$ over all causal operators w, where g_1, g_2, and g_3 are arbitrarily specified bounded operators. In our case we take

$$g_1=[e^{-s}][f^{-1}] \tag{5.38}$$

$$g_2=\begin{bmatrix} -1 \\ \hline -[e^{-s}][f^{-1}] \end{bmatrix} \tag{5.39}$$

and

$$g_3=\begin{bmatrix} -[f] \\ \hline -[e^{-s}] \end{bmatrix} \tag{5.40}$$

in which case the output of the open-loop system is $e=(d,b)$ in the product space constructed from two copies of the (Hilbert) space on which the given system is defined. Now, if we take the a in our open-loop problem to coincide with the given a in the closed-loop optimization problem then the Pythogorean law (in Hilbert space) implies that

$$J=E\|e\|^2=E\|d\|^2+e\|b\|^2. \tag{5.41}$$

As such, our two optimization problems coincide.

Interestingly, an explicit solution has recently been given for the above open-loop optimization problem [9]. Indeed, the optimal causal w is given by

$$w_0=\lambda^{-1}[\lambda^{*-1}g_3^*g_2Q_ag_1^*\theta^{*-1}]_C\theta^{-1} \tag{5.42}$$

where λ and θ are causal, causally invertible operators such that

$$\lambda^*\lambda=g_3^*g_3 \quad \theta\theta^*=g_1Q_ag_1^*. \tag{5.43}$$

Q_a is the covariance for the stochastic processes a,$[\ ]_C$ denotes the causal part of an operator, and "$*$" denotes the adjoint operator. To apply this general theory to our example we represent the adjoint operation when applied to a transfer function by $g(s)^*=g(-s)$ which coincides with the classical adjoint on the $j\omega$ axis. Of course, the memoryless multiplication operators, $[f]$ and $[f^{-1}]$, are self adjoint. Finally, since a is the stochastic process generated from white noise by passing it through the filter $(s+2)/(s+1)$

$$Q_a=\left[\frac{(s+2)}{(s+1)}\right]\left[\frac{(s+2)}{(s+1)}\right]^*=\frac{(s+2)(s-2)}{(s+1)(s-1)}. \tag{5.44}$$

First, we calculate λ and θ via

$$\lambda^*\lambda=\left[\begin{array}{c|c} -f & -e^s \end{array}\right]\begin{bmatrix} -f \\ \hline -e^{-s} \end{bmatrix}=f^2+1 \tag{5.45}$$

and

$$\theta\theta^* = [e^{-s}][f^{-1}]\left[\frac{(s+2)(s-2)}{(s+1)(s-1)}\right][f^{-1}][e^{s}]$$
$$= [f_{-1}^{-1}][e^{-s}]\left[\frac{(s+2)(s-2)}{(s+1)(s-1)}\right][e^{s}][f_{-1}^{-1}]$$
$$= [f_{-1}^{-1}]\frac{(s+2)(s-2)}{(s+1)(s-1)}[f_{-1}^{-1}]. \quad (5.46)$$

Here $f_{-1}(t) = f(t-1)$ and we have used the properties of the delay and predictor to obtain the equalities $[e^{-s}][f^{-1}] = [f_{-1}^{-1}][e^{-s}]$ and $[f^{-1}][e^{s}] = [e^{s}][f_{-1}^{-1}]$. Of course, the exponential transfer functions commute with the rational transfer functions allowing the cancellation of the exponential terms in (5.46). From (5.45) and (5.46) one may now readily obtain the required causal, causally invertible λ and θ operators in the form

$$\lambda = \lambda^* = \sqrt{f^2+1} \text{ and } \lambda^{-1} = \lambda^{*-1} = \frac{1}{\sqrt{f^2+1}} \quad (5.47)$$

while

$$\theta = [f_{-1}^{-1}]\left[\frac{(s+2)}{(s+1)}\right],\ \theta^* = \left[\frac{(s-2)}{(s-1)}\right][f_{-1}^{-1}],$$
$$\theta^{-1} = \left[\frac{(s+1)}{(s+2)}\right][f_{-1}], \text{ and } \theta^{*-1} = [f_{-1}]\left[\frac{(s-1)}{(s-2)}\right]. \quad (5.48)$$

The next step in evaluating (5.42) is to compute the term in the bracket, i.e.,

$$\lambda^{*-1}g_3^*g_2Q_ag_1^*\theta^{*-1}$$
$$= \frac{1}{\sqrt{f^2+1}}\left[\begin{array}{c|c} -f & -e^{s} \end{array}\right]\left[\begin{array}{c|c} \dfrac{-1}{[-e^{-s}][f^{-1}]} & \\ \hline & \end{array}\right]$$
$$\cdot\left[\frac{(s+2)(s-2)}{(s+1)(s-1)}\right][f^{-1}][e^{s}]\cdot[f_{-1}]\left[\frac{(s-1)}{(s-2)}\right]$$
$$= \frac{1}{\sqrt{f^2+1}}[f+f^{-1}]\left[\frac{(s+2)(s-2)}{(s+1)(s-1)}\right]$$
$$\cdot[f^{-1}][f]\frac{(s-1)}{(s-2)}[e^{s}]$$
$$= [f^{-1}]\sqrt{f^2+1}\left[\frac{(s+2)}{(s+1)}\right][e^{s}] \quad (5.49)$$

whose causal part must now be computed. Recalling that the memoryless term factors through the causal part bracket [9] it suffices to compute the causal part of the time-invariant system with transfer function

$$g(s) = \left[\frac{(s+2)}{(s+1)}e^{s}\right]. \quad (5.50)$$

Taking the inverse Laplace transform we obtain the impulse response of this system in the form

$$g(t) = \delta(t+1) + \bar{e}^{(t+1)}U(t+1) \quad (5.51)$$

where δ is the Dirac delta function and U is the unit step function. Now, the causal part of $g(t)$ is obtained by setting $g(t)$ to zero for t less than zero; hence

$$[g(t)]_C = g(t)U(t) = \bar{e}^{(t+1)}U(t) = \frac{1}{e}e^{-t}U(t) \quad (5.52)$$

or equivalently

$$[g(s)]_C = \frac{1}{e(s+1)}. \quad (5.53)$$

Multiplying through by the memoryless factor from (5.49) we then obtain

$$[\lambda^{*-1}g_3^*g_2Q_ag_1^*\theta^{*-1}]_C = [f^{-1}]\sqrt{f^2+1}\ \frac{1}{c(s+1)} \quad (5.54)$$

and finally

$$w_0 = \lambda^{-1}[\lambda^{*-1}g_3^*g_2Q_ag_1^*\theta^{*-1}]_C\theta^{-1}$$
$$= \frac{1}{\sqrt{f^2+1}}[f^{-1}]\sqrt{f^2+1}$$
$$\cdot\frac{1}{e}\left[\frac{1}{(s+1)}\right]\left[\frac{(s+1)}{(s+2)}\right][f_{-1}]$$
$$= \frac{1}{e}[f^{-1}]\frac{1}{(s+2)}[f_{-1}] \quad (5.55)$$

which is surprisingly simple given the complexity of the derivation.

Substituting the expression of (5.55) into the formula of Theorem 3 now yields an expression for our optimal compensator and the input–output mapping for the resultant feedback system in the form

$$c = e[f_{-1}^{-1}](s+2) - 1 \quad (5.56)$$

and

$$h_{y_2u_1} = \frac{-e^{-s}}{e(s+2)} + 1. \quad (5.57)$$

Note that $h_{y_2u_1}$ is stable, as required, even though both p and c are unstable.

VI. Nonlinear Feedback Systems

From an algebraic point of view the fundamental difference between linear and nonlinear systems is the fact that nonlinear systems fail to satisfy the right-distributive property, $x(y+z) = xy + xz$. They do, however, satisfy all of the other axioms for a ring with identity including the left-distributive property $(y+z)x = yx + zx$. As such, one can attempt to extend the preceding development to non-

linear systems by carrying it out in left-distributive rings, G and H [23]. Indeed, if we define a right coprime fractional representation for a system g in a left-distributive ring G relative to $\{G,H,I,J\}$ precisely as we did in Section II the fundamental properties 1, 2, and 3 go through without modification.

Property 1N: Let $g=n_r d_r^{-1}$ be a right coprime fractional representation of g in $\{G,H,I,J\}$ where G and H are left-distributive rings with identity. Let n_r and d_r have a common right factor $r\in H$, i.e., $n_r=x_r r$, $d_r=y_r r$ for some $x_r\in H$ and $y_r\in H$. Then r has a left inverse in H.

Property 2N: Let $g=n_r d_r^{-1}$ be a right coprime fractional representation of g in $\{G,H,I,J\}$ where G and H are left-distributive rings with identity. Let $g=x_r y_r^{-1}$ be a second (not necessarily coprime) right fractional representation of g in $\{G,H,I,J\}$; then there exists r in H such that

$$x_r=n_r r \quad \text{and} \quad y_r=d_r r. \tag{6.1}$$

Property 3N: Let $g=n_r d_r^{-1}$ with $n_r\in H$ and $d_r\in I$ where G and H are left-distributive rings with identity.

a) If $d_r\in J$, then $g\in H$.

b) If $g=n_r d_r^{-1}$ is a *right coprime fractional representation* of g in $\{G,H,I,J\}$, then $g\in H$ implies $d_r\in J$.

With the aid of property 3N one can do a complete analysis of a nonlinear feedback system $h_{yu}=p(1+p)^{-1}=n_r d_r^{-1}$ where $n_r d_r^{-1}$ is a right coprime fractional representation of h_{yu}. Indeed, h_{yu} is well defined in G if and only if $d_r\in I$ and it is well defined in H if and only if $d_r\in J$. Note, however, that we cannot construct our fractional representation for h_{yu} from a fractional representation for p since the verification that such a representation is coprime appears to require right-distributivity [see (3.4) and (3.5)].

The right coprime fractional representation plays a special role in the *nonlinear* case because $h_{yu}=p(1+p)^{-1}$ holds, whereas $h_{yu}=(1+p)^{-1}p$ does not (even though the latter formula is true for the *linear* case). As such, those results on the analysis of feedback systems which assume a left coprime fractional representation theory fail as does the design theorem since it simultaneously employs both left and right coprime fractional representations. We believe, however, that these results should hold, at least in part, for nonlinear systems with an appropriate modification of the theory. In particular, since the rings G and H are asymmetric we believe that asymmetric concepts of left and right coprimeness will be required to achieve this end.

VII. Conclusions

Although several of our examples are characterized by a deep analytic structure the key to our fractional representation approach to feedback system design is the algebraic nature of the main results. Indeed, the entirety of our modeling, analysis, and synthesis theory was derived with no more sophisticated mathematics than addition, multiplication, subtraction, and inversion. As such, it applies to essentially any class of linear systems and by proper choice of the rings G and H the results are applicable to a variety of systems problems.

Although we believe that the present work represents the first attempt at the formulation of an axiomatic fractional representation theory for systems which may be matched to the feedback system analysis and synthesis problems of interest the work owes much to a number of recent results on the input–output theory of linear systems. The use of a fractional representation theory for multivariate systems, though implicit in a number of classical results, was popularized by Rosenbrock's polynomial matrix fractions [19]. Interestingly, however, Rosenbrock's goal was apparently to permit the powerful analytic and arithmetic theory available for polynomial matrices to be applied to rational matrices whereas the present fractional representation theory is motivated by the desire to formulate a representation theory for systems which is closed under inversion. Over the years numerous generalizations of the polynomial matrix fraction concept have been formulated for distributed systems [4], [5], [13], [21], and multidimensional systems [9],[24] while partial extensions to the time-varying and nonlinear cases have appeared in a number of unpublished reports [11],[22].

For any type of fractional representation theory to be meaningful it must be identified with an appropriate coprimeness concept. Indeed, the key to the present formulation is the use of the algebraic coprimeness concept of (2.4) in lieu of the more classical common factor criterion. Such a criterion has previously been applied by one of the authors in a study of fractional representations for distributed system [4] and was also shown to be the strongest of several possible coprimeness criteria for multidimensional systems by Youla and Gnavi [26]. Of course, it is well known as one of the several equivalent criteria for coprimeness in the polynomial matrix fraction theory [16],[19].

The feedback system analysis theorems of Section III are motivated by the now classical theorems for determining the stability of a multivariate feedback system in terms of its polynomial matrix fraction representation [10]. Moreover, the system synthesis theorem is an outgrowth of the feedback system stabilization theorem of Youla *et al.* [24],[25]. Indeed, the present work began with an attempt to give a simple proof of this most powerful analytic theorem and developed through several stages of generalization and simplification into its present form. Finally, the optimization theory used in Example 4 represents the generalization [9] to an operator theoretic setting of a result originally developed by Youla *et al.* in the frequency domain for use in conjunction with their stabilization theorem [24],[25].

Acknowledgment

The authors wish to thank an anonymous reviewer for numerous useful and perceptive comments.

REFERENCES

[1] W. Averson, "Interpolation problems in nest algebras," *J. Functional Analysis*, vol. 20, pp. 208–233, 1975.
[2] Bourbaki, *Commutative Algebra*. Paris: Hermann, 1972.
[3] C. I. Byrnes, private communication.
[4] F. M. Callier, and C. A. Desoer, "An algebra of transfer functions for distributed linear time-invariant systems," *IEEE Trans. Circuits Syst.*, vol. CAS-25, pp. 651–662, 1978.
[5] ——, "Open-loop unstable convolution feedback systems with dynamical feedback," *Automatica*, vol. 12, pp. 507–518, 1976.
[6] ——, "Dynamic output feedback stabilization of a control system," presented at Amer. Math. Soc. 761st Meeting, Charleston, Nov. 1978 (abstract appeared in *Notices AMS*, vol. 25, p. A-665, 1978).
[7] ——, "Stabilization, tracking and disturbance rejection in linear multivariable distributed systems," in *Proc. 17th IEEE Conf. Decision and Control*, San Diego, Jan. 1979, p. 513.
[8] ——, "Stabilization, tracking and disturbance rejection in multivariable convolution systems," Tech. Memo. UCB/ERL M78/83, Dec. 1978.
[9] R. M. DeSantis, R. Saeks, and J. J. Tung, "Basic optimal estimation and control problems in Hilbert space," *Math. Syst. Theory*, to be published.
[10] C. A. Desoer and M. Vidyasagar, *Feedback Systems: Input-Output Properties*. New York: Academic, 1975.
[11] C. A. Desoer, R. W. Liu, J. Murray, and R. Saeks, "On stabilization of feedback systems," unpublished notes, 1978.
[12] R. G. Douglas, *Banach Algebra Techniques in Operator Theory*. New York: Academic, 1972.
[13] B. Francis, "The multivariable servomechanism problem from the input–output viewpoint," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 322–328, 1977.
[14] M. L. J. Hautus and M. Heymann, "Linear feedback—An algebraic approach," *SIAM J. Contr. Optimiz.*, vol. 16, pp. 83–105, 1978.
[15] N. T. Hung and B. D. O. Anderson, "Triangularization for the design of multivariable control systems," *IEEE Trans. Automat. Contr.*, to be published; also in *Proc.* 1978 *IEEE CDC Conf.*, pp. 795–800, Jan. 1973.
[16] E. I. Jury, *Inners and Stability of Dynamic Systems*. New York: Wiley, 1974.
[17] M. B. Morf, B. Levy, and S. Y. Kung, "New results in 2-D system theory, part I: 2-D polynomial matrices, factorization and coprimeness," *Proc. IEEE*, vol. 65, pp. 861–872, 1977.
[18] A. S. Morse, "System invariants under feedback and cascade control," in *Mathematical Systems Theory*, Udine, 1975, G. Marchesini and S. K. Mitter, Eds. Heidelberg: Springer-Verlag, 1976.
[19] H. H. Rosenbrock, *State-Space and Multivariable Theory*. London: Nelson-Wiley, 1970
[20] R. Saeks, *Resolution Space, Operators, and Systems*. Heidelberg: Springer-Verlag, 1973.
[21] M. Vidyasagar, "On the input–output stability of multivariable linear time-invariant feedback systems containing unstable systems," *IEEE Trans. Circuits Syst.*, to be published.
[22] ——, unpublished notes, 1978.
[23] J. C. Willems, *Analysis of Feedback Systems*. Cambridge, MA: MIT Press, 1971.
[24] D. C. Youla, J. J. Bongiorno, and H. A. Jabr, "Modern Wiener–Hopf design of optimal controllers—Part I," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 3–15, 1976.
[25] ——, "Modern Wiener–Hopf design of optimal controllers—Part II," *IEEE Trans. Automat. Contr.*, vol. AC-21, pp. 319–338, 1976.
[26] D. C. Youla and G. Gnavi "Notes on *n*-dimensional system theory," unpublished notes, 1978.
[27] O. Zariski and P. Samuel, *Commutative Algebra*, vol. I. Princeton, NJ: Van Nostrand, 1958 p. 46, exp.
[28] S. MacLane and G. Birkhoff, *Algebra*. New York: MacMillan, 1971, p. 154.
[29] J. J. Bongiorno and D. C. Youla, "On the design of single-loop single-input single-output feedback control systems in the complex frequency domain," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 416–423, 1977.
[30] C. A. Desoer and W. S. Chan, "The feedback interconnection of lumped linear time-invariant systems," *J. Franklin Inst.*, vol. 300, pp. 335–351, 1975.
[31] F. M. Callier and C. A. Desoer, "Simplification and clarification on the paper 'An algebra of transfer function for distributed linear time-invariant systems,'" *IEEE Trans. Circuits Syst.*, to be published.
[32] L. Cheng and J. B. Pearson, "Frequency domain synthesis of multivariable linear regulators," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 3–15, 1978.
[33] P. J. Antsaklis and J. B. Pearson, "Stabilization and regulation in linear multivariable systems," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 928–930, 1978.

C. A. Desoer (S'50–A'53–SM'57–F'64), for a photograph and biography, see page 196 of the April 1980 issue of this TRANSACTIONS.

Ruey-Wen Liu (S'53–M'59) was born in Kiangsu, China. He received the B.S., M.S., and Ph.D. degrees in 1954, 1955, and 1960, respectively, from the Department of Electrical Engineering, University of Illinois, Urbana.

Since 1960, he has been with the University of Notre Dame, and is currently a Professor in the Department of Electrical Engineering. He received a National Science Foundation Fellowship at the Stanford University in the Summer of 1962. He was a Visiting Associate Professor from 1965–1966 (and a Visiting Professor from 1977 to 1978) at the University of California, Berkeley. He also held visiting professorships at the National Taiwan University, Republic of China, in the Spring of 1969, the Universidad de Chile, Santiago de Chile, in the Summer of 1970, and the Institute of Mathematics, Academia Sinica, Republic of China, in the Summers of 1976 and 1978. His main interest has been in the area of nonlinear circuits and systems. Currently, he is the Chairman of the Technical Committee on Nonlinear Circuits and Systems of the IEEE Circuits and Systems Society. His other interests include policy analysis of socioeconomic systems and fault diagnosis of electronic circuits. He was the Cochairman of the NSF Workshop on Nonlinear Circuits and Systems in January 1980. He is the Co-Guest Editor of the special issue on Socio-Economic Systems of the *International Journal on Policy Analysis and Information Systems*. He is also the Guest Editor of the special issue on Nonlinear Circuits and Systems of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS.

Dr. Liu is a member of Sigma Xi and the New York Academy of Sciences.

John Murray was born in Galway, Ireland, on August 8, 1947. He received the B.Sc. and M.Sc. degrees from University College, Cork, Ireland, in 1969 and 1970, respectively, and the Ph.D. degree from the University of Notre Dame, Notre Dame, IN, in 1974, all in mathematics.

He is currently with the Department of Electrical Engineering, Texas Tech University, Lubbock. His principal research interests are in the areas of several complex variables, multidimensional system theory, and time-varying systems.

Richard Saeks (S'59–M'65–SM'74–F'77) was born in Chicago, IL, in 1941. He received the B.S. degree in 1964, the M.S. degree in 1965, and the Ph.D. degree in 1967 from Northwestern University, Evanston, IL, Colorado State University, Fort Collins, and Cornell University, Ithaca, NY, respectively, all in electrical engineering.

He is presently Paul Whitfield Horn Professor of Electrical Engineering and Mathematics at Texas Tech University, Lubbock, where he is involved in teaching and research in the areas of fault analysis, circuit theory, and mathematical system theory.

Dr. Saeks is a member of the American Mathematical Society, the Society for Industrial and Applied Mathematics, and Sigma Xi.

System Diagnosis - A New System Problem

by

R. Liu
C. Lin
A. Deng
V. Raman

SYSTEM DIAGNOSIS - A NEW SYSTEM PROBLEM

R. Liu
C. Lin
A. Deng
V. Raman

Department of Electrical Engineering

University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

The problem of system diagnosis is presented. Its applications and recent results are discussed.

I. INTRODUCTION

A Large-Scale Dynamical System (LSDS) can be described as one which is either impossible or impractical to be represented by a single composite system equation, and is to be solved as such. Therefore, it is better for a large-scale system to be represented as an interconnection of components or subsystems. One such representation is presented in Section II.

Fault diagnosis is a typical problem of large-scale dynamical systems. It is a problem of studying the (large) deviations of the overall system caused by the deviations of the subsystems (sometimes simply circuit elements). It is not practical to formulate this problem from a single composite system equation.

The fault diagnosis of analog systems at present is an art rather than a science. For example, it is not clear how the concept of "fault" can be defined. On the other hand, the system diagnosis problem, while related to the fault diagnosis problem, can be well defined. This is done in Section III.

It turns out that the application of system diagnosis problem is not restricted to fault diagnosis problems. It also has applications, for example, to the modeling of socio-economic problems. A discussion of the applications of system diagnosis problems is given in Section IV. Finally, a discussion of the recent results is presented in Section V.

II. A MODEL OF LINEAR TIME-INVARIANT *LSDS*

Let an LSDS be comprised by a finite number of subsystems, each of which is a linear time-invariant dynamical system of the type

$$\begin{aligned} \dot{x}_i &= A_i x_i + B_i a_i \\ b_i &= C_i x_i + D_i a_i \end{aligned} \qquad (1)$$

$i = 1,2,\ldots,k$, where a_i, b_i, x_i are the input vector, the output vector, and the state vector of each of the components. They are not necessarily of the same dimension. The matrices A_i, B_i, C_i, and D_i are compatible to these vectors. These components can be put into a single composite component equation.

$$\begin{aligned} \dot{x} &= Ax + Ba \\ b &= Cx + Da \end{aligned} \qquad (2a)$$

where $x = \mathrm{col}(x_i)$ and a and b are similarly defined, and $A = \mathrm{diag}(A_i)$ and B, C, and D are similarly defined. For definiteness, let $x \in R^n$, $a \in R^m$, and $b \in R^r$, and A, B, C, and D be constant real matrices of compatible dimensions.

The connection of the input and output terminals of these subsystems with the input terminals u and the output terminals y of the entire system is assumed to have the form

$$\begin{bmatrix} a \\ y \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} \triangleq L \begin{bmatrix} b \\ u \end{bmatrix} \qquad (2b)$$

Here, $u \in R^p$ and $y \in R^q$ and L and R are constant real matrices with compatible dimensions.

The model of LSDS defined by (2) is a linear and time-invariant one and it can be represented in the forms of block diagrams in Fig. 1. Two special classes of LSDS are of special interest. Consider a linear integrating circuit. Each of its n-port subnetworks can be represented by (1) and *the KVL and KCL equations* have the form of (2b). As such, a_i and b_i are the port voltages and the port currents. Therefore, any linear integrating circuit can be represented by an LSDS model. Consider a block-diagram connection of linear time-invariant subsystems (1). Since every block-diagram connection can be represented by (2b), the above system can also be represented by an LSDS model. Therefore, the LSDS model (2) includes both network-type and system-type formulations.

In this paper, we only consider linear time-invariant LSDS's which can be represented by (2).

When the composite component equation (2a) is represented by a transfer function

$$b = Z(s)a \qquad (3)$$

where

$$Z(s) = C(sI - A)^{-1}B + D \qquad (4)$$

It becomes the component connection model in the frequency domain, considered by Saeks et al [1-4]. Combining (3) with (2b), we have

$$y = S(Z)u \qquad (5)$$

where

$$S(Z) = L_{22} + L_{21}(1 - ZL_{11})^{-1}ZL_{12} \qquad (6)$$

which is used for the purpose of system diagnosis.

Trick et al [5] considered the special case when (2b) represents KVL and KCL equations. When fundamental cut-set equations and fundamental loop equations are used, Eq. (2b) becomes,

$$\begin{bmatrix} v_{ct} \\ i_t \end{bmatrix} = \begin{bmatrix} 0 & F \\ -F^T & 0 \end{bmatrix} \begin{bmatrix} i_{ct} \\ v_t \end{bmatrix} \qquad (7)$$

where v_{ct}, i_{ct} are the branch voltages and branch currents in the cotree and v_t and i_t are those in the tree [6]. Comparing (7) with (2b), we have

$$L = -L^T \tag{8}$$

and consequently,

$$[v_{ct}^T \; i_t^T] \begin{bmatrix} i_{ct} \\ v_t \end{bmatrix} = [i_{ct} \; v_t] \, L \begin{bmatrix} i_{ct} \\ v_t \end{bmatrix} = 0 \tag{9}$$

The last equality follows from (9) that L is anti-symmetric. Eq. (9) is a generalization of Tellegen Theorem since it is derived from a more weaker condition (8) instead of (7).

Navid and Willson, Jr., [7] considered the case of DC-network, i.e., the connection equation again has the form (7) and furthermore, the composite component equation (2a) or (3) has the form

$$b = Da \tag{10}$$

with the condition, among others, that each row of D has one and only one non-zero element.

We have presented the mathematical model (2a) and (2b) and its relation to other models. Next, we want to define the system diagnosis problem.

III. SYSTEM DIAGNOSIS PROBLEMS

With the model (2a) and (2b), the system diagnosis problem can be stated:

The System Diagnosis Problem: Given L and n (dimension of x), the problem is to determine the component parameters (A,B,C,D) from input/output measurements.

Let us now formalize the problem. Let us assume that there exists a state equation for the component connection model (2a) and (2b)* of the form:

$$\begin{aligned} \dot{z} &= Fz + Gu \\ y &= Hz + Ju \end{aligned} \tag{11}$$

where $z \in R^{\bar{n}}$, $\bar{n} \leq n$. The transfer function is given by

$$H(s) = H(sI - F)^{-1}G + J \tag{12}$$

Let the vector p denote the component parameters (A,B,C,D) to be determined. Its relation to the transfer function is given by

$$H(s) = h(s, p) \tag{13}$$

where h is determined by the system structure (L, n).

Since by methods of system identification techniques [9,10], the transfer function $H(s)$ can be obtained from proper input/output measurements, the problem of system diagnosis is reduced to the problem of the determination of p from $H(s)$.

Definition 1. The parameters p of an LSDS are said to be diagnosable is p can be determined from $H(s)$, i.e., h is injective.

The same problem can be formulated in the time-domain. It turns out that the state equation (F,G,H,J) of (11) cannot be uniquely determined from input/output measurements. Only Markov parameters can be uniquely determined.

Assertion 1 [11]. Two minimal (controllable and observable) linear time-invariant state equation representation (F,G,H,J) and $(\bar{F},\bar{G},\bar{H},\bar{J})$ with state space of the same dimension n are realization of the same transfer function matrix $H(s)$ if and only if

$$J = \bar{J} \tag{14}$$

and

$$H F^i G = \bar{H} \bar{F}^i \bar{G}, \quad i = 0,1,\ldots,2n-1. \tag{15}$$

The Markov parameters m_i

$$m_0 \triangleq J \tag{16}$$

$$m_i \triangleq H F^{i-1} G \quad i=1,\ldots,2n \tag{17}$$

are well defined functions of p. Therefore, we can write

$$m_i = F_i(p) \;, \quad i = 0,1,\ldots,2n \tag{18}$$

or

$$m = F(p) \tag{19}$$

where $m = \text{col.}(m_i)$ and $F = \text{col.}(F_i)$.

Assertion 2. The parameters p of the LSDS are diagnosable if and only if F is injective.

Therefore, the basic problem of system diagnosis is to study the invertability of h (the frequency domain) or that of F (the time domain). Let us consider an example.

Example 1. Consider the two-stage RC ladder as shown in Fig. 2. The composite component equation (2a) is given by

$$\left[\begin{array}{c} \frac{d}{dt} v_C \\ \hline v_C \\ \hline i_G \end{array}\right] = \left[\begin{array}{c|cc} 0 & D & 0 \\ \hline I & 0 & 0 \\ \hline 0 & 0 & G \end{array}\right] \left[\begin{array}{c} v_C \\ \hline i_C \\ \hline v_G \end{array}\right] \tag{20}$$

where

$$v_C = \text{Col.}(v_{C_1}, v_{C_2})$$

and similarly for i_C, v_G, i_G, and

$$G = \text{diag}(G_1, G_2)$$

$$D = \text{diag}(D_1, D_2)$$

where $D_i = 1/C_i$, $i = 1,2$.

The connection equation (2b) is given by

$$\left[\begin{array}{c} v_G \\ \hline i_C \\ \hline i_s \end{array}\right] = \left[\begin{array}{cc|cc|c} 0 & -1 & 0 & 1 & \\ 1 & -1 & 0 & 0 & \\ \hline 1 & -1 & 0 & & 0 \\ 0 & 1 & & & 0 \\ \hline 1 & 0 & 0 & 0 & 0 \end{array}\right] \left[\begin{array}{c} i_G \\ \hline v_C \\ \hline v_s \end{array}\right] \tag{21}$$

Equations (20) and (21) have a state equation representation

---

* The n.a.s. condition for the existence of state equation for (2a) and (2b) is given by Singh and Liu [8].

$$\frac{d}{dt}v_C = \begin{bmatrix} -G_1D_1 & -G_2D_1 & G_2D_1 \\ G_2D_2 & & -G_2D_2 \end{bmatrix} v_C + \begin{bmatrix} G_1D_1 \\ 0 \end{bmatrix} v_s \quad (22)$$

$$i_s = [-G_1 \quad 0]v_C + [G_1]v_s \quad (23)$$

The transfer function of the above state equation is given by

$$H(s) = \frac{s^2G_1 + sG_1G_2(D_1 + D_2)}{s^2 + s(G_2D_2 + G_2D_1 + G_1D_1) + G_1G_2D_1D_2} \triangleq h(s, p) \quad (24)$$

The Markov parameter of the above state equation is given by

$$m_0 = G_1$$
$$m_1 = -G_1{}^2D_1$$
$$m_2 = -(G_1{}^3D_1{}^2 + G_1{}^2G_2D_1{}^2)$$
$$m_3 = G_1{}^2D_1{}^3(G_1 + G_2)^2 + G_1{}^2G_2{}^2D_1{}^2D_2$$

which has the form

$$m = F(p) \quad (25)$$

where $p = (G_1, G_2, D_1, D_2)$ for both (24) and (25). Note that m_4 is not needed since we only have four parameters to be determined.

The parameter p is diagnosable if and only if the function h or F is invertible. In other words, we want to solve for p from either (24) when the transfer function $H(s)$ is measured, or from (25) when the Markov parameter m is measured.

Note that the system diagnosis is an extension of system identification. The purpose of system identification is to determine either the coefficients of a transfer function or the Markov parameters from input/output measurements. The purpose of system diagnosis is to determine the component parameters. The latter are usually the ones to be designed.

Please note that the functions h and F are nonlinear. Therefore, the system diagnosis problem is a nonlinear problem even if the original system is linear.

Finally, the above problem can be modified to be more flexible. For example, we may know some of the component parameters and want to determine the rest of them. The system diagnosis problem can also be extended to nonlinear systems.

IV. APPLICATIONS

1. Fault Diagnosis

Let Ω be the set of component parameters p for which the system is fault-free, i.e.,

1) if $p \in \Omega$, then the LSDS is fault-free, and
2) if $p \in \Omega^C$, then the LSDS is at fault.

When the LSDS is at fault, the fault diagnosis problem is to determine which components of p causing $p \in \Omega^C$. Clearly, if p can be determined and Ω is clearly defined, the test whether or not $p \in \Omega^C$ can be easily carried out. Therefore, if a LSDS is diagnosable, it is also fault diagnosable. However, the diagnosability is sometimes too strong for the problem of fault diagnosis, because it is not necessarily to determine the values of p in order to test whether or not $p \in \Omega$. Furthermore, the cost of the evaluation of the values of p is high, and accuracy is so poor that the calculated values of p become useless [see Section V]. Therefore, the art of fault diagnosis is to find ways of testing whether or not $p \in \Omega$ with little calculation of the values of p. Nevertheless, the diagnosability does provide the insight for searching such an art and the design of test points.

A special issue on Automatic Analog Fault Diagnosis [12], edited by S. D. Bedrosian, contains much information and many references on this topic.

2. Network Synthesis

Given a prescribed transfer function $H^o(s)$, a network synthesis problem is to find a component connection model, (2a) and (2b), so that its transfer function $H(s)$ equals the prescribed one $H^o(s)$. If the connection L is predetermined, then the rest of the synthesis problem is to determine the parameter values p which yield the prescribed transfer function $H^o(s)$. This is precisely the problem of system diagnosis. Note that the state-of-the-art of network synthesis is mainly on single-input single-output systems, while the system diagnosis problems are not restricted to SISO problems.

3. Modeling of Nonlinear Circuit Models

The circuit shown in Fig. 3 can be used to model the electrical behaviors of the nerve membrane*. Chua found that the Potassium conductance g_K and the sodium conductance g_{Na} should be modeled as a nonlinear time-invariant memristive 1-ports [13] instead of the linear time-variant elements in the original model. They have the state equation form:

$$\dot{n} = f(n, v_K)$$
$$i_K = G_K(n)v_K$$

for g_K and

$$\dot{m} = f_1(m, v_{Na})$$
$$\dot{h} = f_2(h, v_{Na})$$
$$i_{Na} = G_{Na}(m, h)v_{Na}$$

for g_{Na}. The state variables are n and (m, h) respectively. The explicit expressions of the nonlinear functions are given in [13]. One of the problems is to identify the parameter values in the circuit model given in Fig. 3. This problem can be considered as the system diagnosis problem of a nonlinear system.

4. Modeling of Socio-Economical Systems

Two forms of models are often used by economists, the structural form and the reduced form [14]. This can best be illustrated by an example. Consider the economic model:

---

*Hodgkin and Huxley were awarded the Nobel Prize in 1963 for their development of this circuit model.

$$C(t) = \alpha_0 + \alpha_1[Y(t) - T(t)]$$

$$I(t) = \beta_1 Y(t-1) + \beta_2 R(t) \qquad (S)$$

$$Y(t) \equiv C(t) + I(t) + G(t)$$

where

C = Consumption
I = Investment
Y = National Income
G = Government Expenditure on Goods and Services
T = Taxes on Income
R = Government Regulator

The last equation is an identity. The above set of equations can be solved to yield,

$$C(t) = \frac{\alpha_0}{1-\alpha_1} + \frac{\alpha_1\beta_1}{1-\alpha_1} Y(t-1) + \frac{\alpha_1\beta_2}{1-\alpha_1} R(t) + \frac{\alpha_1}{1-\alpha_1}[G(t)-T(t)]$$

$$I(t) = \beta_1 Y(t-1) + \beta_2 R(t) \qquad (R)$$

$$Y(t) = \frac{\alpha_0}{1-\alpha_1} + \frac{\beta_1}{1-\alpha_1} Y(t-1) + \frac{\beta_2}{1-\alpha_2} R(t) + \frac{1}{1-\alpha_1} G(t) - \frac{\alpha_1}{1-\alpha_1} T(t)$$

The above equations give the explicit input (T,R,G) and output (C,I,Y) relation. The first set of equations is called the structural form and the second set of equations, the reduced form. Clearly, the reduced form is like the transfer function representation. The structural form is like the component-connection model with the third equation (the identity equation) being the connection equation. Consequently, the system identificiation techniques can be used to identify the coefficients of the reduced form, while the system diagnosis technique can be used to identify the coefficients of the structural form.

V. DISCUSSION OF RECENT RESULTS

The system diagnosis problem can be classified into three aspects: (A) the problem of diagnosability, (B) the computational problem, and (C) the design of test points.

(A) Diagnosability

Three representative results on the condition of diagnosability will be presented.

Sen and Saeks [3] have considered LSDS which can be represented by (6)

$$g(s,p) \triangleq S(Z(s,p)) = L_{22} + L_{21}[I - Z(s,p)L_{11}]^{-1} Z(s,p) L_{12} \qquad (6)$$

By measuring g at multiple frequencies $s_1, s_2, \ldots, s_n$, the above equation can be expressed in the matrix form

$$\begin{bmatrix} g(s_1, p) \\ g(s_2, p) \\ \vdots \\ g(s_n, p) \end{bmatrix} = G(p) \qquad (26)$$

which is a function of p. Clearly, p is diagnosable if and only if G is invertible.

Trick et al [5] considered the diagnosis problem of electrical networks, with known nominal component parameters. By means of Tellegen's Theorem and adjoint networks, the following equation can be obtained for a linear network:

$$M \cdot \Delta p = \Delta Q \qquad (27)$$

where M and ΔQ are measurable, and Δp is the difference between the nominal values and the actual values of p. Therefore, p is diagnosable if and only if M is invertible. Note that Eq. (26) is in general nonlinear while Equation (27) is linear. The latter is achieved at the expense of more test points.

Navid and Willson, Jr., [7] considered the diagnosis problem for resistive networks. In this case Eq. (6) takes the form

$$g(R) = L_{22} + L_{21}(I - R\,L_{11})^{-1} R\,L_{12}$$

where R is the resistive matrix. By taking advantage of the fact that L_{11}, L_{12}, L_{21} and L_{22} arrived from a network graph, they were able to find an elegant condition for the invertability of g. This condition depends on only the graph, not the component values.

(B) Computational Problems

As it was discussed in Section III, the system diagnosis problem is in general, nonlinear, even when the system is linear. When the dimension of p is large, the computational cost will be high and the results will be inaccurate. The following example will illustrate the point.

Example 2 [15]. Consider a 4-stage RC-ladder as shown in Fig. 4. The transfer function of the ladder is given by

$$Y(s) = \frac{b_0 s^4 + b_1 s^3 + b_2 s^2 + b_3 s}{a_0 s^4 + a_1 s^3 + a_2 s^2 + a_3 s + 1} \triangleq \frac{N_1(s)}{D_1(s)}$$

The component values can be calculated from the coefficients a_i's and b_i's by a continue-fraction expansion [16] as

$$R_i = \lim_{s\to\infty} \frac{D_i(s)}{N_i(s)}$$

$$C_i = \lim_{s\to\infty} \frac{N_i(s)}{D_{i+1}(s)}$$

where

$$N_i(s) = N_{i-1}(s) - C_{i-1} D_i(s)$$

$$D_i(s) = D_{i-1}(s) - R_{i-1} N_{i-1}(s)$$

Now, we first arbitrarily chose the parameter values to be

$$\begin{aligned} R_1 &= 2.2 & C_1 &= 0.015 \\ R_2 &= 47 & C_2 &= 0.470 \\ R_3 &= 8.2 & C_3 &= 0.010 \\ R_4 &= 1.0 & C_4 &= 0.022 \end{aligned} \qquad (28)$$

The exact values of the coefficients can be calculated:

$$\begin{aligned} a_0 &= 0.001\,1506188 & b_0 &= 5.977554E\text{-}4 \\ a_1 &= 0.249\,35516 & b_1 &= 0.09526598 \\ a_2 &= 7.3770808 & b_2 &= 0.492064 \\ a_3 &= 25.0158 & b_3 &= 0.517 \end{aligned} \qquad (29)$$

We now calculate the component values by the continue-fractional expansion method. If the exact values of coefficients (29) are used and if there is no computational error, the component values given in (28) should be completely recovered. Two cases have been examined. When the exact values of (29) are used, the calculated component values are completely agreeable with that of (28). This indicates the error due to computation is small. Now, when significant value of (29) is reduced to seven, the calculated component value becomes,

$$\begin{aligned} \hat{R}_1 &= 2.2 & \hat{C}_1 &= 0.015 \\ \hat{R}_2 &= 47.005 & \hat{C}_2 &= 0.70919 \\ \hat{R}_3 &= -0.018072 & \hat{C}_3 &= -0.23768 \\ \hat{R}_4 &= 9.1305 & \hat{C}_4 &= 0.030482 \end{aligned}$$

which have significant deviation from (28). In fact, $\hat{R}_3$ and $\hat{C}_3$ are negative! In reality, one can hardly measure the transfer function coefficient to 7-digits accuracy. This example shows that even p is diagnosable but it may not be computable in realistic situations.

Let us now re-examine the RC-ladder and diagnose it from another way. Observe the time-domain equations for 2-stage RC-ladder (25), i.e.,

$$\begin{aligned} m_0 &= G_1 \\ m_1 &= -G_1^2 D_1 \\ m_2 &= -(G_1^3 D_1^2 + G_1^2 G_2 D_1^2) \\ m_3 &= G_1^3 D_1^3 (G_1 + G_2)^2 + G_1^2 G_2^2 D_1^2 D_2 \end{aligned}$$

Note that G_1 can be solved from the first equation. After G_1 is obtained, D_1 can be solved from the second equation. After (G_1, D_1) are obtained, G_2 can be solved from the third equation. Finally, after (G_1, D_1, G_2) are obtained, D_2 can be solved from the fourth equation. Observe further, that each time, the equation to be solved is a linear one. Therefore, the above set of nonlinear equations is called sequentially linear. It can be shown that for any n-stage RC-ladder, Eq. (25) is sequentially linear. As a consequence, both the computational cost is reduced and accuracy is improved tremendously. Liu and Visvanathan [17] have provided a sufficient condition for a system whose diagnosis equation is sequentially linear. However, this condition may be too strong because it does not include the n-stage RC-ladder, although the latter motivated the study.

(C) Design of Test Points

If component parameter p of a LSDS is not diagnosable, how can one make it diagnosable? The obvious answer is that we need more measurements, or more test points. How many and what location? These are the questions faced by those who design the test points. This process at present is more of an art than a science. There are two results. Sen and Saeks [3] provide a test point selection algorithm. Visvanathan and Liu [17] provide a design procedure so that the diagnosis equation (25) is sequentially linear.

ACKNOWLEDGEMENT

This paper was supported in part by the Office of Navel Research under Grant N00014-78-C-0444.

