

NON-EUCLIDEAN FUNCTIONAL ANALYSIS AND ELECTRONICS

BY J. WILLIAM HELTON¹

1. Introduction. We shall describe a structure built from several components—functional analysis, a few symmetric spaces, a Lie group over a function field, and Nevanlinna-Pick interpolation theory—all fixed on a framework of engineering motivation which determines the relationship between them. A large branch of functional analysis concerns linear spaces of analytic functions called H^p spaces and linear operators on them. Here we describe an analogous study for some sets of functions which while not linear have a very rigid structure patterned on that of the Poincaré disk. There are several basic results in linear H^p theory which have good “non-Euclidean” analogs. One is the classical Szegő theorem which computes the $L_2(d\mu)$ distance of a function f to H^2 . Actually our non-Euclidean result is closer to a theorem of Nehari which computes the supremum norm distance of an f in L^∞ to H^∞ . Another is the Beurling-Lax-Halmos theorem from which such things as the existence of Wiener-Hopf factorizations and the F. and M. Riesz theorem immediately follow.

The rigidity of a geometry on a space is expressed in terms of the group of isometries on that space. These are thought of as rigid structure preserving motions and an early notion of geometry championed by Felix Klein was to specify a space, a group of motions on it, and then to study invariants of these motions. That viewpoint seems highly appropriate for the physical situations we shall encounter.

Recall that the Poincaré disk is the unit ball

$$\mathbb{B}\mathbf{C} = \{z: |z| < 1\}$$

of \mathbf{C} along with the linear fractional transformations

$$\mathcal{G}_g(s) = (\alpha s + \beta)(\kappa s + \gamma)^{-1}$$

of $\mathbb{B}\mathbf{C}$ onto itself. Each \mathcal{G}_g in this group of rigid motions has a coefficient matrix g which satisfies

$$(1.0) \quad g^* \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix};$$

This is an expanded version of an hour address delivered in January 1980 at the American Mathematical Society in San Antonio, Texas; received by the editors October 16, 1981.

1980 *Mathematics Subject Classification*. Primary 35D55, 30E05, 32M15, 46J15, 47A45, 47A65, 47A68, 47B35, 47B38, 47B50, 94C05; Secondary 22E50.

¹Partially supported by the National Science Foundation.

in other words g is in the Lie group commonly denoted by $U(1, 1)$. The natural metric on this space

$$(1.1) \quad p(z_1, z_2) = \operatorname{arctanh} \left| \frac{z_1 - z_2}{1 - \bar{z}_1 z_2} \right|$$

is invariant under these linear fractional maps and is called the Poincaré metric. That p is invariant plays a major role; that it is a metric never enters our proofs.

A function space analog of $\mathfrak{B}\mathbb{C}$ is $\mathfrak{B}L^\infty$, the set of all functions defined on the unit circle Π which have supremum norm less than one. The “rigid motions” Π are simply the linear fractional maps

$$(1.2) \quad \mathfrak{G}_g(S)(e^{i\theta}) = (\alpha(e^{i\theta})S(e^{i\theta}) + \beta(e^{i\theta}))(\kappa(e^{i\theta})S(e^{i\theta}) + \gamma(e^{i\theta}))^{-1}$$

with functions as coefficients. The coefficient matrix must take values² $g(e^{i\theta})$ which are in $U(1, 1)$ for almost all θ . While there are many natural invariant metrics such as

$$P^q(F, G) = \left[\int_0^{2\pi} p(F(e^{i\theta}), G(e^{i\theta}))^q d\theta \right]^{1/q}$$

which we could place on $\mathfrak{B}L^\infty$, we concentrate on

$$P^\infty(F, G) = \sup_\theta p(F(e^{i\theta}), G(e^{i\theta}))$$

since our electronics problems force us to it.

Not any functions in L^∞ will do for engineering purposes, in fact mostly rational functions arise (see footnote 6). Consequently the field

$$\mathfrak{R} \triangleq \{ \text{functions defined on } \Pi \text{ which have a rational continuation to } \mathbb{C} \}$$

plays a special role. Usually we shall work with $\mathfrak{R} \cap \mathfrak{B}L^\infty$ and the Lie group $U(1, 1)$ over the field \mathfrak{R} ; denote these by $\mathfrak{R}\mathfrak{B}L^\infty$ and $\mathfrak{R}U(1, 1)$. Actually much of this article could be viewed as a study of the Lie group $\mathfrak{R}U(m, n)$. However, there is a major point of departure from the route one would expect of a purely group theoretic study. It occurs because the subset

$$\mathfrak{B}H^\infty \triangleq \{ \text{functions analytic on the unit disk whose supremum on the disk is less than one} \}$$

of $\mathfrak{B}L^\infty$ is very important. In particular the only mappings g in $\mathfrak{R}U(1, 1)$ which occur physically have the property

$$(1.3) \quad \mathfrak{G}_g: \mathfrak{B}H^\infty \rightarrow \mathfrak{B}H^\infty.$$

²Actually to guarantee that \mathfrak{G}_g maps $\mathfrak{B}L^\infty$ into the open ball in L^∞ one must have $\|\mathfrak{G}_g(0)\|_{L^\infty} < 1$ in addition to $g(e^{i\theta}) \in U(1, 1)$. Henceforth we always assume this nondegeneracy condition.

The inverse of a g with this property usually doesn't have the property, so we are not dealing with a group. A constraint³ on g which guarantees⁴ (1.3) is

$$(1.4) \quad g(z) * \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} g(z) \leq \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$$

for all $|z| < 1$. The set of g 's in $\mathfrak{R}U(1, 1)$ satisfying (1.4) will be designated by $\mathfrak{R}U^+(1, 1)$ and this will be called the *physical subsemigroup* of $\mathfrak{R}U(1, 1)$. We ultimately give $\mathfrak{R}U^+(n, n)$ much attention and find that it is tightly bound to classical interpolation theory, modern commutant lifting theory, and kindred topics.

We could, of course, work with spaces of matrix valued rather than scalar valued functions. Let $M_{m,n}$ denote the set of $m \times n$ matrices and $\mathfrak{B}M_{m,n}$ its unit ball. Linear fractional transformations of the form (1.2) with matrix⁵ coefficients $\alpha, \beta, \kappa, \gamma$ map $M_{m,n}$ into $M_{m,n}$ generically. Furthermore, if the coefficient matrix $g = \begin{pmatrix} \alpha & \beta \\ \kappa & \gamma \end{pmatrix}$ satisfies the matrix version of (1.0) then \mathfrak{G}_g is a biholomorphic map on $\mathfrak{B}M_{m,n}$. The group $U(m, n)$ of these g constitutes the "rigid motions" for $\mathfrak{B}M_{m,n}$. There are infinitely many $U(m, n)$ invariant metrics on $\mathfrak{B}M_{m,n}$ and the most natural one for our electronics problems is ironically the first to have been discovered. It is called the Carathéodory metric and its definition is a little more complicated than (1.1); we denote it by $p(\cdot, \cdot)$. Definitions of the function spaces $\mathfrak{B}L^\infty(M_{m,n})$, $\mathfrak{B}H^\infty(M_{m,n})$, the group $\mathfrak{R}U(m, n)$, the semigroup $\mathfrak{R}U^+(m, n)$, and the metric P^∞ follow the pattern of the scalar case exactly. Most of this exposition is done in terms of these objects and matrix valued functions. A list of their precise definitions appears in the notation guide at the end of this section.

The goal of this paper is to present the prominent features of $\mathfrak{R}U(m, n)$, $\mathfrak{R}U^+(m, n)$ and their action on the set $\mathfrak{B}L^\infty(M_{m,n})$ of functions endowed with the "Poincaré" metric P^∞ . An outline of the main structure is

(a) *There is an explicit formula for computing the distance of a rational function F in $\mathfrak{B}L^\infty(M_n)$ to $\mathfrak{B}H^\infty(M_n)$.* This is a "non-Euclidean" analog of the Szegő and the Nehari distance theorems in linear H^p theory.

(b) *A closed subspace \mathfrak{N} of $H^2(C^{m+n})$ has the form*

$$\mathfrak{N} = gH^2(C^{m+n})$$

for a matrix function g analytic on the disk with boundary values $g(e^{i\theta})$ in $U(m, n)$ a.e. if and only if \mathfrak{N} is invariant under multiplication by z and also meets some mild nondegeneracy conditions. The space \mathfrak{N} is cofinite dimensional if and only if g is rational. This is a direct generalization of the Beurling-Lax-Halmos theorem (which is after all just the first statement with $n = 0$). An equivalent formulation amounts to *characterization of sets of the form $\mathfrak{G}_g(\mathfrak{B}H^\infty(M_n))$.* This formulation might be thought of as a "non-Euclidean"

³By $A \leq B$ when A and B are selfadjoint matrices on \mathbf{C}^n we mean $(Ax, x) \leq (Bx, x)$ for all x in \mathbf{C}^n .

⁴Conversely any rational g which satisfies (1.3) must equal an element g^+ of $\mathfrak{R}U^+(1, 1)$ up to a multiple (from \mathfrak{R}).

⁵The appropriate dimensions are $\alpha \in M_{m,m}$, $\beta \in M_{m,n}$, $\kappa \in M_{n,m}$, $\gamma \in M_{n,n}$.

analog of the Beurling-Lax-Halmos theorem. The surprising thing is that this contains most of classical interpolation theory.

(c) *The orbits*

$$O_S \stackrel{\Delta}{=} \{ \mathcal{G}_g(S) : g \in \mathfrak{R}U(n, n) \}$$

of $\mathfrak{R}U(n, n)$ have a nice explicit characterization (at least when S is a constant function.) Note the main difficulty is that \mathfrak{R} is not algebraically closed.

(d) Let E be a submanifold of $\mathfrak{B}L^\infty(M_n)$. The P^∞ closest point G in E to a given F in $\mathfrak{B}L^\infty(M_n)$ has the property that a P^∞ geodesic from F to G hits G in a direction "orthogonal" to the tangent space of E at G . Here orthogonal means in a standard Banach space sense. This generalizes a simple result in Banach space theory which is basic in approximation theory.

Typically one section of this paper is devoted to each of these subjects. Their locations are listed as a

TABLE OF CONTENTS

§2 Is an elementary exposition of the circuit theory which motivated this enterprise.

§3 Contains the generalized Beurling-Lax-Halmos theory along with a survey of modern (Nevanlinna-Pick type) interpolation theory. Most of interpolation theory is a simple and direct application of this theorem.

§4 Solves the P^∞ distance to $\mathfrak{B}H^\infty$ question and several close relatives of it.

§5 (a) Treats orbits of $U(n, n)$.

(b) Concerns the semigroup C of linear fractional maps \mathcal{G}_g of $\mathfrak{B}M_n$ into itself and shows that there are many $U(n, n)$ invariants i in addition to p which are diminished by g in C ; that is

$$i(\mathcal{G}_g(z_1), \mathcal{G}_g(z_2)) \leq i(z_1, z_2)$$

§6 Applies §§3, 4, 5 to the circuit problems in §2.

§7 Lists (mostly qualitative) physical conclusions.

§8 Epilogue.

There is a little redundancy in the paper to make various sections independent. To skip the circuit theory go from here directly to §§3, 4, 5, then maybe §6. For a self-contained account of modern interpolation go directly from here to §3. For an exposition of circuits read the rest of §§1 and 2. For circuit theoretic conclusions read §1 then §7, and maybe part of §6. The approximation theorem (d) above will not be presented in this article. It holds for many metrics other than P^∞ and can be found in [Bdr].

For many pages the reader has seen vague allusions to electronics. The time has come to be a little more specific. An electronic box having two terminals on its face can be thought of as a function S in $\mathfrak{R}H^\infty$. An energy conserving box with four terminals corresponds to a matrix function g in $\mathfrak{R}U^+(1, 1)$. Wire the two terminals of the first box to two of the four terminals of box two; this

leaves two free terminals which, of course, amounts to a new two-terminal circuit (see Figure 1.1). This new circuit must correspond to a new function S' in \mathfrak{R} and miraculously the rule for computing S' is simply to apply the linear fractional map \mathcal{G}_g with coefficient matrix g to S !

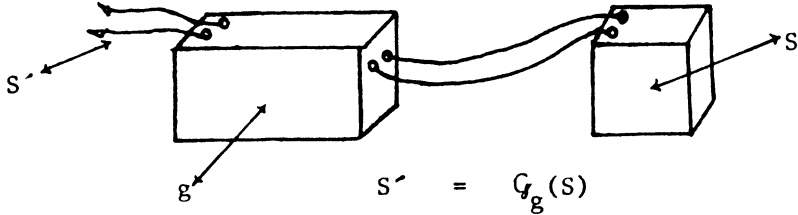


FIGURE 1.1

Thus the orbit O_S^+ of S under $\mathfrak{R}U^+(1, 1)$ corresponds to all circuits which can be made from a fixed circuit S by the basic construction in Figure 1.1.

There in a nutshell is why we are forced to study $\mathfrak{R}H^\infty$ and the action of $\mathfrak{R}U^+(1, 1)$ on it. Even from this primitive starting point it takes little imagination to suspect that a thorough mathematical study of these subjects is justified and that any basic mathematical structure one finds will be physically important. Probably the best way to give a mathematician an idea of what is involved in several areas of old-fashioned linear circuit theory is simply to list the mathematical questions which arise. The word "arise" is probably too weak since these mathematical subjects actually *constitute* various areas of circuit theory. We should emphasize that the most naive considerations in circuit theory force one to an overwhelming collection of mathematical questions. So in this article we restrict attention to situations involving power gain and power transfer (and omit consideration of phase, of voltage or current gain, and of network sensitivity). Even the narrow topic of power gain leads to a large enough class of problems that we can only describe some of the most basic ones. We begin with passive (energy dissipating) circuit theory.

A basic type of problem in classical passive filter design can be summarized mathematically as:

Given S_1 and S_2 in $\mathfrak{R}\mathfrak{B}H^\infty$ find a Σ in the orbit $O_{S_1}^+$ of S_1 under $\mathfrak{R}U^+(1, 1)$ with pointwise Poincaré distance $p(\Sigma(e^{i\theta})^*, S(e^{i\theta}))$ equal to a prescribed function $F(e^{i\theta})$ for all θ .

Here $*$ denotes the complex conjugate. Subdivisions of the question correspond to various branches of the classical theory of filter design.

(i) What is the supremum norm smallest function F which can be obtained this way? (The problem of broadband impedance matching.)

(ii) Which functions F can be obtained this way? (The problem of gain equalization.)

(iii) Take $S_1 = 0$ and $S_2 = 0$ and add the constraint that Σ be rational of (prescribed) low order and approximate a given $F(e^{i\theta})$ with the functions $p(\Sigma(e^{i\theta}), 0)$ so obtained. (The insertion loss approximation problem; includes the first step in Butterworth, Tchebychef, elliptic filter design.)

(iv) Do (iii) not with the full semigroup $\mathfrak{R}U^+(1, 1)$ but with a subsemigroup having prescribed generators. Using different subsemigroups corresponds to

allowing different components in construction of the circuits (includes RC , LC , Fujisawa ladder filter design.)

Amplifiers provide another large class of power transfer problems. Typically one wants to design a (stable) amplifier with maximum power gain or prescribed power gain. Again optimization of a particular function over an orbit O_S^+ is the central issue. For example in (the small signal analysis of) a single transistor amplifier the transistor corresponds to a large S in $H^\infty(M_2)$, the amplifier corresponds to $\mathcal{G}_g(S)$, and its gain (at the worst behaved operating frequency) is

$$\inf_{\theta} |(\mathcal{G}_g(S))_{21}|$$

where M_{ij} means the ij th entry of the matrix M . Thus finding the amplifier with greatest gain over all frequencies amounts to maximizing

$$(1.5) \quad \inf_{\theta} |\Sigma_{21}|$$

over Σ in O_S^+ . Practical amplifiers must be “stable” which adds the constraint that Σ be chosen from (a slightly smaller class than) $H^\infty(M_2) \cap O_S^+$.

Although the discussion has centered on $\mathcal{R}U(n, n)$, $\mathcal{R}U^+(n, n)$ and their orbits, the semigroups

$$C(n, n) = \{g \in \text{GL}(2n): g(z) \text{ satisfies inequality (1.4) for all } |z| = 1\}$$

and

$$C^+(n, n) = \{g \in \mathcal{R}\text{GL}(2n): g(z) \text{ satisfies (1.4) for all } |z| \leq 1\}$$

of $\mathcal{R}\text{GL}(2n)$ are also important, because for a passive circuit g lies in $\mathcal{R}C^+(n, n)$. Typically a simple amplifier consists of a transistor together with passive circuitry which compensates for undesirable properties of the transistor. For the situation in the previous paragraph this says we should not work with the orbit O_S^+ of S but the orbit CO_S^+ of S under $\mathcal{R}C^+(n, n)$. This appears to give rise to a whole new class of problems, but fortunately it does not because the physically important functions have the same optimum over O_S^+ as over the larger set CO_S^+ . Mathematically this is related to the old principle that the Poincaré metric is contracted by analytic maps of $\mathcal{B}\mathbb{C}$ into itself and more is said in §6. At any rate we may restrict our attention to $\mathcal{R}U^+(n, n)$ and O_S^+ for analyzing most power transfer problems.