VI. REFERENCES

[1] R. Saeks, S.P. Singh and R.W. Liu, "Fault Isolation Via Components Simulation", IEEE Trans. Circuit Theory, Vol. CT-19, pp. 634-640, Nov. 1972.

[2] M.N. Ransom and R. Saeks, "Fault Isolation with Insufficient Measurements", IEEE Trans. Circuit Theory, Vol. CT-20, pp. 416-417, 1973.

[3] N. Sen and R. Saeks, "Fault Diagnosis for Linear Systems via Multifrequency Measurements", IEEE Trans. Circuits and Systems, Vol. CAS-26 pp. 457-465, 1979.

[4] H.S.M Chen and R. Saeks, "A Search Algorithm for the Solution of Multifrequency Fault Diagnosis Equations", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 589-594, 1979.

[5] T.N. Trick, W. Mayeda and A.A. Sakla, "Calculation of Parameter Values from Node Voltage Measurements", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 466-474, 1979.

[6] S. Seshu and M.B. Reed, Linear Graphs and Electrical Networks, Addison Wesley, 1961.

[7] N. Navid and A.N. Willson, Jr., " A Theory and an Algorithm for Analog Circuit Fault Diagnosis", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 440-457, 1979.

[8] S.P. Singh and R. Liu, "Existence of State Equation Representation of Linear Large-Scale Dynamical Systems", IEEE Trans. Circuit Theory, Vol. CT-20, pp. 239-246, May 1973.

[9] D. Graupe, Identification of Systems, Krieger Publishing Co., 1976.

[10] P. Eykhoff, System Identification, John Wiley & Sons, 1977.

[11] B.L. Ho and R.E. Kalman, "Effective Construction of Linear State Variable Models from Input-Output Data", Proc. 3rd Ann. Allerton Conf. Circuit and System Theory, pp. 449-459, 1965.

[12] S.D. Bedrosian, (editor), Special Issue on Automatic Analog Fault Diagnosis, IEEE Trans. on Circuits and Systems, July 1979.

[13] L.O. Chua, "Device Modeling Via Basic Nonlinear Circuit Elements", Special Issue on Nonlinear Circuits and Systems, Nov. 1980, (To appear).

[15] J. Johnston, Econometric Methods, McGraw-Hill Book Company, 1970.

[15] C. Lin, "A Comparison of Frequency-domain Approach and Time-domain Approach: A Case Study", Technical Report EE-802, University of Notre Dame, May 1980.

[16] L. Weinberg, Network Analysis and Synthesis, McGraw-Hill Book, 1962.

[17] R. W. Liu and V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part I-Theory, IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 490-496, July 1979.

[18] V. Visvanathan and R.W. Liu, "Sequentially Linear Fault Diagnosis: Part II-Design", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 558-564, July 1979.

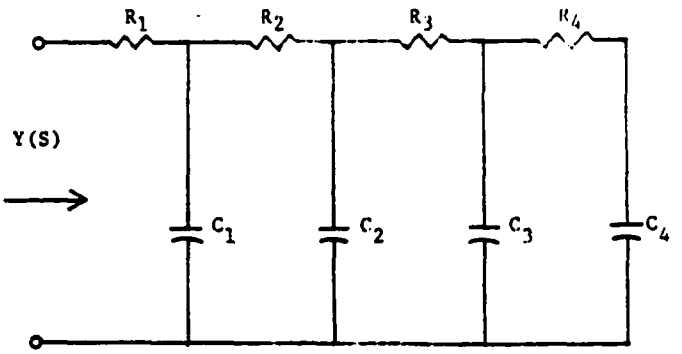


Fig. 4

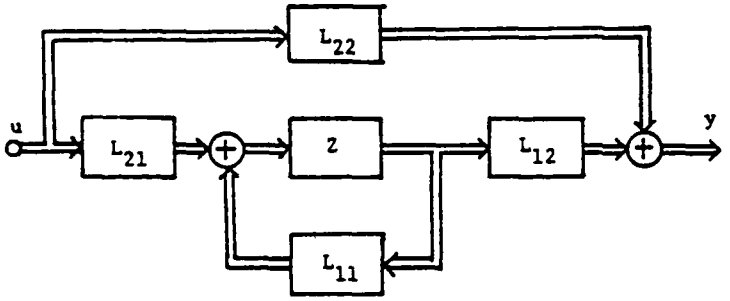


Fig. 1

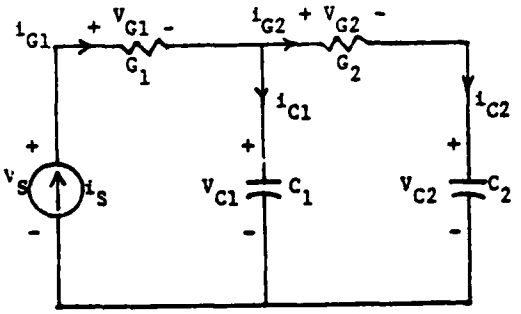


Fig. 2

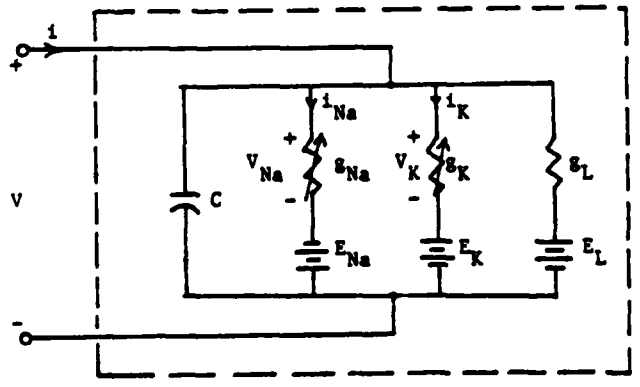


Fig. 3

A Comparison of Frequency-Domain Approach and Time-Domain Approach

by

C. Lin

A COMPARISON OF FREQUENCY-DOMAIN APPROACH AND TIME-DOMAIN APPROACH: A Case Study of Fault Analysis of Analog Circuits

by

Chen-Shang Lin

Technical Report No. EE-802

May 23, 1980

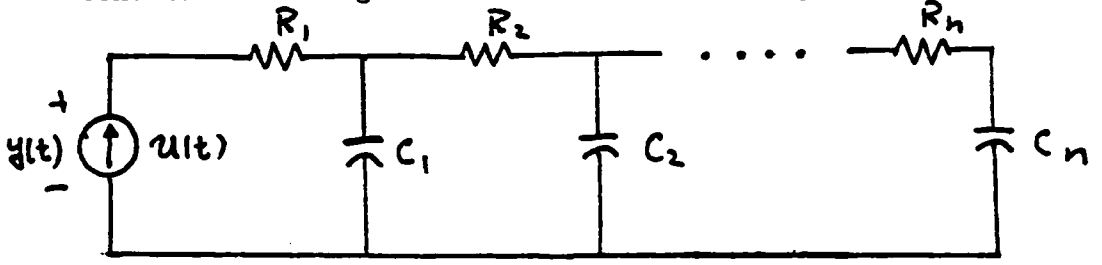
This report was supported in part by the Office of Navel Research under Grant N0014-78-C-0444.

I. Introduction

There have been considerable efforts expended in analog fault analysis. Most of them [1], [2], [3] employ frequency-domain approach, i.e., diagnosing faulty components from measured transfer function, while few [5] use time-domain approach to isolate faults by means of Markov parameters. Theoretically, both approaches are still under development and all seem feasible. It is the purpose of this report to compare these two approaches numerically by simulation on RC ladders. In this example, it is shown that the time-domain approach is far better than the frequency-domain approach.

II. Simulation and Results

Consider an n-stage RC ladder as shown in Figure 1.



The impedance $Z(s)$ of this RC-ladder has the form

$$Z(s) = \frac{a_0 s^n + a_1 s^{n-1} + \ldots + a_n}{s^n + b_1 s^{n-1} + \ldots + b_n}$$

Since there are only $(2n+1)$ coefficients to be determined, we need only to measure the impedance $Z(s_i)$ at $(2n+1)$ different sampling frequencies, s_i, $i = 1,2,\ldots,(2n+1)$.

Once $Z(s)$ is obtained, it can be expressed as

$$Z(s) = R_1 + \cfrac{1}{C_1 s + \cfrac{1}{\vdots \; R_i + \cfrac{1}{C_i s + \cfrac{1}{\vdots \; R_n + \cfrac{1}{C_n s}}}}}$$

and values of components can be calculated as

$$R_1 = \lim_{s\to\infty} \frac{N_1(s)}{D_1(s)}$$

$$C_1 = \lim_{s\to\infty} \frac{D_2(s)}{N_1(s)} \qquad (1)$$

$$R_i = \lim_{s\to\infty} \frac{N_i(s)}{D_i(s)}$$

$$C_i = \lim_{s\to\infty} \frac{D_{i+1}(s)}{N_i(s)} \qquad i = 2, 3, \ldots, n$$

where

N_1 = numerator polynomial of Z

D_1 = denominator polynomial of Z

$N_i = N_{i-1}(s) - C_{i-1} \cdot D_i(s)$

$D_i = D_{i-1}(s) - R_{i-1} \cdot N_{i-1}(s)$

On the other hand, the ladder has a state equation expression

$$\dot{\underline{x}} = A\underline{x} + Bu$$

$$y = C\underline{x} + Du$$

where $\underline{x}$ are the capacitor voltages, u the terminal current, y the terminal voltage and

$$\left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right] = \left[\begin{array}{ccccc|c} -\frac{1}{R_2C_1} & \frac{1}{R_2C_1} & & 0 & & \frac{1}{C_1} \\ +\frac{1}{R_2C_2} & \left(\frac{-1}{R_2C_2} & -\frac{1}{R_3C_2}\right) & \frac{1}{R_3C_2} & & 0 \\ & \ddots & \ddots & \ddots & \frac{1}{R_nC_{n-1}} & \vdots \\ & 0 & & \frac{1}{R_nC_n} & \frac{-1}{R_nC_n} & 0 \\ \hline 1 & 0 & \cdots & & 0 & R_1 \end{array}\right] \qquad (2)$$

The Markov parameters are given by

$$\begin{aligned} m_0 &= D \\ m_1 &= CB \\ m_2 &= CAB \\ &\vdots \\ m_k &= CA^{k-1}B \\ &\vdots \\ m_{2n-1} &= CA^{2n-2}B \end{aligned} \qquad (3)$$

which can be measured by a method developed by Liu and Suen [5].

Once the Markov parameters, m_i, $i = 0,1,\ldots,2n-1$, are obtained, the circuit parameters R's and C's can be solved from the simultaneous equations(2) and (3).

In this simulation, it assumes no measurement error for both Z(s) and m_i's . They are exact. We want to find the numerical error generated by solving (1) and (3).

Ladders of four and six stages were chosen, transfer functions and A, B, C, D parameters were calculated using nominal values of components. Then, as a way of comparison, the significant digits of coefficients of transfer function and entries of A, B, C, D were reduced before we performed the manipulation by these two methods. The results are listed in Tables 1 and 2.

It is clear from the tables that, as significant digits decrease, the estimated values of frequency-domain method stray away from nominal values gradually, then collapse abruptly at a certain point and become unrealizable, i.e., some of the values become negative. On the contrary, the results of the time-domain method remain about the same order of accuracy as parameters of state equation.

The discrepancies may be due to the following reasons:

1) The frequency-domain approach deals with computations of complex numbers while the time-domain approach deals with computations of real numbers.

2) The given circuit is sequentially-linear for the time-domain approach [5,6]. This can be demonstrated by the 4-stage RC ladder. Solving (2) and (3), we have

$$m_0 = R_1$$

$$m_1 = \frac{1}{C_1}$$

$$m_2 = -\frac{1}{R_2 C_1^2}$$

$$m_3 = \frac{1}{R_2^2 C_1^2}\left(\frac{1}{C_1} + \frac{1}{C_2}\right)$$

Therefore, R_1, C_1, R_2, C_2 can be solved sequentially by a set of <u>linear</u> equations,

$$R_1 = m_0$$

$$C_1 = \frac{1}{m_1}$$

$$R_2 = -\frac{1}{m_2 C_1^2}$$

$$\frac{1}{C_2} = m_3 R_2^2 C_1^2 - \frac{1}{C_1}$$

References

[1] R. Saeks, S.P. Singh and R.W. Liu, "Fault Isolation via Components Simulation", IEEE Transaction on Circuit Theory, Vol. CT-19, No. 6, pp. 634-640, November, 1972.

[2] N. Sen and R. Saeks, "Fault Diagnosis for Linear Systems Via Multifrequency Measurements," IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 457-465, 1979.

[3] T.N. Trick, W. Mayeda and A.A. Sakla, "Calculation of Parameter Value from Node Voltage Measurements," IEEE Trans. on Circuits and Systems, Vcl. CAS-26, pp. 466-474.

[4] R. Liu and L.C. Suen, "Minimal Dimension Realization and Identifiability of Input/Output Sequence," IEEE Trans. on Automatic Control, April 1977.

[5] R. Liu and V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part I-Theory," IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 490-496, 1979.

[6] V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part II-The Design of Diagnosable Systems," IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 558-564, 1979.

IV. Conclusion

This simulation strongly suggests that time-domain approach is more data-tolerant than frequency-domain approach in the sense that no sudden breakdown occurs and component values can be estimated with reasonable accuracy when the measurement is not accurate enough or where the noise must be taken into consideration. Thus, though it is still too early to have definite conclusion, time-domain approach seems to be a more promising method in attacking fault diagnosis problem.

Simulation on 6-Stage RC Ladder

	Time-Domain Method						Frequency-Domain Method					
	R_1	R_2	R_3	R_4	R_5	R_6	R_1	R_2	R_3	R_4	R_5	R_6
	C_1	C_2	C_3	C_4	C_5	C_6	C_1	C_2	C_3	C_4	C_5	C_6
Nominal Values	15	220	39	75	8.2	100	15	220	39	75	8.2	100
	.0022	.001	.015	.022	.0047	.01	.0022	.001	.015	.022	.0047	.01
Significant Digits 10	15	220	39	75	8.2	100	15	220	39	75	8.206	99.994
	.0022	.001	.015	.022	.0047	.01	.0022	.001	.015	.022004	.004696	.01
8	15	220	39	75	8.2	100	15	220	39	75.027	8.9232	99.25
	.0022	.001	.015	.022	.0047	.01	.0022	.001	.015	.022436	.0042666	.0099976
7	15	220	39	75	8.2	100	15	220	39.003	74.992	8.0302	100.18
	.0022	.001	.015	.022	.0047	.01	.0022	9.9994E-4	.015	.021883	.0048168	.010001
6	15	220	39	74.999	8.1999	100	15	219.98	39.023	77.659	-.67597	106.2
	.0022	.001	.015	.022	.0047001	.0099997	.0022	.0010004	.015015	-.022917	.049583	.010019
4	15	219.99	38.988	74.908	8.1865	100.16						
	.0022002	.0010001	.015011	.022036	.0047074	.0099507						
2	15	214.29	37.815	75.63	8.6555	120.64						
	.0022222	.001037	.015256	.021107	.0044222	.0073993						

Table 2: In the frequency-domain approach, the calculated values of C_4 and R_5 become negative when the significant digits reduced to 6. No such discrepancies in the time-domain approach.

Simulation on 4-stage RC Ladder

	Time-Domain Method				Frequency-Domain Method			
	R_1	R_2	R_3	R_4	R_1	R_2	R_3	R_4
	C_1	C_2	C_3	C_4	C_1	C_2	C_3	C_4
Nominal Values	2.2	47	8.2	1	2.2	47	8.2	1
	.015	.47	.01	.022	.015	.47	.01	.022
Significant Digits / 10	2.2	47	8.2	1	2.2	47	8.2	1
	.015	.47	.01	.022	.015	.47	.01	.022
8	2.2	47	8.2	1	2.2	47	1.3587	7.7603
	.015	.47	.01	.022	.015	.46597	.0056867	.030347
7	2.2	47	8.2	1	2.2	47.005	-.018072	9.1305
	.015	.47	.01	.022	.015	.70919	-.23768	.030482
6	2.2	47	8.2	1				
	.015	.47	.01	.022				
4	2.2	47.017	8.2044	1.0003				
	.014999	.46982	.0099962	.021997				
2	2.2	47.857	8.4454	1.0569				
	.014925	.46434	.0096776	.020555				

Table 1: In the frequency-domain approach, the calculated values of R_3 and C_3 become negative when the significant digits are reduced to 7. No such discrepancies in the time-domain approach.

An Alternative Consideration in Singular Linear State Estimation

by

Y.F. Huang
E. Fogel
J.B. Thomas

AN ALTERNATIVE CONSIDERATION ON SINGULAR LINEAR STATE ESTIMATION

Y.F. Huang
Dept. of Elec. Eng. and Computer Sci.
Princeton Univ.
Princeton, NJ 08544

E.Fogel
Dept. of Elec. Eng.
Univ. of Notre Dame
Notre Dame, IN 46556

J.B. Thomas
Dept. of Elec. Eng. and Computer Sci.
Princeton Univ.
Princeton, NJ 08544

Abstract

The problem of designing an optimal state estimator for a linear, discrete-time system with a singular noise covariance matrix is considered. In this article, this problem is cast as a constrained optimization problem and the approach appears to be more direct. Solution to this optimization problem gives a reduced-order optimal state estimator.

I. INTRODUCTION:

In a linear stochastic system, the output measurement may be only partially noise corrupted. Although, in practice, one may argue that there exist no noise-free measurements, it is quite possible that some of the measurements are noise corrupted while the others are relatively accurate. Under the Gaussian assumption, this implies that the noise covariance matrix has both large and small eigenvalues, which easily leads to numerical difficulties in the implementation of the Kalman filter. It is convenient in this case to model these more accurate measurements as noise-free entities.

The study of this problem dates back to the work of Bryson [1], for continuous-time systems, and that of Brammer [2], for discrete-time systems. Kwakernaak [3] and Anderson [4] discussed this problem as a singular linear state estimation problem; however, no explicit solutions were given. Tse and Athans [5] derived a rather complicated "observer-estimator" which is essentially an extension of the Luenberger observer [6]. Later, Yoshikawa [7] gave a simpler derivation for minimum-order optimal state estimators. More recently, Fairman [8] proposed a "hybrid estimator" which features "coordinatization" and achieved a reduced-order optimal estimator.

The main feature of the approach used in this paper is the following: After a proper similarity transformation, the state variables are decomposed into two parts, one to be estimated by a reduced-order filter and the other to be recovered exactly from the noise-free measurements. Then the dynamic equation of the latter part of the state equation is considered a constraint on the optimal estimation of the other part of the states. Hence the state estimation problem in this case is cast as a constrained optimization problem, which leads to a reduced-order optimal state estimator.

II. PROBLEM FORMULATION

A linear, discrete-time stochastic system can be described by the following equations

$$x(k+1)=A(k)\,x(k)+B(k)\,u(k),\quad k=0,1,2,\ldots \quad (1)$$

$$y(k) = C(k)\,x(k) + v(k),\quad k=1,2,\ldots \quad (2)$$

where $x(\cdot)\in R^n$, $u(\cdot)\in R^p$, and $y(\cdot)\in R^m$. To further specify the problem, the following assumptions are made:

(i) $x(0)$, $u(0)$, $u(1),\ldots,v(1)$, $v(2)$, ..., are independent random vectors with the following statistics

$E[x(0)]=x_0$ $\qquad$ $E[x(0)x^T(0)]=V_{x_0}$

$E[u(k)]=0\ \forall k$ $\qquad$ $E[u(k)u^T(k-i)]=V_u(k)\,\delta(i)\ \forall k,i$

$E[v(k)]=0\ \forall k$ $\qquad$ $E[v(k)v^T(k-i)]$

Presented at 23rd Midwest Symposium on Circuits and Systems, August 1980; to be published in the Proceedings of the Conference

$$= V_v(k)\,\delta(i) \quad \forall k, i$$

$$E[u(k)v^T(i)] = 0_{p\times m} \quad \forall k, i \qquad E[u(k)x^T(0)] = 0_{p\times n} \quad \forall k$$

$$E[v(k)x^T(0)] = 0_{m\times n} \quad \forall k$$

where $u^T(\cdot)$ and $v^T(\cdot)$ denotes the transpose of vectors $u(\cdot)$ and $v(\cdot)$, respectively, and $\delta(\cdot)$ deontes the Kronecker delta.

(ii) For any k, the $V_v(k)$ is a non-negative definite matrix with rank m_1, where $m_1 < m$. Under this assumption, implementation of the standard Kalman filter involves inversion of a matrix which may be singular. Tse and Athans [5] proposed an observer-estimator of order $n-m_2$ which performs as well as higher order estimators, where $m_2 \triangleq n-m_1$.

(iii) For any k, the C(k) is of full rank, i.e. every element of the output measurement is independent of the others.

The objective here is to design an optimal state estimator of order $n-m_2$. Without loss of generality, one can assume that

$$v(k) = \begin{bmatrix} v_1(k) \\ \hline 0 \end{bmatrix} \begin{matrix} \}m_1 \\ \}m_2 \end{matrix}$$

where $v_1(k) \in R^{m_1}$, and thus the covariance matrix of v(k) can be written as

$$V_v(k) = \left[\begin{array}{c|c} V_{v_1}(k) & 0_{m_1\times m_2} \\ \hline 0_{m_2\times m_1} & 0_{m_2\times m_2} \end{array}\right]$$

where $V_{v_1}(k)$ is strictly positive definite.

It is easy to see that there exists a non-singular matrix Q(k), such that the transformation

$$z(k) = Q(k)\,x(k) \tag{3}$$

yields the following state and measurement equations

$$\begin{bmatrix} z_1(k+1) \\ \hline z_2(k+1) \end{bmatrix} = \left[\begin{array}{c|c} \tilde{A}_{11}(k) & \tilde{A}_{12}(k) \\ \hline \tilde{A}_{21}(k) & \tilde{A}_{22}(k) \end{array}\right] \begin{bmatrix} z_1(k) \\ \hline z_2(k) \end{bmatrix} + \begin{bmatrix} \tilde{B}_1(k) \\ \hline \tilde{B}_2(k) \end{bmatrix} u(k) \tag{4}$$

and

$$\begin{bmatrix} y_1(k) \\ \hline y_2(k) \end{bmatrix} = \left[\begin{array}{c|c} \tilde{C}_{11}(k) & \tilde{C}_{12}(k) \\ \hline 0_{m_2\times n_1} & \tilde{C}_{22}(k) \end{array}\right] \begin{bmatrix} z_1(k) \\ \hline z_2(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ \hline 0 \end{bmatrix} \tag{5}$$

where

$$\tilde{A}(k) = Q(k+1)\,A(k)\,Q^{-1}(k)$$

$$\tilde{B}(k) = Q(k+1)\,B(k)$$

and

$$\tilde{C}(k) = C(k)\,Q^{-1}(k)$$

$z_1(\cdot) \in R^{n_1}, z_2(\cdot) \in R^{m_2}, y_1(\cdot) \in R^{m_1}, y_2(\cdot) \in R^{m_2}$, and $n_1 \triangleq n-m_2$. $\tilde{A}(\cdot)$, $\tilde{B}(\cdot)$ and $\tilde{C}(\cdot)$ are partitioned accordingly. Moreover, since C(k) is of full rank, $\tilde{C}_{22}(k)$ is invertible. Hence there exists a one-to-one correspondence between the state $z_2(k)$ and output $y_2(k)$, namely

$$z_2(k) = C_{22}^{-1}(k)\,y_2(k) \tag{6}$$

and thus in the state equation, (4), only $z_1(\cdot)$ must be estimated. The dynamic equations for $z_1(\cdot)$ and $z_2(\cdot)$ are

$$z_1(k+1) = \tilde{A}_{11}(k)z_1(k) + \tilde{A}_{12}(k)z_2(k) + \tilde{B}_1(k)u(k) \tag{7}$$

$$z_2(k+1) = \tilde{A}_{21}(k)z_1(k) + \tilde{A}_{22}(k)z_2(k) + \tilde{B}_2(k)u(k) \tag{8}$$

It is obvious from (7) and (8) that $z_1(k)$ and $z_2(k)$ are mutually dependent; therefore the estimation of $z_1(k)$ does depend on the dynamic behavior of $z_2(k)$. Thus the filtering problem becomes that of finding an optimal $\hat{z}_1(k+1|k+1)$ subject to (7) and constrained by (8), where $\hat{z}_1(k+1|k+1)$ denotes the estimate of $z_1(k+1)$ given measurements up to time k+1. Note, from (6) and (7), that the state $z_2(k)$ can be regarded as a deterministic input in the Kalman filtering problem.

3. THE REDUCED-ORDER OPTIMAL STATE ESTIMATOR

In this section, the optimal estimator for $z_1(k)$ is derived where the performance measure is the trace of the error covariance matrix. Defining the vector s(k) as

$$s(k) \triangleq z_1(k) - P(k)\,z_2(k) \;;\; P(k) \in R^{n_1\times m_2} \tag{9}$$

from (5), (7) and (8), one obtains

$$s(k+1)=F(k)s(k)+G(k)z_2(k)+M(k)u(k) \quad (10)$$

and

$$y_1(k)=H(k)s(k)+N(k)z_2(k)+v_1(k) \quad (11)$$

where

$$F(k) = \tilde{A}_{11}(k) - P(k+1)\,\tilde{A}_{21}(k) \quad (10.a)$$

$$G(k) = \tilde{A}_{12}(k) - P(k+1)\,\tilde{A}_{22}(k) + F(k)\,P(k) \quad (10.b)$$

$$M(k) = \tilde{B}_1(k) - P(k+1)\,\tilde{B}_2(k) \quad (10.c)$$

$$H(K) = \tilde{C}_{11}(k) \quad (11.a)$$

and

$$N(k) = \tilde{C}_{12}(k) + \tilde{C}_{11}(k)\,P(k) \quad (11.b)$$

Notice that P(k), as defined in the above equations, can be viewed as the Lagrange multiplier in the standard constrained optimization problem. Now, the problem of estimating $z_1(k)$ is replaced by that of estimating s(k) given the measurements $\{y_1(1), y_1(2), \ldots, y_1(k)\}$ and states $\{z_2(1), z_2(2), \ldots, z_2(k)\}$. From (9), it is obvious that

$$\hat{s}(k|k) = \hat{z}_1(k|k) - P(k)\,z_2(k) \quad (12)$$

and

$$V_{\tilde{s}}(k|k) = V_{\tilde{z}_1}(k|k) \quad (13)$$

where $\hat{s}(k|k)$ denotes the estimate of s(k) conditioned on input-output measurements up to time k, and

$$\tilde{s}(k|k) \triangleq \hat{s}(k|k) - s(k)$$

$$\tilde{z}_1(k|k) \triangleq \hat{z}_1(k|k) - z_1(k)$$

$$V_{\tilde{s}}(k|k) \triangleq E[\tilde{s}(k|k)\tilde{s}^T(k|k)]$$

The unbiased linear estimator of s(k) is given by the following n_1th-order filter

$$\hat{s}(k+1|k+1)=[I-K(k+1)H(k+1)]F(k)\hat{s}(k|k) + K(k+1)[y_1(k+1)-N(k+1)\,z_2(k+1) - H(k+1)G(k)z_2(k)] \quad (14)$$

Hence the error quantity $\tilde{s}(k|k)$ propagates as

$$\tilde{s}(k+1|k+1)=[I-K(k+1)H(k+1)]F(k)\tilde{s}(k|k) + [I-K(k+1)H(k+1)]M(k)u(k) - K(k+1)v_1(k+1) \quad (15)$$

and the error covariance matrix $V_{\tilde{s}}(\cdot|\cdot)$ is given by

$$V_{\tilde{s}}(k+1|k+1)=[I-K(k+1)H(k+1)]\Gamma(k+1)[I-K(k+1)H(k+1)]^T + K(k+1)V_{v_1}(k+1)K^T(k+1) \quad (16)$$

where

$$\Gamma(k+1) \triangleq F(k)V_{\tilde{s}}(k|k)F^T(k)+M(k)V_u(k)M^T(k) \quad (17)$$

Observe that $\Gamma(k+1)$ is essentially the one-step prediction error covariance matrix [9]. An optimal estimator is taken to be an estimator which minimizes the trace of the error covariance matrix. Therefore, it is left to minimize $\mathrm{Tr}[V_{\tilde{s}}(k+1|k+1)]$ with respect to $[K(k+1)\vdots P(k+1)]$. Notice that here K(k+1) plays the role of standard Kalman gain while P(k+1) is the Langrangian of the optimization problem. Minimizing $\mathrm{Tr}[V_{\tilde{s}}(k+1|k+1)]$ with respect to K(k+1) yields

$$K^*(k+1) = \Gamma(k+1)\,H^T(k+1)\,R^{-1}(k+1) \quad (18)$$

where $R(\cdot)$ is the positive-definite symmetric matrix given by

$$R(k+1) = H(k+1)\Gamma(k+1)H^T(k+1)+V_{v_1}(k+1) \quad (18.a)$$

Observe that (15)-(18) are identical to the formulation of the standard Kalman filter [9]. However, in this case, it is further required to optimize the state estimator with respect to the choice of P(k+1); *i.e.* minimize $\mathrm{Tr}[V_{\tilde{s}}(k+1|k+1)]$ with respect to P(k+1). Let the optimal P(k+1) which minimizes $\mathrm{Tr}[V_{\tilde{s}}(k+1|k+1)]$ be denoted by P*(k+1). Then it can be shown that

$$P^*(k+1) \in \Sigma^*(k+1) \quad (19)$$

where $\Sigma^*(k+1)$ is the set given by

$$\Sigma^*(k+1) = \{P:\ P\,\Lambda_1(k)=\Lambda_2(k),\ P \in R^{n_1 \times m_2}\} \quad (19.a)$$

$$\Lambda_1(k) = \tilde{A}_{21}(k)V_{\tilde{s}}(k|k)\tilde{A}_{21}^T(k)+\tilde{B}_2(k)V_u(k)\tilde{B}_2^T(k) \quad (19.b)$$

$$\Lambda_2(k) = \tilde{A}_{11}(k)V_{\tilde{s}}(k|k)\tilde{A}_{21}^T(k)+\tilde{B}_1(k)V_u(k)\tilde{B}_2^T(k) \quad (19.c)$$

Notice that, when $\Lambda_1(k)$ is non-singular, P*(k+1) is given by

$$P^*(k+1) = \Lambda_2(k)\,\Lambda_1^{-1}(k) \quad (19.d)$$

The set $\Sigma^*(k+1)$ will be discussed in the next section.

Once P*(k+1) is found, the matrices F(k), G(k), and M(k) can be specified and denoted by F*(k), G*(k), M*(k), respectively, by substituting P*(k+1) and P*(k) into (10). Similarly, H*(k+1), $\Gamma^*(k+1)$ and R*(k+1) can be obtained from (11.b), (17), and (18.a).

All in all, the reduced-order optimal

state estimator is formulated by the following equations

$$\hat{s}(k+1|k+1) = [I-K^*(k+1)H(k+1)]F^*(k)\hat{s}(k|k) + K^*(k+1)[y_1(k+1) - H^*(k+1) z_2(k+1) - H(k+1)G\hat{e}(k) z_2(k)] \quad (20.a)$$

$$\hat{z}_1(k+1|k+1) = \hat{s}(k+1|k+1) + P^*(k+1)\, z_2(k+1) \quad (20.b)$$

$$z_2(k+1) = \tilde{C}_{22}^{-1}(k+1)\, y_2(k+1) \quad (20.c)$$

$$V_{\tilde{z}}(k+1|k+1) = [I-K^*(k+1)H(k+1)]\Gamma^*(k+1) \quad (20.d)$$

$$K^*(k+1) = \Gamma^*(k+1)H^T(k+1)[H(k+1)\Gamma^*(k+1) H^T(k+1) + V_{v_1}(k+1)]^{-1} \quad (20.e)$$

$$\Gamma^*(k+1) = F^*(k)V_{\tilde{z}}(k|k)F^{*T}(k) + M^*(k)V_u(k)$$
$$M^{*T}(k) = [I:-P^*(k+1)]\,[\tilde{A}(k) V_{\tilde{z}}(k|k)\tilde{A}^T(k) + \tilde{B}(k)V_u(k)\tilde{B}^T(k)]\,[I:-P^*(k+1)]^T \quad (20.f)$$

where

$$V_{\tilde{z}}^*(k|k) = \left[\begin{array}{c:c} V_{z_1}^*(k|k) & 0_{m_1 \times m_2} \\ \hdashline 0_{m_2 \times m_1} & 0_{m_2 \times m_2} \end{array}\right]$$

due to the fact that $z_1(k)$ is exactly measurable for every k. Also, the error covariance matrix $V_{\tilde{z}}^*(k+1|k+1)$ is given by

$$V_{\tilde{z}}^*(k+1|k+1) = [(\Gamma^*(k+1))^{-1} + H^T(k+1) V_{v_1}(k+1) H(k+1)]^{-1} \quad (20.g)$$

and $P^*(k+1)$ is specified by Eqs. (19).

4. COMMENTS ON THE REDUCED-ORDER OPTIMAL ESTIMATOR

The formulation for the optimal state estimator derived in last section, (19)-(20), is identical to that of the "hybrid estimator" given in [8], except that in [8] a deterministic input is inserted to the system dynamics. However, the approach here is more straightforward and it is clearer here that the choice of P(k+1) is crucial to the optimality of the estimator. It can be seen that the general coordinate transformation discussed in [8] is split into two coordinate transformations: one which depends on the system output matrix H only and one which depends on the matrices P_{11} and P_{12}. It can be seen that the similarity transformation $Q(\cdot)$ of (3) is equivalent to $J^{-1}(\cdot)$ defined in (14) of [8], and thus the state variable z defined in (3) can be regarded as equivalent to ζ defined in (14) of [8]. Furthermore, the variable s defined in (9) can be similarly related to the variable ξ_1 in (20) of[8].

The relations in (19) which govern the choice of $P^*(k+1)$ are vital to the understanding of the optimal state estimator, and thus deserves some detailed discussion. First, notice that

$$\mathcal{N}(\Lambda_1(k)) \subset \mathcal{N}(\Lambda_2(k))$$

where $\mathcal{N}(\Lambda_1(k))$ and $\mathcal{N}(\Lambda_2(k))$ denote the null spaces of $\Lambda_1(k)$ and $\Lambda_2(k)$, respectively.

The following observation is thus made:

Observation The set $\Sigma^*(k+1)$ given in (19.a) is a non-empty set. Moreover, if $\Lambda_1(k)$ is singular, any member $P(k+1) \in \Sigma^*(k+1)$ yields the same estimator performance.

Now, according to the value of $\Lambda_1(k)$, the following special cases are of interest:

Case 1: $\Lambda_1(k) = 0_{m_2 \times m_2}$. In this case, $\Lambda_2(k) = 0_{n_1 \times m_2}$ and thus

$$\Sigma^*(k+1) = R^{n_1 \times m_2}$$

This case is possible if (8) does not contain any information pertaining to the estimation of $z_1(k)$; for example, if $\tilde{A}_{21}(k) = 0_{m_2 \times n_1}$ and $\tilde{B}_2(k) = 0_{m_2 \times p}$. An extreme example for this case is that $m_2 = 0$, i.e. all measurements are noise corrupted. In this condition, the estimator presented in Section 3 is identical to the standard full-order Kalman filter whose performance is independent of the choice of P(k+1).

Case 2: $\Lambda_1(k) \geq 0_{m_2 \times m_2}$; i.e. $\Lambda_1(k)$ is a singular non-zero matrix. In this case, only some components of $z_2(k)$ contain information about $(u(k), z_1(k))$. Thus the similarity transformation discussed in Section 2 can be redefined so as to isolate only those elements of $z_2(k)$ which constitute a constraint on $(u(k), z_1(k))$. Hence the Lagrange multiplier P(k+1) that should be considered is an element in $R^{(n-r)\times r}$, where $r < m_2$. Alternatively, any member in $\Sigma^*(k+1)$ can be used in the filter realization.

Case 3: $\Lambda_1(k)$ is positive-definite. This condition can be fulfilled when $V_u(k)$ is positive-definite for any $k = 0,1,2,\ldots$. In this case, $\Sigma^*(k+1)$ contains one and

only one element $P^*(k+1)$, which is given by (19.d).

When $P^*(\cdot)$ is uniquely specified (Case 3), one can compare the error covariance matrix given by (20.g) with that obtained for arbitrary $P(\cdot)$ and observe the same expression for $V_{\tilde{z}}(k+1|k+1)$. The difference is that $\Gamma^*(k+1)$ of (20.f) has the following property

$$\mathrm{Tr}[\Gamma^*(k+1)] < \mathrm{Tr}[\Gamma(k+1)]$$

where $\Gamma(\cdot)$ is obtained from non-optimal $P(\cdot)$.

Finally, the implementation of this estimator should be initiated as follows:

$$\hat{x}(0|0) = E[x(0)] = x_0$$

i.e.

$$\hat{z}_1(0|0) = E[z_1(0)]$$

therefore

$$V_{\tilde{z}}(0|0) = \left[\begin{array}{c|c} V_{z_1}(0) & 0_{m_1 \times m_2} \\ \hline 0_{m_2 \times m_1} & 0_{m_2 \times m_2} \end{array}\right] = Q(0) V_{x_0} Q^T(0)$$

and

$$P^*(0) = 0_{n_1 \times m_2}$$

5. CONCLUSION

A reduced-order optimal state estimator for a linear, discrete-time system associated with a singular noise covariance matrix has been derived in this paper. The main idea in this derivation is to cast this singular state estimation problem as a constrained optimization problem. The estimator derived here is fundamentally the same as that derived by Fairman [8]. The major differences are: the approach here is more straightforward, the optimality of the estimator is more explicitly exposed and, furthermore, the possibility of nonuniqueness of $P^*(\cdot)$ is discussed here.

It is worth mentioning that the estimator given here requires lower order matrix inversion than the standard full-order Kalman filter does in the singular case; thus the computational efficiency is improved. This estimation procedure can be applied similarly to smoothing and predicting problems or systems with colored noise.

REFERENCES

[1] A.E. Bryson, Jr. and D.E. Johansen, "Linear filtering for time-varying systems using measurements containing colored noise," IEEE Trans. Automat. Contr., vol.AC-10, pp.4-10, Jan. 1965.

[2] K.G. Brammer, "Lower order optimal linear filtering of non-stationary random sequence," IEEE Trans. Automat. Contr., vol. AC-13, pp.198-199, April 1968.

[3] H. Kwakernaak and R. Sivan, Linear Optimal Control Systems. Wiley Interscience, 1972.

[4] B.D.O. Anderson and J.B. Moore, Optimal Filtering. Prentice-Hall, 1979.

[5] E. Tse and M. Athans, "Optimal minimal-order observer-estimators for discrete linear time varying systems," IEEE Trans. Automat. Contr., vol. AC-15, pp.416-426, Aug. 1970.

[6] D.G. Luenberger, "Observing the state of a linear system," IEEE Trans. Mil. Elec., vol. MIL-8, pp.74-80, April, 1964.

[7] T. Yoshikawa, "Minimal-order optimal filters for discrete-time linear stochastic system," Int. J. of Contr., vol. 21, pp.1-19, Jan. 1975.

[8] F.W. Fairman, "Hybrid estimators for discrete-time stochastic systems," IEEE Trans. Sys. Man., and Cyber., vol. SMC-8, pp.849-854, Dec. 1978.

[9] J.S. Meditch, Stochastic Optimal Linear Estimation and Control. New York: McGraw-Hill Book Co., Inc., 1969.

ACKNOWLEDGEMENT:

The authors wish to express their appreciation to Dr. R.W. Liu for his valuable comments during the course of this work. This research is supported in part by the U.S. Army Research Office under contract DAAG 29-79-C-0024, and in part by the Office of Naval Research under contract N00014-78-C-0444.

Reduced-Order Optimal State Estimator for Linear Systems with Partially Noise Corrupted Measurement

by

E. Fogel
Y.F. Huang

Reduced-Order Optimal State Estimator for Linear Systems with Partially Noise Corrupted Measurement

ELI FOGEL AND Y. F. HUANG

Abstract—**The problem of reduced-order optimal state estimation for linear systems with singular noise covariance matrix is studied. It is shown that the optimal estimator is somewhat different from the Kalman filter.**

The state estimator problem in the singular case can be cast as a constrained optimization problem. Solving this optimization problem yields the truly optimal estimator. The estimator derived here is of the form of the hybrid estimator of Fairman [7]. However, the derivations here are somewhat more direct.

I. Introduction

The problem of state estimation with partially noise corrupted measurements is of practical importance since it is often the case that certain measurements are significantly more accurate than other. This introduces the necessity to construct the state estimator assuming a *singular* (almost singular) *noise covariance matrix* (henceforth, referred to as the singular estimation problem.) Furthermore, it is well known that if the measurement noise is colored, the estimation problem can be formulated as a singular estimation problem [1], [2]. Surprisingly enough, with all the attention given to this problem, standard text (e.g., [1],[2]) and an abundance of papers (e.g., [3]–[5]) present a variety of suboptimal rather than optimal solutions to this classical problem.

Manuscript received March 24, 1980. This work was supported in part by the U.S. Office of Naval Research under Grant N00014-78-C-0444.

The authors are with the Department of Electrical Engineering University of Notre Dame, Notre Dame, IN 46556.

In this correspondence we derive the *true* optimal reduced-order state estimator for a linear system with singular noise covariance matrix. The results reported here are by no means new. Fairman [7] has derived the same estimator using his concept of "optimal coordination." However, Fairman's procedure is somewhat indirect and, thus, the optimality of his hybrid estimator is not transparent.

The derivation of the optimal estimator here is based on the observation that the singular state estimation problem can be reduced to a simple *constrained optimization problem* which yields a Kalman-type estimator.

The correspondence is organized as follows. The problem is formally formulated and in Section III its solution is presented. In Section IV the state estimator, as presented here, is discussed.