While these engineering problems don't seem close to the theory we described, the discrepancy is not actually so great. We shall ultimately see (§6) that the power transfer problem (i) reduces to finding the Poincaré distance of S^* to $\mathcal{R}H^\infty$ and so it is completely solved by result (a). Problem (ii) is settled by an easy generalization of result (a) which is mentioned but not belabored in this paper. Amplifier problems amount to a strict mathematical generalization of this P^∞ distance problem. For example, a transistor with $S_{12} = 0$ is called *unilateral* and a very common gain optimization problem for a unilateral transistor actually is equivalent to the passive filter problems (i) and (ii) which (a) solves. The approximation problems (iii) and (iv) are not related to the main mathematical theme of this paper and were included only for general information.

The present state of affairs is that the machinery described herein and some extensions of it [B-H4] convert about any power gain problem to an interpolation problem. The simplest physical problems correspond to old fashioned Nevanlinna-Pick interpolation and so are completely solved. More difficult physical power transfer problems correspond to vast extensions of Nevanlinna-Pick theory developed in the early 1970's and so are solved. Analyzing the full stable amplifier as in (1.5) requires a direct extension of interpolation theory of the 1970's and so is far from completion. Thus considerable progress has been made on power transfer problems, but more is needed to fully understand some extremely basic situations in electronics. Possibly profound extensions of today's interpolation theory will someday settle all of the problems we have raised. However, it is more likely that mathematical efforts in this area will take a qualitative turn. Although explicit solutions have been the main goal so far this will ultimately become too difficult as it is in most subjects. Fortunately, the theory here seems well suited to deriving general properties which optimal amplifiers and power transfer circuits must have.

From the mathematical point of view it is worth mentioning that while we have focused on $U(m, n)$ it is natural to develop a similar theory for other classical Lie groups. Such work is in progress. Several of these other Lie groups have physical significance.

Thanks are certainly due to J. Ball, L. Chua, H. Tan and my students J. Allen, E. Fletcher and D. Schwartz for their comments on this manuscript.

NOTATION GUIDE

Some common symbols are:

H^2 (resp. $\overline{H^2}$) = the functions in L^2 of the circle Π whose negative (resp. positive) Fourier coefficients are zero.

H^∞ = the uniformly bounded functions in H^2 .

H_l^∞ = the functions of the form $\frac{1}{\varphi} H^\infty$ where φ is some Blaschke product having at most l zeroes inside the disk. Loosely speaking these are functions in H^∞ except for possibly having l poles inside the disk. For matrix valued H_l^∞ the function φ is naturally a matrix valued Blaschke product.

$M_{m,n}$ = the $m \times n$ matrices; they map $\mathbf{C}^n \rightarrow \mathbf{C}^m$.

$M_n = M_{n,n}$.

\mathfrak{R} = functions on Π with a rational continuation to the complex plane.

\mathfrak{B} (resp. $\overline{\mathfrak{B}}$) = a prefix meaning the open (resp. closed) unit ball.

$$\mathfrak{B}L^\infty = \{f \in L^\infty: \|f\|_\infty < 1\}.$$

Combinations of these symbols while cumbersome looking are very easy to interpret. For example, $\mathfrak{R}\mathfrak{B}H^\infty(M_{m,n})$ is just the set of rational functions in the open unit ball of $m \times n$ matrix valued H^∞ .

The main groups and semigroups which occur involve an $m + n \times m + n$ matrix valued function g and the inequality

$$(1.4) \quad g(z)^* \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} g(z) \leq \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

Let $GL(k)$ denote the invertible $k \times k$ matrices and $\mathfrak{R}GL(k)$ denote the group of matrix-valued g with entries in \mathfrak{R} and values $g(e^{i\theta})$ in $GL(k)$ a.e. Then

$$\begin{aligned} \mathfrak{R}U(m, n) &= \{g \in \mathfrak{R}GL(m+n): \text{for which } g(e^{i\theta}) \text{ satisfies (1.4)} \\ &\quad \text{with equality and } \|\mathfrak{G}_g(0)\|_{L^\infty} < 1\}. \\ \mathfrak{R}U^+(m, n) &= \{g \in \mathfrak{R}U(m, n): \text{satisfying (1.4) for all } |z| < 1\}. \\ \mathfrak{R}C(m, n) &= \{g \in \mathfrak{R}GL(m, n): \text{for which } g(e^{i\theta}) \text{ satisfies (1.4)} \\ &\quad \text{and } \|\mathfrak{G}_g(0)\|_{L^\infty} < 1\}. \\ \mathfrak{R}C^+(m, n) &= \{g \in \mathfrak{R}C(m, n): \text{satisfying (1.4) for all } |z| < 1\}. \end{aligned}$$

The notation for the orbit of S :

under $\mathfrak{R}U(n, n)$ is O_S ;

under $\mathfrak{R}U^+(n, n)$ is O_S^+ ;

under $\mathfrak{R}C(n, n)$ and $\mathfrak{R}C^+(n, n)$ is CO_S and CO_S^+ .

Also in §2 one finds a group $\Pi \times SL_2(R)$, isomorphic to $U(1, 1)$ and some spaces $\mathcal{P}\mathcal{C}$ and $\mathcal{P}^+\mathcal{C}$ biholomorphic to $\mathfrak{B}L^\infty$ and $\mathfrak{B}H^\infty$. Since these appear only in §2 we won't put these definitions here in our reference list. Always m^t denotes the transpose of the matrix m and m^* denotes its conjugate transpose.

A useful convention is that lower case roman letters usually denote scalars or matrices, while functions are (usually) capital letters. The one prominent exception is that g typically denotes an element of the group $U(m, n)$ or $\mathfrak{R}U(m, n)$.

2. An exposition of classical filter design for mathematicians. This exposition begins generally, but ultimately emphasizes issues of electronic power transfer. An even more classical subject is circuit synthesis (building a circuit to meet given specifications) and this is treated in a fine article by Effimov and Potapov [E-P]. The basic circuit texts [C-L], [D-K], [K-R], [W] give a variety of viewpoints. An excellent book on the related subject of linear systems is [K].

a. *Connecting boxes.* The mathematical study of electric circuits is rife with linear fractional maps both of the usual and of very general sorts. This is true for a simple reason. Think for the moment of a box which takes in a vector

$$\vec{i} = \begin{pmatrix} i_1 \\ \vdots \\ i_n \end{pmatrix}$$

and then puts out a vector

$$\vec{o} = \begin{pmatrix} o_1 \\ \vdots \\ o_n \end{pmatrix}.$$

At the moment we are considering boxes in the abstract (and the vectors \vec{i} and \vec{o} do not necessarily have the interpretation of voltages and currents). We shall

say the box is linear provided it acts as a linear function, namely, there is a matrix M so that

$$(2.1) \quad Mi = \vec{o}.$$

Such a box with n -inputs and outputs is called an n -port. Suppose that we want to connect two boxes in order to obtain a new one; to make things simple we consider the simple case where one box is a two port $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the other is a one port m and we connect the boxes as in the figure to obtain a one port q .

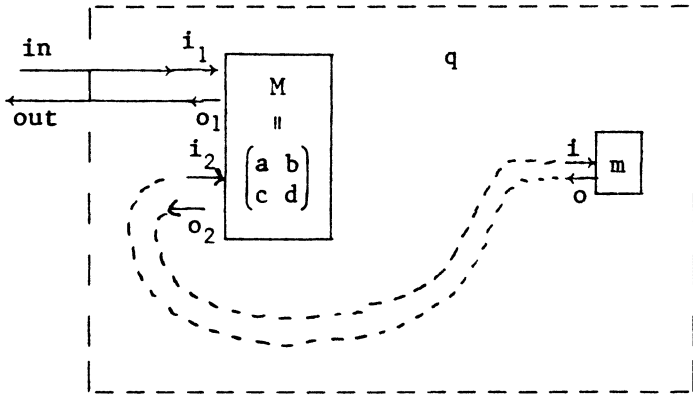


FIGURE 2.0

What is q in terms of M and m ? The definitions of M and m say

$$ai_1 + bi_2 = o_1, \quad ci_1 + di_2 = o_2, \quad mi = o$$

while connecting m to M dictates $o_2 = i$ and $i_2 = o$. Substitution gives

$$ai_1 + bmi = o_1, \quad ci_1 + dmi = i.$$

Use the second equation to eliminate i from the first equation and obtain $o_1 = \mathfrak{F}_M(m)i_1$ where $\mathfrak{F}_M(m)$ is defined to be

$$(2.2) \quad \mathfrak{F}_M(m) = a + bm(1 - dm)^{-1}c.$$

Thus $q = \mathfrak{F}_M(m)$ and one can think of this type of network connection (called *cascade connection*) as the action of a linear fractional map \mathfrak{F}_M on the number m ; in particular, if inputs and outputs are vectors rather than numbers then a, b, c, d, m become vector space operators and the formulas still hold (subject to invertibility of $1 - dm$); now \mathfrak{F}_M maps an operator m to another operator.

Naturally there are several conventions one could use in associating a matrix to an n -port and consequently several equivalent (except for degeneracies) formalisms for describing boxes and their connections. While the formalism just presented is the most intuitively appealing there is another one which is handier for computation and so more commonly used in both engineering and mathematics. It is usually called the *chain formalism* and we now present it for 2-ports.

Suppose that a given 2-port takes inputs $\begin{pmatrix} in_1 \\ in_2 \end{pmatrix}$ to outputs $\begin{pmatrix} out_1 \\ out_2 \end{pmatrix}$. Associate to it a matrix N , called its chain matrix, defined by

$$N \begin{pmatrix} in_2 \\ out_2 \end{pmatrix} = \begin{pmatrix} out_1 \\ in_1 \end{pmatrix}.$$

The computational advantage becomes quickly apparent when we consider the *cascade connection* of two two-ports N_1 and N_2 . This type of connection is illustrated in (and essentially defined by) Figure 2.1. Note that in the figure a 2-port is drawn (entirely for convenience) as having one port coming out one side and the other port coming out the other while in Figure 2.0 both ports come out to the left. The equations actually defining the connection are

$$in_2 = out_3 \quad \text{and} \quad out_2 = in_3$$

while the chain matrix N for the total cascade circuit satisfies $N \begin{pmatrix} in_4 \\ out_4 \end{pmatrix} = \begin{pmatrix} out_1 \\ in_1 \end{pmatrix}$. Clearly $N = N_1 N_2$ where $N_1 N_2$ indicates ordinary composition of the mappings, and this simple composition law for chain matrices is obviously what motivated this otherwise peculiar way of associating matrices with 2-ports.

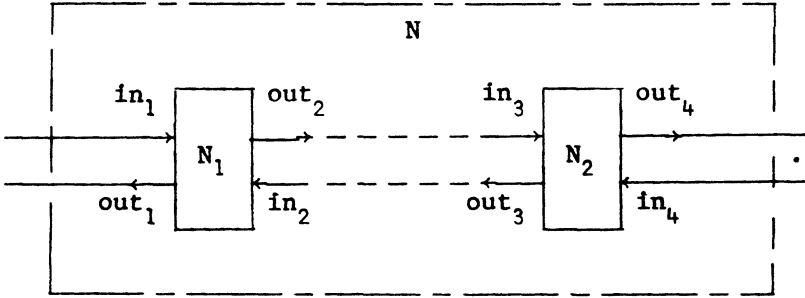


FIGURE 2.1

Another basic (easily checked) fact about the chain formalism is that the cascade connection (as in Figure 2.0) of a 2-port having chain matrix $N = \begin{pmatrix} \alpha & \beta \\ \kappa & \gamma \end{pmatrix}$ with a one-port corresponding to m (according to (2.1)) gives a 1-port corresponding to

$$(2.3) \quad \mathcal{G}_N(m) = (\alpha m + \beta)(\kappa m + \gamma)^{-1}.$$

So we see that the chain formalism gives linear fractional maps in the conventional form.

b. *Electric circuits driven by direct current.* (i) *The impedance matrix.* Suppose we have a (1-port) electric circuit which we think of as a box with two wires sticking out. If we connect the circuit to a battery after a very very brief moment of chaos we observe that a current i flows steadily through each wire and there is a voltage v across the wires (which does not change with time). Now one could test the characteristics of the circuit by connecting batteries of various different strengths and thereby obtain a family of current voltage pairs $\langle i, v \rangle$. The circuit will be called linear provided v is proportional to i , that is,

provided there is a real number z , called the *impedance* of the circuit, such that $v = zi$. From the viewpoint of an observer the behavior of a circuit (with respect to steady current flows) is completely described by z .

A similar procedure can be used to describe most circuits with more than 2 wires. To describe it we will only consider circuits which have an even number $2n$ of wires sticking out. We think of the wires as occurring in n pairs and each pair of wires is assumed to have the property that in the course of normal operation the current flowing in one wire *equals* the current flowing out the other. This seems like a very restrictive setup but by the artifice of appending ‘virtual wires’ one can describe most circuits this way (see, for example, the forthcoming example of a transistor). The situation is best described by Figure 2.2.

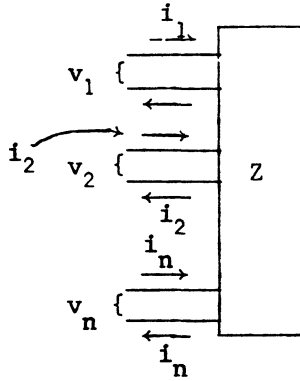


FIGURE 2.2

To associate a matrix with the circuit simply tie one battery to each pair of wires thereby creating some current flow

$$\vec{i} = \begin{pmatrix} i_1 \\ \vdots \\ i_n \end{pmatrix}$$

and a collection of voltages

$$\vec{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

For a linear circuit there is a matrix Z , called its impedance matrix, which satisfies $Z\vec{i} = \vec{v}$. Most classical textbook circuits behave linearly; many practical circuits do not.

There are obviously many other ways to identify a circuit with a matrix and practically all of them are used from time to time. For example, if Z is the impedance matrix for a circuit, $Y \stackrel{\Delta}{=} Z^{-1}$ is called its *admittance* and $S = (1 - Z)(I + Z)^{-1}$ is called the *scattering* matrix for the circuit. One can define

Y and S directly without ever actually mentioning Z itself and do circuit design in terms of them. There is yet another identification and we will use it heavily in this paper. To a 2-port assign \mathfrak{C} by the rule

$$\mathfrak{C} \begin{pmatrix} v_2 \\ -i_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ i_1 \end{pmatrix}.$$

It is called the chain matrix for the circuit.

(ii) *Power consumption.* The instantaneous power consumed by an n -port circuit with current i flowing in and voltage \vec{v} across its wires is

$$P = (\vec{v}, \vec{i})_{R^n} \stackrel{\Delta}{=} v_1 i_1 + v_2 i_2 + \cdots + v_n i_n.$$

So the instantaneous power consumed by a circuit whose impedance matrix is Z is

$$P = (Z\vec{i}, \vec{i})_{R^n}.$$

Thus a circuit driven by direct current conserves energy, dissipates energy or creates energy according to whether or not Z and its transpose Z^T satisfy:

$$\begin{aligned} \text{lossless} & \quad (P = 0) \Leftrightarrow Z = -Z^T; \\ \text{passive} & \quad (P \geq 0) \Leftrightarrow Z + Z^T \geq 0; \\ \text{strictly active} & \quad (P < 0) \Leftrightarrow Z + Z^T < 0. \end{aligned}$$

Note an active circuit must be plugged into the wall (or have batteries) but for our purposes we completely ignore this external connection in our account of what happens at wires coming into the box.

(iii) *Circuit connections.* Simply take the convention that the currents are inputs and voltages are outputs. Then circuit connections are described mathematically in §2(a). For example, let $n = 2$ in Figure 2.2 and connect the second pair of wires to the two leads emanating from a one port with impedance z ; the result is a circuit with two wires coming out. Its impedance (by the calculation surrounding Figure 2.0) is $\mathfrak{F}_Z(z)$. One can check (but probably shouldn't) that if the two port Z is lossless (passive), then the linear fractional map \mathfrak{F}_Z takes the right half plane R.H.P. onto (into) itself. Passive 1-ports correspond to the positive real axis $[0, \infty)$ and this clearly maps into itself.

To describe connections in the chain formalism one simply adopts the seemingly arbitrary convention that

$$\text{in}_2 = v_2, \quad \text{out}_1 = v_1, \quad \text{out}_2 = -i_2, \quad \text{in}_1 = i_1.$$

This immediately makes the general chain matrix N of §2(a) the same as \mathfrak{C} , the specific circuit theoretic one. Also the circuit theoretic cascade connection is a special case of the general one. For example, let z be the impedance of a one port $z i = v$; cascade it with the two port described by \mathfrak{C} : that says $i = -i_2$ and $v = v_2$. This is the same as $z(\text{out}_2) = \text{in}_2$ as required for the general cascade rule.

(iv) *Power transfer.* Suppose we have a source of electricity and a one port with impedance m which the source powers. How much power does the circuit

m receive? The first step is to give the conventional description of a power source. An *ideal voltage source* (voltage v) is a box with two wires sticking out; a voltage meter across the wires reads v volts no matter what you connect the wires to. To see that this model is deficient suppose that a mischievous soul connects our source to a heavy copper bar of resistance $r \approx 0$. The power consumed by the bar is $ri^2 = v^2/r$. Now r is roughly 0 so if we insist that our source maintains a voltage v of moderate size the power consumed by the bar is roughly ∞ . That makes the electric bill too high even for nature so what happens in practice is that as the demand on a power source increases the voltage falls. Thus a *voltage source* is usually described as an ideal voltage source in series with a resistor of resistance r called the *internal impedance* of the source (see Figure 2.3). In this section we will analyze what happens when a source of this type is connected (as in Figure 2.4) to a one-port with impedance z which is passive ($z > 0$).

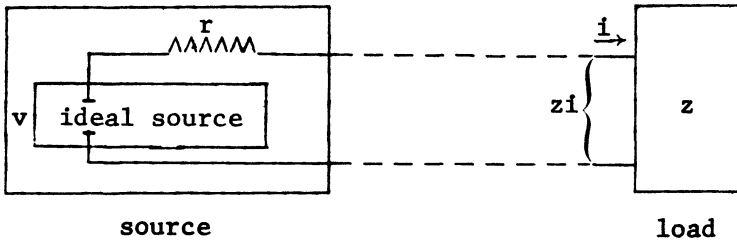


FIGURE 2.3

To compute the power P consumed by the load note

$$\text{ideal voltage} = v = ri + zi = (r + z)i$$

so

$$P = (zi)i = z \left(\frac{v}{r + z} \right)^2 = \frac{z}{(r + z)^2} v^2.$$

As we ultimately shall see what is interesting in most design situations is not the absolute power delivered to the load, but the ratio

$$\frac{\text{power received by load } z}{\text{power available from source}} \triangleq \frac{P_z}{P_s}.$$

The power available from the source P_s is always taken to mean

$$\max_z P_z = \max_z \frac{z}{(r + z)^2} v^2$$

which by calculus is $P_s = rv^2/(2r)^2 = v^2/4r$. Thus

$$\frac{P_z}{P_s} = \frac{4zr}{(r + z)^2} = 1 - \left(\frac{r - z}{r + z} \right)^2$$

which we note is independent of v (the strength of the ideal source). The business part of the power ratio is $\delta(r, z) = |(r - z)/(r + z)|$ and we call it

the *power mismatch* between source r and load z , since if it is small (big) the power transfer ratio is high (low). Note that $\operatorname{arctanh} \delta(r, z)$ is the classical R.H.P. Poincaré distance between the real numbers r and z . Thus *computing the power mismatch between r and z is equivalent to computing the Poincaré distance between r and z .*

(v) *Summary.* Passive one ports correspond to positive numbers. Power mismatch corresponds to R.H.P. Poincaré distance restricted to $(0, \infty)$. The cascade connection of a lossless 2-port with a passive one port corresponds to a linear fractional transformation (L.F.T.) of the R.H.P. which maps $(0, \infty) \rightarrow (0, \infty)$.

c. *Electric circuits driven by a sinusoidally varying current.* (i) *General theory.* In the previous subsection we studied circuits which were powered by a source whose output does not vary with time. Now we consider a power source which varies sinusoidally with time; say the current it produces is $i_0 \sin \omega t$. Connect it to a (linear) one-port. You will observe that after a very brief moment of chaos the voltage across the wires entering the 1-port is

$$v(t) = v_0 \sin(\omega t + \theta).$$

The voltage function has the same frequency but a different phase. Now we all recognize that the easiest way to compute with waves like this is in complex notation. Let $j = \sqrt{-1}$. One could think (loosely) of a complex current $i(t)$ in the wires and corresponding complex voltage $v(t) = e^{j\theta} v_0 i(t) / i_0$ across them. For a linear circuit the ratio v_0 / i_0 and θ are independent of i_0 and one defines the impedance of the circuit at frequency ω to be the complex number

$$z(\omega) = \frac{v_0}{i_0} e^{j\theta}.$$

For example, if $\omega = 0$ there is no phase shift and we get $z(0) = v_0 / i_0$; precisely the situation in §2(b). As before an n -port still has an $n \times n$ impedance matrix $Z(\omega)$ associated with it, only now the entries may be complex numbers.

Now §2(b) extends in a straightforward way to sinusoidally varying currents. We now list the generalization and mention that someone with a serious interest in the physical issues should ultimately consult a standard text since what we have said is abbreviated beyond the point of being a physically solid explanation. Let Z^* denote the conjugate transpose of the matrix Z . *The power consumed by an n -port is*

$$P = \operatorname{Re}(\vec{v}, \vec{i})_{C^n}.$$

An n -port is

lossless	if and only if $\operatorname{Re} Z(\omega) = Z(\omega) + Z(\omega)^* = 0$,
passive	if and only if $\operatorname{Re} Z(\omega) \geq 0$,
strictly active	if and only if $\operatorname{Re} Z(\omega) < 0$,

or for 2-ports the chain matrix $\Psi = \begin{pmatrix} \alpha & \beta \\ \kappa & \gamma \end{pmatrix}$ is

lossless	if and only if $\exists \lambda$ with $ \lambda = 1$ such that
	$\lambda\alpha, -i\lambda\beta, i\lambda\gamma$, and $\lambda\kappa$ are real and $\det \lambda\Psi = 1$.

Cascade connection of a 2-port whose impedance Z (resp. chain matrix is Ψ)

with a 1 port having impedance z gives a 1-port with impedance $\mathfrak{F}_z(z)$ (resp. $\mathfrak{G}_\Psi(z)$). If Ψ is lossless, then $\mathfrak{G}_\Psi(z)$: R.H.P. $\xrightarrow{\text{onto}}$ R.H.P. For one ports the *power transfer ratio* called the *transducer power gain* (T.P.G.) for a source with impedance z_1 and load with impedance z_2 is $1 - \delta(z_1, z_2)^2$ where the *power mismatch* δ is

$$(2.4) \quad \delta(z_1, z_2) = \left| \frac{z_1 - \bar{z}_2}{z_1 + \bar{z}_2} \right|.$$

That is, the *mismatch is essentially the Poincaré distance between z_1 and \bar{z}_2 .*

Thus the basic setup looks like it comes from the beginning of an introductory complex variables course. In fancier language the study of circuits operating at one fixed frequency and built by cascades of lossless 2-port circuits is just a study of the Lie group $\Pi \times \text{SL}_2(\mathbb{R})$ where $\text{SL}_2(\mathbb{R})$ is the usual

$$\text{SL}_2(\mathbb{R}) = \left\{ \kappa = \begin{pmatrix} l & m \\ n & p \end{pmatrix} : l, m, n, p \text{ real and } \det \kappa = 1 \right\},$$

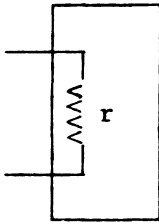
and Π is the circle group. This is obscured slightly by the fact that engineers use different conventions; in particular they work not on the U.H.P. but on the R.H.P. Soon we shall summarize the structure just described in terms of $\Pi \times \text{SL}_2(\mathbb{R})$, the U.H.P., and other familiar mathematical objects. Then we give numerous examples. First we mention that instead of working with L.F.T.'s on the U.H.P. or R.H.P. we could work with L.F.T.'s on the unit disk. This is accomplished with the usual mapping $z \rightarrow (1 - z)(1 + z)^{-1} = s$ of the R.H.P. onto the disk. As already mentioned engineers call s the scattering parameter for the circuit with impedance z . Naturally the transformation group $\Pi \times \text{SL}_2(\mathbb{R})$ on the U.H.P. is converted to a transformation group on the disk by composing with the mapping $im \rightarrow z \rightarrow s$; one finds that this Lie group is $U(1, 1)$. We shall now summarize the engineering, the $\text{SL}_2(\mathbb{R})$, and the $U(1, 1)$ (scattering) conventions. The conversion rule is

	ENG.	MATH.	
		$\Pi \times \text{SL}_2(\mathbb{R})$	$U(1, 1)$
Impedance matrix	z	$m = iz$	$s = (1 + im)(1 - im)^{-1}$
Chain matrix	$\Psi = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$	$g = \begin{pmatrix} 1 & 0 \\ 0 & -i \end{pmatrix} \Psi \begin{pmatrix} 1 & 0 \\ 0 & i \end{pmatrix} = \begin{pmatrix} a & ib \\ -ic & d \end{pmatrix}$	$g = \frac{1}{2} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \Psi \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$
$\sqrt{-1}$	j	i	i

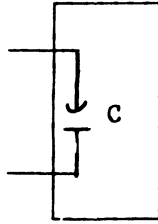
The key structure is

CIRCUIT	$\Pi \times \text{SL}_2(\mathbb{R})$	$U(1, 1)$
passive 1-port	m in U. H. P.	s in disk
lossless 2-port	g in $\Pi \times \text{SL}_2(\mathbb{R})$	g in $U(1, 1)$
cascade of 2 and 1-port	$\mathfrak{G}_g(m)$	$\mathfrak{G}_g(s)$
cascade of a 2-port and a 2-port	$g = g_1 g_2$	$g = g_1 g_2$
power mismatch source m_1 and load m_2	$p(-\bar{m}_1, m_2)$ -Poincaré distance on U.H.P.	$p(\bar{s}_1, s_2)$ -Poincaré distance disk

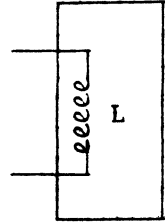
(ii) *Examples and common procedures.* We begin by listing the impedances $m(\omega)$ of common 1-ports operating at frequency ω . We always use our *mathematics* SL_2 convention and not the usual engineering one.



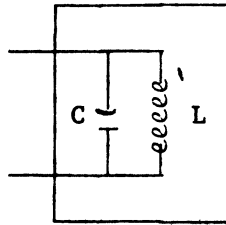
resistor
 $m(\omega) = ir$



capacitor
 $m(\omega) = \frac{1}{\omega C}$



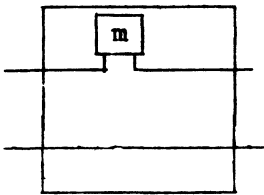
inductor
 $m(\omega) = -\omega L$



tank circuit
 $m(\omega) = \frac{\omega L}{1 - \omega^2 LC}$

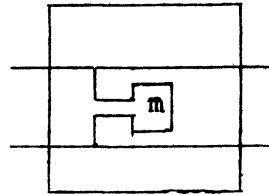
FIGURE 2.4

The two most common two ports are simply made from a 1-port m . They and their chain matrices are



series

$$e(m) \triangleq \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$



parallel

$$h\left(\frac{1}{m}\right) \triangleq \begin{pmatrix} 1 & 0 \\ \frac{1}{m} & 1 \end{pmatrix}$$

There are others:

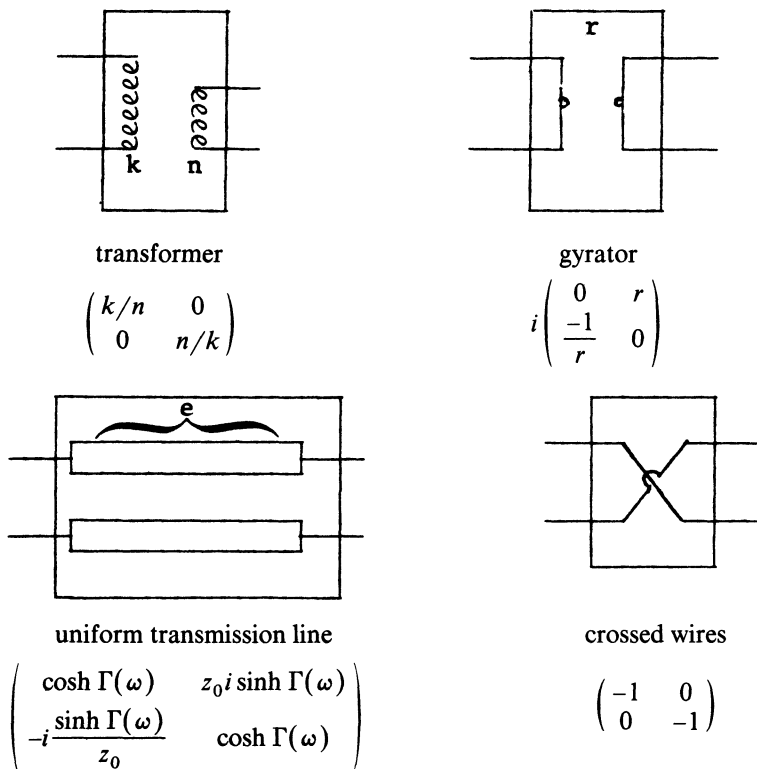


FIGURE 2.5

Note that the chain matrices of everything but the gyrator are actually in $SL_2(R)$ when the circuits are lossless.

Though a transistor has three wires coming out it is (almost) always regarded as a 2-port (see Figure 2.6). Unfortunately it is nonlinear and so for use in an amplifier it is embedded in a circuit which has the effect of linearizing its nonlinear chain function near a certain point. The resulting linear chain matrix can be a fairly general element of $GL_2(C)$. This is what is used in studying the effect of the amplifier on small signals (like the ones you want to amplify).

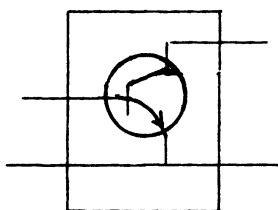
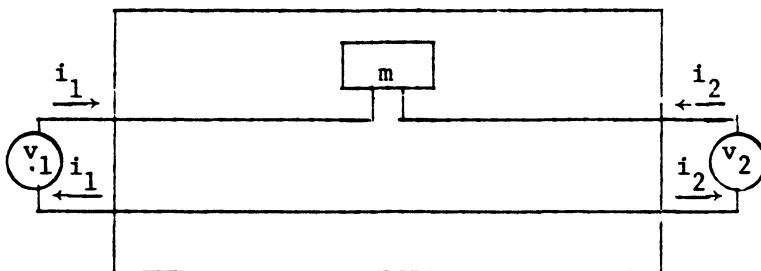


FIGURE 2.6 A transistor as a 2-port.

Now we work some trivial exercises to insure that the reader has the basic ideas locked in.

(A) Let's compute the chain matrix for the 'series' two port in Figure 2.5. One connects a voltage generator to each pair of wires thereby forcing voltages v_1 and v_2 across them as in the figure



Expanded Figure 2.5 'series'.

This sets up a flow of current in the circuit. The relationship between the current and voltage is easily seen to be

$$i_1 = -i_2, \quad v_1 = v_2 + mi_1.$$

That is,

$$\begin{pmatrix} v_1 \\ i_1 \end{pmatrix} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_2 \\ -i_2 \end{pmatrix}.$$

So the chain matrix Ψ is the one indicated.

(B) *Exercise in Lie groups.* At a given frequency ω is it possible to build a lossless two port whose chain matrix is any g in $SL_2(\mathbb{R})$ using only

(1) a ladder of inductors and capacitors (Figure 2.7)

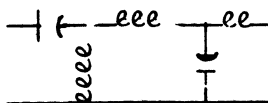


FIGURE 2.7

(2) a capacitor or inductor in series with a transformer and a shunt capacitor or inductor (Figure 2.8).

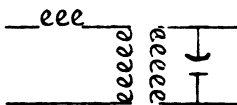


FIGURE 2.8

SOLUTION. (1) Yes, by elementary algebra $SL_2 = MNMN$. This actually is a standard SL_2 lemma.

(2) The Bruhat decomposition says MAN is dense in SL_2 so one gets most of them this way; to get them all you also need $M \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} AN$. For the less scholarly simple algebra will do.

Many manipulations engineers commonly do in this context are surprisingly like what one would expect on purely mathematical grounds. A compelling illustration is the special collection of electrical engineering programs which Hewlett Packard puts on magnetic cards and sells for use with its HP67-97 hand calculators. So instead of inflicting a few pages of something theoretical like a junior level engineering text on the reader I will summarize the instruction manual for H.P.'s market tested collection of programs (called E.E.Pacl). This gives an interesting perspective.