II. Problem Formulation

The standard Markov–Gauss model is considered, i.e., the state equation is given by

$$x(k+1)=A(k)x(k)+B(k)u(k) \qquad k=0,1,2,\cdots \tag{1}$$

and the measurement equation is

$$y(k)=C(k)x(k)+v(k) \qquad k=1,2,\cdots \tag{2}$$

where $x(\cdot)\in\mathcal{R}^n$, $u(\cdot)\in\mathcal{R}^p$, and $y(\cdot)\in\mathcal{R}^m$. $x(0)$, $u(0)$, $u(1),\cdots$, $v(1)$, $v(2),\cdots$, are independent Gaussian random vectors with the following statistics:

$$E[x(0)]=x_0 \qquad \operatorname{cov}[x(0)]=\Sigma \tag{3}$$

$$E[u(i)]=0 \qquad E[u(i)u'(j)]=V_u(i)\delta(i-j) \quad \forall i,j \tag{4}$$

$$E[v(i)]=0 \qquad E[v(i)v'(j)]=V_v(i)\delta(i-j) \quad \forall i,j \tag{5}$$

$$E[u(i)v'(j)]=0 \qquad E[u(i)x'(0)]=0 \quad E[v(i)x'(0)]=0 \tag{6}$$

where $v'(j)$ denotes the transpose of the vector $v(j)$. It is assumed that $V_v(i)$ are nonnegative definite matrices with rank $m-m_1$, where $m_1<m$. In this case, the standard Kalman filter solution involves the inversion of a matrix which may be singular. Furthermore, as shown in [1], an observer–estimator of order $n-m_1$ may be constructed which performs as well as the higher order state estimators.

Thus, our objective is to construct the optimal $n-m_1$ state estimator. With no loss of generality [1], we assume that

$$v(k)=\begin{bmatrix} v_1(k) \\ \cdots \\ 0 \end{bmatrix} \tag{5a}$$

where $v_1(k)\in\mathcal{R}^{m-m_1}$ and, thus, the covariance matrix of $v(k)$ can be written as

$$V_v(k)=\begin{bmatrix} \overbrace{V_{v_1}(k)}^{m-m_1} & \overbrace{0}^{m_1} \\ \cdots & \cdots \\ 0 & 0 \end{bmatrix} \begin{matrix} \}m-m_1 \\ \\ \}m_1 \end{matrix} \tag{5b}$$

where $V_{v_1}(k)$ is strictly positive definite. It is simple to see that there exists a nonsingular matrix $Q(k)$ such that the transformation $Z(k)=Q(k)x(k)$ yields the following state and measurement equations

$$\begin{bmatrix} Z_1(k+1) \\ \cdots \\ Z_2(k+1) \end{bmatrix}=\begin{bmatrix} \tilde{A}_{11}(k) & \vdots & \tilde{A}_{12}(k) \\ \cdots & \cdots & \cdots \\ \tilde{A}_{21}(k) & \vdots & \tilde{A}_{22}(k) \end{bmatrix}\begin{bmatrix} Z_1(k) \\ \cdots \\ Z_2(k) \end{bmatrix}+\begin{bmatrix} \tilde{B}_1(k) \\ \cdots \\ \tilde{B}_2(k) \end{bmatrix}u(k) \tag{6}$$

$$\begin{bmatrix} y_1(k) \\ \cdots \\ y_2(k) \end{bmatrix}=\begin{bmatrix} \tilde{C}_{11}(k) & \vdots & \tilde{C}_{12}(k) \\ \cdots & \cdots & \cdots \\ 0 & \vdots & \tilde{C}_{22}(k) \end{bmatrix}\begin{bmatrix} Z_1(k) \\ \cdots \\ Z_2(k) \end{bmatrix}+\begin{bmatrix} v_1(k) \\ \cdots \\ 0 \end{bmatrix} \tag{7}$$

where

$$\tilde{A}(k)=Q(k+1)A(k)Q^{-1}(k)$$

$$\tilde{B}(k)=Q(k+1)B(k)$$

$$\tilde{C}(k)=C(k)Q^{-1}(k).$$

0018-9286/80/1000-0994$00.75

and where $Z_1(\cdot)\in\mathcal{R}^{n-m_1}$, $Z_2(\cdot)\in\mathcal{R}^{m_1}$, $y_1(\cdot)\in\mathcal{R}^{m-m_1}$, $y_2(\cdot)\in\mathcal{R}^{m_1}$ and the matrices $\tilde{A}$, $\tilde{B}$, and $\tilde{C}$ are partitioned, accordingly. Furthermore, we assume with no loss of generality, that $C(k)$ is of full rank and, thus, $\tilde{C}_{22}(k)$ is invertible. This implies that

$$Z_2(k)=C_{22}^{-1}(k)y_2(k) \tag{8}$$

and, thus, in the state model (6) and (7) only $Z_1(\cdot)$ must be estimated. To emphasize this point, we rewrite (7) as two equations

$$Z_1(k+1)=\tilde{A}_{11}(k)Z_1(k)+\tilde{A}_{12}(k)Z_2(k)+\tilde{B}_1(k)u(k) \tag{9}$$

and

$$Z_2(k+1)=\tilde{A}_{21}(k)Z_1(k)+\tilde{A}_{22}(k)Z_2(k)+\tilde{B}_2(k)u(k). \tag{10}$$

The filtering problem is, thus, that of finding the optimal $\hat{Z}_1(k+1/k+1)$ subject to the dynamic equation (9) and the constraint in (10). Note that the unmeasurable quantities in (9) are $Z_1(k)$ and $u(k)$. These quantities are constrained by equality (10). The vector $Z_2(k)$ can be handled as a deterministic input in the Kalman filtering problem.

In the next section, we derive the optimal observer for $Z_1(k)$, where the performance measure is the trace of the error covariance matrix.

III. The Optimal State Estimator

Defining the variable $S(k)$

$$S(k)\triangleq Z_1(k)-P(k)Z_2(k);\qquad P(k)\in\mathcal{R}^{(n-m_1)\times m_1}, \tag{11}$$

we obtain from (7), (9), and (10),

$$S(k+1)=F(k)S(k)+G(k)Z_2(k)+M(k)u(k) \tag{12a}$$

$$y_1(k)=H(k)S(k)+N(k)Z_2(k)+v_1(k) \tag{12b}$$

where

$$F(k)=\tilde{A}_{11}(k)-P(k+1)\tilde{A}_{21}(k) \tag{12.1}$$

$$G(k)=\tilde{A}_{12}(k)-P(k+1)\tilde{A}_{22}(k)+F(k)P(k) \tag{12.2}$$

$$M(k)=\tilde{B}_1(k)-P(k+1)\tilde{B}_2(k) \tag{12.3}$$

$$H(k)=\tilde{C}_{11}(k) \tag{12.4}$$

$$N(k)=\tilde{C}_{12}(k)+\tilde{C}_{11}(k)P(k). \tag{12.5}$$

We note that $P(k)$ can be viewed as the Lagrange multiplier in the standard constrained optimization problem. Now, the problem of estimating $Z_1(k)$ is replaced by that of estimating $S(k)$ given the measurements $\{y_1(1), y_1(2), \cdots, y_1(k)\}$. Obviously, from the definition of $S(k)$ in (11) we have

$$\hat{S}(k/k)=\hat{Z}_1(k/k)-P(k)Z_2(k) \tag{13}$$

and

$$V_{\tilde{S}}(k/k)=V_{\tilde{Z}}(k/k) \tag{14}$$

where

$$\tilde{S}(k/k)\triangleq\hat{S}(k/k)-S(k)$$

$$\tilde{Z}_1(k/k)\triangleq\hat{Z}_1(k/k)-Z_1(k)$$

Any unbiased estimator of $S(k)$ is given by the $(n-m_1)$th order filter

$$\hat{S}(k+1/k+1)=[I-K(k+1)H(k+1)]F(k)\hat{S}(k/k)+K(k+1)\cdot[y_1(k+1)-N(k+1)Z_2(k+1)-H(k+1)G(k)Z_2(k)] \tag{19}$$

$$\hat{S}(k+1/k+1)=[I-K(k+1)H(k+1)]F(k)\hat{S}(k/k)-[I-K(k+1)H(k+1)]M(k)u(k)+K(k+1)v_1(k+1). \tag{19a}$$

To obtain the optimal estimator, we have to minimize the trace of the error covariance matrix

$$V_{\tilde{S}}(k+1/k+1)=[I-K(k+1)H(k+1)]\Gamma(k+1)\cdot[I-K(k+1)H(k+1)]'+K(k+1)V_{v_1}(k+1)K'(k+1) \tag{20}$$

where

$$\Gamma(k+1)=F(k)V_{\tilde{S}}(k/k)F'(k)+M(k)V_u(k)M^T(k) \tag{21}$$

with respect to $[K(k+1), P(k+1)]$. We note that $K(k+1)$ plays the role of the Kalman gain, whereas $P(k+1)$ is the Lagrangian of the opumization problem. Furthermore, we note that $\Gamma(k+1)$ is the one-step prediction error covariance matrix $V_{\tilde{S}}(k+1/k)$. Minimizing tr $[V_{\tilde{S}}(k+1/k+1)]$ yields

$$K^*(k+1)=\Gamma^*(k+1)H'(k+1)\mathbf{R}^{-1}(k+1)^* \tag{22}$$

where $\mathbf{R}(\cdot)$ is the positive definite matrix

$$\mathbf{R}(k+1)=H(k+1)\Gamma(k+1)H'(k+1)+V_v(k+1). \tag{22a}$$

The optimal $P(k+1)$ is denoted by $P^*(k+1)$. $\mathbf{R}^*(\cdot)$ and $\Gamma^*(\cdot)$ in (22) denote $\mathbf{R}(\cdot)$ and $\Gamma(\cdot)$ of (22a) and (21), respectively, expressed as functions of the optimal $P^*(k+1)$. Optimization with respect to $P(k+1)$ yields

$$P^*(k+1)\in\mathcal{P}^*(k+1) \tag{23a}$$

where $\mathcal{P}^*(k+1)$ is the set

$$\mathcal{P}^*(k+1)=\left\{P\in\mathcal{R}^{(n-m_1)\cdot m_1}:\ P\Lambda_1(k)=\Lambda_2(k)\right\} \tag{23b}$$

$$\Lambda_1(k)=\tilde{A}_{21}(k)V_{\tilde{S}}(k/k)\tilde{A}_{21}'(k)+\tilde{B}_2(k)V_u(k)\tilde{B}_2'(k) \tag{23c}$$

$$\Lambda_2(k)=A_{11}(k)V_{\tilde{S}}(k/k)A_{21}'(k)+B_1(k)V_u(k)B_2'(k). \tag{23d}$$

The relations in (23) are vital to the understanding of the optimal filter and, thus, merit some discussion. Since $\mathcal{N}(\Lambda_1(k))\subset\mathcal{N}(\Lambda_2(k))$, where $\mathcal{N}(A)$ denotes the null space of A, the following observation is made.

Observation: $\mathcal{P}^*(k+1)$ is not an empty set. Furthermore, if $\mathcal{P}^*(k+1)$ is not a singleton ($\Lambda_1(k)$-singular), any member $P(k+1)\in\mathcal{P}^*(k+1)$ yields the same filter performance.

The invariance of the performance over the set $\mathcal{P}^*(k+1)$ can easily be checked via consideration of the error covariance matrix as in (25) below.

According to the value of $\Lambda_1(k)$, the following special cases are of interest.

Case 1: $\Lambda_1(k)=0$. Note that in this case $\Lambda_2(k)=0$ and, thus, $\mathcal{P}^*(k+1)=\mathcal{R}^{(n-m_1)\cdot m_1}$. This case is possible if (10) does not contain any information pertaining to the estimation of $Z_1(k)$. For example, if $\tilde{A}_{21}(k)=0$ and $\tilde{B}_2(k)=0$, etc.

Case 2: $\Lambda_1(k)>0$ ($\Lambda_1(k)$ is a singular nonzero matrix.) In this case, only some components of $Z_2(k)$ contain information on $(u(k), Z_1(k))$. In this case, the transformation $Q(k)$ discussed in the previous section can be defined so as to isolate only those elements in $Z_2(k)$ which constitute a constraint on $(u(k), Z_1(k))$ and, thus, the Lagrange multiplier $P(k)$ that should be considered is in $\mathcal{R}^{(n-r)\cdot r}$ where r is the number of such constraints $r<m_1$. Alternatively, any member in $\mathcal{P}^*(k+1)$ can be used in the filter realization.

Case 3: $\Lambda_1(k)>0$. In this case $P^*(k+1)$ is uniquely given by

$$P^*(k+1)=\Lambda_2(k)\Lambda_1^{-1}(k). \tag{24}$$

Thus, the optimal filter equations are given by

$$\hat{S}(k+1/k+1)=[I-K^*(k+1)H(k+1)]F^*(k)\hat{S}(k/k)+K^*(k+1)[y_1(k+1)-N^*(k+1)Z_2(k+1)-H(k+1)G^*(k)Z_2(k)] \tag{25a}$$

$$\hat{Z}_1(k+1/k+1)=\hat{S}(k+1/k+1)+P^*(k+1)Z_2(k+1) \tag{25b}$$

$$Z_2(k+1)=\tilde{C}_{22}^{-1}(k+1)y_2(k+1) \tag{25c}$$

$$V_2^*(k+1/k+1)=[I-K^*(k+1)H(k+1)]\Gamma^*(k+1)$$
$$=\Gamma^*(k+1)-\Gamma^*(k+1)H'(k+1)[H(k+1)\Gamma^*(k+1)\cdot H'(k+1)+V_v(k+1)]^{-1}H(k+1)\Gamma^*(k+1)$$
$$\doteq\left[(\Gamma^*(k+1))^{-1}+H'(k+1)V_v(k+1)H(k+1)\right]^{-1} \quad (25d)$$

where the equality $\doteq$ is meaningful only if the inverse exists.

$$\Gamma^*(k+1)=F^*(k)V_2^*(k/k)F^*(k)'+M(k)V_u(k)M^*(k)'$$
$$=\left[I \vdots -P^*(k+1)\right]\left[\tilde{A}(k)V_2^*(k/k)\tilde{A}'(k)+\tilde{B}(k)V_u(k)\tilde{B}'(k)\right]\begin{bmatrix} I \\ \cdots \\ -P^*(k+1)\end{bmatrix} \quad (25e)$$

where

$$V_2^*(k/k)=\begin{bmatrix} V_{Z_1}^*(k/k) & \vdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \vdots & 0\end{bmatrix}$$

due to the fact that Z_2 is exactly measurable.

Remarks:

1) Comparing the error covariance in (25d) with that obtained for an arbitrary $P(\cdot)$, we observe the same expression for $V(k+1/k+1)$ as function of $\Gamma(k+1)$. The difference is that $\Gamma^*(k+1)$ of (25e) satisfies $\Gamma^*(k+1)<\Gamma(k+1)$ where $\Gamma(k+1)$ obeys (25e) with nonoptimal $P(\cdot)$. (See, for example, expression (36) of [3].)

2) $\Gamma^*(k+1)$ of (25e) represents the covariance of the prediction error $\tilde{Z}_1(k+1/k)$ as can be easily proved. Thus, our optimization may be viewed as two-step optimization:

Step 1: finding the optimal predictor,

Step 2: finding the optimal filter.

The optimal gain $K^*(k+1)$ is given by

$$K^*(k+1)=\Gamma^*(k+1)H^T(k+1)\cdot\left[H(k+1)\Gamma^*(k+1)H'(k+1)+V_{v_1}(k+1)\right]^{-1} \quad (26a)$$

and the Lagrangian satisfies

$$P^*(k+1)\in\mathcal{P}^*(k+1) \quad (26b)$$

where $\mathcal{P}^*(k+1)$ is given in (23). Equations (25)–(26) are initiated by the values

$$\hat{S}(0/0)=E[S(0)] \quad \text{or} \quad \hat{Z}_1(0/0)=E[Z_1(0)]$$
$$P^*(0)=0$$
$$V_{Z_1}(0/0)=V_{Z_1}(0).$$

Remarks:

1) If all the measurements are noise corrupted, i.e., $m_1=0$, our filter reduces to the standard Kalman filter.

2) If all measurements are noise free, i.e., $m_1=m$, our filter performs exactly as the "optimal minimal-order observer" of [2] as can be seen from the comparison of our error covariance matrix [(25d)–(25e)] to theirs [(42) in [4]]

The optimal state estimator of (26) is depicted in Fig. 1.

IV. Discussion

As stated earlier, the state estimator derived here is fundamentally the same as that derived by Fairman. The major differences are: 1) our approach is direct and, thus, the optimality of the estimator is guaranteed; and 2) in Fairman's paper, a filter equivalent to ours is presented, but the possibility of nonunique $P^*(\cdot)$ is overlooked [cases 1)–2) of the previous section].

The reader might wonder about the results here as compared to those found in literature. The difference is in the optimal choice of the matrix

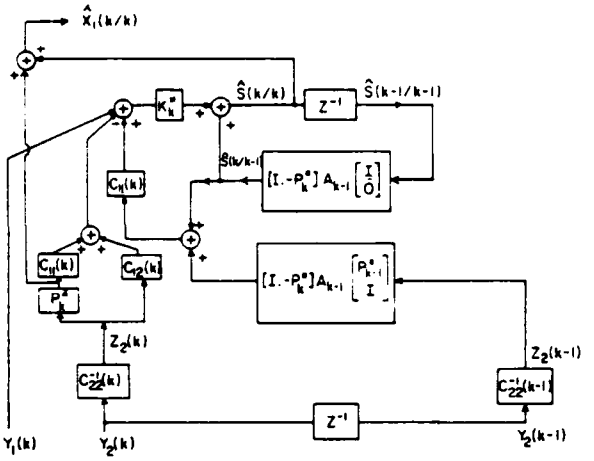


Fig. 1. The reduced-order optimal estimator.

$P(\cdot)$, i.e., $P^*(\cdot)$. Note that the innovations are white regardless of the choice of $P(\cdot)$, which led to the erroneous conclusion that the optimality is achieved by considering $Z_2(\cdot)$ to be a deterministic input and subsequently ignoring (10).

Finally, it should be noted that in constructing an optimal predictor or an optimal smoother, the same procedure as presented in Section III may be utilized. Furthermore, the idea of solving the singular state estimation as constrained optimization problem extends straightforwardly to the continuous time case.

References

[1] H. Kuakernack and R. Sivan, *Linear Optimal Control Systems*. New York: Wiley, 1972.

[2] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[3] E. Tse and M. Athans, "Optimal minimal-order observer-estimator for discrete time-varying systems," *IEEE Trans. Automat. Contr.*, vol. AC-15 1970.

[4] C. T. Leondes and L. M. Novak, "Optimal minimal order observers for discrete time systems—A unified theory," *Automatica*, vol. 8, 1972.

[5] A. E. Bryson and D. E. Johansen, "Linear filtering for time varying systems using measurements containing colored noise," *IEEE Trans. Automat. Contr.*, vol. AC-10, 1965.

[6] F. W. Fairman, "Reduced order state estimation for discrete time stochastic systems," *IEEE Trans. Automat. Contr.*, vol. AC-22, 1977.

[7] F. W. Fairman, "Hybrid estimators for discrete time stochastic systems," *IEEE Trans. Syst. Man, Cyber.*, vol. SMC-8, 1978.

[8] F. W. Fairman, "Optimal observers for a class of continuous linear time-varying stochastic systems," *IEEE Trans. Automat. Contr.*, vol. AC-22, 1977.

[9] F. W. Fairman, "On stochastic observer estimators for continuous time systems," *IEEE Trans. Automat. Contr.*, vol. AC-22, 1977.

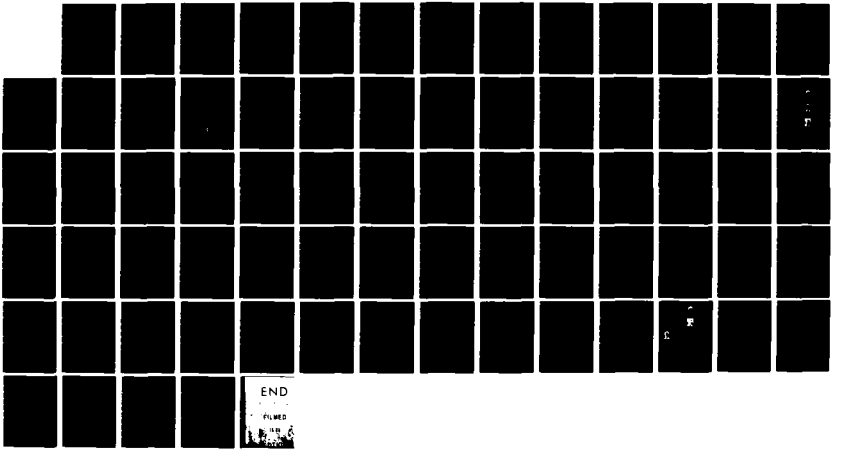
Spectral Analysis of Previewing Controllers

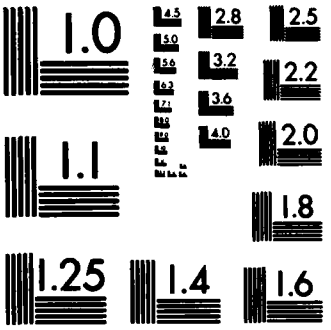
by

E. Fogel
K. McGill

AD-A133 491 ANALOG FAULT DIAGNOSIS OF LARGE-SCALE ELECTRONIC CIRCUITS(U) NOTRE DAME UNIV IN DEPT OF ELECTRICAL ENGINEERING R LIU 01 AUG 83 N00014-78-C-0444 2/2

UNCLASSIFIED F/G 14/2 NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Spectral Analysis of Previewing Controllers

ELI FOGEL AND KEVING McGILL

Abstract—**Analytical expressions for the performance of optimal and suboptimal control policies in trajectory tracking problems are derived. Special attention is devoted to the finite preview problem. In the time-invariant infinite-duration case correspondence of the performance indices for finite energy, finite power, and stationary stochastic trajectories, all with the same spectral density, is established. Using this correspondence, the performance of different control policies is studied in the frequency domain. Simulation results demonstrate the facility of the theory of this paper.**

I. Introduction

One of Bellman's major contributions to optimal control theory is his principle of optimality [2]. This principle implies that to optimally steer a system along a trajectory the entire trajectory must be available. Unfortunately, most practical controllers can only preview the immediate future of the trajectory and must rely on statistical information to characterize the trajectory beyond their preview. A common example of a previewing controller is the driver of an automobile, and we refer to a previewing controller as a driver, and to a driver with complete trajectory information as a prescient driver.

While control policies for previewing controllers have been presented in the literature, little attention has been paid to evaluating their performance. Tomizuka and Whitney [7], [8] have derived the optimal driver under the assumption that the trajectory is generated by a known linear system driven by Gaussian noise. They also give a rule of thumb for calculating the preview length which results in good performance. Miller [4] has presented a simple, well-performing driver with no *a priori* knowledge of the trajectory's statistics. Fogel [3] has evaluated the difference between the control values of the optimal driver and the optimal prescient driver. However, until now, simulation has been the only way to evaluate the performance of any of these drivers.

The purpose of this paper is to derive analytical expressions for driver performance. In Section II arbitrary trajectories are considered in the time-varying case. An integral expression for the performance is derived which depends explicitly on the trajectory and the preview length but is independent of the course of the plant state for a broad class of linear drivers. This result employs the concept of performance penalty which was used by Wernersson to solve the related optimal pursuit problem [10]. As an example, this result is used to optimize a driving policy.

In Section III the expected performance in following a stochastic trajectory of known spectral density is considered. The problem is restricted to the infinite-duration time-invariant case to allow analysis in the frequency domain. The key result is that the expected performance in following a stochastic trajectory may be calculated by considering a much simpler deterministic trajectory with the same spectral density.

Using this result, the performances of the optimal prescient driver, of Miller's driver, and of Tomizuka's driver are calculated in the frequency domain. If the spectral density can be identified as that of the output of a linear system driven by white noise, the performance can be related to the solution of algebraic Riccati equations.

The results in Sections II and III are stated for continuous time. The corresponding discrete time relations are reported in the Appendix. In Section IV computer simulation results are discussed. These simulations demonstrate the validity of the theory and provide some insight into it. Section V concludes the paper.

II. The Driving Problem and Performance Evaluation

The difference between a controller's actual performance and the performance of the optimal prescient controller may be thought of as the controller's penalty. The optimal prescient performance is known from the optimal tracking problem, and a remarkably simple, but less well-known, expression for penalty has been presented by Wernersson [10]. In this section, these ideas are applied to the time-varying driving problem.

The plant to be controlled is described by its state equation

$$\dot{x}(t)=A(t)x(t)+B(t)u(t); \qquad x(t_0)=x_0 \tag{2.1a}$$

where $x(t)\in R^n$, $u(t)\in R^r$, and $A(t)$ and $B(t)$ are matrices of appropriate dimensions. The system matrices are assumed to be time varying throughout this section although for notational convenience the time dependence will often be suppressed.

The control policy is assumed to be of the form

$$u(\cdot)=u(t, x(t), \{y_T(\tau): \tau\leq(t+T)\})$$

where $Y_T=\{y_T(t): t\in[t_0, t_f], y_T(t)\in R^q\}$ is the trajectory to be followed, and T is the preview length. Linear control policies of the form

$$u(t, x(t), Y_T)=-K(t)x(t)+\int_{t_0}^{t_f}H(t,\tau)y_T(\tau)\,d\tau \tag{2.2}$$

will be denoted by the pair $(K(t), H(t,\tau))$. $H(t,\tau)$ can be considered the impulse response matrix of a noncausal "trajectory filter," and the preview constraint means that $H(t,\tau)=0\ \forall\tau>t+T$. The state trajectory of the plant for a particular $u(\cdot)$, t_0, x_0, and Y_T will be denoted by

$$X(t_0, x_0, Y_T, u(\cdot))=\{x(t; t_0, x_0, Y_T, u(\cdot)): t\in[t_0, t_f]\}.$$

The control objective is to minimize the performance index

$$J(t_0, x_0, Y_T, u(\cdot))=\int_{t_0}^{t_f}L(t, x(t; t_0, x_0, Y_T, u(\cdot)), Y_T, u(\cdot))\,dt \tag{2.3a}$$

where

$$L(t, x(t), Y_T, u(\cdot))=\|y_p(t)-y_T(t)\|^2_{Q(t)}+\|u(t, x(t), Y_T)\|^2_{R(t)} \tag{2.3b}$$

$$y_p(t)=C(t)x(t); \qquad y_p(t)\in R^q \tag{2.1b}$$

and $Q(t)>0$ and $R(t)>0$.

Manuscript received September 14, 1979; revised March 25, 1980. Paper recommended by A. Manitius, Chairman of the Optimal Systems Committee. This work was supported in part by the U.S. Office of Naval Research under Grant N00014-78-C-0444.

E. Fogel was with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556. He is now with the Charles Stark Draper Laboratory, Cambridge, MA 02139.

K. McGill was with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556. He is now with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305.

0018-9286/80/1000-0959$00.75

The characteristic which distinguishes the driving problem from the familiar tracking problem of optimal control theory (see, e.g., [1]) is that only a subset of Y_T is available to the controller. If $T > t_f - t_0$ the problems are identical and the controller will be called prescient.

The optimal prescient control policy is known from the tracking problem [1] to be

$$u^0(t, x(t), Y_T) = -R^{-1}B'P(t)x(t) + R^{-1}B'\xi(t) \qquad (2.4a)$$

where

$$\dot{P}(t) = -PA - A'P + PBR^{-1}B'P - C'QC; \qquad P(t_f) = 0 \quad (2.4b)$$

$$\dot{\xi}(t) = [-A' + PBR^{-1}B']\xi - C'Qy_T; \qquad \xi(t_f) = 0 \quad (2.4c)$$

Notice that $\xi(t)$ is independent of $X(t_0, x_0, Y_T, u^0(\cdot))$, and that $P(t)$ is independent of both $X(t_0, x_0, Y_T, u^0(\cdot))$ and Y_T. $u^0(\cdot)$ is of the form of (2.2) and will be denoted $u^0(\cdot) = (K^0(t), H^0(t, \tau))$, where $K^0(t) = R^{-1}B'P(t)$ and $H^0(t, \tau)$ is the impulse response of the trajectory filter whose state equation is (2.4c). This filter is purely anticipatory, that is, $H^0(t, \tau) = 0 \ \forall \tau < t$. The performance index for this driver is presented in a new way in the following lemma.

Lemma 1:

$$J(t_0, x_0, Y_T, u^0(\cdot)) = \|x_0 - P^{-1}(t_0)\xi(t_0)\|^2_{P(t_0)} + \int_{t_0}^{t_f} \|y_T(t) - CP^{-1}(t)\xi(t)\|^2_{Q(t)}\, dt. \quad (2.5)$$

Thus the optimal performance index is the sum of a term which can be made to vanish by a proper choice of x_0 and a term which is independent of $X(t_0, x_0, y_T, u^0(\cdot))$.

Proof: The proof employs the technique used by Wernersson in [10]. The familiar expression for $J(t_0, x_0, Y_T, u^0(\cdot))$ is, from [1],

$$J(t_0, x_0, Y_T, u^0(\cdot)) = x_0'P(t_0)x_0 - 2x_0'\xi(t_0) + \zeta(t_0) \qquad (2.6a)$$

where

$$\dot{\zeta}(t) = \xi'BR^{-1}B'\xi - y_T'Qy_T; \qquad \zeta(t_f) = 0. \qquad (2.6b)$$

Let

$$\gamma(t) = J(t, P^{-1}(t)\xi(t), Y_T, u^0(\cdot)) = \zeta(t) - \xi'(t)P^{-1}(t)\xi(t). \quad (2.7)$$

Notice that from (2.6a)

$$\gamma(t) = \min_x J(t, x, Y_T, u^0(\cdot))$$

and so since from (2.3) $J(t_f, x, Y_T, u(\cdot)) = 0 \ \forall x$ (assuming Y_T is well behaved at t_f), it is clear that $\gamma(t) \to 0$ as $t \to t_f$ even though $P^{-1}(t_f)$ is not defined. Therefore, from (2.6) and (2.7),

$$J(t_0, x_0, Y_T, u^0(\cdot)) = \|x_0 - P^{-1}(t_0)\xi(t_0)\|^2_{P(t_0)} - \int_{t_0}^{t_f} \gamma(t)\, dt.$$

Differentiating (2.7) and substituting the appropriate expressions from (2.4b), (2.4c), and (2.6b), establish the lemma.

The above ideas are extended to suboptimal control policies in the following lemma which defines the performance penalty paid by a control policy and states the remarkable expression for its evaluation which was first presented by Wernersson in [10].

Lemma 2:

$$J_{pen}(t_0, x_0, Y_T, u(\cdot)) \triangleq J(t_0, x_0, Y_T, u(\cdot)) - J(t_0, x_0, Y_T, u^0(\cdot)) \qquad (2.8a)$$

$$= \int_{t_0}^{t_f} \|u(t, x(t; x_0, t_0, Y_T, u(\cdot)), Y_T) - u^0(t, x(t; x_0, t_0, Y_T, u(\cdot)), Y_T)\|^2_R\, dt. \qquad (2.8b)$$

This idea is illustrated in Fig. 1. The proof is given after the following corollary.

Corollary 2.1: The penalty paid by the control policy $u(\cdot) = (K^0(t), H(t, \tau))$ is independent of $X(t_0, x_0, Y_T, u(\cdot))$ and is given by

$$J_{pen}(t_0, x_0, Y_T, u(\cdot)) = \int_{t_0}^{t_f} \left\| \int_{t_0}^{t} [H(t, \tau) - H^0(t, \tau)] y_T(\tau)\, d\tau \right\|^2_R dt. \qquad (2.8c)$$

Proof: In this proof the notation $x(t; t_0, x_0, Y_T, u(\cdot))$ will be abbreviated to $x(t; \cdot)$. The method used is similar to that of Wernesson in [10]. Let

$$\rho(t) \triangleq J(t, x(t; \cdot), Y_T, u^0(\cdot)) = x'(t; \cdot)P(t)x(t; \cdot) - 2\xi'(t)x(t; \cdot) + \zeta(t)$$

From (2.6), $\rho(t_0) = J(t_0, x_0, Y_T, u^0(\cdot))$ and $\rho(t_f) = 0$, and so from (2.8a) and (2.2a)

$$J_{pen}(t_0, x_0, Y_T, u(\cdot)) = \int_{t_0}^{t_f} [L(t, x(t; \cdot), Y_T, u(\cdot)) + \dot{\rho}(t)]\, dt.$$

Differentiating $\rho(t)$, substituting the appropriate expressions from (2.1a), (2.4b), (2.4c), and (2.6b), and combining with (2.3b), establishes Lemma 2. The corollary follows easily from Lemma 2 and (2.2).

Example—The Optimal Limited Preview Driver: To demonstrate the power of Lemma 2, the optimal limited preview control policy with no *a priori* knowledge of the trajectory characteristics will now be derived. Assuming a control policy of the form $(K(t), H(t, \tau))$, the penalty is, from (2.8b) and (2.2),

$$J_{pen}(t_0, x_0, Y_T, (K, H)) = \int_{t_0}^{t_f} \left\{ \left\| [K(t) - K^0(t)] x(t; t_0, x_0, Y_T, (K, H)) + \int_{t_0}^{\min(t_f, t+T)} [H(t, \tau) - H^0(t, \tau)] y_T(\tau)\, d\tau - \int_{\min(t_f, t+T)}^{t_f} H^0(t, \tau) y_T(\tau)\, d\tau \right\|^2_R dt. \qquad (2.10)$$

With no *a priori* information on $\{y_T(\tau): \tau > t+T\}$ at time instant t, a "reasonable" choice of the control policy is

$$K(t) = K^0(t)$$

$$H(t, \tau) = \begin{cases} H^0(t, \tau) & \tau < t+T \\ 0 & \tau > t+T \end{cases}. \qquad (2.11)$$

In statistical terms, assuming that the trajectory is a martingale process is compatible with the assumption of no *a priori* information on the trajectory characteristics. Namely, we assume

$$E[y_T(\tau)/\sigma_{t+T}] = 0 \qquad \forall \tau > t+T \qquad (2.12)$$

where σ_{t+T} is the σ-algebra induced by the measurements $\{y(\tau) = \tau < t+T\}$. Now using (2.10) one obtains

$$E[J_{pen}(\cdot)] = E\left[\int_{t_0}^{t_f} \left\{ \left\| [K(t) - K^0(t)] x(t) + \int_{t_0}^{\min(t+T, t_f)} [H(t, \tau) - H^0(t, \tau)] y_T(t)\, d\tau \right\|^2_R - 2\left([K(t) - K^0(t)] x(t) + \int_{t_0}^{\min(t+T, t_f)} [H(t, \tau) - H^0(t, \tau)] y_T(\tau)\, d\tau \right)' R \cdot E\left[\int_{\min(t+T, t_f)}^{t_f} H^0(t, \tau) y_T(\tau)\, d\tau / \sigma_{t+T} \right] + E\left[\left\| \int_{\min(t+T, t_f)}^{t_f} H^0(t, \tau) y_T(\tau)\, d\tau \right\|^2 / \sigma_{t+T} \right] \right\} dt \right] \qquad (2.13)$$

where the identity $E[\int f(t)\,dt] = E\int E[f(t)/a_{t+T}]\,dt$ has been used. Revoking the martingale assumption in (2.12) the second term in (2.13) nullifies to yield

$$\min_{K(\cdot),\,H(\cdot)} E[J_{pen}(\cdot)] = E\int_{t_0}^{t_f}\left\|\int_{\min(t+T,\,t_f)}^{t_f} H^0(t,\tau)y_T(\tau)\,d\tau\right\|_R^2 dt$$

where the *optimal* control policy is (2.11). Note that the martingale-type assumption amounts to little statistical information of the stochastic process $\{y_T(\cdot)\}$. Further justification to the choice of the policy (2.10) is given in the remarks following Example 2 of the following section.

The results obtained above are of academic interest. However, their generality does not allow their application to the solution of practical problems such as the design of the preview length.

To obtain further insight into the finite preview problem, we attach some structural properties to the trajectory to be followed. Particularly, we consider the infinite duration problem, with time-invariant plant and spectral information on the trajectory.

III. Spectral Analysis of the Performance Index

Since the actual trajectory is rarely available *a priori*, it is important to be able to evaluate a driver's expected performance over an ensemble of possible trajectories. This section considers a simple but useful class of ensembles, namely stationary stochastic processes characterized by spectral density. The problem is restricted to the infinite duration, time invariant case and is analyzed in the frequency domain. The main result is that the performance depends upon the spectral density in the same way for stochastic as for deterministic processes. This result is used to calculate the expected performance of the optimal prescient driver, of a driver which ignores the trajectory's spectral density (Miller's driver), and of a driver which uses knowledge of the trajectory's spectral density optimally (Tomizuka's driver).

The *following assumptions* are made in this section.

Assumption (A1) A, B, C, Q, and R are constant matrices.

Assumption (A2) The plant is completely controllable and observable.

Assumption (A3) $t_0 = -\infty$ and $t_f = +\infty$.

Assumption (A4) $x_0 = 0$.

Assumption (A5) The control policy is of the form $u(\cdot) = (K, H(t-\tau))$, where K is constant, $A - BK$ is stable, and $\int_{-\infty}^{\infty}\|H(\eta)\|^2\,d\eta < \infty$.

In this section t_0 and x_0 will be dropped as arguments of J and x.

The following three classes of trajectory are considered in this section.

I. Finite energy, i.e., $\int_{-\infty}^{\infty}\|y_T(t)\|^2\,dt < \infty$.

II. Finite power, i.e., $\langle\|Y_T\|^2\rangle < \infty$, where

$$\langle Y\rangle \triangleq \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} y(t)\,dt.$$

III. Zero-mean finite-variance wide-sense stationary stochastic process, i.e., $E\{y_T(t)\} = 0$; $E\{y_T(t+\tau)y_T'(t)\} = R_y(\tau) < \infty$.

The performance index $J(Y_T, u(\cdot))$ must be defined differently depending on the class of Y_T as shown in Table I. Table I also defines the spectral density of Y_T for each class. Using the appropriate definitions of performance index and spectral density, the following theorem applies to all three classes of trajectory.

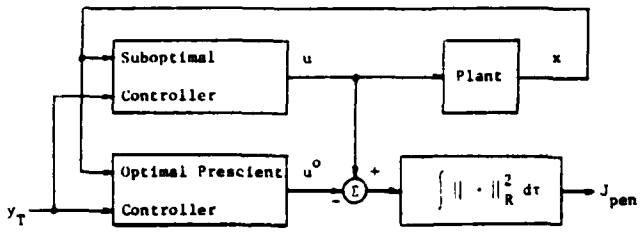


Fig. 1. J_{pen} as the energy out of a system driven by the trajectory.

Theorem 1: Under the assumptions (A1) through (A5), the performance index of a control policy $u(\cdot)$ following a trajectory Y_T whose spectral density is $\tilde{S}_y(\omega)$ is

$$J(Y_T, u(\cdot)) = \frac{1}{2\pi}\operatorname{tr}\int_{-\infty}^{\infty}\left(\tilde{S}_y^{1/2}(\omega)\right)^*\tilde{\Gamma}'(\omega)\tilde{S}_y^{1/2}(\omega)\,d\omega \qquad (3.4)$$

where $(\tilde{S}_y^{1/2}(\omega))^*\tilde{S}_y^{1/2}(\omega) = \tilde{S}_y(\omega)$, the gain function $\tilde{\Gamma}(\omega) = \tilde{G}_u(\omega)R\tilde{G}_u^*(\omega) + \tilde{G}_e(\omega)Q\tilde{G}_e^*(\omega)$, and $\tilde{G}_u(\omega)$ and $\tilde{G}_e(\omega)$ are defined in the proof.

Proof: Under the assumptions of time invariance and stability made in this section, stable transfer functions exist between y_T and u and between y_T and $y_T - y_P$. Calling these transfer functions $\tilde{G}_u(\omega)$ and $\tilde{G}_e(\omega)$ respectively, it is a straightforward exercise in harmonic analysis to establish (3.4) for each class of trajectory. Notice that assumption (A5) is necessary to eliminate any contribution to J from the initial condition x_0 for trajectories of class I.

Remarks: This theorem is important because it allows the performance of drivers to be examined in the frequency domain, and because it allows the performance index of a driver following a trajectory of class II or III to be calculated by evaluating the performance index of the driver following a trajectory of class I. In the following examples, trajectories of class III will be modeled, as is common, as the output of the following stable time-invariant "trajectory generator" driven by white noise

$$\dot{x}_T(t) = A_T x_T(t) + B_T w(t); \qquad x_T(-\infty) = 0 \qquad (3.5a)$$

$$y_T(t) = C_T x_T(t) \qquad (3.5b)$$

where $\tilde{S}_w(\omega) = I$. The corresponding class I trajectory of the same spectral density is the impulse response of the trajectory generator, namely

$$y_{IR}(t) = \begin{cases} 0 & t < 0 \\ C_T e^{A_T t}B_T & t > 0 \end{cases}. \qquad (3.5c)$$

Example 1: The Optimal Prescient Driver: It is known from [1] that under assumptions (A1) through (A3), the optimal prescient control policy is $u^0(\cdot) = (K^0, H^0(t-\tau))$, where

$$K^0 = R^{-1}B'P_{11} \qquad (3.6a)$$

$$H^0(t-\tau) = \begin{cases} R^{-1}B'e^{-\Lambda'(t-\tau)}C'Q & \tau > t \\ 0 & \tau < t \end{cases} \qquad (3.6b)$$

$$\Lambda = A - BR^{-1}B'P_{11}. \qquad (3.6c)$$

TABLE I

Class	I. Finite Energy	II. Finite Power	III. Stochastic	
Performance Index	$J = \int_{-\infty}^{\infty} L\,dt$	$J = \langle L\rangle$	$J = E\{\langle L\rangle\}$	(3.1)
Spectral Density	$\int_{-\infty}^{\infty} y_T(t)y_T'(t+\tau)\,dt \leftrightarrow \tilde{S}(\omega)$	$\langle y_T(t+\tau)y_T'(\tau)\rangle \leftrightarrow \tilde{S}(\omega)$	$E(\langle y_T(t+\tau)y_T'(\tau)\rangle) \leftrightarrow \tilde{S}(\omega)$	(3.2)

where $X(t) \leftrightarrow \tilde{X}(\omega)$ denotes the Fourier transform pair

$$\tilde{X}(\omega) = \int_{-\infty}^{\infty} X(t)e^{-j\omega t}\,dt \qquad (3.3a)$$

$$X(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{X}(\omega)e^{j\omega t}\,d\omega. \qquad (3.3b)$$

P_{11} is the steady-state solution of (2.4b) and satisfies the Riccati equation given in (3.7). Thus, Λ is stable. If Y_{IR} is generated by (3.5) the optimal control may be stated

$$u^0(t, x(t), Y_{IR}(t)) = -K^0 x(t) + R^{-1}B'P_{12}x_T(t)$$

where $P_e \triangleq \begin{bmatrix} P_{11} & -P_{12} \\ -P_{12}' & P_{22} \end{bmatrix}$ satisfies the Riccati equation

$$0 = -P_e A_e - A_e' P_e + P_e B_e R^{-1} B_e' P_e - C_e' Q C_e \tag{3.7a}$$

with

$$A_e = \begin{bmatrix} A & 0 \\ 0 & A_T \end{bmatrix}; \quad B_e = \begin{bmatrix} B \\ 0 \end{bmatrix}; \quad C_e = [C : -C_T]. \tag{3.7b}$$

The performance index of the optimal prescient driver is given in the following lemma.

Lemma 3:

$$J(Y_T, u^0(\cdot)) = \operatorname{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\tilde{S}_y^{1/2}(\omega)\right)^* \cdot \left\|I - CP_{11}^{-1}(-j\omega I - \Lambda')^{-1}C'Q\right\|_Q^2 \tilde{S}_y^{1/2}(\omega)\, d\omega. \tag{3.8a}$$

Furthermore, if Y_T is generated by (3.5), then

$$J(Y_T, u^0(\cdot)) = B_T'\left[P_{22} - P_{12}'\hat{P}_{11}P_{12}\right]B_T \tag{3.8b}$$

where

$$\hat{P}_{11}\Lambda' + \Lambda\hat{P}_{11} = -BR^{-1}B'. \tag{3.8c}$$

Proof: Applying Parseval's formula to the integral in (2.5) yields (3.8a). Note that the first term in (2.5) is zero since (3.6b) guarantees $\xi(-\infty) = 0$ for a finite energy Y_T. (3.8b) is obtained by considering the suboptimal control policy $u(\cdot) = (K^0, 0)$ following $y_{IR}(t)$. Clearly $x(t; Y_{IR}, (K^0, 0)) = 0$ and $u(t, x(t; Y_{IR}, (K^0, 0)), Y_{IR}) = 0$ $\forall t$. Therefore (2.8c) and (2.2) together with (3.5c) and (3.6b) yield

$$J_{\text{pen}}(Y_{IR}, (K^0, 0)) = \int_{-\infty}^{0}\left\|\int_0^{\infty} H^0(t-\tau)y_{IR}(\tau)\, d\tau\right\|_R^2 dt + \int_0^{\infty}\left\|\int_t^{\infty} H^0(t-\tau)y_{IR}(\tau)\, d\tau\right\|_R^2 dt$$

$$J(Y_{IR}, (K^0, 0)) = \int_0^{\infty}\|y_{IR}(t)\|_Q^2\, dt.$$

The optimal performance is then given by (2.8a)

$$J(Y_{IR}, u^0) = J(Y_{IR}, (K^0, 0)) - J_{\text{pen}}(Y_{IR}, (K^0, 0)).$$

Substituting expressions (3.5c) and (3.6b) into the penalty and suboptimal performance yields (3.8b). Details are given in [5].

Note that the performance calculations are made over the entire interval $(-\infty, +\infty)$, although the impulse is not applied to generate y_{IR} until $t = 0$. The significance of this result is that it gives a lower bound on the performance index of any driver for the infinite-duration problem. Furthermore, it should be noted that for class III trajectories, the control (3.6) may not yield $J(\cdot)$ as in (3.8) for a particular realization of the input $\{w(t)\}$. However, the performance of the lemma will be observed on the average. In the case of second-order ergodicity, the actual performance will be that of (3.8) for almost all realizations of the white input to the trajectory generator.

The performance penalty of suboptimal policies can be expressed in the frequency domain as in the following lemma.

Lemma 4: Under (A1) through (A5) the penalty paid by the control policy $u(\cdot) = (K^0, H(t-\tau))$ is

$$J_{\text{pen}}(Y_T, u(\cdot)) = \operatorname{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left\|\left[\tilde{H}(\omega) - \tilde{H}^0(\omega)\right]\tilde{S}_y^{1/2}\right\|_R^2 d\omega. \tag{3.9}$$

Proof: This result follows easily by applying Parseval's formula to (2.8c). This lemma will be used in the following examples of suboptimal drivers.

Example 2—Miller's Driver [4]: Miller [4] has proposed the time-invariant version of the optimal limited preview driver with no *a priori* knowledge of the trajectory characteristics, namely $u^M(\cdot) = (K^0, H^M(t-\tau))$, where

$$H^M(t-\tau) = \begin{cases} H^0(t-\tau) & \tau \le t+T \\ 0 & \tau > t+T \end{cases} \tag{3.10}$$

The performance penalty of Miller's controller is given in the following lemma.

Lemma 5:

$$J_{\text{pen}}(Y_T, u^M(\cdot)) = \operatorname{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\tilde{S}_y^{1/2}(\omega)\right)^* \cdot \left\|R^{-1}B'e^{\Lambda' T}(-j\omega I - \Lambda')^{-1}C'Q\right\|_R^2 \tilde{S}_y^{1/2}(\omega)\, d\omega. \tag{3.11a}$$

Furthermore, if Y_T is generated by (3.5), then

$$J_{\text{pen}}(Y_T, u^M(\cdot)) = B_T'\left[P_{12}'e^{\Lambda' T}\hat{P}_{11}e^{\Lambda' T}P_{12} + \hat{P}_{22}\right]B_T \tag{3.11b}$$

where

$$\hat{P}_{22}A_T + A_T'\hat{P}_{22} = -P_{12}'e^{\Lambda' T}BR^{-1}B'e^{\Lambda' T}P_{12}. \tag{3.11c}$$

Proof: (3.11a) is obtained by transforming (3.6b) and (3.10) and substituting into (3.9). (3.11b) is obtained by using (3.5c), (3.6b), and (3.10) to write (2.8c) as

$$J_{\text{pen}}(Y_T, u^M(\cdot)) = \int_{-\infty}^{-T}\left\|\int_0^{\infty} H^0(t-\tau)y_{IR}(\tau)\, d\tau\right\|_R^2 dt + \int_{-T}^{\infty}\left\|\int_{t+T}^{\infty} H^0(t-\tau)y_{IR}(\tau)\, d\tau\right\|_R^2 dt$$

and then substituting (3.5c) and (3.6b).

Remarks: In the frequency domain, lack of knowledge about the trajectory characteristics means that energy must be considered equally probable at all frequencies, i.e., that $\tilde{S}_y(\omega) = I$. That $u^M(\cdot)$ minimizes $J_{\text{pen}}(Y_T, u(\cdot))$ for $\tilde{S}_y(\omega) = I$, among all control policies of preview length T, is clear if Parseval's formula is applied to (3.9) to give

$$J_{\text{pen}}(Y_T, u(\cdot)) = \operatorname{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty}\|\tilde{H}(\omega) - \tilde{H}^0(\omega)\|_R^2\, d\omega = \operatorname{tr}\int_{-\infty}^{-T}\|H^0(t)\|_R^2\, dt + \operatorname{tr}\int_{-T}^{\infty}\|H(t) - H^0(t)\|_R^2\, dt.$$

Example 3—Tomizuka's Driver [7]–[9]: Tomizuka and Whitney [7]–[9] have derived the optimal driver assuming that the stochastic trajectory generator as given in (3.5) is known to the controller. Their control law is

$$u^T(t, \cdot) = u^M(t, \cdot) + R^{-1}B'e^{\Lambda' T}P_{12}\hat{x}_T(t+T) \tag{3.12}$$

where $\hat{x}_T(t+T)$ is the best estimate of the state of the trajectory generator given $\{y_T(\tau): \tau \le t+T\}$. $u^T(\cdot)$ is the best estimate of $u^0(\cdot)$, as seen by writing the last term of (3.12) as

$$R^{-1}B'e^{\Lambda' T}P_{12}\hat{x}_T(t+T) = \int_{t+T}^{\infty} H^0(t-\tau)\hat{y}_T(\tau)\, d\tau$$

where

$$\hat{y}_T(\tau) = C_T e^{A_T(\tau - t - T)}\hat{x}_T(t+T), \qquad \tau > t+T.$$

Lemma 6: If $C_T = I$, Tomizuka's driver pays the penalty

$$J_{\text{pen}}(Y_T, u^T(\cdot)) = \operatorname{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\tilde{S}_x^{1/2}(\omega)\right)^*\left\|R^{-1}B'e^{\Lambda' T} \cdot \left[(-j\omega I - \Lambda')^{-1}C'Q - P_{12}\right]\right\|_R^2 \tilde{S}_x^{1/2}(\omega)\, d\omega. \tag{3.13a}$$

Furthermore, if X_T is generated by (3.5) and $x_T(t+T)$ is measurable,

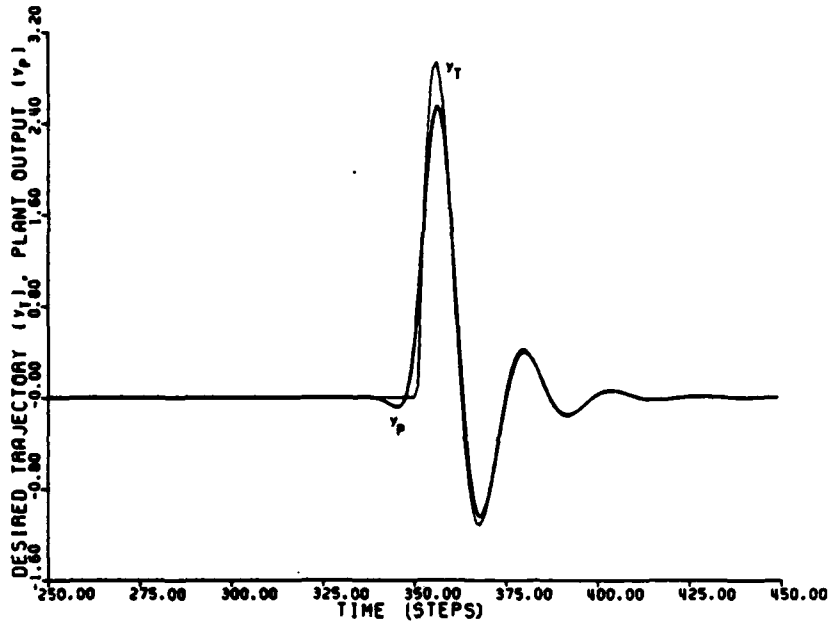


Fig. 2. The optimal prescient controller following the impulse response of the trajectory generator.

then regardless of C_T

$$J_{pes}(X_T, u^T(\cdot)) = B_T' P_{12}' e^{\Lambda^T T} \hat{P}_{11} e^{\Lambda^* T} P_{12} B_T. \tag{3.13b}$$

Proof: Since $\hat{x}_T(t+T) = x_T(t+T) = y_T(t+T)$, the last term in (3.12) may be included as a delta function inside the convolution integral in (2.2). The proof is then similar to that of Lemma 5. Details are given in [5].

Remark: Applying Parseval's formula to (3.9) yields

$$J_{pes}(Y_T, u(\cdot))$$

$$= \int_{-\infty}^{\infty} \left\| \int_t^{\infty} H^0(t-\tau) x_T(\tau)\, d\tau - \int_t^{t+T} H(t-\tau) x_T(\tau)\, d\tau \right\|_R^2 dt.$$

Clearly in this case the optimal choice for $H(t-\tau)$ is

$$H^T(t-\tau) = H^M(t-\tau) + \delta(t+T-\tau) \int_{t+T}^{\infty} H^0(t-\tau) \hat{x}_T(\tau)\, d\tau.$$

IV. Simulation Results

The purpose of this section is to report the results of computer simulations which demonstrate the theory developed earlier, and to provide some insight into the driving problem in the frequency domain. The equations used here are the discrete time versions of the relations derived above, and are given in the Appendix. The plant which is studied is described by the following system matrices:

$$A = \begin{bmatrix} 0.0 & 1.0 \\ -0.99 & 1.99 \end{bmatrix} \quad B = \begin{bmatrix} 0.0 \\ 1.0 \end{bmatrix} \quad C = [1.0 \quad 0.1] \tag{4.1}$$

and unless otherwise noted, $Q = 0.1$ and $R = 2$.

The correspondence between the stochastic and deterministic problems is illustrated by Figs. 2 and 3 for the following trajectory generator:

$$A_T = \begin{bmatrix} 0.0 & 1.0 \\ -0.85 & 1.78 \end{bmatrix} \quad B_T = \begin{bmatrix} 0.0 \\ 1.0 \end{bmatrix} \quad C_T = [1.0 \quad 0.1]. \tag{4.2}$$

In Fig. 2 the trajectory y_T is the impulse response of the trajectory generator; in Fig. 3 y_T is the response of the trajectory generator to a white noise sequence. In both figures, y_p is the output of the plant under optimal prescient control. Notice the similarities between the frequency contents of the two trajectories and between the tracking errors.

Actual driving performances were computed by direct simulation for several realizations of the stochastic trajectory generated via (4.2). Table I compares the average performances observed over ten runs of 1000 steps with the expected performances calculated from (A3.11b) and (A3.13b) for Miller's driver and Tomizuka's driver, respectively, as a function of the preview length. The averages agree quite well with the expected values.

Several general trends can be seen in Table II. Tomizuka's driver outperforms Miller's driver at each preview length, and always improves its performance as the preview length is increased. Miller's driver, on the other hand, sometimes performs more poorly with an increased preview length (compare $K=8$ and $K=9$) due to its ignorance of the trajectory spectrum.

It is insightful to examine the gain functions $\tilde{\Gamma}$ of the various drivers. As shown by (3.4), $\tilde{\Gamma}$ indicates how heavily the trajectory energy at each frequency is weighted in the calculation of the overall performance. Thus, designing a previewing controller may be thought of as specifying a $\tilde{\Gamma}$, subject to the preview constraint, which is close to $\tilde{\Gamma}^0$ in some sense.

$\tilde{\Gamma}^0$ is plotted for several values of the control cost R in Fig. 4. As might be expected, for a given R the optimal prescient driver is better able to follow the low-frequency components of the trajectory. Notice that the upper limit on $\tilde{\Gamma}^0$ is Q, which corresponds to the complete inability of the driver to track a particular frequency. As R increases, the range of untrackable frequencies grows.

$\tilde{\Gamma}^T$ is plotted, for several values of preview length, along with the trajectory spectral density for two different trajectory generators in Figs. 5 and 6. The trajectory generator for Fig. 5 is

$$A_T = 0.9 \qquad B_T = 1. \qquad C_T = 0.042 \tag{4.3}$$

whose spectrum is very narrow. As a consequence, $\tilde{\Gamma}^T$ is very close to $\tilde{\Gamma}^0$ at low frequencies where the trajectory's energy is concentrated, whereas it is far from $\tilde{\Gamma}^0$ at the less important higher frequencies. On the other hand, the trajectory generator for Fig. 6,

$$A_T = 0.6 \qquad B_T = 1. \qquad C_T = 0.138 \tag{4.4}$$

has a fairly wide spectrum which forces $\tilde{\Gamma}^T$ to fit $\tilde{\Gamma}^0$ over a wider frequency range. As the trajectory spectrum becomes flat ($A_T \to 0$), Tomizuka's driver becomes identical to Miller's driver, whose $\tilde{\Gamma}^M$ is plotted in Fig. 7 for several previewing lengths.

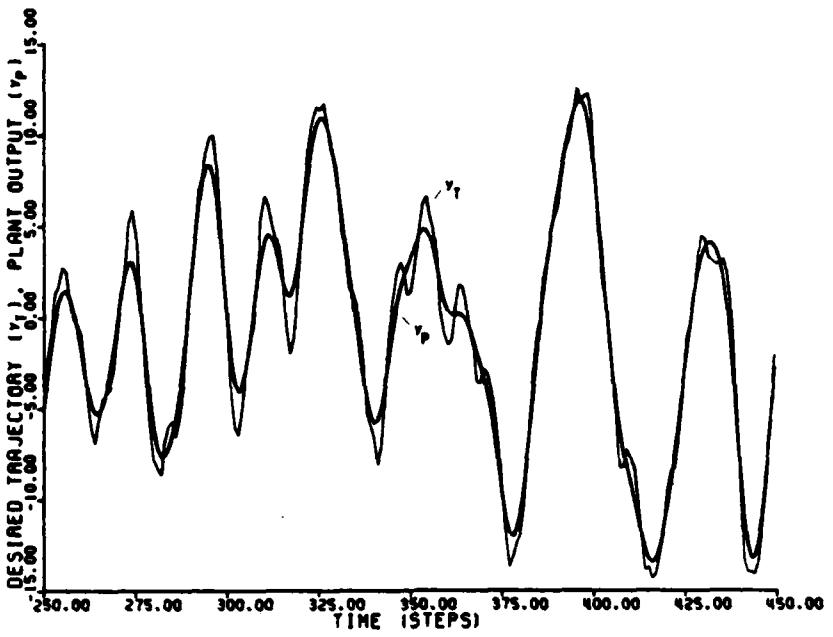


Fig. 3. The optimal prescient controller following a portion of a stochastic trajectory.

TABLE II
DRIVER PERFORMANCES AS OBSERVED IN SIMULATIONS AND AS EXPECTED

	Miller's Driver		Tomizuka's Driver	
Preview Length	Average	Expected	Average	Expected
1	8.00611	7.91545	1.60977	1.59611
2	5.84367	5.81273	1.10136	1.09644
3	3.53969	3.53901	0.85766	0.85799
4	1.96089	1.97025	0.76561	0.77052
5	1.16446	1.16121	0.74756	0.74820
6	0.84626	0.84436	0.74619	0.74557
7	0.76057	0.75840	0.74766	0.74553
8	0.75399	0.74926	0.74902	0.74443
9	0.75307	0.75271	0.74296	0.74276
10	0.75576	0.75233	0.74451	0.74135

Note that if the trajectory energy is concentrated around the frequency $\omega=1.2$, Miller's driver performs better with preview length $K=1$ than $K=2$, a phenomenon observed also in Table II at $K=8$, $K=9$ for the trajectory 4.2. On the other hand, if the trajectory spectrum is available, Tomizuka's driver always improves its performance with increase in the preview length. Finally, it should be noted that Tomizuka's driver may perform more poorly than Miller's if the actual trajectory spectrum differs significantly from that for which it was designed.

V. Concluding Remarks

Expressions for driver performance have been presented for both the time-varying and the time-invariant preview control problems. Of particular importance is the result that a driver's expected performance over an ensemble of trajectories generated by a linear system driven by white noise is just the driver's actual performance following the impulse response of the trajectory generator. This result comes about because the quadratic performance index is expressible in the frequency domain and because the impulse response has the same spectral composition as the stochastic process.

The basic approach of this paper has significance beyond the performance evaluation of control policies. The duality of the linear quadratic control problem and least squares estimation indicates that the same approach can be utilized to evaluate the design of suboptimal estimators. For example, the evaluation of the "goodness" of a finite AR or MA model representing an ARMA model can be based on these ideas.

The problem investigated here entails closed form solutions since the trajectory was parameterized by time. An open issue which deserves attention is the study of self-paced controllers [6], which are more realistic models of the human driver.

Appendix
The Discrete Time Case

The basic relations discussed in the paper are listed here for the discrete time driving problem, keyed for ease of reference to their continuous counterparts in Sections II and III.

The system to be controlled and the performance index are

$$x(k+1)=A(k)x(k)+B(k)u(k); \qquad x(k_0)=x_0 \tag{A2.1a}$$

$$y_p(k)=C(k)x(k) \tag{A2.1b}$$

$$J(k_0, x_0, Y_T, u(\cdot))=\sum_{k=k_0}^{k_f-1}\left\{\|y_p(k)-y_T(k)\|^2_{Q(k)}+\|u(k)\|^2_{R(k)}\right\} + \|y_p(k_f)-y_T(k_f)\|^2_{Q(k_f)} \tag{A2.3}$$

and the notation $(K(k), H(k,i))$ represents the linear control policy

$$u(k, x(k), Y_T)=-K(k)x(k)+\sum_{j=k_0}^{k_f} H(k,i)y_T(i). \tag{A2.2}$$

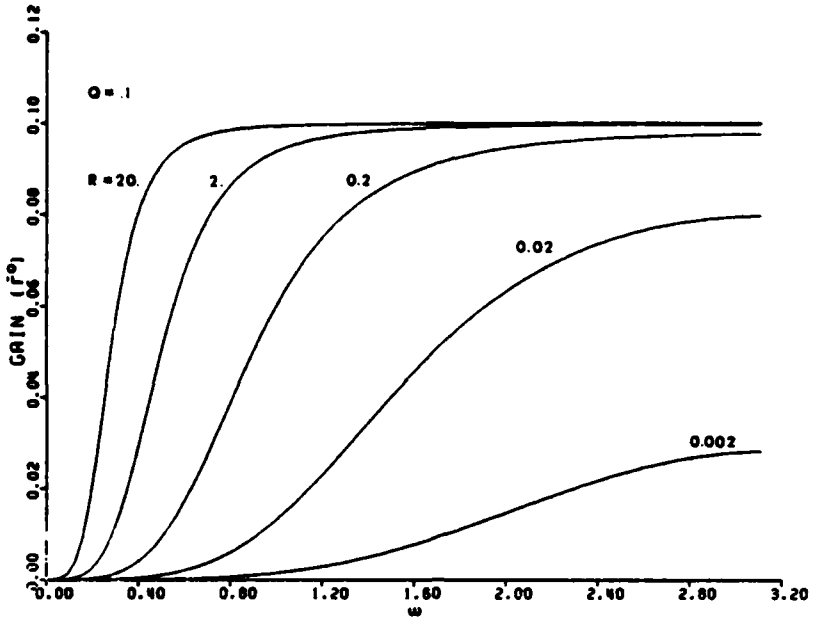


Fig. 4. The gain function of the optimal prescient driver.

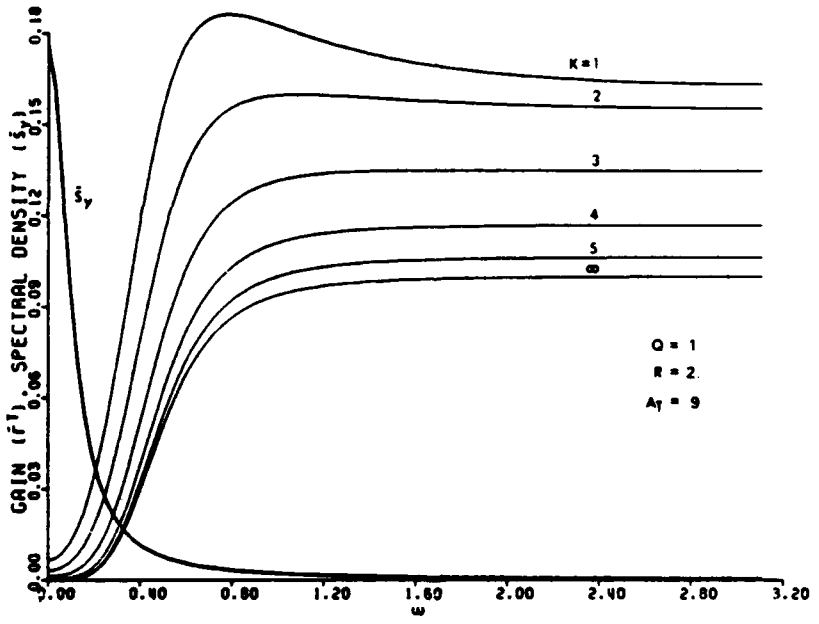


Fig. 5. The gain function of Tomizuka's driver for a trajectory with a narrow spectrum.

The optimal prescient driver and its performance are given by

$$u^0(k, x(k), Y_T) = -\hat{R}^{-1}[B'P(k+1)Ax(k) - B'\xi(k+1)] \tag{A2.4a}$$

$$J(k_0, x_0, Y_T, u^0(\cdot)) = \sum_{k=k_0}^{k_f-1} \|y_T(k) - CP^{-1}(k)\xi(k)\|^2_{Q(k)} \tag{A2.5}$$

where

$$P(k) = A'[P(k+1) - P(k+1)B\hat{R}^{-1}B'P(k+1)]A + C'QC;$$

$$\cdot P(k_f) = C'QC \tag{A2.4b}$$

$$\xi(k) = [A' - A'P(k+1)B\hat{R}^{-1}B']\xi(k+1) + C'Q\,y_T(k);$$

$$\cdot \xi(k_f) = C'Q\,y_T \tag{A2.4c}$$

$$\hat{R}(k) = R + B'P(k+1)B$$

$$\hat{Q}(k) = Q + QC[P - C'QC]^{-1}C'Q = [Q^{-1} - CP^{-1}C']^{-1}.$$

The penalty paid by a suboptimal driver is

$$J_{pen}(k_0, x_0, Y_T, u(\cdot)) = \sum_{k=k_0}^{k_f-1} \|u(k,\cdot) - u^0(k,\cdot)\|^2_{\hat{Q}(k)}. \tag{A2.8}$$

All the above relations are valid for time-varying system parameters, even though for brevity the time argument has sometimes been omitted.

The spectral analysis of Section III may be carried out using the discrete Fourier transform

$$X(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{X}(e^{j\omega})e^{j\omega k}\,d\omega \leftrightarrow \tilde{X}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} X(k)e^{-j\omega k}. \tag{A3.3}$$

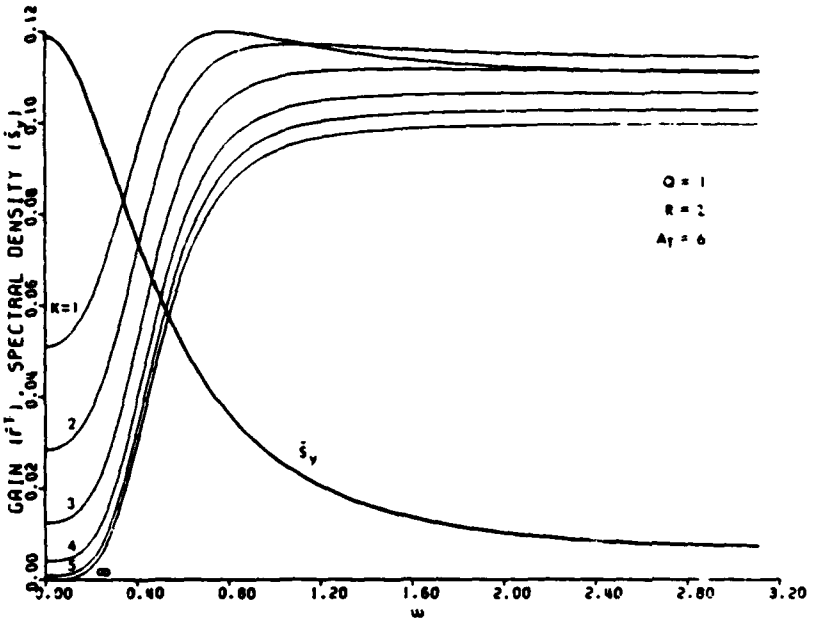


Fig. 6. The gain function of Tominaka's driver for a trajectory with a wide spectrum.

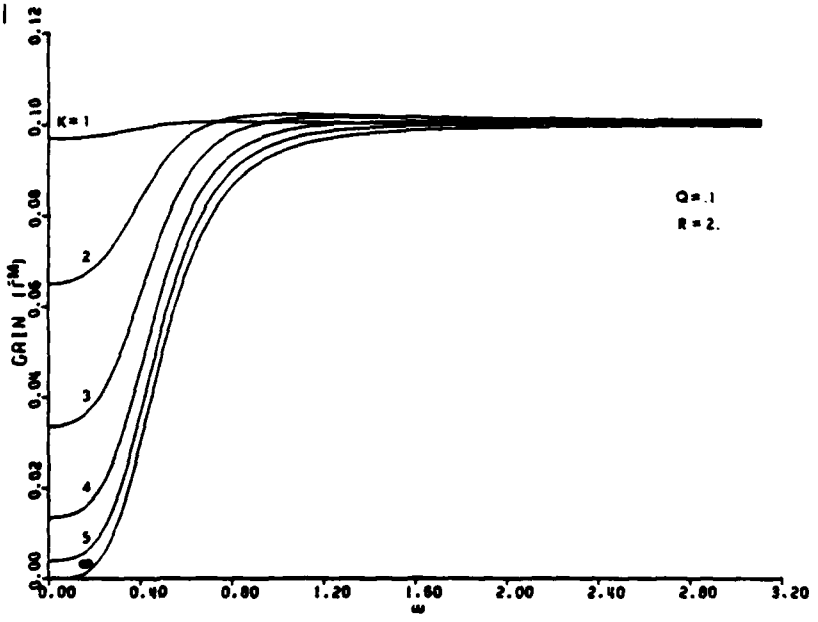


Fig. 7. The gain function of Miller's driver.

With the obvious modifications, Theorem 1 is valid for discrete time.

The time-invariant optimal prescient driver $u^0(\cdot)=(K^0, H(k-i))$, its gain, and its performance following the trajectory generator are

$$K^0=\hat{R}^{-1}B'P_{11} \tag{A3.6a}$$

$$H^0(k-i)=\begin{cases}\hat{R}^{-1}B'(\Lambda')^{-(k-i)-1}C'Q & k-i<0\\ 0 & k-i\geq 0\end{cases} \tag{A3.6b}$$

$$\hat{\Gamma}^0(e^{j\omega})=\left\|I-CP_{11}^{-1}[e^{-j\omega}I-\Lambda']^{-1}C'Qe^{-j\omega}\right\|_Q^2 \tag{A3.8a}$$

$$J(Y_T,u^0(\cdot))=B_T'\left[P_{22}-P_{12}'\hat{P}_{11}P_{12}\right]B_T \tag{A3.8b}$$

where

$$\begin{bmatrix}P_{11} & -P_{12}\\ -P_{12}' & P_{22}\end{bmatrix}\triangleq P_e=A_e'\left[P_e-P_eB_e\hat{R}^{-1}B_e'P_e\right]A_e+C_e'QC_e \tag{A3.7a}$$

$$\Lambda=A-B\hat{R}^{-1}B'P_{11}A \tag{A3.6c}$$

$$\hat{P}_{11}-\Lambda\hat{P}_{11}\Lambda'=B\hat{R}^{-1}B' \tag{A3.8c}$$

and A_e, B_e, and C_e are as defined in (3.7b).

Miller's driver for preview length K, $u^M(\cdot)=(K^0, H^M(k-i))$, its gain, and its penalty following the trajectory generator are

$$H^M(k-i)=\begin{cases}H^0(k-i) & i<k+K\\ 0 & i>k+K\end{cases} \tag{A3.10}$$

$$\hat{\Gamma}^M(e^{j\omega})=\hat{\Gamma}^0(e^{j\omega})+\left\|\hat{R}^{-1}B'(\Lambda')^K(e^{-j\omega}I-\Lambda')^{-1}C'Q\right\|_k^2 \tag{A3.11a}$$

$$J_{pen}(Y_T,u^M(\cdot))=B_T'\left[P_{12}'\Lambda^K\hat{P}_{11}(\Lambda')^KP_{12}+\hat{P}_{22}\right]B_T \tag{A3.11b}$$

where

$$\hat{P}_{22}-A_T'\hat{P}_{22}A_T=A_T'P_{12}'\Lambda^KB\hat{R}^{-1}B'(\Lambda')^KP_{12}A_T. \tag{A3.11c}$$

Tomizuka's driver, its gain when $C_T = I$, and its penalty when $x_T(t+T)$ is measurable, are

$$u^T(\cdot) = (K^0, H^M(k-i)) + \hat{R}^{-1}B'(\Lambda')^K P_{12} A_T \hat{x}_T(k+K) \quad \text{(A3.12)}$$

$$\tilde{\Gamma}^T(e^{j\omega}) = \tilde{\Gamma}^0(e^{j\omega}) + \left\| \hat{R}^{-1}B'(\Lambda')^K \left[(e^{-j\omega}I - \Lambda')^{-1} C'Q - P_{12}A_T \right] \right\|_R^2 \quad \text{(A3.13a)}$$

$$J_{\text{pen}}(X_T, u^T(\cdot)) = B_T' P_{12}' \Lambda^K \hat{P}_{11} (\Lambda')^K P_{12} B_T. \quad \text{(A3.13b)}$$

References

[1] M. Athans and P. L. Falb, *Optimal Control*. New York: McGraw-Hill, 1966.
[2] R. Bellman, *Dynamic Programming*. Princeton, NJ: Princeton Univ. Press, 1957.
[3] E. Fogel, "The optimal driver—An approach to tracking problems," in *Proc. 1st Symp. on Mini and Microcomputers in Control*, San Diego, CA, 1979.
[4] R. A. Miller, "On the finite preview problem in manual control," *Int. J. Syst. Sci.*, vol. 7, pp. 667–672, June 1976.
[5] K. McGill, "Spectral analysis of the performance of previewing controllers," M. S. thesis, Univ. Notre Dame, Notre Dame, IN, 1979.
[6] T. B. Sheridan, "Three models of preview control," *IEEE Trans. Human Factors Electron.*, vol. HFE-7, pp. 91–102, June 1966.
[7] M. Tomizuka, "Optimal continuous finite preview problem," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 362–365, June 1975.
[8] M. Tomizuka and D. E. Whitney, "Optimal discrete finite preview problems," *Trans. ASME J. Dyn. Syst., Meas., Contr.*, vol. 97, pp. 319–325, Dec. 1975.
[9] ——, "The human operator in manual preview tracking," *Trans. ASME J. Dyn. Syst., Meas., Contr.*, vol. 98, pp. 407–413, Dec. 1976.
[10] A. Wernersson, "On pursuit and feedback in optimal stochastic control—Explicit control laws," *Inform. Sci.*, vol. 7, pp. 29–48, 1974.

A Fault Directory Approach to Analog Fault Analysis - A Case Study

by

R. Liu
C. Lin

FAULT DIRECTORY APPROACH - A CASE STUDY*

C. Lin
R. Liu

Department of Electrical Engineering
University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

The perception of the fault directory approach to analog circuit diagnosis is introduced by a case study: the diagnosis of a switching voltage regulator.

INTRODUCTION

Fault Diagnosis of analog circuits has become an active research area recently (1-9). A good survey of this subject can be found in (10,11). Most of the methods (1-9) considered the circuits as interconnections of discrete components even though there are clearly-defined functional blocks in the circuit. This is because these methods depend on the assumption that each component can be characterized by some parameters. This assumption may not be valid for functional blocks. Also these methods require measurements be taken in a purely DC or purely AC test. This restriction may hamper the diagnosis capability to certain circuits, e.g., a switching circuit.

Taking the above problems into account, a new approach, the fault directory approach, which is conceptually different from the previous methods, will be presented. The first step of this method is to divide a circuit into functional blocks. Then a set of fault logic-equations is set up based on the descriptions and the operating conditions of these blocks. From the above equations, the fault directory is established. Note that the operating conditions of each functional block are taken into consideration in the fault directory method. A functional block can work properly only when its operating conditions are met. Therefore, any practical fault diagnosis method should take operating conditions into consideration.

To show the versatility of this method, we will present a case study on a practical switching voltage regulator. This particular circuit is chosen because it is a realistic circuit and it has features which are not considered by previously mentioned methods:

(1) It is a hybrid circuit, consisting of both analog and digital subcircuits.
(2) It contains nonlinear elements whose nominal characteristics may not be precisely described.
(3) It has switching subcircuits.

Note that this circuit cannot be fault diagnosed by the methods based on a purely DC or purely AC test.

In the following section, each functional block of the circuit will be described in detail. The perception of the fault directory approach will be introduced by the diagnosis of this circuit.

A CASE STUDY

In this section we want to show, step by step, how a fault directory can be established to isolate faulty functional blocks.

1. The Switching Voltage Regulator Circuit

The circuit under study is a switching voltage regulator as shown in Fig. 1. In the circuit, the node-voltages 1-6 are measureable. V_I is the input voltage, V_O the output voltage, V_R the reference voltage and V_C the control voltage. The purpose of this circuit is to regulate the output voltage V_O so that it is close to the reference voltage V_R.

This is a slightly modified circuit from an actual circuit design. Note that it contains feedback path, logic circuits, switching circuits, and a transformer. The purpose of this section is not to develop a diagnosis algorithm for this circuit, but to demonstrate the applicability of our approach. The method presented here can be applied to the original circuit, and will lead to the same answer. But more explanation would be needed.

2. Functional Blocks and Operating Conditions

In the first step, the circuit is decomposed into four functional blocks, as shown by the dashed line in Fig. 1. The result is a functional block circuit shown in Fig. 2. These functional blocks are chosen because of their specific functions in the circuit which will be described in the following paragraphs.

Before specifying the functions of the functional blocks, some notations have to be introduced.

Notations

1. V^+ denotes a positive voltage significantly greater than zero.
2. V^- denotes a negative voltage significantly smaller than zero.
3. V^0 denotes a voltage very close to zero.
4. I^+, I^- and I^0 are similarly specified.
5. $\uparrow$ denotes an increase in voltage or current.
6. $\downarrow$ denotes a decrease in voltage or current.

With the above notations, we will specify the functions of the functional blocks B1, B2, B3, B4 as well as L, R1, and R2.

*This paper was supported in part by the Office of Naval Research under Grant N0014-78-C-0444.

CH1635-2/81/0000-0239 $00.75 © 1981 IEEE

Block B1

1. Operating Conditions
   a) V_1 is not stuck at a fixed voltage by the load.
2. The Function Table

V_o	V_1
$<V_R$	V_1^+
$>V_R$	V_1^o

Resistor R1

1. Operating Conditions
   a) $V_2 = V_2^o$ at all times.
2. The Function Table

V_1	I_2
V_1^+	I_2^+
V_1^o	I_2^o

Block B2

1. Operating Conditions
   a) $V_2 = V_2^o$ at all times.
   b) V_3 is not stuck at a fixed voltage by the load.
2. The Function Table

V_C	I_2	V_3	I_3
V_C^o	I_2^+	V_3^o	>0
	I_2^o	>0	I_3^o
V_C^+	X	>0	X

Resistor R2

1. Operating Conditions
   a) $V_4 = V_4^+$ at all times.
2. The Function Table

V_3	I_3	I_4
V_3^o	I_3^+	$I_4^+=I_3^+$
V_3^+	I_3^o	$I_4^o=I_3^o$

Block B3

1. Operating Conditions
   a) $V_4 = V_4^+$ at all times.
   b) V_5 is not stuck at a fixed voltage by the load.
2. The Function Table

I_4	V_5
I_4^+	$V_5^+=V_1$
I_4^o	$V_5^o=0$

Inductor L

1. Operating Conditions
   a) $0 < V_6 < V_1$ at all times.
2. The Function Table

V_5	V_L	I_6
V_1	>0	$\uparrow$
0	<0	$\downarrow$

Block B4

1. Operating Conditions
   a) $0 < V_6 < V_1$ at all times.
   b) V_o is not stuck at a fixed voltage by the load.
2. The Function Table

I_6	V_o
$\uparrow$	$\uparrow$
$\downarrow$	$\downarrow$

If all the operating conditions are satisfied and all functional blocks are connected as shown in Fig. 2, the sequence of events are given as follows:

1) $V_C = V_C^o$

V_o	V_1	I_2	V_3	I_3	I_4	V_5	V_L	I_6	V_o
$<V_R$	V_1^+	I_2^+	V_3^o	I_3^+	I_3^+	$V_5^+=V_1$	>0	$\uparrow$	$\uparrow$
$>V_R$	V_1^o	I_2^o	V_3^+	I_3^o	I_3^o	$V_5^o=0$	<0	$\downarrow$	$\downarrow$

Concluding from the first and last column, if $V_o < V_R$, then V_o increases and if $V_o > V_R$, then V_o decreases. Therefore, V_o will be regulated at V_R.

2) $V_C = V_C^+$

In this case, it is easy to see from the functions of B2 that V_o will decrease to zero and the circuit no longer works like a voltage regulator.

Note that for each functional block we have specified operating conditions as well as a functional table. This stems from the need of specifying operating conditions for practicle functional blocks. Consideration of the operating conditions is essential in the fault diagnosis, since a deviation from the functional table may be caused by either the failure of the functional block itself or by a change of the operating conditions.

3. Fault Logic-Equations

The fault logic-equation for each block is based on the assumption that the functional block is fault-free if and only if its operation conditions are satisfied and it functions as described by its funcional table. The set-up of these fault logic-equations will now be illustrated.

Note first that the failure of operating conditions may be caused by a fault of other functional blocks. Such cases are listed in the following observations:

Observations

1. If R1 is not shorted and $V_2=V_2^o$, then V_1 cannot be stuck by load.
2. If R2 is neither opened nor shorted and $V_4=V_4^+$, then V_3 cannot be stuck by the load.
3. If $0<V_6<V_1$ then L is neither opened or shorted.
4. If L is not shorted then V_5 cannot be stuck by the load.

These obsevations will be used to set up fault logic-equations.

The other operating conditions will not be put into the fault logic-equations:

A) $V_2 = V_2^o$
B) $V_4 = V_4^+$
C) $0 < V_6 < V_1$

D) V_o is not stuck by the load.

The first step of fault diagnosis is to verify the above conditions. If they are satisfied, then the fault logic-equations can be set up as follows:

(E1) $\hat{f}(B1) = f(B1) \times f(R1 \text{ not shorted})$
(E2) $\hat{f}(B2) = f(B2) \times f(R1) \times f(R2)$
(E3) $\hat{f}(B3) = f(B3) \times f(R2)$
(E4) $\hat{f}(B4) = f(B4) \times f(L)$

where x is a logical AND operator, and $\hat{f}$ and f have the following meaning:

$\hat{f}(B1)$ has value 1 if B1 is estimated to be fault-free,
0 if B1 is estimated to be faulty,
f(B1) has value 1 if B1 is actually fault-free,
0 if B1 is acutally faulty.

The others are similarly defined. These equations described the relationship from the actual fault status of the subcircuits and the status of the subcircuits concerning with the operating condition to the estimated status of the subcircuits. Take (E1) for instance; it means that if R1 is not shorted and B1 is fault-free, then the test of B1 will conclude fault-free. The reason is as follows. If R1 is not shorted and $V_2=V_2^o$ by previous test, then V_1 is not stuck by the load, i.e., the operating condition of B1 is satisfied. The test of B1 by its function table, therefore, will yield the true state of fault of B1. Note that if R1 is shorted, a test of B1 will yield B1 at fault even though it is fault-free.