Several programs amount to automating the solution of some typical homework exercise in an introductory complex variables course.

PROGRAM 1. You select any number of elements of $SL_2(\mathbb{C})$ of the form $e(a_j)$ or $h(b_k)$ with a_j and b_k real or positive imaginary numbers. The program will multiply them together for you in any order which you choose. This allows you to evaluate how a ladder circuit will behave at a fixed frequency. For cultural interest the instruction manual's description is reproduced in Figure 2.9.

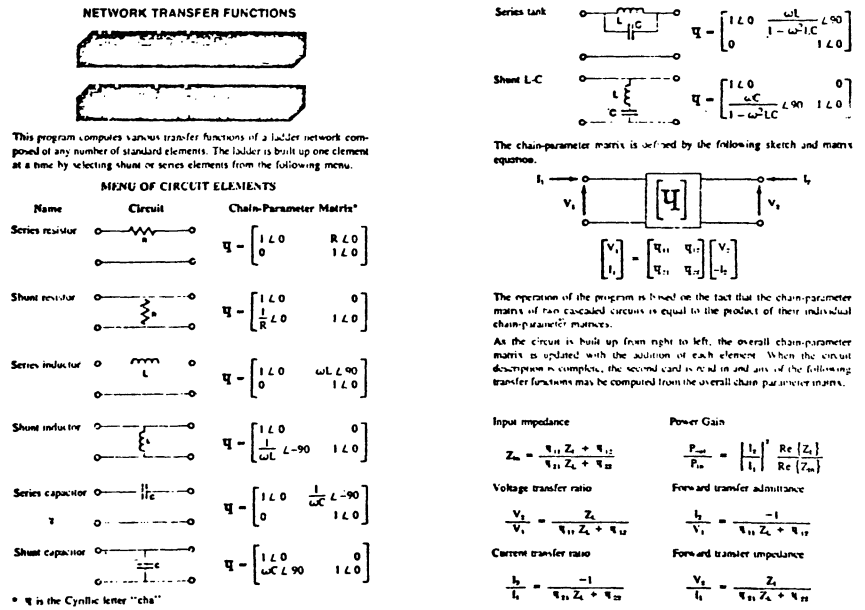


FIGURE 2.9

PROGRAM 2. Given m_1 and m_2 in the U.H.P. the program computes a g in $SL_2(\mathbb{R})$ so that $\mathcal{G}_g(m_1) = m_2$. This tells one how to build an energy conserving circuit g which transfers maximum power from source with internal impedance m_1 to load \bar{m}_2 .

PROGRAM 16. Computes the set of all points at a given Poincaré distance from a given point. This is set in the unit disk rather than the half plane. This is used in computing the gain of a transistor amplifier. See Figure 2.10.

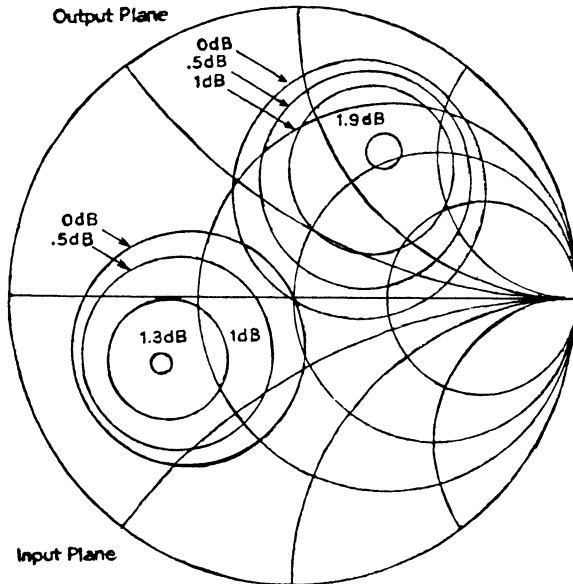


FIGURE 2.10

A couple of programs do parameter conversions. For example, PROGRAM 6. Computes the Cayley transform T' of a 2×2 matrix T , that is

$$T' = (I + T)(I - T)^{-1}.$$

Now we get to programs which aren't mathematical exercises of this flavor.

PROGRAMS 13, 14, 15. Compute the g in $SL_2(R)$ (or some entries in it) for a transmission line at high frequency. The line is given in several different ways, e.g. spacing for lines, dielectric constant for (miniaturized) microwave waveguides.

Four programs remain. Two are described in the next section, one does Fourier series and the other analyzes a ladder of resistors.

d. *Circuits driven by a mixture of sine waves: Problems.* The theoretical design problems described in the preceding section were trivial. That is because the objective was to design a circuit which behaved in a desired way at *one fixed* frequency. An example of such circuitry is the power transmission system to your house. It operates and is designed to operate at 60 cycles/sec. If the frequency suddenly shifted to 120 cycles, the power company's circuitry would behave peculiarly, since the characteristics of some of its components would change and the design doesn't allow for that.

Many circuits don't have the luxury of operating at one fixed frequency. For example, a telephone must faithfully process various mixtures of frequencies ranging between 300 and 3500 cycles/sec. Obviously, this circuitry is designed to meet specifications which are given at every ω between 300 and 3500 cycles/sec. Many mathematical problems in the design of such 'broadband' circuits are extremely hard. In this subsection we describe the setup and basic power transfer problems.

Note that all circuits listed in §2(c) except for the transmission line have impedance and chain functions which are rational in ω . *In this article we restrict attention to rational functions*; thus we cannot include transmission lines in our study. This restriction is more technical than real, because there is a very general class of functions called pseudomeromorphic which are appropriate for transmission lines and which behave mathematically about like the rational functions for the purposes in this article (see footnote 6, §3). Their use is well understood in the context of theoretical engineering (see [A, deW, D-H, F, H6]) and attention to them here would be a distraction with little benefit.

We are now interested in impedances and chain matrices as functions (of ω). The impedance function m of a physically interesting 1-port has the striking property that it has *no poles in the U.H.P.* The reason is that if m has a pole at $x_0 + iy_0$ in the U.H.P. then when one feeds signal $e^{-\gamma_0 t} \sin x_0 t$ into the circuit the circuit puts out an infinite signal; in practice the circuit burns out. Such a circuit is called *unstable* and unstable circuits are anathema. The classical passive circuits are all stable.

Passivity forces m to lie in the function space analog of the upper half plane

$$\mathfrak{P}\mathcal{C} = \{f \text{ rational: } \text{Im } f(\omega) \geq 0 \text{ for } \omega \text{ real}\}$$

and stability forces m to lie in a subset

$$\mathfrak{P}\mathcal{C}^+ = \{f \in \mathfrak{P}\mathcal{C}: f \text{ continues analytically to the U.H.P.} \\ \text{and } \text{Im } f(p) \geq 0 \text{ for } p \in \text{U.H.P.}\}$$

of $\mathfrak{P}\mathcal{C}$. Lossless two ports obviously correspond to a certain subset \mathcal{L} of $\Pi \times \text{SL}_2$ over the field of rational functions with real values on the real axis. The physically obvious property that cascading a two-port in \mathcal{L} with any passive one port yields a passive one-port corresponds to the mathematical statement

$$(2.5) \quad \mathcal{G}_g \text{ maps } \mathfrak{P}\mathcal{C}^+ \text{ into } \mathfrak{P}\mathcal{C}^+ .$$

This property determines \mathcal{L} projectively, but rather than completing the precise mathematical description of \mathcal{L} we change coordinates and describe a transform of it.

Instead of continuing the discussion with this SL_2 formalism we shall switch now to the scattering formalism. While the SL_2 ‘impedance formalism’ is more appealing to the physical intuition and so might be more appropriate here, the reader should be able to mentally convert back and forth and switching now to the conventions used through the rest of the paper improves the transition to entirely mathematical considerations. We shall make two transforms. The first has been described. It is the map $m \rightarrow (1 - im)(1 + im)^{-1} = s$ which takes the R.H.P. to Δ . The second is on the variable ω ; it is $i\omega \rightarrow (1 - i\omega)/(1 + i\omega) = e^{i\theta}$ which takes the R.H.P. to the disk. We refer to θ somewhat loosely as the frequency variable even though it is a transform of the true frequency variable ω . For example a function $m(\omega)$ in $\mathfrak{P}\mathcal{C}$ transforms to $s(e^{i\theta})$,

$$s(e^{i\theta}) = (1 + im(\omega))(1 - im(\omega))^{-1},$$

a function in $\mathfrak{R}\overline{\mathfrak{B}}L^\infty$ and vice versa. Also $\mathfrak{P}\mathcal{C}^+$ becomes $\mathfrak{R}\overline{\mathfrak{B}}H^\infty$ under the transformation and the key property (2.4) of lossless two ports becomes

$$\mathcal{G}_g: \mathfrak{R}\overline{\mathfrak{B}}H^\infty \rightarrow \mathfrak{R}\overline{\mathfrak{B}}H^\infty.$$

Now a g in the semigroup $\mathfrak{R}U^+(1, 1)$ of $\mathfrak{R}U(1, 1)$ defined in the introduction has this property (and g with the property must equal φh for some φ in \mathfrak{R} and h in $\mathfrak{R}U^+(1, 1)$). In fact $\mathfrak{R}U^+(1, 1)$ equals the (scattering) chain matrices of the lossless 2-port networks. The proof is a straightforward algebra exercise.

The extension of all this to n -ports is easy and we list the outcome since it sets the conventions we use through the rest of the paper.

$$\begin{aligned} \text{passive } n\text{-port} &\leftrightarrow S \in \mathfrak{R}\overline{\mathfrak{B}}H^\infty(M_n), \\ \text{lossless 2-port cascade on } S &\leftrightarrow \mathcal{G}_g(S) \text{ with } g \in \mathfrak{R}U^+(n, n), \\ \text{passive 2-port cascade} &\leftrightarrow \mathcal{G}_g(S) \text{ with } g \in \mathfrak{R}\mathcal{C}^+(n, n). \end{aligned}$$

In this article we are interested primarily in power transfer problems and now we list some basic ones. Suppose we are given an n -port passive circuit with scattering matrix function $S(e^{i\theta})$ and a k -port $Q(e^{i\theta})$. The problem is to find a lossless $(k + n)$ -port with matrix $g(e^{i\theta})$ so that the circuit in Figure 2.11 delivers a certain amount of power to the load S at each frequency $e^{i\theta}$.

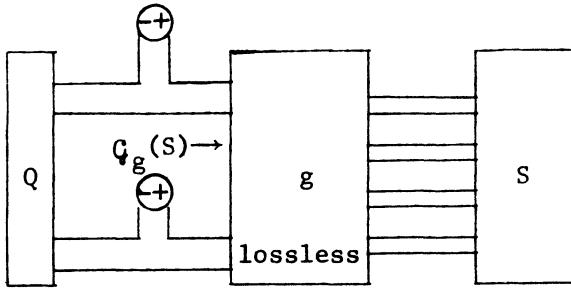


FIGURE 2.11

The discussion so far has prepared us to state the problem mathematically when $k = n = 1$ and this case fully demonstrates the problem. So g is in $\mathfrak{R}U(1, 1)$.

The cascade of g with S has scattering matrix $\mathcal{G}_{g(e^{i\theta})}(S(e^{i\theta}))$ and the power consumed by this circuit at a frequency $e^{i\theta}$ equals the power consumed by S at $e^{i\theta}$, since g is lossless. The mismatch between the source Q and $\mathcal{G}_g(S)$ at frequency $e^{i\theta}$ is

$$\delta(\overline{Q(e^{i\theta})}, \mathcal{G}_{g(e^{i\theta})}S(e^{i\theta})).$$

Thus as we sweep through all g the possible power mismatches sweep through all

$$\delta(\overline{Q(e^{i\theta})}, \Sigma(e^{i\theta}))$$

where $\Sigma \in O_S^+$, the orbit of S under the semigroup $\mathcal{R}U^+(1, 1)$. Two problems are

(i) (BROADBAND IMPEDANCE MATCHING). *What is the largest power transferable uniformly over all frequencies? Which g attains it? That is minimize*

$$\text{minimize}_{\Sigma \in O_S^+} \left\| \delta \left(\overline{Q(e^{i\theta})}, \Sigma(e^{i\theta}) \right) \right\|_{L^\infty}.$$

(ii) *Which power transfer ratios $G(e^{i\theta})$ are obtainable from this kind of circuit?* There are several variations on (ii) which are fundamental problems of classical filter design called *insertion loss* problems:

$$\text{Restrict } Q(e^{i\theta}) = S(e^{i\theta}) = 0.$$

(iii) Restrict g to have rational entries of prescribed (low) order. In other words there is a crude upper bound on the complexity of the lossless coupling circuit. Given a frequency interval $a \leq \theta \leq b$ and a function Γ on it find a g which gives power transfer ratio as close to Γ as possible.

(iv) Suppose you can only build g by cascading circuits from a (small) subset $\mathcal{U} \subset \mathcal{R}U^+(1, 1)$. For example, the technology or cost restricts the parts you can use. Thus g must equal a product $g_1 g_2 \cdots g_k$ of g_j 's from \mathcal{U} ; let $\mathcal{S}_{\mathcal{U}}$ denote the semigroup generated by \mathcal{U} . Which Γ are obtainable with a g in $\mathcal{S}_{\mathcal{U}}$? Given an obtainable Γ write the corresponding g as a product of generators.

The mathematical problems stated in the introduction for passive circuits derive directly from these physical problems. Actually, the mathematical version of part (iii) of the insertion loss problem which engineers attack is a compromise of what we have stated.

There is another broad class of power transfer problems. One is, given an amplifying device (e.g. a transistor) use it to build an amplifier with prescribed gain. The amplifier consists of the given device together with any passive circuitry the designer chooses to build. The problem is to design the best amplifier. There are several different types of amplifying devices and several basic configurations for the total amplifiers. A general model for the most common (e.g. a transistor) amplifier is

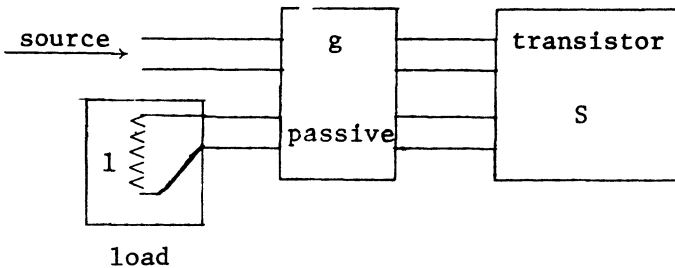


FIGURE 2.12

One is given 2×2 scattering matrix S , and wishes to find the 4×4 chain matrix g of the passive circuit which maximizes the gain. By gain one means

the power transferred from source to load. At 'frequency θ ' the gain for the amplifier in Figure 2.12 is

$$|\mathcal{G}_g(S)_{21}(e^{i\theta})|^2.$$

A worst case analysis requires one to make the inf of this expression over all frequencies θ as big as possible. If we are permitted to use any lossless g this is precisely problem (1.5) of the introduction. One must use an extension of §5(b) to show that one gets the same answer for passive g as lossless g . A further consideration is that typically one wants the gain to be constant over the frequency range where you operate the circuit.

A more restricted but common form of transistor amplifier looks like the one in Figure 2.12 except that the equalizer circuit g is decoupled (see Figure 2.15). That is, g does not mix input and output channels; the source feeds through g directly into the transistor and the transistor output feeds solely to the amplifier output. The gain of such an amplifier (see Figure 100 [HP2] for complete derivation) is

$$(2.6) \quad G = |S_{21}|^2 \frac{1 - |\Gamma_1|^2}{|1 - S_{11}\Gamma_1|^2} \frac{1 - |\Gamma_2|^2}{|1 - S_{22}\Gamma_2|^2}$$

whenever the transistor's S -matrix has $S_{12} = 0$ (i.e., is unilateral). Here Γ_1 and Γ_2 are functions determined by g and range through $\mathfrak{B}H^\infty$. The gain optimization problem is to maximize (2.6) over all Γ_1 and Γ_2 in $\mathfrak{B}H^\infty$.