4. A Fault Directory

From these fault logic-equations, the fault truth-value table (Table 1) can be easily obtained for the case that at most one of functional block is at fault. As can be seen in Table 1, there is a pattern of estimated status of B1, B2, B3, and B4 corresponding to each fault case. By reorganizing Table 1 in terms of these patterns, we have the fault directory (Table 2) which indicates the possible fault location once the pattern is obtained from measurements and simple estimations.

From Table 2, it can be seen that for the single fault case, the fault can be isolated into one of four groups: B1, (B2,R1), B3, and (B4,L). Once a group is isolated to be at fault, the fault directory approach can then be applied to this group. The process can be repeated so that a hierarchical system can be set up for diagnosis purposes. This concludes the demonstration of the fault directory approach.

5. Conclusion

The fault directory approach for analog circuits is similar to the fault dictionary approach for digital circuits. It requires minimum computing for users. This is the major advantage of this method.

It has demonstrated that it has the capability to diagnose circuits which consist of nonlinear components, switching circuits and analog-digital hybrid circuits.

Functional blocks, rather circuit components or elements, are used as the basis for our method. This seems to be a natural approach for fault diagnosis.

REFERENCES

1) R.S. Berkowitz, "Conditions for Network-element-value Solvability", IRE Trans. Circuit Theory, Vol. CT-9, pp. 24-29, Mar. 1962.

2) S. Seshu and R. Waxman, "Fault Isolation in Conventional Linear Systems-A Feasibility Study," IEEE Trans. Rel., Vol. R-15, pp. 11-16, May 1965.

3) R. Saeks, S.P. Singh, and R.W. Liu, "Fault Isolation via Component Simulation", IEEE Trans. Circuit Theory, Vol. CT-19, pp. 634-640, Nov. 1972.

4) T.N. Trick, W. Mayeda, and A.A. Sakla, "Calculation of Parameter Values from Node Voltage Measurements", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 466-474, July 1979.

5) N. Navid and A.N. Willson, Jr., "A Theory and an Algorithm for Analog Circuit Fault Diagnosis", IEEE Tans. Circuits and Systems, Vol. CAS-26, pp. 440-457, July 1979.

6) N. Sen and R. Saeks, "Fault Diagnosis for Linear Systems via Multifrequency Measurements", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 457-465, July 1979.

7) J. Lee and S.D. Bedrosian, "Fault Isolation Algorithm for Analog Electronic Systems Using the Fuzzy Concept", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 518-522, July 1979.

No. of Faults	f(B1)	f(B2)	f(B3)	f(B4)	f(R1)	f(R2)	f(L)	f(R1not shorted)	$\hat{f}(B1)$	$\hat{f}(B2)$	$\hat{f}(B3)$	$\hat{f}(B4)$
0	1	1	1	1	1	1	1	1	1	1	1	1
	1	1	1	1	0	1	1	0	0	0	1	1
	1	1	1	1	1	1	0	1	1	1	1	0
	1	1	1	1	1	0	1	1	1	0	0	1
	1	1	1	1	0	1	1	1	1	0	1	1
1	1	1	1	0	1	1	1	1	1	1	1	0
	1	1	0	1	1	1	1	1	1	1	0	1
	1	0	1	1	1	1	1	1	1	0	1	1
	0	1	1	1	1	1	1	1	0	1	1	1

Table 1. Fault Truth-Value Table

Pattern				Faults	
$\bar{T}(B1)$	$\bar{T}(B2)$	$\bar{T}(B3)$	$\bar{T}(B4)$	Single	Double
1	1	1	1		
1	1	1	0	B4,L	(B4,L)
1	1	0	1	B3	
1	0	1	1	B2,R1	(R1,B2)
0	1	1	1	B1	
1	1	0	0		(B3,L),(B3,B4)
1	0	1	0		(R1,L),(B2,L),(B4,R1),(B2,B4)
0	1	1	0		(B1,L),(B1,B4)
1	0	0	1	R2	(R1,R2),(B2,R2),(B3,R2),(B2,B3),(B3,R1)
0	1	0	1		(B1,B3)
0	0	1	1	R1	(R1,L),(R1,R2),(B1,R1),(B2,R1),(B1,B2)
1	0	0	0		(R2,L),(B4,R2)
0	1	0	0		
0	0	1	0		(B4,R1)
0	0	0	1		(B1,R2)
0	0	0	0		

Table 2. Fault Directory

8) R. Liu and V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part I-Theory", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 490-495, July 1979.

9) V. Visvanathan and R. Liu, "Sequentially Linear Fault Diagnosis: Part II-Design", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 558-564, July 1979.

10) R. Saeks and S.R. Liberty, Eds., Rational Fault Analysis, Marcel Dekker, 1977.

11) P. Duhamel and J.-C. Rault, "Automatic Test Generation Techniques for Analog Circuits and Systems: A Review", IEEE Trans. Circuits and Systems, Vol. CAS-26, pp. 411-440, July 1979.

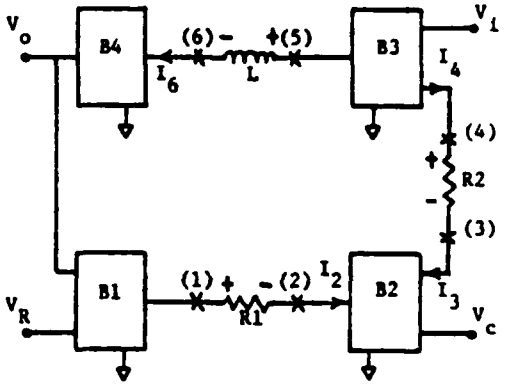


Fig. 2. The Functional Block Circuit

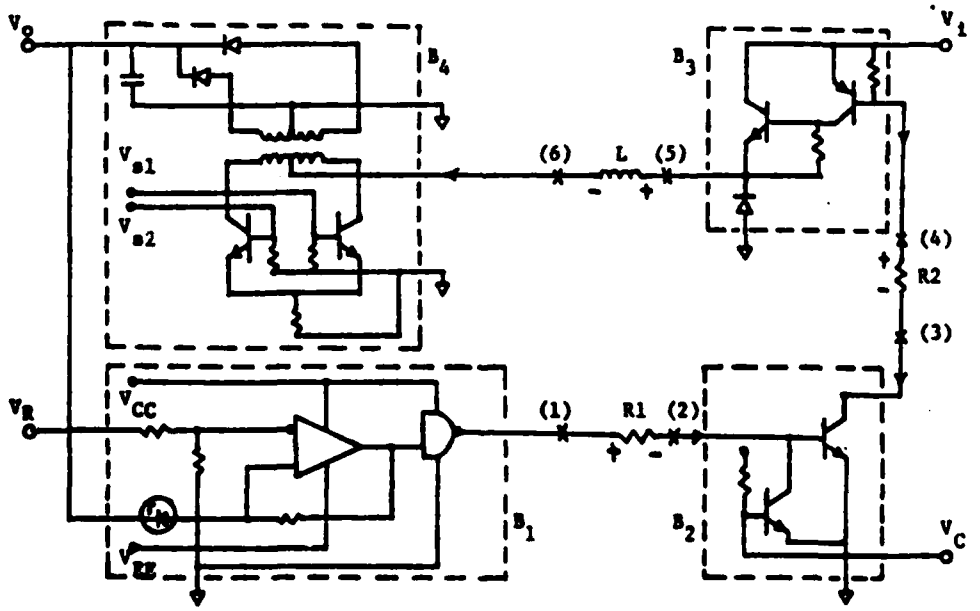


Fig. 1. A Switching Voltage Regulator

Fault Diagnosis in Electronic Circuits

by

R. Liu
R. Saeks

別　刷

計　測　と　制　御

昭和　　年　　第　　巻　　第　　号

（P.　　～P.　　）

社団法人 計測自動制御学会

[Survey]

[計測と制御
Vol. 20, No. 2
(昭和56年2月)]

Fault Diagnosis in Electronic Circuits

R. Saeks* and R. Liu**

During the past quarter century the engineering community has been witness to tremendous strides in the art of electronics design. The graphical algorithms of the previous generation have given way to the modern CAD package, the breadboard has been subsumed by the simulator. Indeed, even the universal building block has become a reality. To the contrary electronics maintenance has changed little since the day of the vacuum tube, remaining the responsibility of the experienced technician with scope and multimeter. As such, our ability to design a complex electronic circuit is quickly out-distancing our ability to maintain it. In turn, the price reductions which have accompanied modern electronics technology have been paralleled by increasing maintenance and operations costs. Indeed, many industries are finding that the life cycle maintenance costs for their electronic equipment now exceed their original capitol investment.

Given the above, it is quickly becoming apparent that the electronics maintenance process, like the design process, must be automated. Unfortunately, the 50 years of progress in circuit theory, on which our electronics design automation has been predicated, does not exist in the maintenance area. As such, the past decade has witnessed the inauguration of a basic research program to lay the foundations for a theory of electronics maintenance and a parallel effort to develop operational electronic maintenance codes.

Thus far the greatest success has been achieved in the digital electronics area, wherein the finite state nature of the UUT (unit under test) may be exploited[3]. Typically, one assumes that all failures manifest themselves in the form of component outputs which are either "stuck-at-one" or "stuck-at-zero" and/or shorts and opens[5]. Under such an assumption a theory for digital system maintenance has been developed and practical fault diagnosis algorithms are in the formative stages of development. Typically, one hypothesizes some limit on the number of simultaneous faults and then simulates the responses of the UUT to a family of test vectors for each allowed combination of faults. The actual responses of the UUT are then compared with the simulated responses to locate the failure. Although lacking in asthetic appeal the above approach, termed *fault simulation*, is ideally suited for the maintenance environment, wherein, the actual simulation process need only be done once at the factory or a maintenance depot with the simulated response data being distributed via magnetic tape to the various field locations where the actual test actual test is conducted. As such, with the aid of some sophisticated software engineering, this apparently "brute force" approach to the fault diagnosis problem has slowly evolved into a workable concept[4]. Indeed, at the present time a number of automatic test program generators which classify faults, choose test vectors, and carry out the appropriate simulation (often in a parallel processing mode), are commercially available and, as such, the automated maintenance of digital electronic circuits is becoming a reality[4].

Unfortunately, the above described success in the digital world has not been paralleled by progress in the analog world. Indeed, test engineers complain that while 80% of the boards are digital, 80% of their headaches are analog and hybrid. This difficulty arises from a number of characteristics

* Texas Tech University
** University of Notre Dame

of the analog problem which are not encountered in digital circuits. Indeed, in an analog circuit:

(i) there is a continuum of possible failures,
(ii) a component may be "in tolerance" but not nominal,
(iii) complex feedback structures are encountered,
(iv) simulation is slow and costly.
(v) post-fault component characteristics may not be known,
(vi) and a fault in one component may induce an apparent fault in another

Items (i) and (ii) imply that an extremely large number of simulations will be required for analog testing. Items (iii) and (iv) suggest that these simulations will be far more expensive than similar digital simulations. Finally, items (v) and (vi) indicate that the simulation of a post-fault circuit by itself may not be a tractable problem. As such, it is by no means clear that the kind of "brute force" fault simulation algorithm associated with the digital problem will be applicable to the analog or hybrid case.

As an alternative to fault simulation, a number of academic researchers have proposed a variety of "post test" fault diagnosis algorithms, wherein, an "equation solving like" algorithm is used to locate the faulty component given the test data from UUT[2),8)]. Although these algorithms are, in some sense, "smarter" than the simulation algorithms, most of the required computing must be done in the field after the UUT has been tested. Moreover, these computational requirements must be replicated each time a unit fails. As such, the success of such "post test" algorithms is contingent on reducing their computational requirements to a bare minimum. Although no system is yet operational, with the aid of the powerful linear circuit theory developed over the past half century, a computationally efficient solution to the fault diagnosis problem for linear analog circuits appears to be within reach[1),2)]. Unfortunately, no such light exists at the end of the nonlinear tunnel, wherein progress appears to be limited by a "computational complexity/test point" bound.

Not suprisingly, the computational cost of an analog fault diagnosis algorithm is an inverse function of the number of test points at which measurements of the UUT may be made. Indeed, if one lets n be a measure of UUT complexity (which may loosely be taken to be the total number of terminals for all of the circuit components), then if one has access to $0(n)$(注1) test points the fault diagnosis problem can be resolved using linear algorithms[7),10)]. Moreover, by combining such algorithms with the above mentioned linear algorithms, acceptable computational efficiency can be obtained with $0(m)$ test points where m is a measure of the complexity of the "nonlinear subsystem" of the UUT[6),7)]. Although such algorithms can be effective on the typical academic example a "real world" PC (printed circuit) board does not have terminal space for the 20 or 30 test points which are required even for a routine board made up of discrete components and/or SSI (Small Scale Integration). Although the problem can be partially alleviated by making internal measurements with the aid of a "bed-of-nails" tester it has been our experience that such testers cause as many failures as they locate while their applicability to two-sided, multilayer, and coated boards is severely limited. As such, we would like to limit the number of test points to the terminal space available at the edge of a PC board. On the other hand, the UUT complexity, n, increases with the area of the board. As such, the number of test points required by an analog fault diagnosis algorithm should increase at a rate of no greater than $0(n^{1/2})$. A further study of the possible tradeoff between test points and computational cost appears in references 11) and 12).

Unfortunately, all computationally acceptable "post test" algorithms which have thus far been proposed have test point requirements which grow linearly with UUT complexity (assuming that m grows linearly with n). As such, many researchers are looking at the classical fault simulation algorithms with renewed vigor. Indeed, these algorithms have minimal on-line computational costs, while the number of test points employed, can easily be

(注1) $f(n)=0(n)$ means f increases in the order of n; more precisely, $|f(n)| \leq c|n|$ for some $c>0$.

kept below $0(n^{2/1})$. The difficulty lies with the required number of simulations and the development of decision algorithms which will allow us to "interpolate" between simulated data points.

Thus, while the state-of-the-art in digital diagnosis is fast maturing, a serious investigation of analog fault diagnosis problems is only just beginning. Indeed, a satisfactory fault diagnosis code for linear analog circuits has yet to be demonstrated while the nonlinear problem has yet to progress beyond the basic research stage.

References

1) Chen, H.S.M. and R. Saeks: A Search Algorithm for the Solution of the Multifrequency Fault Diagnosis Equations, IEEE Trans. on Circuits and Systems, **CAS-26**, 589/594 (1979)
2) Duhamel, P. and J.C. Rault: Automatic Test Generation Techniques for Analog Circuits—A Review, IEEE Trans. on Circuits and Systems, **25**, 411/439 (1979)
3) Friedman, A.D. and Memon, P.R.: Fault Detection in Digital Circuits, New York, Prentice Hall (1971)
4) Greenbaum, J.R.: Computer-Aided Fault Analysis —*Today, Tomorrow, or Never, in Rational Fault* Analysis (ed. R. Saeks, and S.R. Liberty), New York, Marcel Dekkar, 96/111 (1977)
5) Hayes, J.P.: Modeling Faults in Digital Circuits. in Rational Fault Analysis, (ed. R. Saeks, and S.R. Liberty), New York, Marcel Dekker, 78/95 (1977)
6) Hsieh, M.: Ph. D. Dissertation, Texas Tech Univ. (1980)
7) Ngo, Q.-D.: M.S. Thesis, Texas Tech Univ. (1980)
8) Plice, W.A.: Automatic Generation of Fault Isolation Tests for Analog Circuit Boards-A Survey, Presnted at ATEX EAST 78, Boston, Sept. (1978)
9) Saeks, R., Singh, S.P. and R.W. Liu: Fault Isolation via Components Simulation, IEEE Trans. on Circuit Theory, **CT-19**, 634/640 (1972)
10) Trick, T.N., Mayeda, W. and A.A., Sakla: Calculation of Parameter Values from Node Measurements, IEEE Trans. on Circuits and Systems, **CAS-26**, 466/474 (1979)
11) R.W. Liu and V. Visvanathan: Sequentially Linear Fault Diagnosis: Part I—Theory, IEEF Trans. on Circuits and Systems, 490/496, July (1979)
12) V. Visvanathan and R.W. Liu: Sequentially Linear Diagnosis: Part II—The Design of Diagnosable Systems, IEEE Trans. on Circuits and Systems, 558/564, July (1979)

Topological Conditions for Single-Branch Fault

by

R. Liu
C. Lin
Z.F. Huang

Topological Conditions for Single-Branch-Fault

C. S. LIN, Z. F. HUANG, AND R. LIU

Abstract—**The testability condition for single branch-fault of an analog network is given. Then several necessary and sufficient conditions for this testability condition are provided and their proofs are shown in detail. These conditions have great applications in the testability design as demonstrated in the examples.**

I. Introduction

LET N (Fig. 1) be a connected, b-branch, $(n+1)$-node, linear time-invariant network. Out of the $(n+1)$ nodes, $(m+1)$ are accessible terminals for both excitation and measurement. Let one of the accessible nodes be the reference node of the measurements denoted by n_0. The following notations are used for N.

- v_m terminal (accessible node) voltage vector with respect to n_0,
- i_m terminal current vector,
- v_b branch voltage vector,
- i_b branch current vector,
- Y branch-admittance matrix,
- M set of all accessible nodes including n_0.

Note that v_m and i_m do not contain the entries of n_0 and each contains m entries. Also note that $i_b = Yv_b$ and Y needs not to be symmetrical. Then in N together with the m measurement branches from $(M-\{n_0\})$ to n_0, we have the following KVL and KCL equations:

$$B_b v_b = -B_m v_m \tag{1a}$$

$$Q_b i_b = -Q_m i_m \tag{1b}$$

where B's and Q's are the submatrices of the loop and cut-set matrices, respectively.

Since N is connected, we can choose a tree in N. For a given tree T, denote the branch voltage vector and branch current vector of those in T (its cotree in N) by $v_t(v_c)$ and $i_t(i_c)$, respectively. Label the branches in T ahead of those in the cotree of T. Then $v_b = [v_t^T : v_c^T]^T$ and (1) can be written as follows:

$$\begin{bmatrix} -F^T & \vdots & +I \\ -E^T & \vdots & 0 \end{bmatrix} \begin{matrix} v_t \\ v_c \end{matrix} = \begin{bmatrix} 0 \\ I \end{bmatrix} v_m \tag{2a}$$

Manuscript received September 2, 1983. 1982.

C. S. Lin was with the University of Notre Dame, Notre Dame, IN. He is now with the Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

Z. F. Huang and R.-W. Liu are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

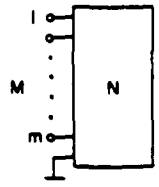


Fig. 1. Network N.

and

$$\left[I \vdots +F \right] \begin{matrix} i_t \\ i_c \end{matrix} = [-E] i_m \tag{2b}$$

where I is an identity matrix and E and F are submatrices of Q_b and Q_m, respectively, w.r.t. the tree T.

Now consider that the network N is perturbed to $(N+\Delta N)$ in the way that Y is perturbed to $Y+\Delta Y$, and the graph remains the same. In the perturbed network, we denote the corresponding voltages, currents, and admittance by $v_m+\Delta v_m$, $i_m+\Delta i_m$, $v_b+\Delta v_b$, and $i_b+\Delta i_b$, respectively.

Given Y and the graph, the purpose of the fault diagnosis is to estimate ΔY from the information i_m, Δv_m, and Δi_m. Note that v_m can be calculated from i_m and N. Without loss of generality, we may assume that

$$\Delta i_m = 0$$

i.e., apply the same i_m to the original network N and the perturbed network $N+\Delta N$.

Under this condition, it can be shown from (2) that in N

$$v_m = +E^T(HYH^T)^{-1}Ei_m \tag{3}$$

and in $N+\Delta N$

$$v_m + \Delta v_m = +E^T(HYH^T)^{-1}[Ei_m - H\Delta Y(v_b+\Delta v_b)] \tag{4}$$

where

$$H \triangleq \left[I \vdots F \right].$$

Subtracting (3) from (4)

$$\begin{aligned} \Delta v_m &= -E^T(HYH^T)^{-1}H\Delta Y(v_b+\Delta v_b) \\ &= Z_{mb} j_b \end{aligned} \tag{5}$$

where

$$Z_{mb} \triangleq +E^T(HYH^T)^{-1}H$$

0098-4094/83/0600-0376$01.00

and

$$j_b \triangleq -\Delta Y(v_b + \Delta v_b). \tag{6}$$

Define that a branch is at fault if and only if the corresponding row of ΔY is nonzero. Then, in (6), the kth entry of j_b is nonzero only if the kth row of ΔY is nonzero or, equivalently, the kth branch is at fault. And the converse is true for almost all practical cases. Taking advantage of this property of j_b, (5) suggests that one can determine the location of nonzero elements of ΔY and therefore the faulty branches provided that the number of accessible nodes is large enough. This discovery was made by Biernacki and Bandler [1], [2] and Sakla *et al.* [3]. The advantages of this approach are that: 1) one can rely only on the parameters of N, i.e., the nominal network, and the measurements to diagnose the faulty network and 2) the computation involved are all linear.

However, despite all these advantages, systematical studies on materializing this approach to locate the faulty branches have been lacking. In the next section, as a demonstration, the diagnosis of the single branch-fault will be discussed in detail and, more importantly, the graphical conditions for placing the test points will also be stated and proved.

II. Single-Fault Testability

In this section, we will study the diagnosis of single branch-fault, or abbreviately single fault, by using (5). For the simplicity of demonstration, only the networks without dependent sources are considered in this paper. Simple conditions on testability will be obtained by taking advantage of a property that network without dependent sources has diagonal branch-admittance matrix. Similar results can also be obtained for the general cases.

To study the diagnosis of networks, we will characterize a network by its capability to be diagnosed by a single test vector i_m. This is given in the following definition.

Definition 1: A network N is said to be single-fault testable if when N is perturbed to $N + \Delta N$, by choosing one appropriate test vector i_m, one will be able to determine from the measurements on accessible nodes M:

(a) whether or not $N + \Delta N$ has more than one branch fault,

(b) if negative, the faulty branch can be uniquely located.

To locate single fault by using (5), each column-vector of Z_{mb} is compared to Δv_m for consistency and the branches whose corresponding column-vectors are in parallel with Δv_m are the potential candidates of the fault. In order to achieve single-fault testability, it is essential to have any two column-vectors of Z_{mb} linearly independent. Let us define this condition as single-fault testability condition. This is a necessary condition for N to be single-fault testable. In fact, it is also sufficient as shown in the following theorem.

Theorem 1: N is single-fault testable if the single-fault testability condition is satisfied. Conversely, if a single test vector is used then it is also necessary.

The proof of the above theorem can be found in [4].

Note that Theorem 1 translates an abstract definition of single-fault testability to a familiar property of a matrix Z_{mb} with the explicit expression which can be further studied. An initial step is to investigate the implication of two column-vectors of Z_{mb} to be linearly independent. However, not to restrict ourselves to only two column-vectors, the linear independence of k column-vectors of Z_{mb} is studied and the result is shown in Theorem 2.

Theorem 2: Let Q be a set of k branches in N and Z_{mq} be the $m \times k$ submatrix of Z_{mb} whose columns correspond to branches in Q. Then Z_{mq} has full column rank, only if there exists a tree T in N such that

(1) T contains Q,

(2) $(T - Q)$ is connected when all accessible nodes in N are shorted together.

The converse is also true for almost all Y.

Proof: (*The necessary part*): As a preliminary step, we want to show that Q cannot contain loops. Let $\tilde{H}$ be the $n \times k$ submatrix of H whose columns correspond to the branches of Q. Then $Z_{mq} = E^T(HYH^T)^{-1}\tilde{H}$. Evidently, Z_{mq} has full column-rank only if $\tilde{H}$ has full column-rank. Since $\tilde{H}$ is a submatrix of a cut-set matrix with the same number of rows, $\tilde{H}$ has full column-rank if and only if Q contains no loop [10]. Consequently, the branches of Q with Z_{mq} full column-rank cannot contain any loop.

Since Q contains no loop, we can choose a tree T in N such that T contains Q. Then label the branches in the following order: branches in Q, branches in $(T - Q)$, and the rest. Corresponding to this labeling, the matrix H has the following from:

$$H = \left[\begin{array}{c|c|c} I_k & 0 & F_1 \\ \hline 0 & I_{(n-k)} & F_2 \end{array}\right] \triangleq \left[\begin{array}{c} H_1 \\ \hline H_2 \end{array}\right] \tag{7}$$

where I_k is an $k \times k$ identity matrix, F_1 and F_2 are matrices of 0, 1, and -1, with dimensions $k \times (b-n)$ and $(n-k) \times (b-n)$, respectively, and

$$H_1 \triangleq \left[I_k \vdots 0 \vdots F_1\right] \qquad H_2 \triangleq \left[0 \vdots I_{(n-k)} \vdots F_2\right]$$

and Z_{mq} can be written as

$$Z_{mq} = E^T(HYH^T)^{-1}\left[\begin{array}{c} I_k \\ \hline 0 \end{array}\right]. \tag{8}$$

From a well-known formula in [7], it can be shown that Z_{mg} has full column-rank if and only if there exists a $k \times n$ submatrix $\bar{E}$ of E such that the matrix

$$W \triangleq \left[\begin{array}{c|c} HYH^T & \begin{array}{c} I_k \\ \hline 0 \end{array} \\ \hline \bar{E}^T & 0 \end{array}\right] \tag{9}$$

is nonsingular. Expanding the determinant of W by $k \times k$ blocks of $\bar{E}^T$, we have

$$\det W = \sum_i \det(\bar{E}^T)_i \cdot \det(H_2 Y H)_i \tag{10}$$

where $(\bar{E}^T)_i$ is a $m \times m$ submatrix of $\bar{E}^T$, $(H_2YH^T)_i$ is a $(n-k)\times(n-k)$ submatrix of (H_2YH^T) associated with $(\bar{E}^T)_i$ in expansion, and the summation is over all $(m \times m)$ submatrices of $\bar{E}^T$. Moreover, any term $\det(H_2YH^T)_i$ in (10) can be further expanded by Binet–Cauchy Formula [6] into

$$\det\left(H_2YH^T\right)_i = \sum_j \det(H_2)_j \cdot \det(Y)_j \cdot \det\left((H^T)_i\right)_j \tag{11}$$

where

- $(Y)_j$ is a $(n-k)\times(n-k)$ principal submatrix of Y and is diagonal,
- $(H^T)_i$ is the $b\times(n-k)$ submatrix of H^T whose columns correspond to those of $(H_2YH^T)_i$ in (10),
- $((H^T)_i)_j$ is the $(n-k)\times(n-k)$ submatrix of $(H^T)_i$ whose rows correspond to the branches in $(Y)_j$,
- $(H_2)_j$ is the $(n-k)\times(n-k)$ submatrix of H_2 whose columns correspond to the branches in $(Y)_j$, and the summation is over all the $(n-k)\times(n-k)$ principal submatrices of Y.

Note that in the above expansion, the assumption of Y being diagonal has been used.

From (10) and (11), it can be seen that W is nonsingular only if there exist a pair of (i, j) such that $\det(\bar{E}^T)_i$, $\det(Y)_j$, $\det((H^T)_i)_j$, and $\det(H_2)_j$ are all nonzero. Since in this theorem only the connection of the branches is concerned, we will examine only the conditions of $\det(\bar{E}^T)_i \neq 0$, $\det((H^T)_i)_j \neq 0$, and $\det(H_2)_j \neq 0$.

First, consider the implication of $\det(H_2)_j \neq 0$. Let B be the set of branches corresponding to the columns of $(H_2)_j$. Then from the definition of H_2 in (7), it can be seen that $B \cap Q = 0$ and $\det(H_2)_j \neq 0$ if and only if the column-vectors of H corresponding to the branches of $B \cup Q$ are linearly independent. Since H is a cut-set matrix, the latter condition is equivalent to that $B \cup Q$ contains no loop in N [10]. From this condition together with the fact that there are n branches in $B \cup Q$, we can conclude that $B \cup Q$ must be a tree T' in N.

Next, consider the implication of $\det(\bar{E}^T)_i \neq 0$ and $\det((H^T)_i)_j \neq 0$ for some i. First, note that the branches corresponding to the rows of $((H^T)_i)_j$ are exactly those in B which together with Q constitute a tree T' in N. Let H_3 be the $n\times(n-k)$ submatrix of H whose columns correspond to those branches in B. Without loss of generality, we may assume that $T' = T$ and, therefore, $H_3^T = [0 : I_{(n-k)}]$. Then we let N' be a graph constructed from N by adding m branches from the reference node n_0 to all the rest of accessible nodes. Denote the set of these m branches by B'. Then E in (2) is the submatrix of the cut-set matrix of N' whose columns correspond to the m branches in B'. Let B'' be the subset of B' whose k branches correspond to the columns of $\bar{E}$. Now, from (10), we can see that $\det(\bar{E})_i \neq 0$ and $\det((H^T)_i)_j \neq 0$ for some i if and only if the matrix

$$\left[\begin{array}{c} H_3^T \\ \hline \bar{E}^T \end{array}\right] = \left[\begin{array}{c|c} 0 & I_{(n-k)} \\ \hline & \bar{E}^T \end{array}\right]$$

has full row-rank, or, equivalently, the matrix $[H_3 : \bar{E}]$ has full column-rank. The latter implies $B \cup B''$ contains no loop. Furthermore, since $B \cup B''$ contains n branches, it can be concluded that $B \cup B''$ forms a tree in N'. In other words, $T' - Q + B''$ is a connected graph. Since $B'' \in B'$, $T' - Q + B'$ is also connected. Thus $T' - Q$ is connected when branches in B' are shorted or, equivalently all the accessible nodes are shorted together.

Therefore, we have shown that Z_{mq} having full column-rank implies the existence of a tree T' which satisfies the conditions in Theorem 2. This proves the necessary part.

(*The sufficient part*): Suppose that there exists a tree T in N satisfying the conditions of Theorem 2. We will prove that Z_{mq} has full column-rank for almost all Y. In the first step, we want to show that this is true for a particular Y. We choose a specific set of branch admittances in the following way: first, set all admittances to 1 for those branches in $(T - Q)$, then set the rest to zero. We further label the branches in the same way as in the necessary part. Then it can be shown

$$\left(H_2YH^T\right) = \left[0 : I_{(n-k)}\right].$$

Note that $[0 : I_{(n-k)}]$ is also the submatrix of H^T whose rows correspond to the branches in $T - Q$. From Condition (2) of Theorem 2, $(T - Q)$ is connected when nodes in M are shorted. That is equivalent to say that $(T - Q + B')$ is connected and contains all nodes in N' from the discussion in the necessary part. We claim that there exists a subset B'' of B' such that $(T - Q + B'')$ is a tree in N'. Suppose that such B'' exists. Then the matrix

$$\left[\begin{array}{c|c} 0 & \bar{E} \\ \hline I_{(n-k)} & \end{array}\right]$$

is nonsingular, where $\bar{E}$ is the $n \times k$ submatrix of E whose columns correspond to B'' in N', because its column-vectors correspond to a tree $(T - Q + B'')$. Therefore

$$W = \left[\begin{array}{c|c|c} H_1YH^T & I_k \\ \hline 0 & I_{(n-k)} & 0 \\ \hline \bar{E}^T & 0 \end{array}\right]$$

is nonsingular and Z_{mq} has full column-rank for this Y. To prove the claim, note that $(T - Q)$ has no loop. Thus each loop in $(T - Q + B')$ must contain one branch in B'. Delete this branch from $(T - Q + B')$. The resultant graph is still connected and may contain loops each of which contains a branch in B'. Hence, the same deleting process may be repeated until there are only n branches left in the final graph which still contains $(T - Q)$ and k branches in B'. Denote those k branches by B''. Then $(T - Q + B'')$ is a tree in N' because $(T - Q + B'')$ is a connected spanning

subgraph of N' with $(n+1)$ nodes and $(T-Q+B'')$ consists of n branches [10].

In the second and final step, we want to show that Z_{mq} has full column-rank for almost all Y. From (10) note that $\det W$ is a polynomial of branch admittances. By a well-known result [11], it can be shown that $\det W$ is nonzero for almost all Y. Therefore, Z_{mq} has full column-rank for almost all Y. This proves the sufficient part. Q.E.D.

Note that the conditions given in Theorem 2 depend only on the graph of N but not on the values of Y. This is quite desirable in testability design since it relieves us of the consideration of the values of Y. However, these conditions must be verified for every two branches. It could be very cumbersome when N is very large. We will study the improvement next.

At this stage, it is essential to introduce some useful notations before the improvement of Theorem 2 can be done. First, it is convenient to have a testing graph N_α of N. N_α is constructed from N by connecting all accessible nodes to a new node α. Then, we would like to generalize the definition of a path. A path is conventionally defined as a series of nodes and branches incident with each other and contains distinct end-nodes. For our purpose, define a single node as a null-path, and a generalized-path (abbrev. g-path) is either a path or a null-path.

Now, we are able to have the two lemmas which will be used in the proof of the next theorem.

Lemma 1: Given N and its associated N_α. Suppose that there are r nodes which separate N_α into two parts; Parts A and B such that Part A contains α and Copart[1] A contains no less than r branches. Then the submatrix of Z_{mb} defined in (5) corresponding to those branches in Copart A has rank less than r.

Proof: First, note that Part B contains no accessible node since all accessible nodes are adjacent to node α. Also notice that the rank of the submatrix of Z_{mb} is exactly the dimension of space spanned by the column vectors in the submatrix. From (5), this is equivalent to the one spanned by the vectors Δv_m generated by the current sources each of which in parallel with a branch is Copart A. By Thevenin Equivalent Theorem, Copart A and those current sources can be represented by a $(r-1)$-port, as shown in Fig. 2. Each port is driven by a voltage source. Since there are only $(r-1)$-ports, by superposition theorem, the space spanned by Δv_m generated from these ports has dimension no larger than $(r-1)$. Therefore, the submatrix of Z_{mb} corresponding to the branches in Copart A has rank less than r. Q.E.D.

Lemma 2: Given N and its associated N_α, the following two statements are equivalent:

(a) There are at least k disjoint[2] paths in N_α between α and any inaccessible node.

[1]Copart A of a graph G associated with Part A is the subgraph of G formed by removing all the nodes in Part A from G.

[2]Two paths are disjoint if they have no common node except the end-nodes. And they are absolutely disjoint if they have no common node at all.

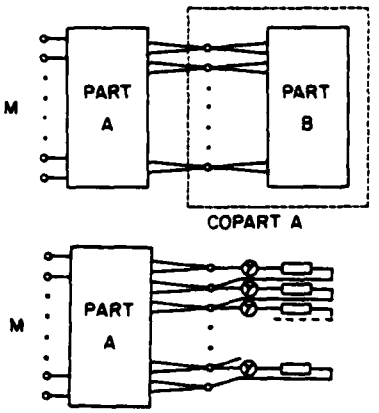


Fig. 2. Network of Lemma 1.

(b) There are at least k absolutely disjoint[2] g-paths in N between M and any k nodes.

Proof (a) → (b): It will be proved by contradiction. Suppose that (b) fails for a set of k nodes. Construct a graph $N_{\alpha\beta}$ from N_α by connecting these k nodes to a new node β. Then by Menger's Theorem, there exists a set R of $(k-1)$ nodes which separate $N_{\alpha\beta}$ into two parts: Part A contains α and Part B contains β. Note that Part B contains no accessible node but does contain at least one inaccessible node. The latter is true because if Part B contains no inaccessible node then β is the only node in Part B and all those k nodes connected to β must be all accessible nodes which automatically satisfy (b) and thus violate the assumption. Let v be such a node. Then v is separated from α by the same R in both N_α and $N_{\alpha\beta}$. Hence by Menger's Theorem, there are at most $(k-1)$ disjoint paths between v and α in N_α. This is a contradiction. Therefore, (a) must imply (b).

(b) → (a): It will also be proved by contradiction. Suppose that (a) fails for an inaccessible node v. Then by Menger's Theorem, there exists a set R of $(k-1)$ nodes which separate N_α into two parts: Part A contains α and Part B contains v. For the k nodes in $R \cup \{v\}$, we construct a graph $N_{\alpha\beta}$ from N_α by connecting all nodes in $R \cup \{v\}$ to a new node β. Then there are only $(k-1)$ disjoint paths between α and β in $N_{\alpha\beta}$. Hence in N, there are only $(k-1)$ absolutely disjoint g-paths between M and $R \cup \{v\}$. This is a contradiction. Therefore, (b) must imply (a). Q.E.D.

Then we can have the following theorem.

Theorem 3. Given a network N which has no parallel branches and its associated N_α, the following statements are equivalent.

(a) N satisfies the single-fault testability condition for almost all values of Y.

(b) There are at least three disjoint paths between α and any inaccessible node in N_α.

(c) The local connectivity between α and any inaccessible node is at least 3 in N_α.

Proof (b) ↔ (c): Directly from Menger's Theorem [7].

(a) → (b): We will prove it by contradiction. Let (b) be false. Then there are at most two nodes which separate N_α into two parts: Part A contains node α and Part B

contains at least one inaccessible node. Since all accessible nodes are adjacent to α, they cannot belong to Part B. If Part A and Part B are separated by one node, then the branches in Copart A has no effect on the measurements on M and henceforth the faults in Copart A cannot be determined. Now suppose that there are two nodes, say a and b, separating Part A and B. Then there are at least two branches in Copart A because N is connected and Copart A contains at least three nodes. Therefore, from Lemma 1, the single-fault testability condition cannot be satisfied. In either case, there is a contradiction. Therefore, Statement (b) must be true.

(b) $\rightarrow$ (a): Let b_1 and b_2 be any two branches in N. Branches b_1 and b_2 cannot be in parallel from the assumption, hence there are at least three nodes incident with b_1 and b_2. From Lemma 2 there are three absolutely disjoint g-paths in N between M and any three nodes. Thus there are three g-paths between M and some three nodes incident with b_1 or b_2. The Statement (a) will be shown by constructing a tree satisfying the two conditions in Theorem 2 based on these three g-paths.

First, consider the case that b_1 and b_2 have a common node. Then a tree T can be constructed by adding enough branches to b_1, b_2, and the three g-paths. Moreover, $T-\{b_1, b_2\}$ is a subgraph of three components each of which contains a g-path and therefore an accessible node. Thus $T-\{b_1, b_2\}$ is still connected when the nodes in M are shorted together. And hence T satisfies the two conditions of Theorem 2.

Next, for the case that b_1 and b_2 have no common node, let each of the three g-paths contain only one node incident with b_1 or b_2. This can be achieved by shortening a path which does not satisfy this condition. Denote these g-paths by P_1, P_2, and P_3. Without loss of generality, assume that b_1 joins P_1 and P_2 and P_3 contains one of the two nodes (v_a, v_b) incident with b_2 and v_c in M. We will obtain the desirable tree by modifying P_1, P_2, and P_3.

There are two sub-cases to be considered separately. The first sub-case is that v_a, b_b and one node of $P_1 \cup P_2$ are in a nonseparable subgraph in N. Then there is a loop that contains all three nodes. In this loop, there are two absolutely disjoint paths each of which contains exactly one node of b_2 and one node in $P_1 \cup P_2$. Denote the path that contains v_a by L_a and the other by L_b. Now in P_3, there exists a sub-g-graph such that it contains v_c and exactly one node, say v_0, in $L_a \cup L_b$. Without loss of generality, we may assume that $v_0 \in L_a$. Then a new g-path P_3' can be formed by the union of the g-path of P_3 and the g-path between v_0 and v_a of L_a. Note that the other node v_b of b_2 is connected to $L_1 \cup L_2$ by the path L_b which is absolutely disjoint from P_3'. Thus a tree T can be obtained by adding enough branches to the union of b_1, b_2, L_b, P_1, P_2, and P_3'. And $T-\{b_1, b_2\}$ must be connected when nodes in M are shorted together because the union of L_b, P_1, P_2, and P_3' which does not contain b_1 or b_2 consists of exactly three components each of which contains at least one node in M and the added branches do not alter this fact. Therefore, T satisfies the conditions of Theorem 2.

The remaining sub-case is that there is a cut-node v_0 between nodes in $P_1 \cup P_2$ and $\{v_a, v_b\}$. From Statement (b), there exist two absolutely disjoint g-paths between $\{v_a, v_b\}$ and M. Let P_4, P_5 be two such g-paths. Because v_0 is a cut-node, either $P_4 \cup P_5$ doesn't contain any node of $P_1 \cup P_2$ or only one of P_4 and P_5 contains nodes of $P_1 \cup P_2$ and it must also contain v_0. In the former case, it can be shown by arguments similar to those of the preceding cases that a tree obtained by adding enough branches to the union of b_1, b_2, P_1, P_2, P_4, and P_5 satisfies the conditions of Theorem 2. In the latter case, without loss of generality, assume that P_4 contains v_0, v_a, and nodes in $P_1 \cup P_2$. Let P_4' be the sub-path of P_4 which contains v_a and exactly one node in $P_1 \cup P_2$. Then it can be seen that the union of b_1, b_2, P_1, P_2, P_4', and P_5 contains no loop. Therefore, a tree in N can be obtained by adding enough branches to the union. And this tree satisfies the conditions of Theorem 2.

Summing up, for any two branches, a tree T can be obtained to meet the conditions in Theorem 2. Thus by Theorem 2, single-fault testability condition is satisfied and Statement (b) is proved. Hence the proof is completed. Q.E.D.

Comparing Theorem 3 to Theorem 2, both of them provide graphical conditions. However, the conditions in Theorem 3 are much simpler to implement and they only have to be verified once for every inaccessible node. The latter advantage becomes more significant when N is large. The applications of Theorem 3 will be illustrated by two examples in the next section.

Examples

To demonstrate the versatility of the above results, two examples will be provided. The first example is to illustrate the usage of Theorem 3 to verify the single-fault testability condition while the second one will extend the usage of Theorem 3 to the choice of test points or, equivalently, additional accessible nodes to satisfy the testability condition.

Example 1: Consider a ladder network N as shown in Fig. 3(a) with three accessible nodes a, b, and c. Construct the associated N_α by connecting these three accessible nodes to a new node α, as shown in Fig. 3(b). It is easy to show that every inaccessible node in N_α has exactly 3 disjoint paths to α. Therefore, by Theorem 1 and 3, N is single-fault testable. In fact, stages can be arbitrarily inserted into N without destroying the testability.