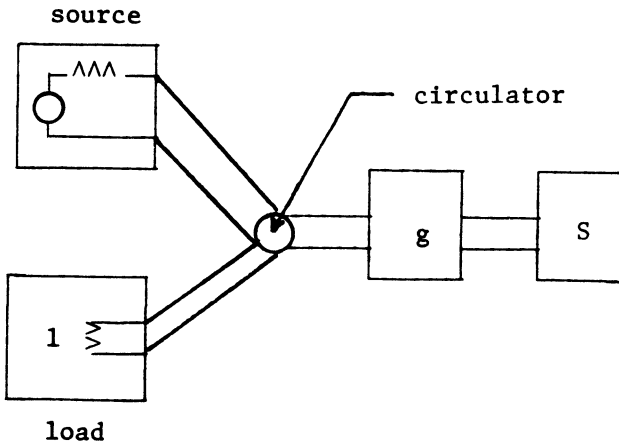


FIGURE 2.13

Another type of amplifier is the reflection type. One is given a one port S and wishes to find a passive g so that there is a large flat gain from source to load in Figure 2.13. The gain of this amplifier at frequency θ is $|\mathcal{G}_g(S)(e^{i\theta})|^2$. At the worst frequency it has gain

$$(2.7) \quad \inf_{\theta} |\mathcal{G}_g(S)(e^{i\theta})|^2.$$

The best gain obtainable from an amplifier based on S is the maximum of

$$\inf_{\theta} \|\Sigma(e^{i\theta})\|$$

over Σ in CO_S^+ . This equals the maximum over O_S^+ . Guaranteeing stability of the amplifier forces us to restrict Σ to $O_S^+ \cap H^\infty$ — actually a slightly smaller set is necessary. We must emphasize that in designing an amplifier stability is an overriding consideration. This imposes a strong constraint on the optimization problems just stated.

As we said in the introduction these amplifier optimization problems give rise to generalizations of the Poincaré distance problems solved in §4. In fact (2.6) is a product of terms which have the form of a power mismatch (2.4) and optimizing the broadband gain of this amplifier amounts to the Poincaré distance problem solved in §4. Also (2.7) without the stability constraint is equivalent to finding the Poincaré distance of the function S^{-1} to $\mathbb{B}H^\infty$. So it is solved. The connections between optimization over orbits and Poincaré distance problems are explained in §6. Also explicit results appear there.

We conclude the section by returning to our favorite collection of examples, the H.P.E.E. Pac I. Programs 16, 17, 18 can be used in design of a transistor amplifier of the form in Figure 2.12 where the channels of the coupling circuit don't interact (see Figure 2.14).



FIGURE 2.14

Fix g_1 ; the programs tell you which g_2 's give an amplifier with prescribed (or maximum) gain; conversely fix g_2 and find g_1 's. The reason there are three programs is that they operate under various assumptions on the transistor M and also they do stability calculations. Program 4 computes voltage and current (rather than power) gain of a given transistor. Program 3 is not of interest here; it concerns biasing of a transistor.

Program 9 addresses the insertion loss problem of type (b). The set \mathcal{R} of allowable elements consists of chain matrices for

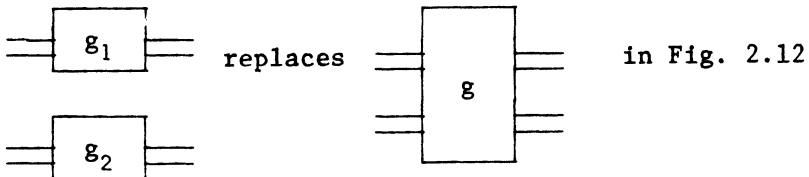


FIGURE 2.15

To use the program feed an interval $[a, b]$ and (roughly) the length of product you are willing to allow. The program computes a g in $\mathcal{S}_{\mathcal{R}}$ which is a product of

n generators so that the power delivered to the load resistor in Figure 2.14 (with a purely resistive source) closely approximates the characteristic function $\chi_{[a, b]}$ of $[a, b]$. Program 10 evaluates the approximating function obtained in Program 9 on a grid of frequencies so the function can be graphed.

That concludes a description of programs in the E.E.Pac for the HP67-97 calculator. All programs have been mentioned here so one sees that this brief description of old-fashioned circuit design is more complete than one might expect. We have indeed explained what is in about $\frac{2}{3}$ of them. The HP41C is larger. The E.E.Pac for it does calculations like the one in Program 1 above for a broad range of frequencies (not just one). Also it analyzes the behavior of more general classes of circuits than mere ladders.

A more serious description of engineering literature on these problems appears in sections where the problems are treated. We wish to thank the Hewlett-Packard Corporation for allowing us to reproduce Figure 2.9 and Figure 2.10.

3. A non-Euclidean Beurling-Lax-Halmos theorem, interpolation and supremum norm approximation. The theory of interpolation could be viewed as part of the study of $\mathcal{R}U^+(m, n)$. This certainly is not how it is treated classically but it seems fair to say that this is the most powerful approach to the subject. Although it is not apparent from the outset interpolation theory essentially amounts to a Beurling-Lax-Halmos theory based on the Lie group $U(m, n)$. From the mathematical viewpoint the goal of this section is to present some basic facts about the Lie group $\mathcal{R}U(m, n)$ and the subsemigroup $\mathcal{R}U^+(m, n)$.

From the physical viewpoint in this and in the next three sections we shall be building a mathematical machine which applies to the design of circuits with prescribed gains. While classical circuits and gain studies were our original motivation subsequently interpolation found its most widespread use in the realm of digital signal processing (see the works of DeWilde, T. Kailath and coworkers). A forthcoming special issue of *Circuits, systems and signal processing* with articles from a conference in Leuven (1981) on rational approximation should emphasize this subject and applications of this type of mathematics to the 'state space reduction' problem of system theory. Recently interpolation techniques have been introduced into control theory by G. Zames and B. Francis [Z-F]. Another application of interpolation by J. Evans and myself concerns a test for pulmonary function [E-H].

The agenda for §3 is

3(a) indicates how the 'non-Euclidean' Beurling-Lax-Halmos theory serves to parameterize all solutions to an interpolation problem.

3(b) surveys classical interpolation theory.

3(c) describes the correspondence between interpolation and supremum norm approximation.

3(d) presents an algebra trick.

3(e) proves one side of Theorem 3.1.

3(f) sketches the rudiments of an $H^2(V)$ theory for V a vector space with signed bilinear form; proves the other side.

3(g) lists some further problems.

A casual reader might stop after §3(c) although §3(f) presents a topic of fairly general interest.

a. *Non-Euclidean Beurling-Lax-Halmos Theorem.* The classical Beurling-Lax-Halmos theorem characterizes subspaces of $L^2(\mathbb{C}^m)$ of the form

$$VH^2(\mathbb{C}^m)$$

where V is in $\mathcal{R}U(m, 0)$ (actually it handles a larger⁶ field ψ as well as \mathcal{R}). It is of crucial importance to us that this generalizes to g in $\mathcal{R}U(m, n)$ and allows one to characterize subspaces

$$gH^2(\mathbb{C}^{m+n})$$

of $L^2(\mathbb{C}^{m+n})$. Actually the key case for us is $m = n$. The linear fractional version of this amounts to characterizing subsets of $\mathcal{B}L^\infty(M_n)$ of the form

$$\mathcal{G}_g(\mathcal{B}H^\infty(M_n)).$$

This subsection takes the linear fractional viewpoint and subsection (f) presents the subspace (Grassmannian) viewpoint.

It turns out that the range of \mathcal{G}_g acting on $\mathcal{B}H^\infty$ always equals a set of functions which satisfy some interpolation conditions. The simplest type of interpolating condition is to require that a function F take given values w_j at given points z_j in the disk. A good convention is to denote the N -tuple of complex numbers $z = \{z_1, z_2, \dots, z_N\}$ by \mathbf{z} , the N -tuple $w = \{w_1, \dots, w_N\}$ by \mathbf{w} , and then to define $\mathcal{T}(\mathbf{z}, \mathbf{w})$ to be the set of functions F meromorphic on the open disk which satisfy

$$F(z_j) = w_j.$$

Here we have assumed that all $|z_j| < 1$. That is, \mathcal{T} is a (big) class of functions meeting a certain ‘interpolation’ constraint.

An illustrative case of the general theorem we eventually derive is

SIMPLEST THEOREM. *For a generic g in $\mathcal{R}U^+(1, 1)$ there exists \mathbf{z}, \mathbf{w} so that*

$$\mathcal{G}_g(\mathcal{R}\mathcal{B}H^\infty) = \mathcal{T}(\mathbf{z}, \mathbf{w}) \cap \mathcal{R}\mathcal{B}H^\infty.$$

Conversely, given any \mathbf{z}, \mathbf{w} for which the set on the right is nonempty there exists a g in $\mathcal{R}U^+(1, 1)$ making the equality hold.

Thus the range of \mathcal{G}_g is the set of all solutions in $\mathcal{R}\mathcal{B}H^\infty$ to a particular interpolation problem. This ultimately allows us to reduce engineering problems to interpolation problems of Nevanlinna-Pick and much more general type. Also it gives a simple parameterization of all solutions to a given interpolation problem.

To state the general result we need a fancier notation for $\mathcal{T}(\mathbf{z}, \mathbf{w}) \cap H^\infty$.

⁶The field ψ of pseudomeromorphic functions is defined to consist of functions F on Π which have meromorphic continuations $F_<$ on $\{|z| < 1\}$ and $F_>$ on $\{|z| > 1\}$. The continuations must have radial limits which agree with F almost everywhere and $F_>$ must be the quotient of two functions bounded and analytic in $\{|z| > 1\}$. Any V in $H^\infty(M_n)$ unitary a.e. on Π has all entries in ψ ; this is because the ‘Schwartz reflection’ $V^*(\frac{1}{z})^{-1}$ of V can be used to continue the entries of V outside the disk. A similar argument applies to $U(m, n)$ and for this reason ψ is the ‘mathematically natural’ field for our study rather than \mathcal{R} . However, we use \mathcal{R} for pedagogical convenience. Functions in ψ occur physically, for example, in connection with transmission lines (Figure 2.5).

Though it looks overly elaborate at first it gives a very condensed formulation of high multiplicity interpolation for matrix valued functions. Let Φ denote the finite Blaschke product whose zeros are precisely z_j and let G be any function in $\mathfrak{T}(\mathbf{z}, \mathbf{w}) \cap H^\infty$. Then

$$G + \Phi H^\infty = \{G + \Phi h : h \in H^\infty\}$$

obviously equals $\mathfrak{T}(\mathbf{z}, \mathbf{w}) \cap H^\infty$.

This last description extends easily to very general matrix valued functions. Define $H_l^\infty(M_n) = \{F \in L^\infty(M_n) : F \text{ extends to a function analytic on the disk except for } l \text{ poles}\}$. Let Φ, Ψ be inner functions in $H^\infty(M_n)$ and let $G \in H^\infty(M_n)$. Define

$$\mathfrak{T}_{\Phi, \Psi, l}(G) = \{G + \Phi H \Psi : H \in H_l^\infty(M_n)\}.$$

Clearly when $n = 1$ the set $\mathfrak{T}_{\Phi, \Psi, l}(G)$ is of the form $\mathfrak{T}(\mathbf{z}, \mathbf{w}) \cap H_l^\infty$. Also these sets have a straightforward though messy description as an interpolation set. For example, suppose $\Phi = I$ and Ψ has zeroes $\Psi(z_j)x_j = 0$ of multiplicity at most 1; set $w_j = G(z_j)x_j$. For convenience set

$$(3.1) \quad \mathfrak{T}(\mathbf{z}, \mathbf{x}, \mathbf{w}) = \{F \in \mathfrak{R}L^\infty(M_n) : F(z_j)x_j = w_j\}.$$

Then clearly

$$\mathfrak{R}\mathfrak{T}_{\Phi, \Psi, 0}(G) = \mathfrak{T}(\mathbf{z}, \mathbf{x}, \mathbf{w}) \cap H^\infty(M_n).$$

When $\Psi = I$ and $\Phi(z_j)^*y_j = 0$, set $G(z_j)^*y_j = w_j$. Then the appropriate interpolating set is

$$\mathfrak{T}^*(\mathbf{z}, \mathbf{y}, \mathbf{w}) = \{F \in \mathfrak{R}L^\infty(M_n) : F(z_j)^*y_j = w_j\}$$

in that

$$(3.1^*) \quad \mathfrak{R}\mathfrak{T}_{\Phi, \Psi, 0}(G) = \mathfrak{T}^*(\mathbf{z}, \mathbf{x}, \mathbf{w}) \cap H^\infty(M_n).$$

At last we can begin to describe the range of \mathfrak{G}_g . It turns out that it consists of all functions satisfying a certain interpolation condition up to a strong type of equivalence which preserves magnitudes. To be precise we say that two matrix functions M and N have *equivalent magnitudes* if and only if there exist rational matrix functions Ω and Ξ with unitary values on Π so that $\Omega M \Xi = N$. Frequently we shall just call the functions M and N *equivalent*. The main theorem is

THEOREM 3.1. *Given g in $\mathfrak{R}U(n, n)$ there exist functions G, Φ, Ψ in $\mathfrak{R}H^\infty(M_n)$ with Φ, Ψ inner and an integer $l \geq 0$ such that*

$$\mathfrak{G}_g(\mathfrak{R}\mathfrak{B}H^\infty(M_n)) \text{ and } \mathfrak{T}_{\Phi, \Psi, l}(G) \cap \mathfrak{R}\mathfrak{B}L^\infty(M_n)$$

have equivalent magnitudes. Conversely, given such G, Φ, Ψ let l be the smallest integer for which this set is nonempty. Then there is a g in $\mathfrak{R}U(n, n)$ for which this relationship holds. Note when $n = 1$ we could have taken either $\Phi = 1$ or $\Psi = 1$ throughout.

The theorem is actually concrete in that all functions and l can be computed explicitly. This emerges in the proof and is essential in getting the main explicit

results of the paper; however, it doesn't matter much to conceptual understanding. The theorem at a greater level of generality can be found in a slightly different form in [B-H4] and [H1] for the forward and converse directions respectively. For $l = 0$ it is in [A-A-K3], also [A-C-F]; for $n = 1$ it is in [A-A-K2]. We sketch a proof in §3(e)(f).

b. *Classical interpolation.* An item left dangling in Theorem 3.1 was how one computes the integer l . This is the subject of interpolation theory. An early interpolation theorem was published by Pick in 1916. It says

THEOREM 3.2 (PICK). *There is a function F in $\mathfrak{J}(\mathbf{z}, \mathbf{w}) \cap \mathfrak{R}\overline{\mathfrak{B}}H^\infty$ if and only if the matrix*

$$\Lambda_{\mathbf{z}}(\mathbf{w}) = \left\{ \frac{1 - w_j \bar{w}_k}{1 - z_j \bar{z}_k} \right\}_{j,k}^N$$

is nonnegative definite. The proof is constructive.

The relationship of this theorem to Theorem 3.1 is that for scalar G , Ψ and $\Phi = 1$ the Theorem 3.2 gives an elegant practical test for determining if $l = 0$. Instead of moving immediately to the general result for arbitrary l we proceed historically.

Carathéodory-Fejer gave a matrix positive definiteness test to determine if a function F in $\mathfrak{R}\overline{\mathfrak{B}}H^\infty$ exists whose derivatives $F^{(r)}(0)$ at the origin equal a given sequence a_0, a_1, \dots, a_n of complex numbers. The general problem of finding an F in $\mathfrak{R}\overline{\mathfrak{B}}H^\infty$ satisfying $F^{(r)}(z_j) = w_j^r$ for $r = 1, \dots, k_j$ can be approached with the classical method of divided differences but wasn't really solved until the advent of the commutant lifting theory of Ando-Nagy-Foiaş-Sarason; see [N-F, Fo]. Sarason [S] originated this operator theoretic approach to interpolation by which it was a straightforward exercise to derive a 'Pick matrix' for the general interpolation problem. This Pick matrix is positive if and only if an interpolating function exists. The explicit computation of this Pick matrix was carried out independently by Rosenblum and Rovnyak [R-R] and the author, §6 [H2].

Takagi studied Carathéodory-Fejer type interpolation not with H^∞ functions but with functions having at most l poles in the disk. Löwner in 1934 analyzed interpolation at z_j satisfying $|z_j| = 1$ rather than $|z_j| < 1$ as the others had.

Interpolation with matrix valued functions was successfully treated first by Nagy and Koranyi [N-K] in 1958. They found the natural matrix generalization of Pick interpolation. Their methods also have substantial success with Löwner interpolation. The commutant lifting approach handles interpolation to any order with $H^\infty(M_n)$ functions readily (but not of Löwner or Takagi type). A forthcoming paper of J. Ball gives the most general result. Ball shows how to write down a Pick matrix for arbitrary order Pick, Löwner, Takagi, etc. Interpolation with matrix valued functions. In particular, given functions Φ, Ψ , and G define a matrix

$$(3.2) \quad \Lambda_{\Phi, \Psi}(G)$$

according to a recipe in [B]. Then

THEOREM 3.3. (BALL). *There is a function F in $\mathfrak{T}_{\Phi, \Psi, l}(G) \cap \mathfrak{R} \overline{\mathfrak{B}} L^\infty(M_n)$ if and only if $\Lambda_{\Phi, \Psi}(G)$ has at most l negative eigenvalues. Moreover, F can be chosen to be an inner function.*

We shall refer to Λ as a ‘Pick’ matrix for the interpolating problem: As an example the ‘Pick matrix’ for the interpolating set $\mathfrak{T}^*(\mathbf{z}, \mathbf{y}, \mathbf{w})$ corresponding to $\mathfrak{T}_{\Phi, l}(G)$ is

$$(3.2a) \quad \Lambda_{\Phi, l}(G) = \left\{ \frac{(y_j, y_k) - (w_j, w_k)}{1 - z_j \bar{z}_k} \right\}_{j, k}^N.$$

Recall from (3.1*) that $G(z_j)^* y_1 = w_j$. Theorem 3.3 completely settles the problem of determining l in Theorem 3.1. We remark that Ball’s proof is an extension of the Nagy-Koranyi method for $l = 0$.