Example 2: Consider a more complicated network N, as shown in Fig. 4(a) with three accessible nodes a, b, and c. Its associated N_α is shown in Fig. 4(b). It can be shown that each of nodes d, e, and f has only two disjoint paths to α. Thus N is not single-fault testable. To achieve testability, there are two remedies. The apparent one is to introduce an additional accessible node. Any one of the nodes d and e can be the choice. Fig. 4(c) shows N_α' with the additional accessible node d which can be shown to achieve testability. Another remedy is to have new branches inserted in the network. This insertion can be either permanent if the network performance can be compromised, or introduced

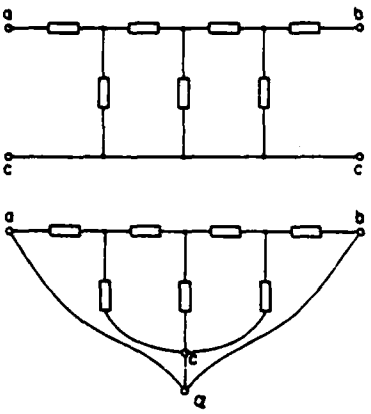


Fig. 3. Network of Example 1.

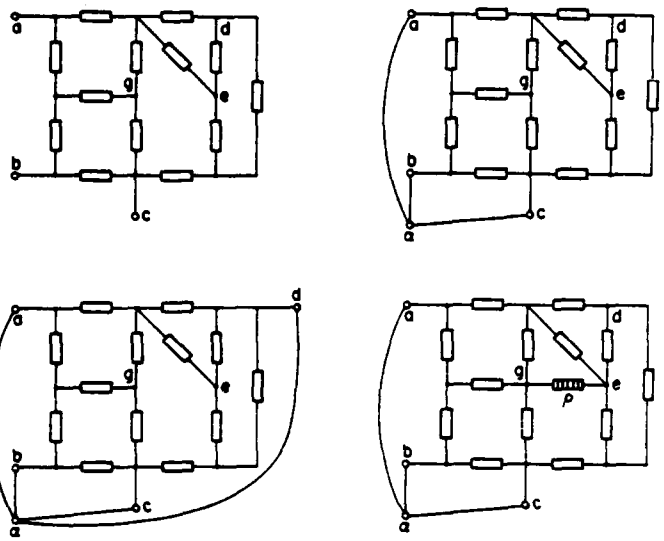


Fig. 4. Network of Example 2.

only during the testing. Fig. 4(d) shows that testability can be achieved by inserting a new branch p between nodes g and e.

REFERENCES

[1] R. M. Biernacki and J. W. Bandler, "Multiple-fault location of analog circuits," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 361–366, May 1981.
[2] R. M. Biernacki and J. W. Bandler, "Fault location of analog circuits," in *IEEE 1980 Int. Symp. Circuits and Systems*, pp. 1082–1086.
[3] A. A. Sakla, E. I. El-Masry, and T. N. Trick, "A sensitivity algorithm for fault detection in analog circuits," in *IEEE 1980 Int. Symp. Circuits and Systems*, pp. 1075–1077.
[4] Z. F. Huang, C. S. Lin, and R. Liu, "Node-fault diagnosis and design of testability," in *20th IEEE Conf. Decision and Control*, pp. 1037–1042, 1981. Also in *IEEE Trans. Circuits Syst.*, vol. CAS-30, pp. 257–265, May 1983.
[5] F. R. Gantmacher, *The Theory of Matrices*, vol. 1, Chelsea, 1974.
[6] S. Barnett and C. Storey, *Matrix Methods in Stability Theory*. New York: Barnes & Noble, 1970.
[7] F. Harary, *Graph Theory*. Reading, MA: Addison-Wesley, 1972.
[8] H. J. Carlin and A. B. Giordand, *Network Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1964.
[9] S. Seshu and M. B. Reed, *Linear Graphs and Electrical Networks*. Reading, MA: Addison-Wesley, 1961.
[10] W. K. Chen, *Applied Graph Theory*. North-Holland, 1971.
[11] W. H. Wonham, *Linear Multivariable Control*. Berlin, Germany: Springer-Verlag, 1974.

C. S. Lin was born in Taiwan in 1951. He received the B.S. degree in electrical engineering from National Taiwan University in 1973, and the M.S. and the Ph.D. degrees from the University of Notre Dame, Notre Dame, IN, in 1979 and 1983, respectively, both in electrical engineering.

Presently, he is Assistant Professor in the Department of Electrical Engineering at National Tsing Hua University, Hsinchu, Taiwan, Republic of China. His current research interest is in the area of analog network fault diagnosis.

Zeng F. Huang was born in Guangxi, China, in October 1943. He graduated in electrical engineering from Polytechnical Institute of Harbin, Harbin, China, in 1967.

He is currently a graduate student at the University of Notre Dame and expects to receive the M.A. degree in January 1983. He worked on power transistor amplifiers and transmitters in Broadcast Equipment Factory, Tianjin, from 1970 to 1973 and on color TV transmitters in Broadcase Equipment Factory, Beijing, from 1974 to 1978. His principal research interests are in the areas of nonlinear systems and fault diagnosis of analog circuits.

Ruey-Wen Liu (S'53–M'59–SM'78–F'81) was born in Kiangsu, China. He received the B.S., M.S., and Ph.D. degrees in 1954, 1955, and 1960, respectively, from the Department of Electrical Engineering at the University of Illinois at Urbana.

Since 1960, he has been with the University of Notre Dame, and is currently a Professor in the Department of Electrical Engineering. He has held visiting professorships at the National Taiwan University, Republic of China, in the spring of 1969, the Universidad de Chile, Santiago de Chile, in the summer of 1970, the Institute of Mathematics, Academia Sinica, Republic of China, in the summers of 1976 and 1978, and University of California at Berkeley, from 1965 to 1966 and from 1977 to 1978. His main interest has been in the area of nonlinear circuits and systems. Currently, he is the Chairman of the Technical Committee on Nonlinear Circuits and Systems of the IEEE Circuits and Systems Society. His other interests include fault diagnosis of analog circuits, feedback system design, estimation theory, and policy analysis. He is the Guest Editor of the Special Issue on Nonlinear Circuits and Systems, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, November 1980. He was the organizer and a Co-Chairman of the NSF-Workshop on Nonlinear Circuits Systems in January 1981. He was a Co-Guest Editor of the special issue on Socio-Economic Systems, *International Journal on Policy Analysis and Information Systems*. Currently, he is an associate editor of the *Journal of Circuits, Systems and Signal Processing*, and the *Journal on Policy Analysis and Information Systems*.

Dr. Liu is a member of Sigma Xi.

Linear Analog Fault Analysis - Theory and Implementation

by

R. Liu
C. Lin
Z.F. Huang

FAULT DIAGNOSIS OF LINEAR ANALOG NETWORKS: A THEORY AND ITS IMPLEMENTATION

C. S. Lin
Z. F. Huang
Ruey-wen Liu

University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

In the fault diagnosis of linear networks, a method which requires only the solving of linear equations has been attracting much attention [1-5]. This particular method was initiated independently by Biernacki and Bandler [1] and Sakla, El-Marsy and Trick [2]. Since then, numerous useful and interesting results have been published in a rather short time. It is the purpose of this paper to provide a summary in this area of research. Furthermore, an example is given to show how this idea can be "stretched" to apply to those cases which normally would not be applicable. This is achieved by use of multiple frequency testing and fault-logic equations.

I. A Theory of Fault Diagnosis

Let N be a connected, b-branch, (n+1)-node, (m+1)-terminal, linear, time-invariant lumped network. Out of the (n+1) nodes, (m+1) are accessible terminals for both excitation and measurement. In this section, the following notations are used for N.

v_m - terminal (accessible node) voltage vector.
i_m - terminal current vector.
v_b - branch voltage vector.
i_b - branch current vector.
Y_b - branch admittance matrix.
Y_n - node admittance matrix.

Note that Y_b and Y_n need not be symmetrical, and

$$i_b = Y_b\, v_b \tag{1}$$

$$Y_n = A\, Y_b\, A^T \tag{2}$$

where A is the incident matrix of the graph.

Now consider that the network N is perturbed to (N+ΔN) in the way that Y_b is perturbed to $Y_b+\Delta Y_b$, and the graph remains the same. In the perturbed network, we denote the corresponding voltages, currents, and admittances by $v_m+\Delta v_m$, $i_m+\Delta i_m$, $v_b+\Delta v_b$, $i_b+\Delta i_b$, and $Y_n+\Delta Y_n$, respectively.

Given Y_b and the graph, the purpose of the fault diagnosis is to locate the nonzero entries of ΔY_b from the information i_m, Δi_m, and Δv_m. Note that v_m can be calculated from i_m and N. Without loss of generality, we may assume that

$$\Delta i_m = 0 \tag{3}$$

i.e., apply same i_m to the original network N and the perturbed netwoek N+ΔN.

Denote Z_n and partition it in the following way:

$$Z_n = Y_n^{-1} = \left[\begin{array}{c} Z_{mn} \\ \hline Z_{\ell n} \end{array}\right] \tag{4}$$

where $Z_{mn} \in C^{mxn}$, $Z_{\ell n} \in C^{\ell xn}$, and $\ell=n-m$. Under this condition, it can be shown that

$$Z_{mb}\, j_b = \Delta v_m \tag{5}$$

where

$$Z_{mb} = Z_{mn} \cdot A \tag{6}$$

and

$$j_b = -\Delta Y_b\, (v_b+\Delta v_b) \tag{7}$$

In the above equation, a row of ΔY_b is entirely zero if and only if the corresponding branch is fault-free. Hence, a nonzero entry of j_b indicates that the corresponding branch is faulty. The converse is also true if the rare possibility of cancellation is disregarded. Since in Eq. (5), both Z_{mb} and Δv_m are known, we can solve the equation for j_b and thus obtain the location of the faulty branches. The usage of Eq. (5) to locate the branch faults is called branch diagnosis [1,2, 4].

Alternatively, Eq. (5) to (7) can be rearranged into the following equations:

$$Z_{mn}\, j_n = \Delta v_m \tag{8}$$

where

$$j_n = A\, j_b. \tag{9}$$

In Eq. (9), an entry of j_n is nonzero only if the corresponding node is incident with at least one faulty branch. The converse is almost always true in practical cases. Define a faulty node as the one incident with at least one faulty branch. Then we can solve Eq. (8) for j_n to locate the faulty nodes. Once the faulty nodes are located, the faulty branches can be determined from Eq. (9). This approach is called node diagnosis [3,5]. Branch diagnosis and node diagnosis are the two major approaches to locate the faults by solving linear equations.

There is a unique problem shared by both approaches, i.e., they all depend on solving the equation with the following form:

CH1845-7/83/0000-1090 $1.00

where the entries of the matrix B and the vector p are known and are in the complex field. Moreover, Eq. (10) in the fault diagnosis usually has more unknown variables than equations. Therefore, the solution is not unique in general. Fortunately, in most practical cases, there are only a small number of branch faults, thus a different but still small number of faulty nodes, in a faulty network. Taking advantage of this fact, we can restrict the set of solutions x to those with only a small number of nonzero entries, and then the solution can be unique. This will be discussed next.

Reconsider Eq. (10),

$$B\,x = p \tag{10}$$

where $x \in C^n$, $p \in C^m$, $B \in C^{m\times n}$ and $n > m$. Let

$$X_k = \{x \in C^n \mid \text{number of non-zero components of } x \text{ not more than than } k\}.$$

We want to investigate the uniqueness of the solution $x \in X_k$.

Definition. The global column-rank of B is said to be r if every combination of r columns of B is linearly independent, and some combination of (r+1) columns of B is linearly dependent.

Let Ω be the range of B. Then the following theorem is given in [3].

Theorem 1. Let $p \in \Omega$. Then Eq. (10) has a unique solution $x \in X_k$ for almost all $p \in \Omega$ if and only if the global column-rank of B is at least k+1.

Note that for fault diagnosis problem, $p \in \Omega$ is guaranteed. When Theorem 1 is satisfied, the solution of (10) can be obtained in the following way. Let the column vectors of B be denoted by

$$B = [b_1, b_2, \ldots, b_n]$$

Let the test matrices T_i be constructed in the following way:

$$T_i = [b_{i_1}, b_{i_2}, \ldots, b_{i_k}, p],\ i = 1,2,\ldots,s,\ s = C(n,k)$$

for every combination of k columns of B. Then,

(A) If

$$\text{Rank } T_i = k + 1,\ i = 1,2,\ldots,s \tag{11}$$

then there is no solution $x \in X_k$.

(B) Otherwise, there is a unique j such that

$$\text{Rank } T_j = k \tag{12}$$

and

$$\text{Rank } T_i = k + 1\ ,\ i \neq j. \tag{13}$$

and the unique solution $x \in X_k$ is given by

$$x = (B_j^T B_j)^{-1} B_j p \tag{14}$$

where $B_j = [b_{j_1}, b_{j_2}, \ldots, b_{j_k}]$.

The above result can now be applied to the node-fault-equation (8) to locate uniquely the faulty nodes if the number of faulty nodes is strictly less than the global column-rank of Z_{mn}. The application to the branch diagnosis needs more elaboration and can be found in [4]. Note that this result can also be applied to any other fault equation having the form of (10).

The remaining big question is that whether or not the theory developed above is robust, i.e., whether or not the theory based on the nominal circuit is still valid for the actual circuit whose parameters always differ from those of the nominal circuit. The answer depends on how the theory is implemented. One such implementation will be presented next. Note that (11), (12), and (13) are equivalent to the question that whether or not p is a linear combination of $B_i = [b_{i_1}, b_{i_2}, \ldots, b_{i_k}]$. The question has an answer in the residual number

$$e^0 = p^T(I - B_i\, B_i^+)p$$

where B_i^+ is the pseudoinverse of B_i. It is known that $e^0 = 0$ if and only if p is a linear combination of column vectors of B_i. It can be shown that this test is robust [7]. This implementation was also suggested in [1].

An important development should also be noted that the global column-rank of Z_{mb} and Z_{mn} can be shown that it depends mainly on the graph of N, not on the circuit parameters [3,4].

So far, we have considered the case with single test signal and with single frequency. In the next section, we will show, by an example, how the theory can be "stretched" when multiple frequencies are used and how a solving of nonlinear equations can be avoided by use of fault-logic equations.

II. Implementation on an Active Filter

Consider a 3-pole low-pass active filter as shown in Fig. 1. We are to locate the single faults (resistors and capacitors) in this network by using the input and output terminals as the only accessible nodes. It can be easily shown that the global column-rank of both Z_{mb} and Z_{mn} is one, and therefore according to the theory the circuit is not one-fault diagnosible. We will show how to use multiple frequency and fault-logic equations to alleviate the problem.

Our algorithm is hierarchical in essence, i.e., the fault is first isolated into a subnetwork from a single test signal and then this faulty subnetwork is further diagnosed by the usage of multiple test frequencies. The algorithm will be described as follows.

Step A: The circuit is first reformulated into a block diagram as shown in Fig. 2. The relationship between the transfer functions, denoted by A_i's and the R's and C's is given as follows:

$$\begin{aligned} A_1 &= R_1 \\ A_2 &= 1/(SC_1) \\ A_3 &= 1/R_2 \\ A_4 &= 1/R_7 \\ A_5 &= -1/(G_3+SC_2) \\ A_6 &= R_5/(R_4R_6SC_3) \end{aligned} \qquad (15)$$

I_i is the input, and V_i and V_o are the outputs. The extra signal sources J_i, i=1,2,...,6, at the output of the blocks A_i is the fault compensator of the block A_i. The meaning of the fault compensator is given in Fig. 3. Note that if A_i is changed to $(A_i+\Delta A_i)$ then $J_i=U_i\Delta A_i$, where U_i is the input to A_i. Assuming that $U_i\neq 0$, we have $J_i=0$ if and only if $\Delta A_i=0$. Our first step is to locate non-zero J_i's from V_i and V_o.

From Fig. 2, it can be shown that

$$\begin{bmatrix} \Delta V_i \\ \Delta V_o \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & A_3A_5A_6A_G & A_5A_6A_G & A_5A_6A_G & A_6A_G & A_G \end{bmatrix} \begin{matrix} J_1 \\ J_2 \\ J_3 \\ J_4 \\ J_5 \\ J_6 \end{matrix} \qquad (16)$$

where $A_G=1/(1-A_4A_5A_6)$. This equation has the form of (10). Therefore Theorem 1 can be applied. Since the last four column-vectors are linearly dependent, we cannot distinguish among (J_3,J_4,J_5,J_6). Let the last four column-vectors be grouped into one. Then the resulting matrix has three columns and whose global column-rank is 2. Therefore, it is one-fault diagnosable, i.e., we may determine whether A_1, A_2 or one of (A_3,A_4,A_5,A_6) is at fault. In view of (15), by use of single signal testing, we can isolate the fault into three groups: R_1, C_1 and the rest. This is established in Table 1a.

Step B: The faulty subnetwork isolated by Step A can be further diagnosed by the usage of multiple test frequencies. First, observe that the only part that requires further diagnosis is the group of (A_3,A_4,A_5,A_6). It can be described by the transfer function of V_0/V_2 where V_0 is the output voltage of the circuit and V_2 is the output of A_2 and which can be calculated if A_2 is fault-free. This can be determined from Step A. The expression of V_0/V_2 is given by

$$\frac{V_0}{V_2} = \frac{B_4}{B_1S+B_2S^2+B_3S^3} \qquad (17)$$

where S is the complex frequency, and

$$\begin{aligned} B_1 &= 1/R_7 \\ B_2 &= (R_4R_6C_3)/(R_3R_5) \\ B_3 &= (R_4R_6C_2C_3)/R_5 \\ B_4 &= 1/R_2 \end{aligned} \qquad (18)$$

The relation between (V_0,V_2) and circuit parameters R's and C's is nonlinear. In order to avoid solving nonlinear equations, we propose to do the following. To save subscripts, denote V_0 by V and V_2 by W.

Assume that $B_i \rightarrow (B_i+\Delta B_i)$, i=1,2,3,4. Exciting the ciruit by k different frequencies, the above equation becomes

$$\begin{bmatrix} S_1V_1 & S_1^2V_1 & S_1^3V_1 & -W_1 \\ S_2V_2 & S_2^2V_2 & S_2^3V_2 & -W_2 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ S_kV_k & S_k^2V_k & S_k^3V_k & -W_k \end{bmatrix} \begin{bmatrix} \Delta B_1 \\ \Delta B_2 \\ \\ \Delta B3 \\ \\ \Delta B_4 \end{bmatrix} = \begin{bmatrix} S_1V_1 & S_1^2V_1 & S_1^3V_1 & -W_1 \\ S_2V_2 & S_2^2V_2 & S_2^3V_2 & -W_2 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ S_kV_k & S_k^2V_k & S_k^3V_k & -W_k \end{bmatrix} \begin{matrix} B_1 \\ B_2 \\ \\ B3 \\ \\ B_4 \end{matrix} \qquad (19)$$

We need to determine non-zero ΔB_i's from (19). Since the R.H.S. is known, Eq. (19) has the form of (10). Therefore, Theorem 1 again can be applied. When the k different frequencies are properly chosen, the global column-rank of the L.H.S. matrix is three. Therefore, we can uniquely determine two non-zero ΔB_i's. From these non-zero ΔB_i's, the faulty circuit parameters can be located from (18) by a logic argument. This can be presented formally as follows.

From (18), a set of fault-logic equation is constructed:

$$\begin{aligned} F(B_1) &= F(R7) \\ F(B_2) &= F(R_3) \vee F(R_4) \vee F(R_5) \vee F(R_6) \vee F(C_3) \\ F(B_3) &= F(C_2) \vee F(R_4) \vee F(R_5) \vee F(R_6) \vee F(C_3) \\ F(B_4) &= F(R_2) \end{aligned} \qquad (20)$$

The first fault-logic equation reads that ΔB_1 is non-zero if and only if ΔR_7 is non-zero. The second fault-logic equation reads that ΔB_2 is non-zero if and only if $\Delta R_3,\Delta R_4,\Delta R_5,\Delta R_6$, or ΔC_3 is non-zero. Similarly for the third and the fourth fault logic eqations. We have assumed that no two faults wll occur in such a way that their effects will cancel each other, which is rare in practice.

Finally, we need to solve the set of fault-logic equations (20). This can be done by use of a truth value table. The solution, together with the solution in Step A, is given in Table 1. An entry 0 means that $\Delta A_i\neq 0$ in Table 1a and $\Delta B_i\neq 0$ in Table 1b. An entry X means $\Delta A_i=0$ in Table 1a and $\Delta B_i=0$ in Table 1b. For example, the second row of Table 1a reads that if $\Delta A_2\neq 0$, and $\Delta A_1=\Delta A_3=\Delta A_4=\Delta A_5=\Delta A_6=0$ then C_1 is faulty. The fifth row of Table 1b reads that if $\Delta B_2\neq 0$, $\Delta B_3\neq 0$, and $\Delta B_1=\Delta B_4=0$ then a fault occurs among (R_4,R_5,R_6,C_5). The reason that the last group of circuit parameters cannot be further distinguished can be easily found from the circuit. They are related as a single product and therefore, they cannot be further diagnosed unless additional test points are introduced.

This implementation has many interesting points:

1. The fault diagnosis decision is based on matching fault patterns (Table 1). It is like the fault dictionary approach for fault diagnosis of digital systems. It is extremely simple from the user's point of view. Such table is called fault directory [6].

2. The post-fault computation is minimal. The only computation required is the calculation of residuals. Its complexity depends on the number of faults, not the size of the network.

3. The use of multiple frequency and fault-logic equation avoids a complex and non-robust computation of solving a set of nonlinear equations.

4. Our computer simulation study [8] shows that this method of implementation is very robust.

III. References

[1] R.M. Biernacki and J.W. Bandler, "Fault Location of Analog Circuits", IEEE 1980 Int'l. Symposium on Circuits and Systems, pp. 1078-1081.

[2] A.A. Sakla, E.I. El-Masry and T.N. Trick, "A Sensitivity Algorithm for Fault Detection in Analog Circuits, IEEE 1980 Int'l. Symposium on Circuits and Systems, pp. 1075-1077.

[3] Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and Design of Testability", Tech Report No. 811, University of Notre Dame, Notre Dame, Indiana, July 27, 1981. Also, 20th IEEE Conference on Decision and Control, pp. 1037-1042, 1981.

[4] C.S. Lin, Ph.D. Dissertation, University of Notre Dame, Notre Dame, IN, 1983.

[5] Y. Togawa and T. Matsumoto, "On the Topological Testability Conjecture for Analog Fault Diagnosis Problems", (submitted for publication).

[6] C. Lin and R. Liu, "Fault Directory Approach - A Case Study", IEEE 1981 Int'l. Symposium on Circuits and Systems Proceedings, pp. 239-242.

[7] L.C. Suen and R. Liu, "Determination of the Structure of Multivariable Stochastic Linear Systems", IEEE Transactions on Automatic Control, pp. 458-464, 1978.

[8] C.S. Lin and R. Liu, "Fault Diagnosis on a Navy Candidate Circuit", Notre Dame Memurandum, March 1982.

A_1	A_2	A_3	A_4	A_5	A_6	Faulty Components
0	X	X	X	X	X	R_1
X	0	X	X	X	X	C_1
X	X	0	0	0	0	Others

Table 1a. Fault Patterns for Step A.

B_1	B_2	B_3	B_4	Faulty Components
0	X	X	X	R_7
X	0	X	X	R_3
X	X	0	X	C_2
X	X	X	0	R_2
X	0	0	X	R_4, R_5, R_6, C_3

Table 1b. Fault Patterns for Step B.

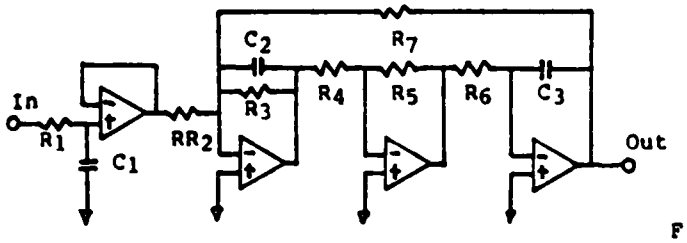


Fig. 1

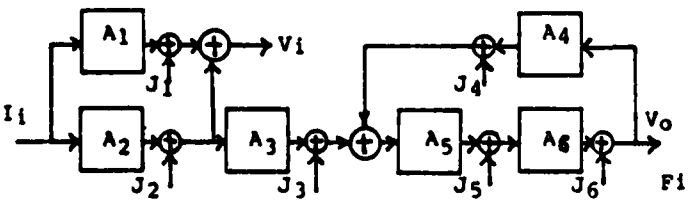


Fig. 2

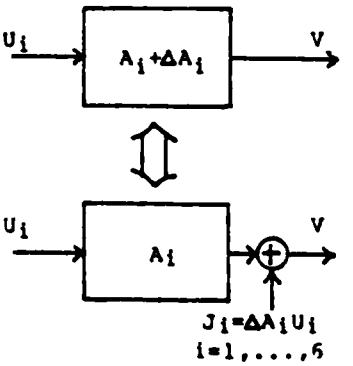


Fig. 3

Analog Fault Diagnosis - A New Circuit Theory

by

R. Liu
C. Lin
Z.F. Huang
L.Z. Hu

ANALOG FAULT DIAGNOSIS: A New Circuit Theory

Ruey-wen Liu
Chen-Shang Lin*
Zheng-Feng Huang
Lai-Zhao Hu

University of Notre Dame
Notre Dame, IN 46556

0. THE BACKGROUND

During the past quarter century, the engineering community has been witness to tremendous strides in the art of electronics design. On the contrary, electronics maintenance has changed little since the day of the vacuum tube. As such, our ability to design a complex electronic circuit is quickly outdistancing our ability to maintain it. In turn, the price reductions which have accompanied modern electronics technology have been paralleled by increasing maintenance and operation costs. Indeed, many industries are finding that the life cycle maintenance costs for their electronic equipment now exceeds their original capitol investment.

Given the above, it is quickly becoming apparent that the electronics maintenance process, like the design process, must be automated. Unfortunately, the 50 years of progress in circuit theory, on which our electronics design automation has been predicated, does not exist in the maintenance area. As such, the past decade has witnessed the inauguration of a basic research program to lay the foundations for a theory of electronics maintenance and a parallel effort to develop operational electronic maintenance codes.

Thus far the greatest success has been achieved in the digital electronics area to the point that commercialized test programs are now readily available. On the other hand, the analog testing is still in its infancy. This is not without reasons.

* National Tsing Hua University, Hsinchu, Taiwan, Republic of China.

For one reason, the analog fault diagnosis had a late start. The research and theory development of digital testing started in the mid 1960's when the large-scale computers were readily available. Not until a decade later did a commercialized test program first become available. On the other hand, it was not until the mid 1970's that the test technology community began to face up to the analog test problem. Indeed, even in a predominantly digital world, analog systems were not disappearing. Analog systems were proving to be among the most unreliable and least readily tested of all electronic systems. Assuming the same speed for the development of digital testing, a commercially available analog testing program would not have been ready until the mid 1980's.

The main reason is that the analog fault diagnosis has inherited certain difficult problems which are not shared by digital fault diagnosis. These will be explained later.

I. INTRODUCTION

One can trace as far back as the early 1960's to find that circuit theorists had an interest in the analog fault diagnosis problem [1,2]. However, there was only sparse interest thereafter [3-9]. It was not until 1977 that a collection of papers [10] appeared, and a special issue followed [11]. In the meantime, the interest among circuit theorists suddenly became active [12-52]. An excellent review paper with extensive references for the pre-1979 period appeared in the special issue [12].

In this paper, only the post-1979 activities will be discussed. The Fault/Tolerance Compensator approach will be extensively discussed because it is probably the most promising approach at the present time, (based on the authors' prejudiced point of view?)

As a start, a simplified analog fault diagnosis problem is presented in Section II and some major issues are discussed in Section III. The Element-Value Solvability Problem is discussed in Section IV. A fault/tolerance compensation model is presented in V. The k-Fault Diagnosis Problem is the ideal case (Section VI) and the tolerance case (Section VII) follows. A conclusion is given in Section VIII.

II. AN INTRODUCTION TO ANALOG FAULT DIAGNOSIS

Consider a simple system

$$\begin{aligned} Ax &= u \\ y &= Cx \end{aligned} \tag{1}$$

where $A \in R^{n\times n}$ is non-singular, $u \in R^n$ the input, $y \in R^m$ the output, $x \in R^n$ the internal variable, and C a selector matrix (each row of C has one and only one entry being 1 and the rest of them 0), which selects certain components of x for measurement. Therefore, m is the number of test points, and n the size of the system. In general, $m << n$.

Suppose that A is perturbed to $(A + \Delta A)$. With the same input u, x and y will be perturbed accordingly, i.e.,

$$\begin{aligned} (A + \Delta A)(x + \Delta x) &= u \\ (y + \Delta y) &= C(x + \Delta x) \end{aligned} \tag{2}$$

We will now pose the first problem.

Problem 1: (Element-Value Solvability Problem)

Can we determine ΔA from the input/output measurements?

The answer to this problem is relatively simple. If all the internal variables can be measured, i.e., $m = n$, then one can determine $(A + \Delta A)$, and hence ΔA, with n independent inputs. The condition $m = n$ is also necessary if no additional information is known about ΔA.

Since the constraint $m << n$ is imposed upon us, we have to ask the next realistic question. Suppose that a perturbation $(y + \Delta y)$ is observed at the output and we want to locate which entries of A have been perturbed. In this case, it is quite realistic to assume that the number of entries of A, say k, were perturbed at an instant when a perturbation was observed at the output is small ($k << n$). In other words, ΔA is a sparse matrix. Furthermore, we need only to determine the location of non-zero entries of A, not their values. We can now pose the second problem.

Problem 2: (k-Fault Diagnosis: The ideal case)

With the constraint $m << n$, and the assumption that ΔA is sparse, can we determine the location of non-zero entries of ΔA from the input/output measurements?

It turns out that if certain conditions are met, we can uniquely determine the non-zero rows of ΔA. (Fortunately, this information is enough for us to locate the faulty circuit elements.) This can be seen as follows.

Subtracting (1) from (2), one obtains

$$CA^{-1} \Delta A(x + \Delta x) = -\Delta y \tag{3}$$

or

$$CA^{-1} Z = -\Delta y \tag{4}$$

and

$$Z = \Delta A(x + \Delta x) \tag{5}$$

Note that since CA^{-1} is fat ($m < n$), the determination of ΔA from (3) is not possible. On the other hand, barring cancellations, it can be seen from (5) that a component of Z is non-zero if and only if the corresponding row of ΔA is non-zero. Therefore, the problem is reduced to the location of non-zero components of Z from (4) when the matrix CA^{-1} is fat. A satisfactory solution to this problem can be found in [30], and it will be presented later.

In the final case, the problem of tolerance will be introduced. In practice, System (1) can not be built precisely according to the specified A, i.e., the nominal A. A specified tolerance usually is also given so that an actual system $(A + \Delta A_t)$ is said to be "good" if ΔA_t is within the tolerance. In this case, the perturbation ΔA will have two components,

$$\Delta A = \Delta A_t + \Delta A_f \tag{6}$$

where ΔA_t is the deviation of actual A from nominal A due to the tolerance, and ΔA_f is due to some large deviation (fault) of some entries of A. It is realistic to assume that ΔA_t is not sparse but the value of each entry is "small" and that ΔA_f is sparse but the values of non-zero entries are "large". Now, we can state the third problem.

Problem 3. (k-Fault Diagnosis: The tolerance case)

With the constraint $m \ll n$, and the assumption that ΔA_t is small and ΔA_f is sparse but non-zero entries are large, can we locate the non-zero entries of ΔA_f from the input/output measurements?

The above problem can be posed in two different kinds of setting.

1. The Problem of Robustness. We may design a testing program (for k-fault isolation) based on the ideal case, and then put the testing program into a simulation test to see if this program is robust under the tolerance case. This is the state-of-the-art at the present time. It turns out that to design a robust testing program is a very tough problem. For one thing, any robust testing program should avoid the inversion of a matrix, or keep it at a minimum.

2. The Fault Decision Problem. This tolerance problem can also be posed as one of decision/detection problem. This can be seen as follows. Let

$$Z_t = \Delta A_t(x + \Delta x)$$

$$Z_f = \Delta A_f(x + \Delta x),$$

then Eq. (4) can be represented in Fig. 1. In this figure, Z_t is generally small while the non-zero entries of Z_f are generally large, but sparse. The problem is to locate the non-zero entries of Z_f from Δy. This is a non-conventional decision/detection problem. At the present time, it is still an open problem.

At this point, it is important to notice that when the number of test points is reduced, the nature of the problem changed completely from Problem 1 to Problem 2; and that when the tolerance problem is introduced, the issue of computation becomes more complex.

Finally, note that what has been presented is a simplified analog fault diagnosis problem. In general, the matrix A is a (nonlinear) function of circuit parameters as well as frequencies.

III. IMPORTANT ISSUES OF ANALOG FAULT DIAGNOSIS

In this section, we will discuss two major issues of analog fault diagnosis problems and three important measures of the effectiveness of a testing program.

The two major issues are the tolerance problem and the problem of modeling and simulation of faulty components:

- Tolerance: Possibly the single greatest unknown in the design of an analog testing program is the effect of the tolerances of the "good" component on the performance of a testing program. This tolerance problem has absolutely no counterpart in the digital testing problem. The effect of these tolerances can completely dominate the performance of a testing program. In an analog circuit, unlike digital circuits, the actual values of circuit parameters almost always deviates from the nominal values. Therefore, any analog testing program has to face up to the problem of tolerance problem.

- Modeling and Simulation of Faulty Components: Unlike the digital testing, a complete modeling (and thus simulation) of faulty components is not available for the development of a testing problem. The modes of faulting is too many to encounter. For example, a faulty resistor may have an infinite number of possible resistances (outside of the tolerance). In fact, it can even be nonlinear. A faulty capacitor may have a model of parallel RC. A faulty operational amplifier may have a model of 22 transistors 12 resistors and a capacitor! A good transistor may behave like a faulty one if its bias is switched due to a fault which occurred elsewhere! In fact, in a nonlinear analog environment, we are still in the process of developing viable CAD models for nominal devices, let alone for faulty devices. As such, a thorough test of the performance of a testing program is impossible. Furthermore, each testing program has to be designed based soley on the nominal values of the circuit.

Besides the two major problems mentioned above, there are three important measurements of the effectiveness of a testing program.

- Test Points. Due to the practical restriction that there are usually only a few nodes accessible for measurement and testing, the number of required test points has to be as small as possible.
- Post-Fault Computation. Since the post-fault computation is directly related to the per unit cost, it is important to keep the post-fault computation time short and simple.
- Robustness. This issue has been raised many times; only because it is indeed the dominant issue at hand. We want the testing program to be reliable when the fault/tolerance ratio is small.

IV. THE ELEMENT-VALUE SOLVABILITY PROBLEM

The element-value solvability problem is initiated by R.S. Berkowitz [1] in 1962. This problem is presented in Fig. 2. The network N consists of RLC-elements, possibly with dependent sources. There are m accessible terminals. The problem is to determine the circuit parameters, RLC, from the measurements at the accessible terminals.

It is easy to see that at the terminals one can at most measure the m-port input impedance matrix H(S) and this can be done, for example, by system identification methods. The transfer function matrix is a function M of frequency S and circuit parameter p, i.e.,

$$M(S,p) = H(S)$$

The problem is to solve for p from H(S). In other words, it is a m-port network synthesis problem when the graph and the element-kind is prescribed.

In general, M is a nonlinear function of p. As a rule of thumb, when the number, m, of accessible nodes (test points) increases, M becomes "less nonlinear", and vice versa. This is an important trade-off problem. In this regard, there are two standard results to be quoted constantly as a measuring stick. Trick, Mayeda and Sakla [15] have shown that if all nodes are accessible then M becomes linear and p can be uniquely determined.

On the other hand, for a one-port RC-ladder, their element values can always be determined, regardless of the number of stages. This follows from a well-known RC synthesis theorem. This example shows that there exists a circuit-type that regardless how large the circuit is, the circuit parameters can be determined from only two accessible nodes. However, in this case M becomes "extremely nonlinear". To see this, a 4-stage RC ladder is studied [27]. The continue-fractional expansion method is used to determine the circuit values. A striking result has been found. When the significant digits

used for computation is eight, the answer came out correctly. However, when it is reduced to seven, the answer becomes erroneous. In fact, some values of R and C become negative! This shows that the RC ladder is solvable in theory, but it will yield erroneous conclusions even with a slightest computation error. Therefore, it becomes unreliable.

In between the above two extreme cases of the computation-test point trade-off, the most elegant result is found by Navid and Willson [13]. First, their solvability condition is on the topology not on the element values. Therefore, it can be tested very easily. Second, the number of test points required is not too large, but roughly the square root of the number of branches. Finally, it is computationally tractable, although a set of non-linear equations has to be solved.

A necessary and sufficient condition on local diagnosability of nonlinear circuits is elegantly derived by Visvanathan, Sangiovanni-Vincentelli [39] and Saeks et al. [40]. This is an important contribution because it provides a theoretical limit. It will be most helpful to see if such a condition can be made robust in the presence of tolerance.

In summary, the advantage of the element value solvability method is that it avoids the tolerance problem because all parameters are calculated and can be compared to see if they fall within the tolerance. The method also can be applied to the case when the number of faulty elements is large. The major difficulty with this method is that they have to solve, in general, a set of nonlinear equations every time they are tested. The number of nonlinear equations to be solved is no less than the number of internal parameters, which in general is large. In order to make this method work, we need a computational procedure which is robust for a large number of nonlinear equations.

V. A FAULT/TOLERANCE COMPENSATION MODEL

In this section, a model is introduced which can be used effectively for fault diagnosis problems. This can be simply explained in Fig. 3.

Suppose that the actual admittance (Fig. 3b) of particular branch is deviated from its nominal admittance y (Fig. 3a) by Δy. The deviation may be caused by the tolerance (in this case Δy is small), or by a fault (in this case Δy is large). According to a circuit theory, the deviation can be compensated by a current source j which depends on the deviation Δy and its branch voltage. This is true whether Δy is linear or nonlinear (in this case, a small signal is assumed). The value of j in general is a complex number (even when y is real) and it is "large" if the deviation is caused by a fault, and it is "small" if by a tolerance. Hence, the fault diagnosis problem becomes a detection problem of whether j is large or small.

The same is true for a three-terminal device as shown in Fig. 4. Here Y is a 2x2 admittance matrix. We need two current sources to compensate the deviations. If the device is a "good" transistor, Y is its linear model and the compensators represent the errors caused by the linear approximation to the nonlinear transistor. If the device is a "faulty" operational amplifier, the compensators may represent the large deviations of the gain, or even the deviations caused by its internal transistors and capacitors. In fact, this model can be applied to any three-terminal device or any three-terminal sub-circuit.

In general, an n-terminal device (chip) needs (n-1) current sources to compensate its deviations. When all actual devices are replaced by their compensation model, we have a Fault/Tolerance Compensation Model for the actual circuit. In this model, it consists only of the nominal circuit excited by

actual excitations and by fault/tolerance compensators. Therefore, our problem is to locate the fault compensators.

Consider the circuit in Fig. 2a, with (m + 1) accessible nodes and b branches. Its Fault/Tolerance Compensation Model is given in Fig. 2b, where J_b is the F/T compensator vector. When these two circuits are excited by the same I_m, and the responses V_m and $(V_m + \Delta V_m)$ are measured respectively, it is easy to show that

$$Z_{mb}\ J_b = -\Delta V_m \tag{7}$$

where Z_{mb} is the branch-to-accessible node transfer function matrix, depending only on the nominal circuit.

Equation (7) was first derived by Biernacki and Bandler [37], based on linear perturbations. By use of the Fault/Tolerance Compensation model, we have shown that Eq. (7) is also valid for linear nominal circuit with nonlinear perturbations. In fact, the perturbations can be from R to parallel RC, or from an operational amplifier to a complex circuit of (nonlinear) transistors and capacitors. Therefore, the problem of modeling faulty devices can be avoided as long as its nominal model is linear.

Finally, the compensator J_b has two components:

$$J_b = J_{bF} + J_{bT} \tag{8}$$

where J_{bF} is the fault compensator and J_{bT} is the tolerance compensator. J_{bF} is sparse but non-zero entries are large while J_{bT} is small but unknown to us. The problem becomes the determination of the non-zero entries of J_{bF} from (7) and (8). This will be discussed in the next two sections.

VI. k-FAULT DIAGNOSIS: The Ideal Case

In this section, we assume that the number of accessible nodes m is much smaller than the total number of nodes n, i.e., $m \ll n$. We further assume that the number of faulty branches is much smaller than the total number of

branches, i.e., $k \ll b$. The nominal circuit is linear and it can be represented by Eqs. (7) and (8). Finally, we assume that $J_{bF} = 0$. In other words, we want to locate the non-zero entries of J_b from (7), where J_b is sparse. A solution to this problem can be found in [49], which will be presented as follows.

Consider the following equation:

$$B\,x = p \tag{9}$$

where $x \in C^n$, $p \in C^m$, $B \in C^{mxn}$ and $n > m$. Let

$$X_k = \{x \in C^n \mid \text{number of non-zero components of } x \text{ not exceeding } k\}.$$

We want to investigate the uniqueness of the solution $x \in X_k$.

Definition. The global column-rank of B is said to be r if every combination of r columns of B is linearly independent, and some combination of (r+1) columns of B is linearly dependent.

Let Ω be the range of B. Then the following theorem is given in [49].

Theorem 1. Let $p \in \Omega$. Then Eq. (9) has a unique solution $x \in X_k$ for almost all $p \in \Omega$ if and only if the global column-rank of B is at least k+1.

Note that for fault diagnosis problem, $p \in \Omega$ is guaranteed. When Theorem 1 is satisfied, the solution of (9) can be obtained in the following way. Let the column vectors of B be denoted by

$$B = [b_1, b_2, \ldots, b_n]$$

Let the test matrices T_i be constructed in the following way:

$$T_i = [b_{i_1}, b_{i_2}, \ldots, b_{i_k}, p],\ i = 1,2,\ldots,s;\ s = C(n,k)$$

for every combination of k columns of B. Then,

(A) If

$$\text{Rank } T_i = k + 1,\ i = 1,2,\ldots,s \tag{10}$$

then there is no solution $x \in X_k$.