Note that a selfadjoint matrix Λ has l negative eigenvalues if and only if $M^* \Lambda M$ has l negative eigenvalues for any invertible M . So there are many matrices you might use in Theorem 3.3. For example, another common recipe for building a matrix Λ gives one of the form $I - \mathfrak{H} \mathfrak{H}^*$ where \mathfrak{H} is a certain Hankel matrix. This will soon be discussed in more detail.

More refined results were given by Adamajan-Arov-Krein [A-A-K1, 2] who gave a representation for all solutions to a given interpolation problem (when $l = 0$ or $n = 1$) of the form in Theorem 3.1. Many of these results were obtained independently by D. Clark [C1] when $n = 1$. One side of Theorem 3.1 is a special case of this. They also were the first to write down an explicit matrix test for existence of solutions to the full $n \geq 1$ interpolation problem when $l = 0$. They used a Λ obtained from a Hankel recipe. A highly developed study of these representations when $l = 0$ is [A-C-F]. When $n \geq 1$, $l \geq 0$ the existence of the representation as well as a new proof of Theorem 3.3 (for all but the Löwner case) is given in [B-H4] which will be described in §5(f). Further work on interpolation by Russian functional analysts is [K-P; Fd1, 2; IV1, 2, 3; Nu1, 2]. To keep our explanation simple we have kept all functions rational. This constraint is not essential in most of the results we have cited here.

c. *The sup norm distance to H^∞ .* There is a different viewpoint to interpolation which we shall also use heavily in this paper. It is very good for practical engineering treatment of some problems as we shall see in §4, but does not seem to generalize well to the harder problems in §5. A theorem (due to Nehari in the scalar case when $l = 0$) says

THEOREM 3.4. *Given F in $L^\infty(M_n)$ the L^∞ norm distance of F to $H_F^\infty(M_n)$*

$$\inf_{H \in H_F^\infty(M_n)} \|F - H\|_{L^\infty(M_n)}$$

equals λ , the l th from largest eigenvalue of $(\mathfrak{H}_F \mathfrak{H}_F^)^{1/2}$. The minimum is attained*

when F is rational. Here any continuous spectrum is considered to have ∞ multiplicity and \mathfrak{H}_F means Hankel operator⁷ with generating function F .

THEOREM 3.4'. For F rational, $l = 0$ and $n = 1$ there is a practical algorithm (cf. [Htz]) which uses the eigenvector x for the largest eigenvalue λ_1 of

$$\mathfrak{H}_F \mathfrak{H}_F^* x = \lambda_1^2 x$$

to construct the best approximating H . Also [A-A-K] gives algorithms which hold when $l \neq 0$.

This is equivalent to Theorem 3.3 when F is rational. The idea (essentially due to Nehari) is very simple. In Theorem 3.3 we wanted to find an H of the form

$$(3.3) \quad H = G + \Phi K \Psi$$

where $K \in H_l^\infty(M_n)$ and $\|H\|_{L^\infty(M_n)} \leq 1$. Here G, Φ, Ψ are rational. In other words

$$(3.4) \quad \Phi^{-1} H \Psi^{-1} = F + K$$

where $F = \Phi^{-1} G \Psi^{-1}$ has norm ≤ 1 . So the problem in Theorem 3.3 has a solution H if and only if the distance of the function F to $H_l^\infty(M_n)$ is less than or equal to one. We have seen that the approximation and interpolation problems are equivalent. Now we need some explanation why the answers are equivalent. With some work (which this author has only seen performed in the scalar case) one can surely compute an invertible operator $M: \text{Range } \mathfrak{H}_F \rightarrow \mathbf{C}^\mu$ so that $\Lambda_{\Phi, \Psi}(G) = M(I - \mathfrak{H}_F \mathfrak{H}_F^*) M^*$. Here μ equals the number of interpolating points together with multiplicity. Clearly Λ is positive if and only if $I - \mathfrak{H}_F \mathfrak{H}_F^*$ is positive.

d. *An algebraic rephrasing.* Linear fractional maps obviously can be written in several ways. Here is one which is particularly useful in dealing with $\mathfrak{R}U^+(n, n)$. Presenting it as an entire subsection might be misleading since this was done mainly for convenience. We already mentioned both ways of writing these maps in §2(a), but we need the explicit conversion rule and specific properties of the coefficient matrix; so the purpose here is to list them.

Straightforward algebra allows one to write the linear fractional map $\mathfrak{G}_g(s)$ with $g = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \text{GL}(2n)$ in the form

$$(3.5) \quad \mathfrak{F}_u(s) \stackrel{\Delta}{=} a + bs(1 - ds)^{-1}c.$$

⁷The definition of Hankel matrices goes as follows: Given F on Π denote its j th Fourier coefficient by F_k . Define the Hankel matrix for F to be the infinite matrix

$$\mathfrak{H}_F = \begin{pmatrix} F_{-1} & F_{-2} & F_{-3} \\ F_{-2} & F_{-3} & \dots \\ F_{-3} & \dots & \dots \end{pmatrix}.$$

For F in $L_{M_n}^\infty$ each F_k is of course an $n \times n$ matrix.

Here $u = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If c is invertible, then $\overline{\mathcal{F}}_u = \mathcal{G}_g$ provided

$$(3.6) \quad g = \begin{pmatrix} b - ac^{-1}d & ac^{-1} \\ -c^{-1}d & c^{-1} \end{pmatrix}.$$

Conversely, if γ is invertible $\overline{\mathcal{F}}_u = \mathcal{G}_g$ provided

$$(3.7) \quad u = \begin{pmatrix} \beta\gamma^{-1} & \alpha - \beta\gamma^{-1}\kappa \\ \gamma^{-1} & -\gamma^{-1}\kappa \end{pmatrix}.$$

One can show that g is in $U(n, n)$ [resp. $C(n, n)$] if and only if u is a unitary [resp. $\mathbb{B}M_{2n}$] matrix. This can be used to show that g is in $\mathcal{R}U^+(n, n)$ [resp. $\mathcal{R}C^+(n, n)$] if and only if u is a rational inner function [resp. $\mathcal{R}\mathbb{B}H^\infty(M_{2n})$]. Note c is a uniformly invertible function if and only if $\|\beta\gamma^{-1}\|_{L^\infty} < 1$. Use of the form (3.5) rather than the usual \mathcal{G}_g conventions was developed thoroughly by R. Redheffer (cf. [R]).

e. Proofs—elementary complex variables. The basic structure necessary for the rest of the paper has been presented. In the next two subsections we give some idea of proofs. These subsections can be skipped entirely or read independently because they have a completely different flavor. This subsection proves one side of Theorem 3.1, namely, given a linear fractional map find its range. The proof we give uses elementary complex variables only. The next subsection proves the other side of Theorem 3.1 and Theorem 3.3 and uses a sophisticated (but easy) argument. These two approaches are radically different and from them the reader should get a good perspective. To my tastes the approach in the next section is the most informative and some readers might just as well skip this one.

PROOF OF THEOREM 3.1 (CONVERSE SIDE). This will only be a sketch; a detailed proof is in [H1]. One importance of this proof is that it gives an explicit recipe for constructing G , Φ , and l from a given g in $\mathcal{R}U(1, 1)$ and this is essential to the computational (but not conceptual) part of the paper. To keep matters simple we treat only the $n = 1$ case and refer one to [H1] for the matrix case. Suppose that g is $\mathcal{R}U(1, 1)$. We shall construct an interpolation set $\mathcal{T}_{\Phi, l}(G)$ so that $\mathcal{G}_g(\mathcal{R}\mathbb{B}H^\infty)$ is equivalent to $\mathcal{T}_{\Phi, l}(G) \cap \mathcal{R}\mathbb{B}L^\infty$. We begin the proof by expressing \mathcal{G}_g as $\mathcal{G}_g(S) = (\alpha S + \beta)(\kappa S + \gamma)^{-1}$; then use some algebra (see §3(d)) to rewrite it as

$$\mathcal{G}_g(S) = A + BS(1 - DS)^{-1}C$$

where A, B, C, D are easily expressed via (3.6) and (3.7) in terms of $\alpha, \beta, \kappa, \gamma$. The key is to analyze $\mathcal{N}_0 \stackrel{\Delta}{=} \{S(1 - DS)^{-1} : S \in \mathcal{R}\mathbb{B}H^\infty\}$ and obtain

LEMMA 3.5. *Factor $D = RV^{-1}$ where R and V are in $\mathcal{R}H^\infty$, coprime, and V is inner. Define D_I to be a function in $\mathcal{R}H^\infty$ with the property that $(D_I - D^{-1})/V^2$ is analytic near the zeroes of V in Δ (that is, near the poles of D in the disk). Let l denote the number of poles of D in Δ ; then*

$$\mathcal{N}_0 = \{-D_I + V^2\mathcal{R}H_I^\infty\} \cap \{S(1 - DS)^{-1} : S \in \mathcal{R}\mathbb{B}L^\infty\}.$$

PROOF OF LEMMA. Suppose $M = S(1 - DS)^{-1}$ for S in $\mathfrak{B}\mathfrak{R}H^\infty$. Let $\#$ poles F denote the number of poles of F inside the disk. Since $|DS| < 1$, the winding number of $1 - DS$ is 0. Thus

$$\# \text{ poles}(1 - DS)^{-1} \leq \# \text{ poles}(1 - DS) \leq \# \text{ poles } D = l$$

and we have that $M \in H_l^\infty$.

Moreover, simple algebra gives

$$D^{-1} + M = V^2 R^{-1}(V - RS)^{-1} \stackrel{\Delta}{=} V^2 G.$$

Here G is a function which is analytic on the zero set τ of V (poles of D) provided S does not vanish on τ . For such S the definition of D_l gives $M = -D_l + V^2 G_l$ where G_l is analytic on τ . Thus the poles in Δ of

$$G_1 = V^{-2}(M + D_l)$$

are the same as those of M . So $G_1 \in \mathfrak{R}H_l^\infty$ and we have proved

$$\mathfrak{N}_\infty \stackrel{\Delta}{=} \{S(1 - DS)^{-1} : S \in \mathfrak{R}\mathfrak{B}H^\infty \text{ and } S(z_0) \neq 0 \text{ if } z_0 \in \tau\}$$

is contained in the set given in the conclusion of the lemma. Arbitrary rational functions in $\mathfrak{R}\mathfrak{B}H^\infty$ can be approximated uniformly by ones which never vanish on τ from which one can obtain that \mathfrak{N}_0 is contained in the set required by the lemma. \square

Now we apply the lemma and actually obtain a recipe for G and Φ .

The recipe is: Define β_1 to be the modulus one (on the circle) function which makes $\beta_1 B C V^2$ an H^∞ function with an H^∞ reciprocal and use the lemma to write

$$\beta_1 \mathfrak{G}_g(\mathfrak{R}\mathfrak{B}H^\infty) = \beta_1 A + \beta_1 B C \mathfrak{N}_0 = \{\beta_1 A - \beta_1 B C D_l + \mathfrak{R}H_l^\infty\} \cap \mathfrak{R}\mathfrak{B}L^\infty.$$

Let Φ be the smallest Blaschke product so that $\Phi\beta_1 A$ is in H^∞ . Then

$$(3.8) \quad \Phi\beta_1 \mathfrak{G}_g(\mathfrak{R}\mathfrak{B}H^\infty) = \{G + \Phi\mathfrak{R}H_l^\infty\} \cap \mathfrak{R}\mathfrak{B}L^\infty$$

where $G = \Phi\beta_1(A - B C D_l)$ is in $\mathfrak{R}H^\infty$.

This proves Theorem 3.1 when $n = 1$. \square

One should note two things from this proof. First it is completely constructive. Secondly, the range of \mathfrak{G}_g can be determined exactly (not just up to equivalence).

f. *A $U(m, n)$ Beurling-Lax-Halmos theory.* We now present some rudiments of $H^2(V)$ theory on a vector space V with a signed bilinear form. We do this with the explicit goal of proving Theorem 3.2 which connects the Pick matrix with existence of solutions to an interpolation problem and Theorem 3.1 which parameterizes all solutions to the interpolation problem. We shall see that all of this fits snugly together within the structure of $\mathfrak{R}U(m, n)$.

I think of the earliest related work as Potapov [P] factorization of functions in $\mathfrak{R}U^+(m, n)$ even though no Beurling-Lax-Halmos theorem is directly involved. Later Davis and Foias [D-F] gave a geometric operator model theory based on a signed bilinear form; this too is related. McEnnis pursued [D-F] and set up several preliminaries to a symplectic Beurling-Lax-Halmos theorem.

On the other hand there is an interpolation component to the subject to which [An, S, N-F, D-M-P, A-C-F] made the heaviest contributions. Amongst theoretical engineers P. DeWilde has expressed a need for a symplectic Beurling-Lax-Halmos theory and several papers have dealt with engineering applications of related matrix factorization [D-K, D-V]. This subsection follows [B-H4].

In the previous subsections we took the linear fractional approach to $\Re U(m, n)$. In this subsection we work directly with matrices on \mathbf{C}^{m+n} . The study of $\Re U(m, n)$ (on \mathbf{C}^{m+n}) is essentially no more difficult than the one for $m = n$, so we take arbitrary m and n throughout.

We begin by fixing a decomposition $\mathbf{C}^{m+n} = K_1 + K_2$ of \mathbf{C}^{m+n} into disjoint spaces of dimension m and n respectively. Write vectors x in \mathbf{C}^{m+n} as pairs $\langle x_1, x_2 \rangle$ where x_j is in K_j ; that is $x = x_1 + x_2$. Then any linear map $M: K_2 \rightarrow K_1$ can be identified with its graph

$$\Gamma(M) \stackrel{\Delta}{=} \{ \langle Mx, x \rangle : x \in K_2 \}$$

and conversely most n -dimensional subspaces Γ are the graph of some operator M called the *angle operator* for Γ (frequently, M, I are called affine coordinates for Γ). In this section we will be using the standard decomposition $\mathbf{C}^m \oplus \mathbf{C}^n$ of \mathbf{C}^{m+n} ; so $K_1 = \mathbf{C}^m$ and $K_2 = \mathbf{C}^n$. Then observe g maps a graph $\Gamma(M)$ to another subspace

$$\begin{aligned} g\Gamma(M) &= \left\{ \left\langle \begin{pmatrix} \alpha & \beta \\ \kappa & \gamma \end{pmatrix} \begin{pmatrix} Mx \\ x \end{pmatrix} \right\rangle : x \in K_2 \right\} \\ (3.9) \quad &= \{ \langle (\alpha M + \beta)x, (\kappa M + \gamma)x \rangle : x \in K_2 \} \\ &= \{ \langle \mathcal{G}_g(M)y, y \rangle : y \in K_2 \} = \Gamma(\mathcal{G}_g(M)). \end{aligned}$$

Thus it is the action of g on n -dimensional subspaces which corresponds to the linear fractional map $\mathcal{G}_g(M)$.

Next we impose on $\mathbf{C}^m + \mathbf{C}^n$ the signed bilinear form

$$[x, y]_{\mathbf{C}^{m+n}} = (x_1, y_1)_{\mathbf{C}^m} - (x_2, y_2)_{\mathbf{C}^n}$$

and recall that $U(m, n)$ is just the set of all matrices which preserve it. The statement that $\Gamma(M)$ is *negative*, that is $[x, x]_{\mathbf{C}^{m+n}} \leq 0$ for all $x \in \Gamma(M)$, says

$$(x_2, x_2)_{\mathbf{C}^n} \geq (Mx_2, Mx_2)_{\mathbf{C}^m}$$

for all $x_2 \in P_{\mathbf{C}^n}\Gamma(M)$. That is $\|M\| \leq 1$. In fact every *maximal* negative space Γ has angle operator M defined on all of \mathbf{C}^n which is a contraction. Naturally a g in $U(m, n)$ maps (maximal) negative subspaces to (maximal) negative subspaces.

Everything we do in this section will be set in the space $L^2(\mathbf{C}^{m+n})$ or its subspace $H^2(\mathbf{C}^{m+n})$. The bilinear forms $[\cdot, \cdot]_{\mathbf{C}^{m+n}}$ and $(\cdot, \cdot)_{\mathbf{C}^{m+n}}$ induce bilinear forms $[\cdot, \cdot]$ and (\cdot, \cdot) on these via

$$[f, g] = \frac{1}{2\pi} \int_0^{2\pi} [f(e^{i\theta}), g(e^{i\theta})]_{\mathbf{C}^{m+n}} d\theta$$

and

$$(f, g) = \frac{1}{2\pi} \int_0^{2\pi} (f(e^{i\theta}), g(e^{i\theta}))_{\mathbf{C}^{m+n}} d\theta$$

for f, g in $L^2(\mathbf{C}^{m+n})$. A common operator on L^2 is a multiplication operator $\mathfrak{M}_F: L^2(\mathbf{C}^w) \rightarrow L^2(\mathbf{C}^w)$ defined for a given $L^\infty(M_{r,w})$ function F by

$$[\mathfrak{M}_F f](e^{i\theta}) = F(e^{i\theta})f(e^{i\theta})$$

for any $f \in L^2(\mathbf{C}^w)$. Here $M_{r,w}$ denotes all $r \times w$ matrices. We denote the $[\cdot, \cdot]$ orthogonal complement of a space \mathfrak{N} by \mathfrak{N}' , that is

$$\mathfrak{N}' = \{x: [x, y] = 0 \text{ for all } y \in \mathfrak{N}\}.$$

A subspace \mathfrak{N} is called *nondegenerate* if $\mathfrak{N} \cap \mathfrak{N}' = 0$.

A *full range subspace* \mathfrak{N} of $H^2(\mathbf{C}^w)$ is one with the property that at some $|z_0| < 1$ we get $\{f(z_0): f \in \mathfrak{N}\} = \mathbf{C}^w$. It is easy to check that if this happens at one z_0 it happens at all but an isolated set of z_0 , see [Hs]. Our main theorem in this subsection is

THEOREM 3.6 [B-H4]. *If \mathfrak{N} is a closed nondegenerate full range subspace of $H^2(\mathbf{C}^{m+n})$ which is invariant under the operator $\mathfrak{M}_{e^{i\theta}}$ and if \mathfrak{N}' is finite dimensional, then there is g in $\mathfrak{R}U(m, n) \cap H^\infty(M_{m,n})$ so that*

$$\mathfrak{N} = \text{clos } g H^2(\mathbf{C}^{m+n}).$$

Conversely, any g in $\mathfrak{R}U(m, n) \cap H^\infty(M_{m,n})$ gives rise to such a space \mathfrak{N} .

The moral is simple. If \mathfrak{N} is an $\mathfrak{M}_{e^{i\theta}}$ invariant subspace of $H^2(\mathbf{C}^w)$, it is no shock that there is a function ϕ in $H^\infty(M_w)$ so that $\mathfrak{N} = \text{clos } \phi H^2(\mathbf{C}^w)$. The content of the Beurling-Lax-Halmos theorem is that one can choose ϕ to satisfy the additional algebraic constraint that $\phi(e^{i\theta}) \in U(w)$ a.e. (provided \mathfrak{N} is full range). What Theorem 3.6 says is that there are other ways to use the enormous freedom in the functions ϕ satisfying $\mathfrak{N} = \text{clos } \phi H^2$. For example, if $m + n = w$ and \mathfrak{N} satisfies the very mild algebraic consistency condition of nondegeneracy, then we may take $\phi(e^{i\theta})$ to be in $U(m, n)$.

To help orient the reader we mention that the converse of Theorem 3.6 turns out to correspond to the forward side of the ‘Simplest Theorem’ of §3(a) and vice versa. Now we sketch the idea behind Theorem 3.6.

PROOF OF THEOREM 3.6. The proof follows Halmos’ proof of the Beurling-Lax-Halmos theorem closely. Set $\mathfrak{L} = \mathfrak{N} \cap (\mathfrak{M}_{e^{i\theta}} \mathfrak{N})'$. Then the spaces $\mathfrak{L}_n = \mathfrak{M}_{e^{in\theta}} \mathfrak{L}$ are mutually $[\cdot, \cdot]$ orthogonal. Furthermore, any vector x in \mathfrak{N} which is $[\cdot, \cdot]$ orthogonal to all $\mathfrak{M}_{e^{in\theta}} \mathfrak{L}$ is in

$$\bigcap_{k>0} \mathfrak{M}_{e^{ik\theta}} \mathfrak{N} \subset \bigcap_{k>0} \mathfrak{M}_{e^{ik\theta}} H^2(\mathbf{C}^{m+n}) = \{0\}.$$

Thus

$$(3.10) \quad \mathfrak{L} \boxplus \mathfrak{L}_1 \boxplus \mathfrak{L}_2 \boxplus \dots$$

is dense in \mathfrak{N} . Here the \boxplus around the $+$ emphasizes that the \mathfrak{L}_n are $[\cdot, \cdot]$ orthogonal.

A good example of this decomposition is obtained from taking $\mathfrak{N} = H^2(\mathbf{C}^{m+n})$. Then $\mathfrak{L} =$ all constant functions $\approx \mathbf{C}^{m+n}$ and

$$\mathbf{C}^{m+n} \boxplus \mathbf{C}^{m+n}e^{i\theta} \boxplus \mathbf{C}^{m+n}e^{i2\theta} \boxplus \dots$$

is just the standard Fourier decomposition of $H^2(\mathbf{C}^{m+n})$.

It is easy to check that \mathfrak{L} is finite dimensional and so $[\cdot, \cdot]$ restricted to \mathfrak{L} has signature m_1, n_1 . Thus $\mathfrak{L}, [\cdot, \cdot]$ and $\mathbf{C}^{m_1+n_1}, [\cdot, \cdot]_{\mathbf{C}^{m_1+n_1}}$ are isometric via an isometry $V: \mathbf{C}^{m_1+n_1} \rightarrow \mathfrak{L}$ and this can be used to identify the decomposition (3.10) of \mathfrak{N} with the Fourier decomposition of $H^2(\mathbf{C}^{m_1+n_1})$. The map $\Phi: H^2(\mathbf{C}^{m_1+n_1}) \rightarrow \mathfrak{N}$ accomplishing the identification is

$$\Phi \left(\sum_{l=0}^N y_l e^{il\theta} \right) = \sum_{l=0}^N V(y_l) e^{il\theta}.$$

Clearly it commutes with $\mathfrak{N}_{e^{i\theta}}$ and so it must have the form \mathfrak{N}_g where g is a matrix valued function which continues analytically onto the disk. The full range hypothesis on \mathfrak{N} guarantees that $m_1 = m$ and $n_1 = n$. Since Φ is a $[\cdot, \cdot]$ isometry we get $g(e^{i\theta}) \in U(m, n)$ a.e. We have not used that \mathfrak{N}' is finite dimensional yet. As far as the basic structure goes it is unimportant (see [B-H4] for the general theorem). The role it plays here is to force g to be rational and to be in L^∞ (because \mathfrak{N}' is not just nondegenerate, but topologically nondegenerate). We omit the proof here. \square

PROOF OF THEOREM 3.1. Now we show how this amounts to the side of Theorem 3.1 which is as of yet unproved. Suppose we are given G and Φ in $\mathfrak{R}H^\infty(M_{m,n})$ with Φ inner and that $\Psi = 1$. It was already mentioned that we may take $\Psi = 1$ without loss of generality. We must find a g and l which satisfy

$$\mathfrak{G}_g(\mathfrak{R}\mathfrak{B}H^\infty) = \mathfrak{T}_{\Phi, l}(G) \cap \mathfrak{R}\mathfrak{B}L^\infty(M_{m,n}) = \{G + \Phi H: H \in \mathfrak{R}\mathfrak{B}H_l^\infty(M_{m,n})\}.$$

To build g set

$$\mathfrak{N} = \text{clos}\{\langle (G + \Phi H)h, h \rangle: H \in H^\infty(M_{m,n}), h \in H^2(\mathbf{C}^n)\}.$$

Since G, Φ are rational, the space \mathfrak{N}' is finite dimensional. So Theorem 3.6 applies to give a g in $\mathfrak{R}U(n, n)$ with

$$\mathfrak{N} = gH^2(\mathbf{C}^{m+n}).$$

Now we must show that this is indeed the g which we want. It is slightly more convenient to prove (3.10) with the closed ball $\overline{\mathfrak{B}}$ replacing the open ball \mathfrak{B} throughout. So we shall henceforth work with $\overline{\mathfrak{B}}$.

The notion of angle operator extends directly from \mathbf{C}^{m+n} to $L^2(\mathbf{C}^{m+n})$ with respect to the decomposition of $L^2(\mathbf{C}^{m+n})$ into positive and negative spaces $L^2(\mathbf{C}^m) \oplus L^2(\mathbf{C}^n)$; let P_+ and P_- denote the orthogonal projections of $L^2(\mathbf{C}^{m+n})$ onto $L^2(\mathbf{C}^m)$ and $L^2(\mathbf{C}^n)$. Any negative subspace \mathfrak{N} has an angle operator $M: P_- \mathfrak{N} \rightarrow P_+ \mathfrak{N}$ and if $\mathfrak{N}_{e^{i\theta}} \mathfrak{N} \subset \mathfrak{N}$ this implies $\mathfrak{N}_{e^{i\theta}} M = M \mathfrak{N}_{e^{i\theta}}$. This implies that M is a multiplication operator \mathfrak{N}_F ; we take the natural convention that if \mathfrak{N} is a subspace which is maximal negative within $H^2(\mathbf{C}^{m+n})$, the corresponding F is in $\overline{\mathfrak{B}}H^\infty(M_{m,n})$. By this convention any invariant

negative \mathcal{N} for which $P_- \mathcal{N}$ is a full range space of codimension l in $H^2(\mathbf{C}^n)$ has angle operator \mathfrak{N}_F with

$$F = F_0 \Psi^{-1}$$

where F_0 is in $\overline{\mathfrak{B}}H^\infty(M_{m,n})$ and Ψ is the unitary valued function which (by the traditional Beurling-Lax-Halmos theorem) represents $P_- \mathcal{N}$ as $P_- \mathcal{N} = \Psi H^2(\mathbf{C}^n)$. This is equivalent to saying F is in $\overline{\mathfrak{B}}H^\infty(M_{m,n})$.

To analyze $\mathfrak{G}_g(\overline{\mathfrak{B}}H^\infty(M_n))$ look at its geometric interpretation: The set $\overline{\mathfrak{B}}H^\infty(M_n)$ corresponds to the set of all angle operators of invariant maximal negative subspaces \mathcal{N} of $H^2(\mathbf{C}^{m+n})$. The $[\cdot, \cdot]$ isometry \mathfrak{N}_g maps each such \mathcal{N} to $g\mathcal{N}$, an invariant subspace which is maximal negative in \mathfrak{N} . The map $\mathcal{N} \rightarrow g\mathcal{N}$ written in terms of angle operators as in (3.9) tells us that $\mathfrak{G}_g(\overline{\mathfrak{B}}H^\infty(M_n))$ corresponds to the set of all angle operators for invariant subspaces which are maximal negative in \mathfrak{N} . The special structure of \mathfrak{N} forces these angle operators to have a special form. For simplicity suppose for the moment that each invariant subspace \mathcal{N}_1 which is maximal negative in \mathfrak{N} is also maximal negative in $H^2(\mathbf{C}^{m+n})$. Then the angle operator for \mathcal{N}_1 is \mathfrak{N}_F for a function F with the property: Given h in $P_- \mathcal{N}_1 = H^2(\mathbf{C}^n)$ there is a k in $H^2(\mathbf{C}^n)$ so that

$$Fh = Gh + \Phi k.$$

This says that H defined to be $\Phi^{-1}(F - G)$ satisfies $Hh = k$. Thus $\mathfrak{N}_H: H^2(\mathbf{C}^n) \rightarrow H^2(\mathbf{C}^n)$ or equivalently H lies in $H^\infty(M_{m,n})$. In other words, $F = G + \Phi H$ for $H \in H^\infty(M_{m,n})$, that is F is in $\mathfrak{T}_{\Phi,0}(G) \cap \overline{\mathfrak{B}}L^\infty(M_{m,n})$. The argument reverses to give an \mathcal{N}_1 from such an F . This is exactly the conclusion of Theorem 3.1 when $l = 0$, except we need to show that when \mathcal{N} is cofinite dimensional in $H^2(\mathbf{C}^{m+n})$ the function F is rational. Now F is rational if and only if $\mathcal{N}_1 = g\mathcal{N}$ is cofinite dimensional, but G, Φ rational forces g to be rational and this implies that \mathcal{N}_1 is cofinite dimensional.

To finish our proof for $l = 0$ we must see when the key assumption we made above is valid. That proves to be surprisingly easy and calls for a few definitions. The negative *co-signature* of a negative subspace \mathcal{N} of $H^2(\mathbf{C}^{m+n})$ equals the dimension of $H^2(\mathbf{C}^n) \ominus P_{H^2(\mathbf{C}^n)} \mathcal{N}$, or alternatively equals the codimension of \mathcal{N} in any maximal negative subspace containing it. The *negative signature* of a subspace X is the dimension of a maximal negative subspace of X .

LEMMA 3.7. *Suppose \mathfrak{N} is a closed nondegenerate subspace of $H^2(\mathbf{C}^{m+n})$. Each subspace \mathcal{N} of \mathfrak{N} which is maximal negative in \mathfrak{N} has negative cosignature l in $H^2(\mathbf{C}^{m+n})$, where l is the negative signature of \mathcal{N} .*

PROOF. This is a consequence of very elementary linear algebra. The point is that nondegenerate subspaces have a decomposition, for example, $\mathfrak{N}' = (\mathfrak{N}')_+ \boxplus (\mathfrak{N}')_-$ into $[\cdot, \cdot]$ orthogonal strictly positive and negative subspaces. The negative signature of \mathfrak{N}' equals $\dim(\mathfrak{N}')_-$. If \mathcal{N} is maximal negative in \mathfrak{N} , then $(\mathfrak{N}')_- \boxplus \mathcal{N}$ is maximal negative in $H^2(\mathbf{C}^{m+n})$. \square

The lemma allows us to complete the proof of Theorem 3.1 and what is even better to actually compute explicitly the integer l which occurs in the theorem.

To complete the theorem take $l =$ the negative signature of \mathfrak{N}' . The key assumption in the previous argument amounted to the statement that this equals zero (because of the lemma). However, in general the lemma implies that the angle operator for each \mathfrak{N} invariant and maximal negative in \mathfrak{N} has the form \mathfrak{N}_F where $F \in H_1^\infty$. Again it can be shown that $\Phi^{-1}(F - G) \in H_1^\infty$ and consequently that $F \in \mathfrak{T}_{\Phi, l}(G) \cap \overline{\mathfrak{B}}L^\infty(M_n)$. In other words, the argument given for $l = 0$ carries straight through to the case of general l .

To compute l explicitly from G and Φ is surprisingly simple. We simply compute a basis for \mathfrak{N}' and then look at the Grammian for this basis with respect to the bilinear form $[\cdot, \cdot]$. The number of negative eigenvalues for this Grammian clearly equals l . Let us do this explicitly for the scalar case and Φ a Blaschke product with zeroes of multiplicity one only (say at z_1, z_2, \dots, z_N). Let $w_j = G(z_j)$ and let $k_z(e^{i\theta})$ be the kernel $k_z(e^{i\theta}) = (1/2\pi)(1/(1 - \bar{z}e^{i\theta}))$ which reproduces an H^2 function from its boundary values, i.e.,

$$f(z) = (f, k_z).$$

Then \mathfrak{N}' is the span χ of the functions $k_z \langle 1, \bar{w}_j \rangle$. To see this note that

$$[\langle f, h \rangle, k_{z_j} \langle 1, \bar{w}_j \rangle] = f(z_j) - h(z_j)w_j.$$

Consequently, $\langle f, h \rangle$ is $[\cdot, \cdot]$ orthogonal to χ if and only if $(f - Gh)(z_j) = 0$ if and only if $f = Gh + \Phi k$ for some $k \in H^2$; that is, $\langle f, h \rangle \in \mathfrak{N}$.