(B) Otherwise, there is a unique j such that

$$\text{Rank } T_j = k \quad (11)$$

and

$$\text{Rank } T_i = k + 1 \ , \ i \neq j. \quad (12)$$

Furthermore, the unique solution $x \in X_k$ is given by

$$x = (B_j^T B_j)^{-1} B_j P \quad (13)$$

where $B_j = [b_{j_1}, b_{j_2}, \ldots, b_{j_k}]$.

As a consequence, the k-fault diagnosability depends on the global column-rank of Z_{mb}. Let us now examine the issue of (1) post-fault computation and (2) the number of required test points.

1. Post-Fault Computation. The post-fault computation includes only Eqs. (10-13) It involves the test of the rank of mx(k+1) matrices T_i, and an inversion of a kxk matrix $(B_i^T B_i)$. This is a very good feature since both m and k are small. More importantly, the size of these matrices does not depend on the complexity of the circuit b. The total post-fault computation can be implemented on a mini-computer with computation time ranging from a fraction of a second to the order of a few seconds.

2. Number of Required Test Points. For k-fault diagnosis, it is necessary for the number of required test points $m \geq k + 1$. It is important to note that it again depends mainly on k not on b; the complexity of the circuit.

As an example, assuming the optimal condition $m = k + 1$, for a single-fault diagnosis, we need only two test points, testing the rank of 2x2 matrices and an inverting an 1x1 matrix, regardless of how large b is.

Next, we will investigate how the optimal condition $m = k + 1$ can be achieved. This requires

$$\text{global column-rank } Z_{mb} = m = k + 1 \quad (14)$$

This depends on how the circuit is designed and more importantly how the test points are located. It turns out surprisingly that the global column-rank of Z_{mb} depends mainly on the topology of the circuit, not its element values.

3. Topological Conditions for k-Fault Diagnosibility. First, it can be shown that if there is any internal loop (loops not incident with test nodes) consisting of r branches than the global column-rank of Z_{mb} cannot be greater than (r-1). Therefore, we can diagnose at most (r-2)-faults. Since most circuits have loops consisting of three branches, we can diagnose up to only one-fault. This is a serious limitation. Alternatively, we can work on the node-fault diagnosis equation

$$Z_{mn}\ J_n = -\Delta V_m \tag{15}$$

where $J_n = A\ J_b$ is the node compensator vector, $Z_{mb} = Z_{mn}\ A$ and A the incident matrix. Theorem 1 can now be applied to (15). In this case, a topological condition has been derived independently by Huang, Lin and Liu [49] and Togawa and Matsumoto [51]. First, construct a testing graph G_t from the given graph G by (1) deleting all branches which are incident between two accessible nodes and (2) connecting all accessible nodes, except the reference node, to a new node t.

Theorem 2. Let the network be passive and G be connected. The following three statements are equivalent:

(A) The global column-rank of Z_{mn} is k for almost all branch admittances.

(B) The local connectivity* between the node t and any inaccessible node in G_t is k.

(C) There are at least k independent paths* in G_t from any inaccessible node to the node t.

The topological conditions for the global column-rank of Z_{mb} equal to 2 is given in [50]. A more general topological condition is obtained in [52]. These topological conditions are very useful. (1) They provide a very easy evaluation of the global column-rank of a transfer function matrix. (2) They provide a foundation for the design of diagnosable circuits and (3) They provide the foundation for the design of locations of test points.

VII. k-FAULT DIAGNOSIS: The Tolerance Case

The fault diagnosis equation for the tolerance case is given by (7) and (8)

$$Z_{mb}(J_{bF} + J_{bT}) = -\Delta V_m \tag{16}$$

which can be shown to have the form

$$(Z_{mb} + \Delta Z_{mb})J_{bT} = -\Delta V_m \tag{17}$$

where ΔZ_{mb} is caused by the tolerance. This can be viewed as the same problem as the ideal case except that the matrix Z_{mb} is polluted by the noise ΔZ_{mb}. Therefore, we need robust computational methods for the testing procedures (10-13). A well-known method for the test of the rank of a noisy matrix is the singular value decomposition method [53]. Unfortunately this method is unreliable for the test of the global column-rank of a noisy matrix. There exists examples where the singular value decomposition method fails [55]. In order to circumvent this problem, Suen and Liu have developed a method using residual numbers for the testing of global column-ranks [55]. The same method can also be used for the tests (10-12) [54, 55]. For all the cases we have tried, this method has shown to be very reliable when the ratio fault/tolerance is large.

VIII. CONCLUSION

Two main approaches to analog fault diagnosis has been presented: The element-value solvability problem and the fault/tolerance compensation approach.

The element-value solvability problem is the same as the n-port synthesis problem except that the graph and the element-type have been prescribed. Therefore, it is very much in the interest of circuit theorists. The major advantage of this approach is that the tough problem of tolerance is avoided. The difficulty of this approach is that its performance is limited by a trade-off between the large number of test points required and the complexity of post-fault computation.

The fault/tolerance compensation approach requires a very few number of test points and very simple post-fault computation. However, the number of faulty devices is limited to be small. More importantly, we have to face up to the tolerance problem, although some limited success in this direction has been accomplished.

In summary, in this uncharted area of research we have been struggling in the past to find out what the real problems are, and we begin to see some daylight. It still has a long way to go.

For one thing, the result presented is for single-signal and single-frequency. When multiple-signal and/or multiple-frequency are used, the performance should be improved since more information is gathered for analysis. Some initial studies have been made [14,21,33,36,44, 48]. However, how to choose these signals and frequencies so that the extra information can be used more efficiently and effectively and what the limitations are, are still unknown to us.

The diagnosis of nonlinear circuits are barely touched [20,31,39,40,42]. This important problem is still at large.

The problem of diagnosis of analog/digital hybrid circuit is essentially untouched [32].

The fault diagnosis techniques have other applications. Since it has the capability of locating the faulty devices from available test points, it has the potential to be integrated with manufacture process to do CAM for IC chips. For another application, the single-frequency single-signal method can do on line fault diagnosis. Therefore, it provides an important link for the design of self-repairing systems.

IX. REFERENCES

[1] R.S. Berkowitz, "Conditions for network-element-value Solvability", IRE Trans. Circuit Theory, Vol. CT-9, pp. 24-19, 1962.

[2] S.D. Bedrosian, "On element value solution of single-element-kind networks", Ph.D. dissertation University of Pennsylvania, Dec. 1961.

[3] S. Seshu and R. Waxman, "Fault Isolation in Conventional Linear Systems -A Feasibility Study", IEEE Trans. Rel., Vol. R-15, pp. 11-16, May 1965.

[4] R. Saeks, S.P. Singh and R.W. Liu, "Fault Isolation via Component Simulation", IEEE Trans. Circuit Theory, Vol. CT-19, pp. 634-640, Nov. 1972.

[5] G.D. Martens and J.D. Dyck, "Fault Identification in Electronic Circuits with the Aid of Bilinear Transformations", IEEE Trans. on Reliability, Vol. R-21. pp. 99-104, May 1972.

[6] W.A. Plice, "Techniques for the Automatic Generation of Fault Isolation Tests for Analog Circuits", IEEE 1975 ASSC, pp. 127-130, 1975.

[7] G.C. Temes, "Efficient Methods of Fault Simulation", Proc. of 20th Midwest Symp. on Circuits and Systems, pp. 191-194, August 1977.

[8] T.N. Trick and C.J. Alajajian, "Fault Analysis of Analog Circuits", Proc. of 20th Midwest Symp. on Circuits and Systems, pp. 211-215, 1977.

[9] R. Liu and V. Visvanathan, "Diagnosability of Large-Scale Dynamical Systems", Proc. of 20th Midwest Symposium on Circuits and Systems, Lubbock, 1977, pp. 585-589.

[10] R. Saeks and S.R. Liberty, (Editors) Rational Fault Analysis, Marcel Dekker, 1977.

[11] S.D. Bderosian (Editor), "Special Issue on Automatic Analog Fault Diagnosis", IEEE Trans. Circuits and System, Vol. CAS-26, July 1979.

[12] P. Duhamel and J.C. Rault, "Automatic Test Generation Techniques for Analog Circuits and Systems: A Review", IEEE Trans. on CAS, pp. 411-440, 1979.

[13] N. Navid and A.N. Willson, "A Theory and an Algorithm for Analog Circuits Fault Diagnosis", IEEE Trans. on Circuits and Systems, Vol. CAS-26. pp. 440-457, July 1979.

[14] N. Sen and R. Saeks, "Fault Diagnosis for Linear Systems via Multifrequency Measurements", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 457-465, July 1979.

[15] T.N. Trick, W. Mayeda and A.A. Sakla, "Calculation of Parameter Values from Node Voltage Measurements", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 466-475, July 1979.

[16] R.W. Liu and V. Visvanathan, "Sequentially Linear Fault Diagnosis: Part I-Theory", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 490-495, July 1979.

[17] S. Freeman, "Optimmum Fault Isolation by Statistical Inference", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 505-512, July 1979.

[18] J. Lee and S.D. Bedrosian, "Fault Isolation Algorithm for Analog Electronic Systems Using the Fuzzy Concept", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 518-522, July 1979.

[19] W. Hochwald and J.D. Bastian, "A DC Approach for Analog Fault Dictionary Determination", IEEE Trans. on Circuits and Systems, Vol. CAS-26, pp. 523-528, July 1979.

[20] C.S. Lin and R. Liu, "A Single-Fault Diagnosis Theory". Allerton Conf. on Communication, Control and Computing, pp. 172-178, 1979.

[21] Z.F. Huang and G.M. Shen "The Fault Diagnosis of Analog Network and System", 1980 National Symposium on Circuits and Systems, Guang Zhou, Jan. 6-10, 1980.

[23] A.A. Sakla, E.I. El-Masry and T.N. Trick, "A Sensitivity Algorithm for Fault Detection in Analog Circuits", IEEE 1980 Int'1. Symposium on Circuits and Systems, pp. 1075-1077.

[24] A.A. Sakla, E.I. El-Masry and T.N. Trick, "A Sentivity Algorithm for Fault Detection in Analog Circuits", IEEE 1980 Int'1. Symposium on Circuits and Systems, pp. 1075-1077, 1980.

[25] R.M. Biernacki and J.W. Bandler, "Fault Location of Analog Circuits", IEEE 1980 Int'1. Symposium on Circuits and Systems, pp. 1078-1081.

[26] R.M. Biernacki and J.W. Bandler, "Fault Location of Analog Circuits", IEEE 1980 Int'1. Symposium on Circuits and Systems, pp. 1082-1086, 1980.

[27] R. Liu, C.S. Lin, A. Deng, and V. Raman, "System Diagnosis-A New System Problem", Proc. National Electronics Conf., pp. 280-285, 1980.

[28] R. Saeks and R. Liu, Proc. Workshop on Analog Automatic Test Program Generation, University of Notre Dame, May 26-27, 1981.

[29] R. Saeks and R. Liu, "Fault Diagnosis in Electronic Circuits', J. of Society of Instrument and Control Engineers, Vol. 20, pp. 214-216, Tokyo, Japan, 1981.

[30] Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and Design of Testability", Tech. Report No. 811, University of Notre Dame, Notre Dame, Indiana, July 27, 1981. Also, 20th IEEE Conference on Decision and Control, pp. 1037-1042.

[31] Lin Xie-ting, "Fault Analysis of Nonlinear Analog Circuits", Acta Electronica Sinica, Vol. 9, No. 1, p. 70, 1981.

[32] C.S. Lin and R. Liu, "A Fault Directory Approach to Analog Fault Analysis -A Case Study", Proc. 1980 IEEE International Symposium on Circuits and Systems, pp. 239-242, 1981.

[33] R. DeCarlo and C. Gordon "Tableau Approach to AC-Multifrequency Fault Diagnosis" Proc. 1981 IEEE Int'l Symposium on CAS, pp. 270-273, 1981.

[34] C.C. Wu, A. Sangiovanni-Vercentelli and R. Seaks "A Differential-interpolative Approach to Analog Fault Simulation" Proc. IEEE Int'l. Symposium in CAS, pp. 266-269, 1981.

[35] Z.F. Huang and G.M. Shen "Method of Fault Diagnosis in Linear Systems and Their Diagnosabilities" Proc. 1981 IEEE Int'l. Symposium on Circuits and Systems, 32 pp. 285-288, 1981.

[36] L.M. Roytman and M.N.S. Swamy "Some Properties of Orthonomal Excitations of the Circuit and the Calculation of the Circuit Elements", Proc. 1981 IEEE Int'l. Symposium On CAS, pp. 292-294, 1981.

[37] R.M. Biernacki and J.W. Bandler, "Multiple-Fault Location of Analog Circuits", IEEE Trans. Circuits and Systems, Vol. CAS-28, pp. 361-366, May 1981.

[38] W.A. Plice, "The Model Adjusting Techniques for Analog Fault Detection and Location", Technical Report, Honeywell Inc., System and Research Center, Minneapolis, MN., 1981.

[39] V. Visvanathan and A. Sangiovanni-Vincentelli, "Diagnosability of Non-linear Circuits and Systems -Part I: The dc Case", IEEE Trans. on CAS, pp. 1093-1102, 1981.

[40] R. Saeks, A.S. Vincenteli and V. Visvanathan, "Diagnosability of Non-linear Circuits and Systems Part II: Dynamical Systems", IEEE Trans. Circuits and Systems, pp. 1103-1108, Nov. 1981.

[41] R. Saeks, "Criteria for Analog Fault Diagnosis in Proc. European Conf. Circuit Theory and Design, (The Hague), pp. 75-78, Aug. 1981.

[42] C.C. Wu, K. Nakazima, C.L. Wei and R. Saek, "Analog Fault Diagnosis with Failure Bounds", IEEE Trans. on Circuits and Systems, pp. 271-284, 1982.

[43] Z.F. Huang, C.S. Lin and R. Liu, "Topological Conditions on Multiple-Fault Testability of Analog Circuits", 1982 IEEE International Symposium on Circuits and Systems, 1982, Rome, Italy.

[44] L.M. Roytman, E. Plotkin and M.N.S. Swamy, "Multifrequency Method of Fault Diagnosis in Analogue Circuits", Proc. 1982 IEEE Int'l. Symposium on Circuits and System, 1982, Rome.

[45] J.A. Starzyk and J.W. Bandler, "Model Approach to Multi Fault Location in Analog Circuits", Proc. 1982 IEEE Int'l. Symposium on Circuits and System, 1982, Rome.

[46] T.N. Trick and Y. Li, "Algorithms for the location of Single and Multiple Fault in Analog Circuits with Inaccessible Nodes", Proc. 1982 IEEE Int'l. Large Scale System Symposium, pp. 491-494, Oct. 1982.

[47] R.M. Lin, "DC Fault Diagnosis Using Complimentary Pivot Theory", Proc. 1982 IEEE Int'l. Symposium on CAS, pp. 1132-1135, 1982.

[48] J.W. Bandler, R.M. Biernacki, A.E. Salama and J.A. Starzyk, "Fault Isolation In Linear Analog Circuits Using The L_1 Norm", Proc. 1982 IEEE Int'l. Symposium in CAS, 1982.

[49] Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and A Design of Testability", IEEE Trans. on CAS, (to appear).

[50] C.S. Lin, Z.F. Huang and R. Liu, "Topological Conditions for Single-Branch-Fault", IEEE Trans. on CAS, (to appear).

[51] Y. Togawa and T. Matsumoto, "On the Topological Testability Conjecture for Analog Fault Diagnosis Problems".

[52] C.S. Lin, "Reachability, Excitability, and Testability of Analog Networks", Ph.D. Dissertation, University of Notre Dame, IN. 1983.

[53] G.W. Stewart, Introduction to Matrix Computation, Academic Press, 1973.

[54] L.C. Suen and R. Liu, "Determination of the Structure of Multivariable Stochastic Linear Systems", IEEE Transactions on Automatic Control, pp. 458-464, 1978.

[55] L.C. Suen and R. Liu, "Numerical Rank Determination and Residual Numbers", (to be published).

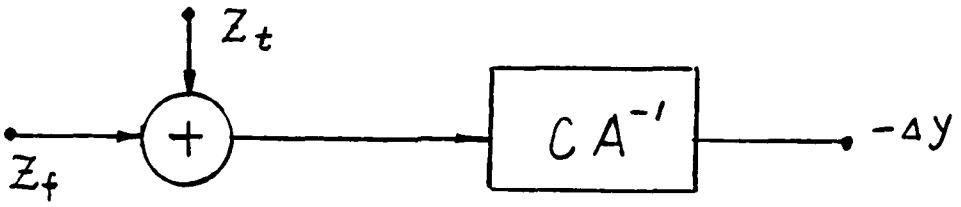


Figure 1.

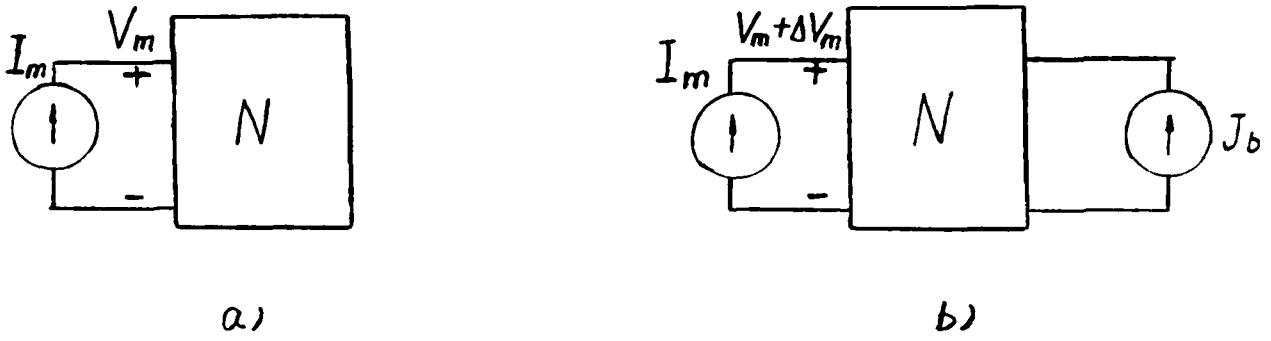
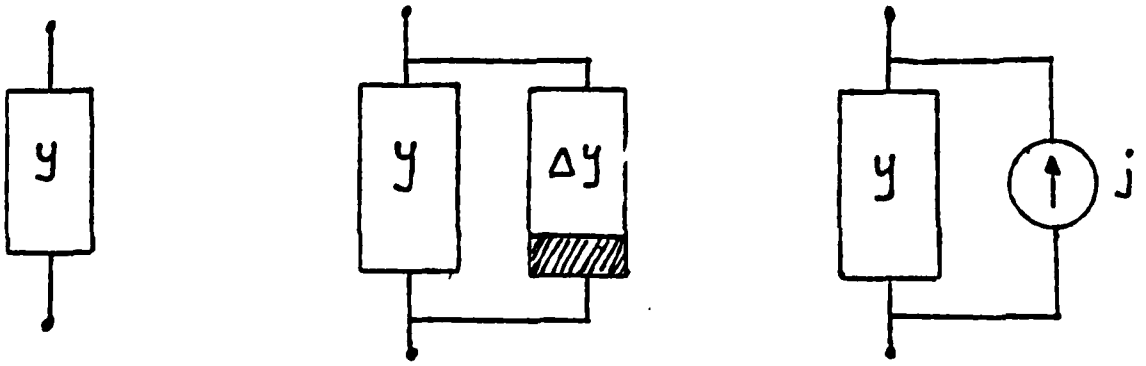


Figure 2.

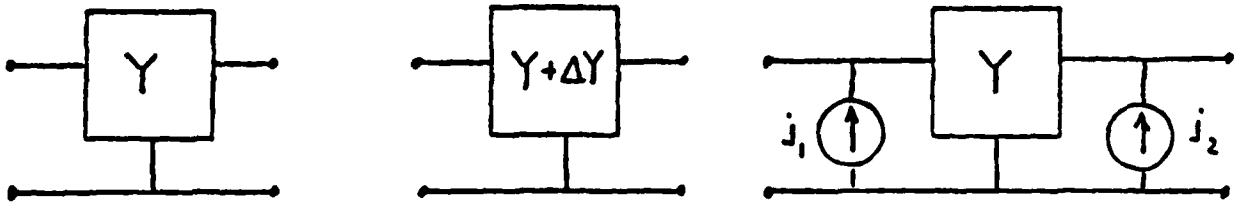


(a) Nominal Admittance

(b) Actual Admittance

(c) The Compensation Model

Figure 3. A Simple Fault/Tolerance Compensation Model.



(a) Nominal Circuit (b) Actual Circuit (c) The Compensation Model

Figure 4. The Fault/Tolerance Compensation Model for a Three-Terminal Device.

Node-Fault Diagnosis and a Design of Testability

by

R. Liu
C. Lin
Z.F. Huang

Node-Fault Diagnosis and a Design of Testability

ZHENG F. HUANG, CHEN-SHANG LIN, AND RUEY-WEN LIU, FELLOW, IEEE

Abstract—A concept of k-node-fault testability is introduced. A sufficient and almost necessary condition for testability as well as the test procedure is presented. This condition is further evolved to a necessary and almost sufficient topological condition for testability. A unique feature of this condition is that it depends only on the graph of the circuit, not on the element values. Based on this condition, a design of testability can be established.

I. Introduction

FAULT DIAGNOSIS of analog networks has become an increasing active research area recently [1]–[11]. A good survey on this subject can be found in [1], [2]. Roughly speaking, analog fault diagnosis is the problem of detecting the faulty analog networks and locating the faulty elements or parameters in those faulty networks. In this paper, we are concerned with only the second part of the problem, namely, the fault location problem.

Recently, two similar methods on fault location of linear networks which involve only linear equations have been proposed by Biernacki and Bandler [9], [10] and Skala *et al.* [11]. These methods are based on checking consistency or inconsistency of certain linear equations which are invariant on faulty elements. And they have been shown to be equivalent in [17]. Since only linear equations are involved, they are computationally appealing. However, as far as testability condition is concerned, none is given explicitly in [11], while only a necessary condition is provided in [10]. This necessary condition is given in terms of network functions (impedance function). Hence it is not easy to apply, especially in the environment of testability design.

In Section II of this paper, instead of port-voltages in [9], [10] or branch-voltages in [11], node-voltages are used as the basis for the derivation of the node-diagnosis equation. In Section III, following the procedure given in [11], a necessary condition for testability is hence derived. The major effort is to show that this necessary condition is also almost sufficient.

The main contribution of this paper is given in Section IV. It is shown that the testability condition which is based on network functions, can be reduced to a condition which depends only on the graph of the circuit not the element values. The simplicity of the topological condition for testability allows us to study (1) the design of the location and the number of test points for a given circuit, and (2) the design of alternative circuits which are testable.

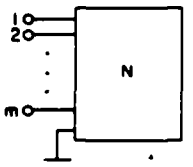


Fig. 1.

For the purpose of the application to branch diagnosis, Theorem 5 of Section V provides a method to determine the faulty branches as well as their deviated admittance values if the node fault testability condition is satisfied.

Examples are given in Section VI.

II. Node-Diagnosis Equations

Let N be a b-branch, $(n+1)$-node, $(m+1)$-terminal, linear, time-invariant, lumped nonreciprocal network (Fig. 1). Out of the $(n+1)$ nodes, $(m+1)$ are accessible terminals for excitation and measurement. Label the nodes in the following way. Let one of the accessible nodes be the reference node. Then label the rest of the accessible nodes $(i=1,\cdots,m)$ ahead of inaccessible nodes $(i=m+1,\cdots,n)$. The following notations are used for N:

- v_m terminal (accessible node) voltage vector,
- i_m terminal current vector,
- v_b branch voltage vector,
- i_b branch current vector,
- v_n node (including accessible node) voltage vector,
- Y_b branch-admittance matrix,
- Y_n node-admittance matrix.

Note that Y_b, Y_n need not be symmetrical and

$$i_b = Y_b v_b \tag{1}$$

$$Y_n = A Y_b A^T \tag{2}$$

where A is the incidence matrix of the graph and Y_n can be obtained from the network directly.

Now consider that the network N is perturbed to $(N+\Delta N)$ in the way that Y_b is perturbed to $Y_b+\Delta Y_b$, and the graph remains the same. In the perturbed network, we denote the corresponding voltages, currents, and admittance by $v_m+\Delta v_m$, $i_m+\Delta i_m$, $v_b+\Delta v_b$, $i_b+\Delta i_b$, $v_n+\Delta v_n$, and $Y_n+\Delta Y_n$, respectively.

Given Y_b and the graph, the purpose of the fault diagnosis is to estimate ΔY_b from the information i_m, Δv_m, and

Manuscript received August 3, 1981; revised April 15, 1982 and December 8, 1982. This work was supported in part by the Office of Naval Research under Contract N00014-78-C-0444.

Z. F. Huang and R-W. Liu are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556.

C-S. Lin was with the University of Notre Dame, Notre Dame, IN. He is now with the National Tsing Hua University, Taiwan, Republic of China.

0098-4094/83/0500-0257$01.00

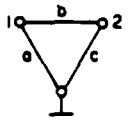


Fig. 2.

Δi_m. Note that v_m can be calculated from i_m and N. Without loss generality, we may assume that

$$\Delta i_m = 0$$

i.e., apply the same i_m to the original network N and the perturbed network $N + \Delta N$.

Denote Z_n and partition it in the following way:

$$Z_n = Y_n^{-1} = \begin{bmatrix} Z_{mn} \\ Z_{ln} \end{bmatrix} \tag{3}$$

where $Z_{mn} \in C^{m \times n}$, $Z_{ln} \in C^{l \times n}$, and $l = n - m$.

Under these conditions, the node-voltage equations for N and $N + \Delta N$ are given by

$$Y_n v_n = \begin{bmatrix} i_m \\ 0 \end{bmatrix} \tag{4}$$

$$(Y_n + \Delta Y_n)(v_n + \Delta v_n) = \begin{bmatrix} i_m \\ 0 \end{bmatrix}. \tag{5}$$

A subtraction of (5) by (4) yields

$$Y_n \Delta v_n = -\Delta Y_n (v_n + \Delta v_n) \triangleq j_n \tag{6}$$

where the RHS is denoted by j_n. Using (3), (6) becomes

$$Z_n j_n = \Delta v_n \tag{7}$$

and, hence,

$$Z_{mn} j_n = \Delta v_m. \tag{8}$$

The equations (6) and (8) are called the *node-diagnosis equations*.

Since Z_{mn} and Δv_m are known, we can first solve for j_n from (8) and then estimate ΔY_n from (6). However, for practical diagnosis problems, n is much larger than m, and, therefore, j_n is not unique. Furthermore, Δv_n is unknown, which creates difficulties for solving (6). However, if we take advantage of a structural property of the network, which will be discussed later, these difficulties can be circumvented. This will be discussed next.

A crucial property of ΔY_n is that it has the same structure as Y_n, i.e., if (i, j)th element of Y_n is zero, then the (i, j)th element of ΔY_n is also zero. The problem of locating the faulty branches becomes the problem of locating the positions of nonzero entries of ΔY_n.

Definition 1: Branch k, $k = 1, 2, \cdots, b$, is said to be fault-free if the kth row of ΔY_b is zero. Node i, $i = 1, 2, \cdots, n$, is said to be fault-free if all branches incident with node i are fault-free. Otherwise, they are faulty.

Note that the number of node faults may be smaller than the number of branch faults. This is illustrated in Fig. 2. There are three branch faults (a, b, c) but only two node faults (1, 2).

It is easy to see from (6) that if node i is fault-free, then the ith component of j_n is zero. However, the converse may not be true. It is possible that node i is at fault but the effect of faulty branches associated with node i may cancel each other to give zero to the ith component of j_n. This possibility is very very small. Henceforth, we will make the following assumption.

Assumption 1: If node i is faulty then the ith component of j_n is nonzero. Under this assumption, node i is faulty if and only if the ith component of j_n is nonzero.

Next, we want to determine those nonzero components of j_n from the node-diagnosis equation (8).

III. k-Node-Fault Testability

We start with a definition of testability.

Definition 2: A network N is said to be *k-node fault testable* if when N is perturbed to $(N + \Delta N)$, one will be able to determine, by choosing appropriate testing signals i_m, from the measurements on accessible nodes:

(a) whether or not N has no more than k node faults,

(b) if affirmative, the faulty nodes can be uniquely located.

Recall that

$$Z_{mn} j_n = \Delta v_m \tag{8}$$

and (with Assumption 1) that node i is faulty if and only if the ith component of j_n is nonzero. Let

$$X_k = \{x \in C^n | \text{ number of nonzero components of } x \text{ not more than } k\}$$

then the perturbed network $N + \Delta N$ has the number of multiple node-faults less than or equal to k if and only if $j_n \in X_k$. Let z_i, $i = 1, 2, \cdots, n$ be the columns of Z_{mn}. We call the condition

$$\text{rank}\left[z_{i_1}, z_{i_2}, \cdots, z_{i_{k+1}}\right] = k + 1, \quad \forall 1 \leq i_1 < i_2 < \cdots < i_{k+1} \leq n \tag{9}$$

the *k-node-fault testability condition*.

Let the *test matrices* be the combinations of k columns of Z_{mn} and Δv_m, i.e.,

$$T_i = \left[z_{i_1}, z_{i_2}, \cdots, z_{i_k}, \Delta v_m\right] \tag{10}$$

$i = 1, 2, \cdots, s$, where $s = C(n, k)$. Then we have the following two theorems.

Theorem 1: Let the k-node-fault testability condition (9) be satisfied.

(a) If

$$\text{rank } T_i = k + 1, \qquad i = 1, 2, \cdots, s \tag{11}$$

then N has more than k-node faults. Conversely, if N has more than k-node faults, then (11) is satisfied for almost all Δv_m.

(b) If (11) is not satisfied, i.e.,

$$\text{rank } T_i < k + 1 \tag{12}$$

for some i, then N has no more than k-node faults for almost all Δv_m. Furthermore, all faulty nodes can be uniquely isolated by (13).

Proof: (a) Conditions (9) and (11) imply that Δv_m cannot be represented by any k columns of Z_{mn}, i.e., (8) has no solution in X_k. Therefore, N has more than k-node faults.

Conversely, if N has more than k-node faults, then $j_n \notin X_k$, i.e., for almost all Δv_m, it cannot be spanned by k columns of Z_{mn}. Therefore, Condition (11) is satisfied for almost all Δv_m.

(b) Let the test in (a) be failed. Therefore, there are at most k-node faults, and (12) is satisfied. It follows from (9) that

$$\text{rank}\left[z_{i_1}, z_{i_2}, \cdots, z_{i_k}\right] = \text{rank}\left[z_{i_1}, z_{i_2}, \cdots, z_{i_k}, \Delta v_m\right] = k$$

for some $\{z_{i_1}, z_{i_2}, \cdots, z_{i_k}\}$. Consequently, (8) has a solution in X_k. Furthermore, in view of Theorem A1 in the Appendix, the solution is unique for almost all Δv_m.

Finally, the faulty nodes can be located as follows. From (8), delete the columns of Z_{mn} and components of j_n which are not associated with $\{z_{i_1}, z_{i_2}, \cdots, z_{i_k}\}$. As a result, we have

$$Z_{mk} j_k = \Delta v_m$$

where $Z_{mk} = [z_{i_1}, z_{i_2}, \cdots, z_{i_k}]$ and $j_k = \text{col}(j_{i_1}, \cdots, j_{i_k})$. Then

$$j_k = \left(Z_{mk}^T Z_{mk}\right)^{-1} Z_{mk}^T \Delta v_m. \tag{13}$$

The nonzero components of j_k indicate the faulty nodes. Q.E.D.

Theorem 2: N is k-node-fault testable if the k-node-fault testability condition (9) is satisfied. Conversely, if a single test-vector is used then condition (9) is also necessary.

Proof: Let (9) be satisfied. Theorem 1 implies that N is k-node-fault testable for almost all Δv_m. Consequently, there exists an i_m so that N is k-node-fault testable. Conversely, if a single test-vector is used, then applying Theorem A1 to (8) one can conclude that condition (9) is also necessary for k-node-fault testability. Q.E.D.

Remark 3.1: The idea of setting the fault-diagnosis equation by an algebraic linear equation, and locating the fault elements by a matrix rank test were presented by Biernacki and Bandler [9] and Sakla, El-Masry and Trick [11]. Some conditions on the matrix rank test for multiple-fault isolation was first presented by Biernacki and Bandler [9], [10]; and also by Trick [24] and Trick and Li [25]. The necessary and almost sufficient condition (Theorem 1) is first given by Huang, Lin, and Liu [18], and also appeared in Trick and Li [25].

Remark 3.2: Biernacki and Bandler applied the matrix rank test method to the branch-diagnosis equation

$$Z_{mb} j_b = \Delta v_m$$

and Trick to a sensitivity equation, which can be shown to be equivalent to the branch-diagnosis equation [26]. The difficulty of using the branch-diagnosis equation is, as pointed out by Trick [25], that multiple *branch* faults usually cannot be uniquely located. On the other hand, multiple *node* fault can be uniquely located. Therefore, we used the node-diagnosis equation instead of the branch-diagnosis equation.

Remark 3.3: The number k of multiple faults can be tested by (11), not pre-assumed as the case in [9]–[11], [24]. Since k is usually not known, this feature has practical importance.

IV. Testability Design

Partition the matrices Y_n and Z_n in the following way:

$$I = Y_n Z_n = [Y_{nm} Y_{nl}] \begin{bmatrix} Z_{mn} \\ Z_{ln} \end{bmatrix}$$

where $m + l = n$. Let the row-vectors of Y_{nl} be denoted by y_i, $i = 1, 2, \cdots, n$, i.e.,

$$Y_{nl} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

An equivalent testability condition can be stated in terms of row-vectors of Y_{nl}.

Theorem 3: The k-node-fault testability condition (9) is satisfied if and only if

$$\text{rank} \begin{bmatrix} y_{j_1} \\ y_{j_2} \\ \vdots \\ y_{j_t} \end{bmatrix} = l, \qquad \forall 1 \leq j_1 < j_2 \cdots < j_t \leq n \tag{14}$$

where $t = n - k - 1$.

The proof of Theorem 3 is given in the Appendix B.

Since usually $n - k - 1 \gg k + 1$ and $l \gg m$, condition (9) is easier to be tested than (14). However, condition (14) is useful for the *design* of testability if condition (9) fails. This is because each element of Y_{nl} has a direct interpretation in the circuit. For example, since Y_{nl} concerns only those branches connected with inaccessible nodes, *the branches among accessible nodes have no effect on the testability*. We will now present a criterion which depends only on the *graph* of the circuit, and thereby provides a means for the design for testability.

We will first consider the network N with no dependent sources. From the graph G of the network N, we construct a testing graph G_t. First, delete all branches which are incident between two accessible nodes. Then, connect all accessible nodes, except the reference node, to a new node t.

Theorem 4: Let G be connected and $k < m$. The following three statements are equivalent.

(a) The k-node-fault testability condition (14) is satisfied for almost all values of branch admittances.

(b) Any inaccessible node and the node t have at least $(k+1)$ local connectivity[1] in G_t.

(c) There are at least $(k+1)$ independent paths[1] from any inaccessible node to the node t in G_t.

[1]The local connectivity of two nonadjacent nodes is the minimum number of nodes separating them, and two paths are independent if they do not have any common node, except at the terminal nodes [16].

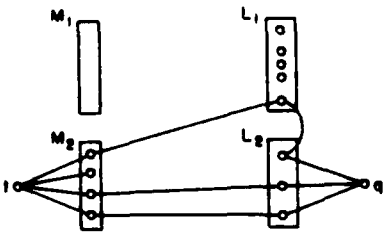


Fig. 3.

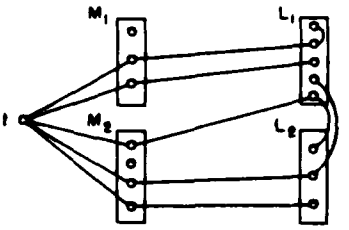


Fig. 4.

Before the theorem is shown, let us prove two lemmas. First, consider any $(n-k-1)\times l$ submatrix Y of Y_{nl}. Let these $(n-k-1)$ rows correspond to the set M_2 of m_2 accessible nodes and the set L_1 of l_1 inaccessible nodes where $m_2 + l_1 = n-k-1$. Then the node equation $Y_n v_n = i_n$ can be rearranged to the form

$$\begin{bmatrix} Y_{m_1m_1} & Y_{m_1m_2} & Y_{m_1l_1} & Y_{m_1l_2} \\ Y_{m_2m_1} & Y_{m_2m_2} & Y_{m_2l_1} & Y_{m_2l_2} \\ Y_{l_1m_1} & Y_{l_1m_2} & Y_{l_1l_1} & Y_{l_1l_2} \\ Y_{l_2m_1} & Y_{l_2m_2} & Y_{l_2l_1} & Y_{l_2l_2} \end{bmatrix} \begin{bmatrix} v_{m_1} \\ v_{m_2} \\ v_{l_1} \\ v_{l_2} \end{bmatrix} = \begin{bmatrix} i_{m_1} \\ i_{m_2} \\ i_{l_1} \\ i_{l_2} \end{bmatrix} \quad (15)$$

and

$$Y = \begin{bmatrix} Y_{m_2l_1} & Y_{m_2l_2} \\ Y_{l_1l_1} & Y_{l_1l_2} \end{bmatrix}. \quad (16)$$

Hence the set M_1 of m_1 accessible nodes and the set L_2 of l_2 accessible nodes are those not associated with the rows of Y.

Note that

$$k+1-m_1 = l_2 \quad (17)$$

and if $m_1 + m_2 = m > k$, then

$$m_2 \geq l_2. \quad (18)$$

The purpose of the following two lemmas is the construction of a subgraph G_0 of G_t, which will be useful in the proof of Theorem 4.

Lemma 1: If (b) of Theorem 4 is satisfied, then there are l_2 disjoint paths connecting the set L_2 and M_2 in G, and no two such paths have any common node.

Proof: From hypothesis (b), the local connectivity in G_t between the node t and any inaccessible node is at least $k+1$. Now, remove all branches from M_1 to the node t. Then the local connectivity between any inaccessible node in L_2 to the node t is at least $k+1-m_1$ which is l_2 by (17). Finally, connect all the nodes in L_2 to a new common node q. The local connectivity between q and t is exactly l_2. Therefore, by Menger's Theorem, there exist l_2 independent paths connecting q and t, as shown in Fig. 3. Consequently the lemma is proved. Q.E.D.

Lemma 2: Let (b) be satisfied. In G_t, there exists a loopless connected subgraph G_0 which contains the node t, all inaccessible node and all the disjoint paths in Lemma 1.

Proof: First, take all the disjoint paths given in Lemma 1 and extend the paths to the node t. This subgraph contains all the nodes in L_2 and the node t. For every node in L_1, by (b), there exists a path from it to the node t. Now, G_0 can be constructed by repeating the following process.

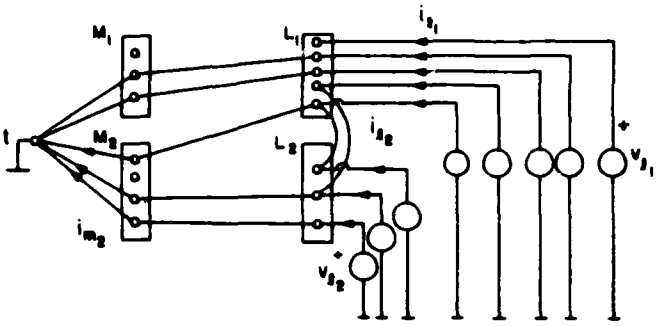


Fig. 5.

Pick a node in L_1, follow a path to node t and stop when it first reaches a node of existing subgraph. Then add all the new branches to the existing graph and repeat the process for the next node in L_1. A typical G_0 is shown in Fig. 4. Q.E.D.

Proof of Theorem 4:

(b) $\Rightarrow$ (a) Consider any $(n-k-1)\times l$ submatrix Y of Y_{nl}. We want to show that Y has full column-rank for almost all branch admittances. In the first step, we want to show that it is true for a particular set of branch admittances. We choose a specific set of branch admittances in the following way: first, set all branch admittances to zero for those branches not in G_0; and then, set other branch admittances to positive values. Therefore, this network is connected, passive, and resistive. Let it be denoted by N_0. We want to show that, for the network N_0, Y has full column-rank.

With N_0, we ground all accessible node and excite all nodes in L_1 and L_2 by voltage sources v_{l_1} and v_{l_2}, as shown in Fig. 5. From (15), we have

$$Y \begin{bmatrix} v_{l_1} \\ v_{l_2} \end{bmatrix} = \begin{bmatrix} i_{m_2} \\ i_{l_1} \end{bmatrix}. \quad (19)$$

If Y is not full column-rank, then there exists nonzero (v_{l_1}, v_{l_2}) such that $(i_{m_2}, i_{l_1}) = 0$. We want to show that this is not possible for N_0. Since N_0 is a connected passive network, and (v_{l_1}, v_{l_2}) are the only sources, it concludes that $(i_{l_1}, i_{l_2}) = 0$ if $(v_{l_1}, v_{l_2}) = 0$. If $i_{l_1} = 0$, we already have a contradiction. It remains to consider the case when $i_{l_1} = 0$ and $i_{l_2} \neq 0$. We want to show that in this case $i_{m_2} \neq 0$. Since $i_{l_1} = 0$ and N_0 is passive resistive circuit, we may delete all branches except the paths given in Lemma 1, as far as KCL equation is concerned. Since all these paths are disjoint, $i_{l_2} \neq 0$ implies that $i_{m_2} \neq 0$, which again is a contradiction. Therefore, $(v_{l_1}, v_{l_2}) = 0$ if and only if $(i_{m_2}, i_{l_1}) =$

0. It concludes from (19) that Y has full column-rank for N_0.

Now in the second and final step, we want to show that Y has full column-rank for almost all branch admittances. Note that Y has full column-rank if and only if $\det(Y^TY) \neq 0$. A key fact is that $\det(Y^TY)$ is a *polynomial* of the branch admittances, and it is not zero for a particular set of admittances, namely N_0. By a well-known theorem [20, 21], it can be concluded that $\det(Y^TY) \neq 0$ for almost all branch admittances. This proves (b) $\Rightarrow$ (a).

(b) $\Rightarrow$ (c): Directly from the Menger's Theorem [16].

(a) $\Rightarrow$ (b): We will prove it by contradiction. Let (b) be false. Then there are k nodes which separate G_t into two parts: Part A contains node t and Part B contains at least one inaccessible node. Since all accessible nodes are adjacent to t, they cannot belong to Part B. Let Part B contain $r (r \geq 1)$ inaccessible nodes. Note that any element y_{ij} of Y_n is zero if and only if nodes i and j are nonadjacent. Since Part A and Part B are separable, the element of Y_n which associate with any nodes of Part A (except t) and any node of Part B is zero. Since there are r inaccessible nodes in Part B, there are a zero-submatrix Y^* of Y_{n_l} whose dimension is $r \times (n-r-k)$. Therefore, from [15], any submatrix of $(n-k-1)\times l$ of Y_{n_l} which contains Y^* cannot have full column-rank. Henceforth, condition (14) cannot be satisfied for such submatrix. Q.E.D.

Let us now extend Theorem 4 to include active elements. First, represent all active elements by a set of dependent current sources each of which is incident to the reference node on one side and to any other node on the other side, and controlled by node voltages. Note that a dependent current source has entry in the submatrix Y_{n_l} if and only if it is controlled by the inaccessible node voltages. In view of the above, we will construct a test graph $\tilde{G}$ from the graph G by the following steps. (1) Delete all dependent sources which are controlled by accessible nodes. (2) If a dependent source is incident with an accessible node i but controlled by an inaccessible node voltage v_j, then delete this dependent source and add a new branch incident with nodes i and j. Note that $\tilde{G}$ and G have the same nonzero entries in Y_{n_l}. It remains to consider these dependent sources which are incident with inaccessible nodes and controlled by inaccessible node voltages. A complete description of this would be too involved for this paper. We will give only necessary and sufficient conditions. Construct G_S from $\tilde{G}$ by deleting all dependent sources and G_N from $\tilde{G}$ by the following step: if there is a dependent source incident with node i and controlled by node voltage v_j, then delete the dependent source and add a new branch incident with nodes i and j.

Corollary 4.1: If G_N does not satisfy (b) or (c) of Theorem 4, then N is not k-node-fault testable. If G_S satisfies (b) or (c), then N is k-node-fault testable for almost all branch admittances.

The proof is straightforward and is omitted.

Remark 4.1: The necessary topological condition of Theorem 4 is first given by [18]. Togawa and Matsumoto [28] have given an independent proof of the sufficient condition. Starzyk and Bandler have studied the case when the topological condition is partially satisfied [27].

V. Branch-Fault Diagnosis

In this section, we will show that branch faults can be isolated if multiple test-signal vectors are used.

Theorem 5: Let N and its adjoint circuit $\hat{N}$ be k-node-fault testable. If $N + \Delta N$ and $\hat{N} + \Delta\hat{N}$ have no more than k faulty nodes, then the faulty branches can be located and their admittances can be determined by a set of m independent vectors i_m.

Proof: Since N is k-node-fault testable and $N + \Delta N$ has no more than k faulty nodes, by Theorem 1, j_n can be uniquely determined for almost every i_m. From (7) and (4),

$$\Delta v_n = Z_n j_n \tag{7}$$

$$v_n = Z_n \begin{bmatrix} i_m \\ 0 \end{bmatrix} \tag{4}$$

we can calculate $v_n + \Delta v_n$ by

$$v_n + \Delta v_n = Z_n \left(\begin{bmatrix} i_m \\ 0 \end{bmatrix} + j_n \right). \tag{20}$$

Hence, ΔY_n can be determined by (6)

$$-j_n = \Delta Y_n (v_n + \Delta v_n) \tag{6}$$

using multiple independent test-signal-vectors i_m in the following way. Take independent i_m's so that

$$I_m = \left[i_{m_1}, i_{m_2}, \cdots, i_{m_m} \right] \tag{21}$$

is nonsingular. Correspondingly, we can measure

$$\Delta V_m = \left[\Delta v_{m_1}, \Delta v_{m_2}, \cdots, \Delta v_{m_m} \right] \tag{22}$$

Then it can be shown that when $\hat{i}_m$ is applied to the perturbed *adjoint* circuit

$$\left(\hat{Y}_n + \Delta\hat{Y}_n \right) = \left(Y_n + \Delta Y_n \right)^T \tag{23}$$

the perturbed voltage can be calculated by

$$\Delta\hat{v}_m = \left(\Delta V_m I_m^{-1} \right)^T \hat{i}_m. \tag{24}$$

With this $\Delta\hat{v}_m$ and $\hat{i}_m$, we can apply the node-fault diagnosis procedure to the adjoint circuit. Since by hypothesis the adjoint circuit is k-node-fault testable and has no more than k faulty nodes, we can identify the nonzero rows of $\Delta\hat{Y}_n$; i.e., k' $(k' \leq k)$ nonzero columns ΔY_n by (23). Once the set of nonzero columns of ΔY_n is identified, we go back to (6). Let S be a column-selector matrix [22] such that $[\Delta Y_n]S$ is an $n \times k'$ matrix with $k' \leq k$ and that $[\Delta Y_n]S$ retains all nonzero columns of ΔY_n. Then (6) becomes

$$-j_n = [\Delta Y_n][v_n + \Delta v_n] = [\Delta Y_n] S S^T [v_n + \Delta v_n].$$

Applying the m independent i_m's in (21), we have correspondingly the matrix equation

$$-J_n = [\Delta Y_n] S S^T [V_n + \Delta V_n]. \tag{25}$$

It remains to show that $S^T[V_n + \Delta V_n]$ has full row-rank. From (5) we have

$$V_n + \Delta V_n = (Z_n + \Delta Z_n) \begin{bmatrix} I_m \\ 0 \end{bmatrix} = (Z_{nm} + \Delta Z_{nm}) I_m.$$

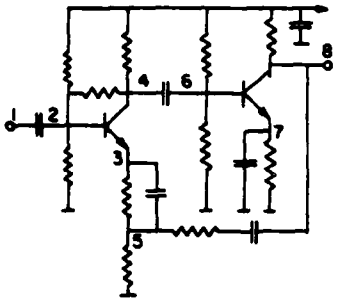


Fig. 6.

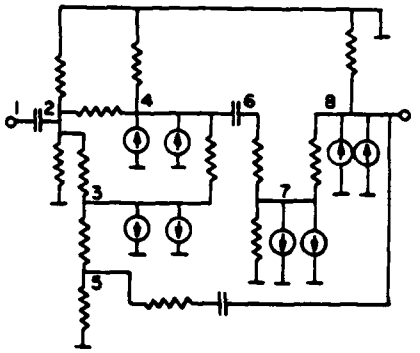


Fig. 7.

So,

$$S^T[V_n + \Delta V_n] = S^T(Z_{nm} + \Delta Z_{nm}) I_m.$$

Since the adjoint network N is k-node-fault testable, $S^T(Z_{nm} + \Delta Z_{nm})$ has full row-rank for almost all branch admittance. Consequently, $S^T[V_n + \Delta V_n]$ has full row-rank for almost all branch admittance. Therefore, the nonzero columns of ΔY_n can be evaluated by

$$[\Delta Y_n]S = -J_n\big([V_n + \Delta V_n]^T S\big) \cdot \big(S^T[V_n + \Delta V_n][V_n + \Delta V_n]^T S\big)^{-1}. \quad (26)$$

From this, the faulty branches can be located and their admittances can be determined. Hence, Theorem 5 is proved. Q.E.D.

Remark 5.1: Note that N and ΔN may not have the *same* graph. Therefore, a short between two open nodes can be detected, although in this case an extra branch between the two open nodes is inserted in the faulty network.

Remark 5.2: If Y_n is symmetrical, then the conditions on the adjoint circuit are automatically satisfied, and only k independent test-signal-vectors i_m are sufficient.

VI. Examples

Example 1: (Testability Design)

For a given circuit, the purpose of testability design is to decide the number and the location of test points such that the circuit is testable. Let's consider a feedback amplifier (Fig. 6). First we get the AC small signal equivalent circuit (Fig. 7). It is reasonable to choose node 1 and 8 to be the test points. By the method described in Corollary 4.1, the test graph G_s and G_n are in Fig. 8. Note from Theorem 4 that the ground branches have no effect on the testability, and hence they are deleted. For this circuit (if we don't consider the self-loops in node 3 and 7), the G_s and G_n are the same. We can easily conclude by condition (c) of Theorem 4 that the circuit is one-node fault testable. Now, suppose that we want it to be 2-node-fault testable. In the G_s, node 5 and 7 are incident with only two branches respectively, and each of them cannot have three independent paths to other nodes. Therefore, nodes 5 and 7 have to be test points. Under this choice, every inaccessible node has three independent paths to the accessible nodes. Hence by Theorem 4 and Corollary 4.1, the circuit is 2-node fault testable for almost all values of the branch admittances.

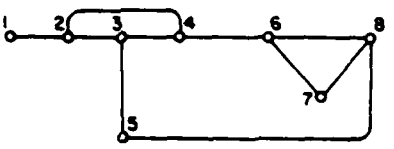


Fig. 8.

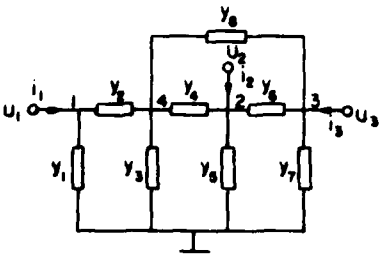


Fig. 9.

In the previous example we described how to choose the test points for a given circuit.

Example 2: (Branch Fault Isolation)

Consider the network N as shown in Fig. 9. Nodes 1, 2, 3 are accessible nodes. Let the nominal values of the circuit of Fig. 9 be $y_i = 1$, $i = 1, 2, \cdots, 6, 8$ and $y_7 = 2$. We can calculate $Z_{mn} = [Z_1, Z_2, Z_3, Z_4]$:

$$Z_{mn} = \begin{bmatrix} 0.593220 & 0.084746 & 0.067797 & 0.186441 \\ 0.084746 & 0.440678 & 0.152542 & 0.169491 \\ 0.067797 & 0.152542 & 0.322034 & 0.135593 \end{bmatrix}.$$

Now assume branches 3 and 8 are perturbed such that $y_3 + \Delta y_3 = 2$, and $y_8 + \Delta y_8 = 2$ and other branches remain the same. Choosing

$$I_m = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

we can measure the voltages on accessible nodes:

$$V_m + \Delta V_m = \begin{bmatrix} 0.5625 & 0.0625 & 0.0625 \\ 0.0625 & 0.4196 & 0.1339 \\ 0.0625 & 0.1339 & 0.2767 \end{bmatrix}$$

and evaluate

$$\Delta V_m = \begin{bmatrix} -0.030720 & -0.0222458 & -0.0052966 \\ -0.0222458 & -0.0210351 & -0.0186138 \\ -0.0052966 & -0.0186138 & -0.0452482 \end{bmatrix}.$$

From Theorem 4, this circuit is 2-node-fault testable. By Theorem 1, check the singularities of the test matrices $[Z_i, Z_j, \Delta v_m]$ for all $1 \leqslant i < j \leqslant 4$. It is found that only $\det[Z_3, Z_4, \Delta v_m] = 0$. Therefore, we can conclude that node 3 and 4 are faulty and calculate the J_n by (13)

$$J_n = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.0625 & -0.008929 & -0.151789 \\ -0.1875 & -0.116071 & -0.026786 \end{bmatrix}.$$

It is seen that the first two rows of J_n are zero and hence the first two rows of ΔY_n are zero. Since ΔY_n is symmetrical, the first two columns of ΔY_n are zero. Now, $S^T[V_n + \Delta V_n]$ can be calculated from (20) and deleting the first two rows, we obtain

$$S^T[V_n + \Delta V_n] = \begin{bmatrix} 0.0625 & 0.1339 & 0.2768 \\ 0.125 & 0.125 & 0.125 \end{bmatrix}.$$

Finally, we can calculate the ΔY_n by (26):

$$[\Delta Y_n]S = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ +1 & -1 \\ -1 & +2 \end{bmatrix}.$$

The diagnosis is two branch faults:

$$\Delta y_8 = 1 \qquad \Delta y_3 = 1.$$

VII. Conclusion

The purpose of node diagnosis is to determine those nodes which are faulty. A node is faulty if any branch incident with it is faulty. Therefore, if faulty nodes are correctly located, all the faulty branches must be incident with these faulty nodes. A network is k-node-fault testable if from some measurements, (1) one can determine whether or not there are k or less faulty nodes; and in addition, (2) if affirmative, the faulty nodes can be uniquely located. Note that in the definition of testability it includes the determination of the number k.

A necessary and almost sufficient condition of k-node-fault testability is given in Theorem 2. It has three features:

1) The minimum value of test points m for which the k-node-fault testability condition (9) to be satisfied is $m = k + 1$. Therefore, it is possible to diagnose k-node fault with $k + 1$ test points, regardless of number of branches and number of nodes.

2) The matrices to be tested in (9) have size only $m \times (k + 1)$, regardless of how large the circuit is.

3) The testability condition depends on the nominal circuit and can be tested before a fault occurs.

If the testability condition is satisfied, procedures to locate the faulty nodes and faulty branches are given in Theorem 1 and Theorem 5, respectively.

If the testability condition is not satisfied, then additional test points are needed. The choice of this additional test points can be greatly simplified by the usage of Theorem 4. The most important feature of this theorem is that it depends only on the graph but not the element values of the network. It is shown that this graphical condition is necessary and almost sufficient for the testability condition.

VI. Appendix

A. The Uniqueness of Solution of Equation $Bx = p$

Recall that (8) has the form

$$Bx = p \tag{A1}$$

where $B \in C^{m \times n}$ with $n \geqslant m$, $x \in C^n$ and $p \in C^m$. For practical problems, $n > m$, and, therefore, the solution of (A1) in general is not unique. However, if we restrict the number of nonzero components of x to be less than m, then the solution could be made unique. Hence, we want to find the solution of (A1) in X_k.

Note that a solution $x \in X_k$ of (8) exists if and only if p is a linear combination of some k columns of B. Henceforth, let the subspaces spanned by the combinations of k column vectors of B be denoted by R_i, $i = 1, 2, \cdots, r$, where $r = C(n, k)$. Let the intersections of combinations of two R_i's be denoted by Q_j, $j = 1, 2, \cdots, q$, where $q = C(r, 2)$. Furthermore, let

$$\Omega = \bigcup_{i=1}^{r} R_i \tag{A2}$$

$$\Omega_e = \bigcup_{j=1}^{q} Q_j. \tag{A3}$$

Evidently, $\Omega_e \subset \Omega$.

Lemma 1: Equation (A1) has a solution $x \in X_k$ if and only if $p \in \Omega$, and the solution is unique in X_k if and only if $p \in \Omega - \Omega_e$.

Proof: A solution $x \in X_k$ of (A1) exists if $p \in R_i$ for some i, i.e., $p \in \Omega$. Furthermore, (A1) has two distinct solutions in X_k if $p \in R_i$ and $p \in R_j$ with $i \neq j$, i.e., $p \in \Omega_e$. Therefore, (A1) has a unique solution in X_k if $p \in \Omega - \Omega_e$. Q.E.D.

Let $b_1, b_2, \cdots, b_n$ be the column vectors of B.

Lemma 2:

$$\dim R_i = k, \qquad i = 1, 2, \cdots, r \tag{A4}$$

and

$$\dim Q_j < k, \qquad j = 1, 2, \cdots, q \tag{A5}$$

if and only if

$$\operatorname{rank}\left[b_{i_1}, b_{i_2}, \cdots, b_{i_{k+1}}\right] = k + 1, \quad \forall 1 \leqslant i_1 < i_2 < \cdots < i_{k+1} \leqslant n \tag{A6}$$

i.e., every combination of $(k + 1)$ columns of B is linearly independent.

This lemma has an important implication as follows. It can be shown that

$$\Omega - \Omega_e = \bigcup_{i=1}^{r} \left[R_i - \bigcup_{\substack{j=1 \\ j \neq i}}^{r} (R_i \cap R_j) \right].$$

In view of (A5), the measure of the second term in the bracket is zero with respect to any k-dimensional measure. Consequently, the measure of Ω_e is zero. In other words, for any bounded distribution of p, the conditional probability $P(p \in \Omega - \Omega_e | p \in \Omega) = 1$. The above discussion can be summarized together with Lemma 1 in the following.

Theorem A1: Let $p \in \Omega$. Then (A1) has a unique solution in X_k for almost all $p \in \Omega$ if and only if condition (A6) is satisfied.

Since the existence of the solution is guaranteed for fault diagnosis problems, condition (A6) becomes the only condition needed for the uniqueness of solutions in X_k.

Proof of Lemma 2: Clearly (A6) implies (A4). It also implies that $R_i \neq R_j$ if $i \neq j$. Therefore,

$$\dim(R_i \cap R_j) < \dim R_i = k$$

which is (A5).

Conversely, let (A4) be satisfied but (A6) be not satisfied; w.l.g., let

$$b_1 = \sum_{i=2}^{k+1} \alpha_i b_i \tag{A7}$$

where $\alpha_i \in C$ and not all α_i's are zero. Let

$$P = \mathrm{Sp}[b_1, b_2, \cdots, b_{k+1}]$$
$$R_1 = \mathrm{Sp}[b_1, b_2, \cdots, b_k]$$
$$R_2 = \mathrm{Sp}[b_2, b_3, \cdots, b_{k+1}]$$

where $\mathrm{Sp}[x_1, \cdots, x_n]$ is the notation for the subspace spanned by $\{x_1, x_2, \cdots, x_n\}$. It follows from (A7) that $\dim P \leq k$, and from (A4) that $\dim R_1 = k = \dim R_2$. Since $P \supset R_1$ and $P \supset R_2$,

$$R_1 = R_2 = P.$$

Therefore, $\dim R_1 \cap R_2 = k$, which contradicts (A5).

Q.E.D.

B. The Proof of Theorem 3

The proof of Theorem 3 follows immediately from Lemma B2 and Theorem 2. Let $P, Q \in C^{n \times n}$; let P and Q be partitioned in the following way:

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$
$$Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \tag{B1}$$

where $P_{11} \in C^{m \times k}$, $Q_{11} \in C^{k \times m}$, $k \leq m$ and $m + l = n$.

Lemma B1: Let $PQ = I$. Then, rank $P_{11} = k$ if and only if rank $Q_{22} = l = n - m$.

Proof: First, consider $m = k$. See [23].

Next, consider the case $m > k$.

Let rank $P_{11} = k$. Without loss of generality, let the first k rows of P_{11} be linearly independent. If not, apply simultaneously row-operations on $[P_{11} \quad P_{12}]$ and column operations on

$$\begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix}$$

then the first $k \times k$ principal submatrix of P is nonsingular. Then the last $(n-k) \times (n-k)$ principal submatrix of Q is nonsingular. Consequently, rank $Q_{22} = l = n - m$.

From Lemma B1, the following is self-evident.

Lemma B2: Let $Z_n, Y_n \in C^{n \times n}$ and $Y_n Z_n = I$. Let

$$Z_n = \begin{bmatrix} Z_{mn} \\ Z_{ln} \end{bmatrix}, \qquad Y_n = [Y_{nm} \quad Y_{nl}].$$

where $Z_{mn} \in C^{m \times n}$, $Y_{nm} \in C^{n \times m}$ and $m + l = n$. Let $r \leq m$. Then every $m \times r$ submatrix of Z_{mn} has rank r if and only if every $(n-r) \times l$ submatrix of Y_{nl} has rank l.

ACKNOWLEDGMENT

The authors want to thank An-Chang Deng for his constructive suggestions to the proof of Theorem 4.

REFERENCES

[1] "Special issue on automatic analog fault diagnosis," *IEEE Trans. Circuits Syst.*, vol. CAS-26, July 1979.

[2] R. Saeks and S. R. Liberty, *Rational Fault Analysis*. Marcel Dekker, 1977.

[3] R. S. Berkowitz, "Conditions for network-element-value solvability," *IRE Trans. Circuit Theory*, vol. CT-9, pp. 24–29, 1962.

[4] S. Seshu and R. Waxman, "Fault isolation in conventional linear systems—A feasibility study," *IEEE Trans. Rel.*, vol. R-15, pp. 11–16, May 1965.

[5] R. Saeks, S. P. Singh and R. W. Liu, "Fault isolation via component simulation," *IEEE Trans. Circuit Theory*, vol. CT-19, pp. 634–640, Nov. 1972.

[6] N. Navid and A. N. Willson, "A theory and an algorithm for analog circuit fault diagnosis," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 440–456, July 1979.

[7] N. Sen and R. Saeks, "Fault diagnosis for linear system via multifrequency measurements," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 457–465, July 1979.

[8] H. M. Merrill, "Failure diagnosis using quadratic programming," *IEEE Trans. Rel.*, vol. R-22, pp. 207–213, Oct. 1973.

[9] R. M. Biernacki and J. W. Bandler, "Fault location of analog circuits," *IEEE 1980 Int. Symp. Circuits and Systems*, pp. 1078–1081.

[10] R. M. Biernacki and J. W. Bandler, "Multiple-fault location of analog circuits," *IEEE Trans. Circuits Syst.*, vol. CAS-28, pp. 361–366, May 1981.

[11] A. A. Sakla, E. I. El-Masry and T. N. Trick, "A sensitivity algorithm for fault detection in analog circuits," *IEEE 1980 Int. Symp. Circuits and Systems*, pp. 1075–1077.

[12] C. Lin and R. Liu, "Fault directory approach—A case study," *IEEE 1981 Int. Symp. Circuits and Systems Proc.*, pp. 239–242.

[13] T. N. Trick, W. Mayeda and A. A. Sakla, "Calculation of parameter values from node voltage measurements," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 466–475, July 1979.

[14] W. A. Plice, "The model adjusting technique for analog fault detection and location," *Tech. Rep.* Honeywell Inc., System and Research Center, Minneapolis, MN.

[15] A. C. Deng, "Rank of large sparse matrix", M.S. thesis, Univ. of Notre Dame, Notre Dame, IN, 1981.

[16] B. Bollobás, *Extremal Graph Theory*. New York: Academic, 1978.

[17] C. Lin, "A study on fault diagnosis," Tech. Rep. 82-2, Univ. of Notre Dame, Notre Dame, IN, 1982.

[18] Z. F. Huang, C. S. Lin, and R. Liu, "Node-fault diagnosis and design of testability," Tech. Rep. 811, Univ. of Notre Dame, Notre Dame, IN, July 27, 1981. Also, *20th IEEE Conf. Decision and Control*, pp. 1037–1042, 1981.

[19] ____, "Topological conditions on multiple-fault testability of analog circuits," *1982 IEEE Int. Symp. Circuits and Systems*, Rome, Italy.
[20] R. W. Liu and Visvanathan, "Sequentially linear fault diagnosis: Part I—Theory," *IEEE Trans. Circuits Syst.*, vol. CAS-26, pp. 490–496, July 1979.
[21] W. M. Wonham, *Linear Multivariable Control*, Berlin, Germany: Springer-Verlag, 1974.
[22] L. C. Suen and R. Liu, "Determination of the structure of multivariable stochastic linear systems," *IEEE Trans. Automat. Contr.* pp. 458–464, 1978.
[23] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
[24] T. N. Trick, "A sensitivity approach to analog circuit fault analysis," *Workshop on Analog Fault Diagnosis*, Notre Dame, IN, May 26–27, 1981.
[25] T. N. Trick and Y. Li, "Algorithms for the location of single and multiple fault in analog circuits with inaccessible nodes," in *Proc. 1982 IEEE Int. Large Scale System Symp.*, pp. 491–494, Oct. 1982.
[26] T. N. Trick, private communication, 1982.
[27] J. A. Starzyk and J. W. Bandler, "Location of fault regions in analog circuits," submitted for publication.
[28] Y. Togawa and T. Matsumoto, "On the topological testability conjecture for analog fault diagnosis problems," submitted for publication.

Zheng F. Huang was born in Guangxi, China, in October 1943. He graduated in electrical engineering from Polytechnical Institute of Harbin, Harbin, China, in 1967. He is currently a graduate student at the University of Notre Dame and expects to receive the M.A. degree in January 1983.

He worked on power transistor amplifiers and transmitters in Broadcast Equipment Factory, Tianjin, from 1970 to 1973 and on color TV transmitters in Broadcase Equipment Factory, Beijing, from 1974 to 1978. His principal research interests are in the areas of nonlinear systems and fault diagnosis of analog circuits.

Chen-Shang Lin was born in Taiwan in 1951. He received the B.S. degree from National Taiwan University in 1973 and the M.S. and Ph.D. degrees from the University of Notre Dame, Notre Dame, IN, in 1979 and 1983, respectively, all in electrical engineering.

He is currently with the National Tsing Hua University, Taiwan, Republic of China. His research interest is in the area of analog network fault diagnosis.

Ruey-Wen Liu (S'53–M'59–SM'78–F'81) was born in Kiangsu, China. He received the B.S., M.S., and Ph.D. degrees in 1954, 1955, and 1960, respectively, from the Department of Electrical Engineering at the University of Illinois at Urbana.

Since 1960, he has been with the University of Notre Dame, and is currently a Professor in the Department of Electrical Engineering. He has held visiting professorships at the National Taiwan University, Republic of China, in the spring of 1969, the Universidad de Chile, Santiago de Chile, in the summer of 1970, the Institute of Mathematics, Academia Sinica, Republic of China, in the summers of 1976 and 1978, and University of California at Berkeley, from 1965 to 1966 and from 1977 to 1978. His main interest has been in the area of nonlinear circuits and systems. Currently, he is the Chairman of the Technical Committee on Nonlinear Circuits and Systems of the IEEE Circuits and Systems Society. His other interests include fault diagnosis of analog circuits, feedback system design, estimation theory, and policy analysis. He is the Guest Editor of the Special Issue on Nonlinear Circuits and Systems, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, November 1980. He was the organizer and a Co-Chairman of the NSF-Workshop on Nonlinear Circuits Systems in January 1981. He was a Co-Guest Editor of the special issue on Socio-Economic Systems, International Journal on Policy Analysis and Information Systems. Currently, he is an associate editor of the Journal of Circuits, Systems and Signal Processing, and the Journal on Policy Analysis and Information Systems.

Dr. Liu is a member of Sigma Xi.

Topological Conditions on Multiple-Fault Testability of Analog Circuits

by

R. Liu
C. Lin
Z.F. Huang

TOPOLOGICAL CONDITIONS ON MULTIPLE-FAULT TESTABILITY OF ANALOG CIRCUITS*

Z. F. Huang
C. S. Lin
and
Ruey-wen Liu

University of Notre Dame
Notre Dame, IN 46556

ABSTRACT

The topological conditions for k-node, k-cut fault, and k-rank branch-set fault testability by single test signal vector in analog circuits are given. These conditions which are both necessary and sufficient are essential in the choice of test points. These conditions also have applications to the design of testable circuits. The outcome of these methods can be further diagnosed by the usage of multiple test signal vectors. This is also discussed and the associated problems are studied.

I. INTRODUCTION

Let N (Fig. 1) be a connected b-branches, (n+1)-nodes, (m+1)-terminals, linear, time invariant, lumped non-reciprocal network (Fig. 1). Out of the (n+1)-nodes, (m+1) are accessible terminals for exitation and measurement. Label the nodes in the following way. Let one of the accessible nodes be the reference node. Then label the rest of the accessible nodes ($i = 1,\ldots,m$) ahead of inaccessible nodes ($i = m+1,\ldots,n$). The following notations are used for N.

v_m - terminal (accessible node) voltage vector.
i_m - terminal current vector.
v_b - branch voltage vector.
i_b - branch current vector.
v_n - node (including accessible node) voltage vector.
Y_b - branch-admittance matrix.
Y_n - node-admittance matrix.
M_o - the set of all accessible nodes.
M - the set of all accessible nodes except the reference node.

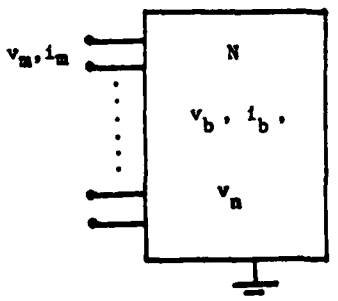


Fig. 1

*This research is supported in part by ONR Contract No. N00014-78-C-0444.

Note that Y_b, Y_n need not be symmetrical, and:

$$i_b = Y_b v_b \tag{1}$$

$$Y_n = A\, Y_b\, A^T \tag{2}$$

where A is the incidence matrix of the graph.

Now consider the network N is perturbed to (N+ΔN) in the way that Y_b is perturbed to $Y_b+\Delta Y_b$, and the graph remains the same. In the perturbed network, we denote the corresponding voltages, currents and admittance by $v_m+\Delta v_m$, $i_m+\Delta i_m$, $v_b+\Delta v_b$, $i_b+\Delta i_b$, $v_n+\Delta v_n$ and $Y_n+\Delta Y_n$ respectively.

Given Y_b and the graph, the purpose of the fault diagnosis is to estimate ΔY_b from the information i_m, Δv_m and Δi_m. v_m can be calculated from i_m and N. Without loss generality, we may assume that

$$\Delta i_m = 0 \tag{3}$$

i.e., apply same i_m to the original network N and the perturbed network N+ΔN.

Denote Z_n and partition it in the following way:

$$Z_n = Y_n^{-1} = \left|\begin{matrix} Z_{mn} \\ Z_{\ell n} \end{matrix}\right| , \tag{4}$$

where $Z_{mn} \in C^{mxn}$, $Z_{\ell n} \in C^{\ell xn}$, and $\ell = n-m$. Under this condition, it can be shown that

$$Z_{mn}\, A(\Delta Y_b)(v_b + \Delta v_b) = \Delta v_m \tag{5}$$

From Eq. (5), assuming the number of nonzero elements in ΔY_b is small, it suggests that one can determine the locations of nonzero elements of ΔY_b and therefore the location of faulty branches provided that the number of accessible nodes is large enough. This discovery was made by Biernacki and Bandler [1,2] and Sakla et al. [3]. The advantages of this approach are that one can rely only on the parameters of N (or the nominal circuit) and the measurements to diagnose the faulty circuit and the computation involved is all linear.

However, there are certain inevitable structures, such as a small loop, in the practical circuits which, if not carefully taken care of, would mess up the results of this approach. In the next section, we will discuss the systematical ways of materializing this approach and, more importantly, provide the graphical conditions for placing the test points.

CH1681-6/82/0000-1152 $00.75 © 1982 IEEE

For the purpose of simplicity, all the results presented here are for the case when N contains no dependent sources. Extensions of necessary conditions and sufficient conditions to include dependent sources are given in [4],[5],[6] and [10].

II. NODE-FAULT, CUT-FAULT, AND BRANCH-SET FAULT DIAGNOSIS

In this section, we will present three different methods of systematically isolating the faults in analog circuits from a single test signal vector. These three methods, node-fault, cut-fault, and branch-set fault diagnosis, represent three different ways of grouping the branches. They will be concisely discussed subsequently and the graphical conditions for placing the test points in each case will be summarized.

First, let us define a faulty branch and a faulty node.

Definition 1: Branch k, k = 1,2,...,b is said to be fault-free if the k-th row of ΔY_b is zero. Node j, j = 1,2,...,n, is said to be fault-free if all branches incident with node j are fault-free. Otherwise they are faulty.

Then, k-node-fault testability can be defined as follows.

Definition 2: A network N is said to be k-node-fault testable if when N is perturbed to (N+ΔN), one will be able to determine, by choosing one appropriate test signal vector i_m, from the measurement on accessible nodes M_o:

(A) whether or not N has no more than k node-faults,
(B) if affirmative, the faulty nodes can be uniquely located.

The following theorem provides a graphical condition for k-node-fault testability. To state the theorem, first let us construct a modified graph G_n from N by deleting all branches incident with the reference node and connecting the rest of accessible nodes to a new node α_n. Then the theorem follows.

Theorem 1 [10]: The following three statements are equivalent:

(1) N is k-node-fault testable w.r.t. M for almost all values of Y_b.
(2) There are at least (k+1) disjoint paths[1] between any inaccessible node and α_n in G_n.
(3) There are at least local (k+1)-connectivity[2] in G_n between any inaccessible node and α_n.

When N is k-node-fault testable, the procedure of locating the faulty nodes is given in [4] and [10].

1. Two paths are said to be disjoint if there are no common nodes except the two end nodes [8].
2. The local connectivity of two nodes is the minimum number of the nodes separating them [9].

A node fault is a way of grouping branches incident with the node. An extension is the group of branches by cuts. A cut is a cut-set or a disjoint union of cut-sets in a graph G [7]. For a given set of t cuts, there is a corresponding cut matrix $H \in R^{txb}$ whose (i,j)th entry $h_{ij} = 1$ if branch j is in cut i and $h_{ij} = 0$ if branch j is not in cut i. A set of basis cuts Q in G is the set of n cuts whose corresponding cut matrix has rank n where n is the number of nodes minus one. Then the following definitions can be stated.

Definition 3: A cut is said to be fault-free if all branches contained in this cut are fault-free. Otherwise, it is faulty.

Definition 4: For a given set of basis cuts Q, a network N is said to be k-cut-fault testable if when N is perturbed to N+ΔN, one will be able to determine, by choosing one appropriate test signal vector i_m, from the measurements on M_o:
(A) whether or not N has no more than k cut-faults in Q.
(B) if affirmative, the faulty cuts can be located uniquely.

With a mild condition, it can be shown [5] that, in a circuit N, k-cut-fault testability for any given set of basis cuts is equivalent to k-node-fault testability, and that the faulty cut can be uniquely determined.

For a given set F of faulty branches, the number of faulty cuts will be different for different choices of set of basis cuts. The following theorem shows how to choose the set of basis cuts so that it contains fewest faulty cuts. First, let F be the set of faulty branches in N and k′ be the number of nodes, excluding the reference node, incident with F. Since F may not be connected, let F consist of r separated parts.

Theorem 2 [5]: For a given F in a circuit N there exists a set of basis cuts Q_F which has only (k′-r) faulty cuts. Furthermore, any other set Q has at least (k′-r) faulty cuts. In other words, (k′-r) is the lowest bound of faulty cuts in any Q.

Remark: Theorem 2 has the following implications. If N is k-node-fault testable, then for any fault pattern F for which (k′-r) ≤ k, there exists Q_F which can uniquely locate the n-(k′-r) fault-free cuts. F will be contained in the remainder of the graph. Since there are no more than C(b,k) number of fault patterns for which (k′-r) ≤ k, we need only to use no more than C(b,k) number of Q's to cover all such fault patterns. The construction of Q_F is given in [5].

Finally, let us consider another method of fault isolation, namely k-rank branch-set diagnosis. A k-rank branch-set is a subgraph of N which consists of (1) k branches containing no loops and (2) all of those branches each of which forms a loop with branches in (1). Since the incident matrix of a k-rank branch-set has rank k, the name follows. Then the following definition can be given.

Definition 5: A k-rank branch-set is said to be

fault-free if all branches contained in this branch-set are fault-free. Otherwise, it is faulty.

Evidently, given a set of faulty branches in N+ΔN, there may be many k-rank branch-sets which all contain these faulty branches. However, there exists a unique k^0-rank branch-set in N+ΔN such that if another k-rank branch-set also contains all the faulty branches then $k > k^0$. Define this k^0-rank branch-set as minimum-rank faulty branch-set. The goal of k-rank branch-set diagnosis is to locate this minimum-rank faulty branch-set. It can be accomplished, if the condition given in Theorem 3 is satisfied, by a procedure similar to that in [1,2,3]. However, because of our grouping of branch-sets, the unnecessary repitition of computions resulting from indistinguishable fault patterns is eliminated. The detail can be found in [6].

Next, we will give the definition of k-rank branch-set fault testability.

Definition 6: A network N is said to be k-rank branch-set fault testable if when N is perturbed to N+ΔN, one will be able to determine, by choosing one appropriate test signal vector i_m, from the measurements on M_o:

(A) whether or not there exists a k-rank branch-set which contains all faulty branches in N,
(B) if affirmative, the minimum-rank faulty branch-set can be uniquely determined. The minimum-rank is no greater than k.

As in the case of k-node-fault testability, there is a graphical condition for k-rank branch-set fault testability. To state the theorem, we need to introduce another modified graph G_b of N. This graph G_b is constructed from N by connecting all the nodes in M_o, i.e., all the accessible nodes including the reference node, to a new node α_b. Notice that the difference in dealing with the reference node here and the case of k-node-fault testability.

Theorem 3 [6]: The following three statements are equivalent:

(1) N is k-rank branch-set fault testable w.r.t. M_o for almost all values of Y_b.
(2) There are at least (k+2) disjoint paths between any inaccessible node and α_b in G_b.
(3) There are at least local (k+2)-connectivity in G_b between any inaccessible node and α_b.

As a conclusion of this section, let us investigate the relationship between k-node-fault testability and k-rank branch-set fault testability. A relationship for a given M_o in N can be summed up in the following two theorems.

Theorem 4 [6]: For any given set M_o of accessible nodes, if N is k-rank branch-set fault testable, then N is k-node-fault testable w.r.t. any reference node in M_o.

The converse of Theorem 4 is not true in general. However, if N satisfies a structural condition, then the converse is also true. This is stated in the following theorem.

Theorem 5 [6]: Suppose that there is an accessible node m_o in M_o, which is adjacent to every inaccessible node. Then k-rank branch-set fault testability is equivalent to the k-node-fault testability if m_o is used as the reference node.

III. MULTIPLE TEST SIGNAL VECTORS

From the discussions in the previous section, it is clear that the fault isolation methods with single test signal vector can only isolate the faults into a subcircuit but are unable to further identify the faults. In this section, we will discuss how to take advantage of the multiple test signal vectors in fault diagnosis. In the following discussion, we only state the theorems concerned with k-rank branch-set diagnosis. The results concerning the other two cases are similar and therefore are omitted.

To begin with, let us introduce some more notations. A set of branches is acyclic if it contains no loops. An acyclic t-branch in N is excitable w.r.t. M_o if the branch voltage vectors of these t branches can be generated to span a t-dimensional space by appropriately applying multiple test signal vectors on M_o. If every acyclic t-branch in N is excitable w.r.t. M_o, then N is said to be t-acyclic-branch excitable. With these definitions, we can state the following theorems [6].

Theorem 6: Suppose that N is k-rank branch-set fault testable and N+ΔN is k-acyclic-branch excitable. Then all the faulty branches can be uniquely determined by k independent multiple test signals.

Theorem 7: N is (k-1)-rank branch-set fault testable if and only if N is k-acyclic-branch excitable.

Theorem 8: If N is k-acyclic-branch excitable and N+ΔN has the same graph as N then N+ΔN is k-acyclic-branch excitable for almost all ΔY_b.

In summary, suppose that N is k-rank branch-set fault testable. Then we can determine whether a set of fault branches can be contained in one of the k-rank branch-sets by a single test signal vector (Theorem 3). Furthermore, if affirmative, we can uniquely determine the minimum-rank branch-set which contains the faulty branches by a single test signal vector (Theorem 3) and we can uniquely determine the faulty branches by a k multiple test signal vectors if the conditions of Theorem 8 are fulfilled.

Similar results can be shown for node-fault and cut-fault diagnosis.

IV. REFERENCES

[1] R.M. Biernacki and J. W. Bandler, "Multiple-Fault Location of Analog Circuits", IEEE Trans. Circuits and Systems, vol. CAS-28, pp. 361-366, May 1981.

[2] R.M. Biernacki and J.W. Bandler, "Fault Location of Analog Circuits", IEEE 1980 Int'l. Symposium on Circuits and Systems, pp. 1082-1086.

[3] A.A. Sakla, E.I. El-Masry and T. N. Trick, "A Sensitivity Algorithm for Fault Detection in Analog Circuits", IEEE 1980 Int'l. Symposium on Circuits and Systems, pp. 1075-1077.

[4] Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and Design of Testability", 20th IEEE Conference on Decision and Control, pp. 1037-1042, 1981.

[5] Z.F. Huang, C.S. Lin and R. Liu, "Cut-Fault Diagnosis of Analog Circuits", (in preparation).

[6] C.S. Lin, Z.F. Huang and R. Liu, "Excitability, Accessibility, Solvability and Diagnosability of Analog Circuits", (in preparation).

[7] W.K. Chen, Applied Graph Theory, North-Holland, 1971.

[8] F. Harary, Graph Theory, Addison-Wesley, 1972.

[9] B. Bollobas, Extremal Graph Theory, Academic Press, 1978.

[10] Z.F. Huang, C.S. Lin and R. Liu, "Node-Fault Diagnosis and Design of Testability", (to appear).

[11] T.N. Trick, "A Sensitivity Approach to Analog Circuit Fault Analysis", Workshop on Analog Fault Diagnosis, University of Notre Dame, April 1980.

Distribution List

Contract N00014-78-C-0444 Final Report

Addressee	# Copies
Scientific Officer Office of Naval Research Code 414 Arlington, VA 22217	2
Admin. Cont. Officer ONRRR The OSU Research Center 1314 Kinnear Road Columbus, OH 43212	1
Director Naval Research Laboratory Attn: Code 2627 Washington, DC 20375	6
Defense Technical Info Center Building 5, Cameron Station Alexandria, VA 22314	12
Mr. Ken Campbell Naval Ocean Systems Center Code 8254 San Diego, CA 92152	1
Capt. Stacy V. Holmes NAVELEX Syscom - 611 Washington, DC 27360	1
Mr. I. Krulac RADC/RBRAC Griffiss Air Force Base New York 13441	1
Mr. Jerry L. Kunert Department of the Navy Naval Air Engineering Center Code 92A3 Lakehurst, NJ 08733	1
Dr. Clifford Lau Office of Naval Research Detachment 1030 E. Green St. Pasadena, CA 91106	1
Mr. Melvin E. Nunn Naval Ocean Systems Center Code 921 San Diego, CA 92152	1
Dr. Randy Shumaker Commander NAVAIR System Command Code AIR 314E Washington, DC 20361	1
Mr. Joel Trimble Office of Naval Research 800 N. Quincy Ave. Arlington, VA 22217	1

There are three branch faults (a, b, c) but only two node faults (1, 2).

almost all Δv_m. Furthermore, all faulty nodes can be uniquely isolated by (13).