The computation of l amounts to finding the negative signature of \mathfrak{N}' . A typical vector in \mathfrak{N}' is $\sum_{j=1}^N \alpha_j k_{z_j}$ and it is negative if and only if

$$0 \geq \sum_1^N \alpha_u \bar{\alpha}_v [k_{z_u} \langle 1, \bar{w}_u \rangle, k_{z_v} \langle 1, \bar{w}_v \rangle] = \sum_1^N \alpha_u \bar{\alpha}_v \frac{1 - \bar{w}_u w_v}{1 - \bar{z}_u z_v},$$

$$0 \geq (\Lambda \alpha, \alpha)_{\mathbb{C}^N},$$

where Λ is the Pick matrix for $\{z, w\}$ and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$. Consequently the biggest $[\cdot, \cdot]$ negative subspace of \mathfrak{N}' has the same dimension as the biggest $(\Lambda, \cdot)_{\mathbb{C}^N}$ negative space of \mathbb{C}^N . So l is the number of negative eigenvalues of Λ ; also \mathfrak{N} is nondegenerate if and only if \mathfrak{N}' is nondegenerate if and only if Λ has no null space. \square

PROOF OF THEOREM 3.3. The last part of the outline for Theorem 3.1 when expanded to the proper generality proves Theorem 3.3 except for the case when the Pick matrix has a null space. A null space forces $\mathfrak{N} \cap \mathfrak{N}' \neq \{0\}$ and the behavior is radically different from the nondegenerate case. For example, when $n = 1$ it is very easy to find an interpolating function: The space $\boxplus_{k \geq 0} \mathfrak{N}_{e^{i\theta}}^k(\mathfrak{N} \cap \mathfrak{N}')$ is $[\cdot, \cdot]$ null and its angle operator is \mathfrak{N}_φ where φ is an inner function which solves the interpolation problem (uniquely). When $n > 1$ the procedure is more complicated; one must factor out this null space and apply the theory in the first part of our proof to the quotient. \square

g. More is needed. So far we have had the pleasure of describing known results. In §6 we shall reduce six different circuit problems labeled (a), (b), (c), (d), (e), and (f) to interpolation problems. The theory and construction listed so far completely handle (a)–(c), but unfortunately fall short of settling (d)–(f).

Thus we are presented with more refined interpolation problems to solve. The nature of these problems is to determine if there is a function f in

$$\mathfrak{T}_{\Phi, \Psi, l}(G) \cap \mathfrak{B}L^\infty(M_n) \cap \mathfrak{A}$$

where \mathfrak{A} is some class of functions. Various classes come up. However, we shall describe only the one produced by the stability constraint in §6(d).

The set \mathfrak{A}_0 which arises in the stability problem (see §6(d)) is

$$\mathfrak{A}_0 \triangleq \{F \in H_l^\infty(M_n) : \det(F(z) - K(z)) \neq 0 \text{ for any } z \text{ in } |z| \leq 1\}.$$

The problem of finding an F in $\mathfrak{T}_{\Phi, \Psi, l}(G) \cap \mathfrak{B}L^\infty(M_n) \cap \mathfrak{A}_0$ is open. The simplest case ($n = 1$ and $l = 0$) is reasonably well understood (see [B-H2]). The idea there is simple; an analytic function F with no zeroes has an analytic logarithm $G = \ln F$ and $F \in \mathfrak{B}H^\infty$ if and only if $\text{Re } G(z) \leq 0$ for all $|z| \leq 1$. Moreover $F(z_j) = w_j$ if and only if $G(z_j) = \ln w_j + 2\pi i n_j$ for some integer n_j . Now interpolation with functions having negative real part is just a transform of a problem settled by Theorem 3.2 and 3.3 (cf. Chapter XI in [D]) and there is a Pick matrix appropriate for such problems. This argument gives

THEOREM 3.8 (SEE [B-H2]). *There is a function F in*

$$\mathfrak{T}(\mathbf{z}, \mathbf{w}) \cap \overline{\mathfrak{B}}H^\infty \cap \mathfrak{A}_0$$

if and only if there exists integers $\mathbf{m} = (m_1, \dots, m_N)$ so that the matrix

$$L_{\mathbf{z}, -\ln \mathbf{w}} + L_{\mathbf{z}, 2\pi i \mathbf{m}}$$

is positive definite. Here

$$L_{\mathbf{z}, \mathbf{v}} \triangleq \left\{ \frac{v_i + \bar{v}_j}{1 - z_i \bar{z}_j} \right\}_1^N.$$

When this holds one can construct an F of the form $F(z) = e^{R(z)}$ where R is rational, imaginary on Π , and has all poles $e^{i\theta_1}, \dots, e^{i\theta_N}$ on Π . This F satisfies $|F(e^{i\theta})| \equiv 1$ but has a discontinuity at each $e^{i\theta}$.

The type of interpolation problem produced by our analysis of the common amplifier in §6(e) is also a very natural generalization of the ones solved in §3(d)(e) but it is unsolved. The nature of it is roughly this. Find a function M in

$$\mathfrak{B}H_l^\infty(M_2) \cap \mathfrak{T}(\mathbf{z}, \mathbf{x}, \mathbf{y})$$

which minimizes

$$\|s.e.v. M(e^{i\theta})M(e^{i\theta})^*\|_{L^\infty}.$$

Here s.e.v. M means the smallest eigenvalue of the matrix M .

4. The Poincaré distance to $\mathfrak{B}H^\infty$. The classical Poincaré metric on the disk is invariant under biholomorphic mappings of the disk onto itself. On higher dimensional domains there are many biholomorphically invariant metrics and this is certainly true of $\mathfrak{B}M_n$. However, one of the better behaved metrics

called the Carathéodory metric (cf. [K, H1]) can be written explicitly as

$$p(h, s) = \operatorname{arctanh} \|(h - s)(s^*h - 1)^{-1}(s^* - h^*)(1 - sh^*)^{-1}\|_{M_n}$$

for $s, h \in \mathfrak{B}M_n$. It is indeed invariant under a map \mathcal{G}_g with g in $U(n, n)$. As we eventually see in §6 this metric involving a sup of eigenvalues is the one best suited to our engineering problems because they involve a worst case analysis rather than averages. Recall that p induces a metric

$$\rho(F, H) = \sup_{\theta} p(F(e^{i\theta}), H(e^{i\theta})).$$

The metric was called P^∞ in the introduction but henceforth we abbreviate this to ρ .

The main issue in this section is the computation of

$$\delta_F = \inf_{H \in \mathfrak{B}H^\infty(M_n)} \rho(F, H)$$

for F in $\mathfrak{R}\mathfrak{B}L^\infty(M_n)$. We shall present several solutions and begin by stating the most polished type of result.

(a) *A solution to the distance problem.* The following results give a clear idea of the answer which is ultimately available. Complete details and proofs are found in [H1, Theorem 5.2 and Corollary 1.1]; see also [H3].

THEOREM 4.1. *If F is in $\mathfrak{R}\mathfrak{B}L^\infty(M_n)$ with l poles outside the unit disk (that is, $F^* \in \mathfrak{B}H_l^\infty(M_n)$) has poles and zeroes at different places,⁸ then the Poincaré distance δ_F of F to $\mathfrak{B}H^\infty(M_n)$ equals its Poincaré distance to $\mathfrak{R}\mathfrak{B}H^\infty(M_n)$ and this equals $\operatorname{arctanh} \sqrt{\lambda}$ where λ is the l th largest solution of the generalized $Ax = \lambda Bx$ where A and B are selfadjoint operators which can (and soon will be) written down explicitly.*

THEOREM 4.1'. *There is an algorithm which uses the solution λ_l and corresponding eigenvectors x from Theorem 4.1 in a straightforward way to compute a ρ -closest point H_0 in $\mathfrak{B}H^\infty(M_n)$ to F . One such closest point H_0 satisfies*

$$p(H_0(e^{i\theta}), F(e^{i\theta})) \equiv \delta_F$$

for all θ . When $n = 1$ the closest point is unique.

These two theorems indicate in general terms the nature of solutions one can get. There are explicit formulas for A and B several of which we list at the end of this section (see §4(c)). To give the flavor of these formulas we state the one which has been most heavily polished. It is for the (physically basic) case where F on Π has a bounded analytic continuation to the exterior of the disk.

THEOREM 4.1'' (COROLLARY 1.1 AND THEOREM 5.5 [H1]). *Suppose $F^* \in \mathfrak{R}\mathfrak{B}H^\infty(M_n)$. Then a choice of operators A and B for which Theorem 4.1 and Theorem 4.1' hold is*

$$A = \mathfrak{I}\mathfrak{C}_{\delta_* F} \mathfrak{I}_{\delta}^{-1} \mathfrak{I}\mathfrak{C}_{\delta_* F}^*, \quad B = \mathfrak{I}_{\delta_*} \checkmark$$

⁸To be precise we say that F satisfies condition N if and only if there is no z in the disk at which both F and F^{-1} have a pole. That F satisfies condition N is a precise statement of the hypothesis of Theorem 4.1.

where

$$\delta = (1 - F^*F)^{-1} \quad \text{and} \quad \delta_* = (1 - FF^*)^{-1}.$$

Here \mathfrak{H}_F and \mathfrak{T}_F denote the Hankel and Toeplitz matrices⁹ with generating function F and \check{H} designates the function defined by $\check{H}(e^{i\theta}) = H(e^{-i\theta})$.

This theorem is precisely Corollary 1.1 [H1]. Its proof when $n = 1$ will be given in §4(c). Instead of giving more detailed formulas we turn now to explaining the idea behind Theorem 4.1 and behind a large class of solutions to the Poincaré distance problem of which Theorem 4.1 is but an example.

(b) *A general class of solutions.* There are many solutions to the Poincaré distance problem and results of the type in Theorem 4.1 and 4.1' are special cases of a more general theory. This more general theory settles the question:

(4.1) Given $F \in \mathfrak{R} \mathfrak{B} L^\infty(M_n)$ and $0 < r < 1$ is there an $H \in \mathfrak{R} \mathfrak{B} H^\infty(M_n)$ such that $\rho(F, H) \leq \operatorname{arctanh} r$? Build such an H .

Since Theorem 4.1 automatically produces the smallest r it actually solves a harder problem than (4.1). So one might hope that there are easier (to prove) solutions to (4.1) than Theorem 4.1. This is indeed the case because as we shall soon see that Theorem 3.4 alone with only a little manipulation settles question (4.1). Also we may take r to be a function, not just a scalar, without complicating any of our procedures.

The key notion needed is that of a disk in function space. Such a disk $\Delta_C^{P,R}$ is defined in terms of functions C, P, R in $L^\infty(M_n)$ where $P(e^{i\theta})^2, R(e^{i\theta})^2$ are uniformly strictly positive; we define $\Delta_C^{P,R}$ to be the set of all $H \in L^\infty(M_n)$ which satisfy

$$(H - C)P^2(H - C)^* \leq R^2.$$

It is easy to check that the set

(4.2) $\{H \in \mathfrak{B} L^\infty(M_n) : \rho(H, F) \leq \operatorname{arctanh} r\}$

which plays the key role in question (4.1) is a disk and to compute¹⁰ its center $C_{F,r}$ and partial radii $P_{F,r}^2$ and $R_{F,r}^2$. Consequently question (4.1) amounts to

⁹Hankel operators were defined in footnote 7. A Toeplitz operator looks like

$$\mathfrak{T}_F = \begin{pmatrix} F_0 & F_1 & F_2 & & \\ F_{-1} & F_0 & F_1 & \cdots & \\ F_{-2} & F_{-1} & F_0 & & \\ & \vdots & & & \end{pmatrix}.$$

¹⁰The computation can be found in §5, footnote 6 of [H3E]. It gives

$$C = [\mathfrak{R} - I]F[F^*\mathfrak{R}F - I]^{-1},$$

$$R^2 = \mathfrak{R}(1 - FF^*)[\mathfrak{R}FF^* - I]^{-1}, \quad P^2 = [F^*\mathfrak{R}F - I](1 - F^*F)^{-1}$$

where we have suppressed the subscripts F, r and have taken

$$\mathfrak{R} = (1 - FF^*)^{-1/2}r(1 - FF^*)^{+1/2}.$$

determining if a particular disk $\Delta_C^{P,R}$ contains a function in $\mathfrak{R}H^\infty(M_n)$. There is a nice test for determining if such an H exists which even settles the more advanced question of when there is an H in $H_l^\infty(M_n) \cap \Delta_C^{P,R}$.

THEOREM 4.2. *The disk $\Delta_C^{P,R}$ contains a function in $H^\infty(M_n)$ if and only if*

$$\mathfrak{K}_C[\mathfrak{X}_{P-2}]^{-1}\mathfrak{K}_C^* \leq \mathfrak{X}_{R^2}.$$

More generally this disk contains a function in $H_l^\infty(M_n)$ if and only if the eigenvalue problem

$$\mathfrak{K}_C[\mathfrak{X}_{P-2}]^{-1}\mathfrak{K}_C^*x = \lambda \mathfrak{X}_{R^2}x$$

has at most l solutions greater than one. If P, R, C are rational, then a rational solution exists whenever any solution exists.

THEOREM 4.2'. *If $P^2, R^2,$ and C are rational, there are in principle several algorithms which use λ_l and the corresponding eigenvector x from Theorem 4.2 to construct a rational H in $\Delta_C^{P,R} \cap H_l^\infty(M_n)$ satisfying*

$$\|(H(e^{i\theta}) - C(e^{i\theta})) * P(e^{i\theta})\|^2 = \|R(e^{i\theta})\|^2$$

for all θ . One such algorithm is given for $n = 1$ and $l = 0$ in [H3, Lemma I.2].

To prove the theorem we need Wiener-Hopf (or in engineering parlance, spectral) factorization: every uniformly positive matrix function W on Π can be factored as $W(e^{i\theta}) = M(e^{i\theta}) * M(e^{i\theta})$ where $M \in H^\infty(M_n)$. Moreover, M can be taken to be an outer function,¹¹ that is a function M in $H^\infty(M_n)$ with M^{-1} in $H^\infty(M_n)$.

PROOF. We assume P^2, R^2, C are rational. Theorem 4.2 follows quickly from Theorem 3.4. One takes α and β to be outer factors of

$$\alpha\alpha^* = P^2, \quad \beta\beta^* = R^2.$$

Then H lies in $\Delta_C^{P,R}$ if and only if $K \stackrel{\Delta}{=} \beta^{-1}H\alpha$ lies in $\Delta_{\beta^{-1}C\alpha}^{I,I}$. Since β and α are invertible outer functions, there exists an H in $\Delta_C^{P,R} \cap H_l^\infty(M_n)$ if and only if $\Delta_{\beta^{-1}C\alpha}^{I,I} \cap H_l^\infty$ is nonempty. In other words, we must determine if

$$\{K \in L^\infty(M_n) : \|K - \beta^{-1}C\alpha\|_{L^\infty} \leq 1\}$$

contains a function in $H_l^\infty(M_n)$. Theorem 3.4 says it does if and only if $\|\mathfrak{K}_{\beta^{-1}C\alpha}\| \leq 1$.

To convert this to the answer which appears in the theorem one uses some Toeplitz and Hankel trickery. The argument for $l = 0$ is this: $\|\mathfrak{K}_{\beta^{-1}C\alpha}\| \leq 1$ if and only if

$$(4.3) \quad \mathfrak{K}_{\beta^{-1}C\alpha}\mathfrak{K}_{\beta^{-1}C\alpha}^* \leq I$$

if and only if

$$\mathfrak{X}_{\beta^{-1}}\mathfrak{K}_C\mathfrak{X}_\alpha\mathfrak{X}_\alpha^*\mathfrak{K}_C^*\mathfrak{X}_{\beta^{-1}}^* \leq I,$$

$$\mathfrak{K}_C(\mathfrak{X}_{\alpha^{-1}*}\mathfrak{X}_{\alpha^{-1}})\mathfrak{K}_C^* \leq \mathfrak{X}_\beta\mathfrak{X}_\beta^*, \quad \mathfrak{K}_C\mathfrak{X}_{P-2}\mathfrak{K}_C^* \leq \mathfrak{X}_{R^2}.$$

¹¹Actually an outer function M is one in $H^\infty(M_n)$ with the property that $MH^2(\mathbf{C}^n)$ is dense in $H^2(\mathbf{C}^n)$. For M^{-1} uniformly bounded on Π , the condition M outer implies M^{-1} is in $H^\infty(M_n)$.

For $l > 0$ similar reasoning prevails. Theorem 4.2' follows from constructions referred to in Theorem 3.4'.

There are many other recipes besides this basic one for determining if $\Delta_C^{P,R} \cap \mathfrak{R}H^\infty(M_n)$ is empty. They are based on combining Theorem 4.2 with a 'coordinate change' in function space. A linear fractional map \mathfrak{G}_g with g in $\mathfrak{R}GL(2n)$ will transform one disk in function space to another. One could work on either disk and by doing this properly get different forms of solutions to the problem. Here we shall describe the linear fractional maps appropriate for the Poincaré distance problem.

Suppose we want to write down a condition for there to exist an H in the disk (4.2) and in $H^\infty(M_n)$. Let \mathfrak{G} be a (rational) Poincaré isometry, that is, $\mathfrak{G} = \mathfrak{G}_g$ for some g in $\mathfrak{R}U(n, n)$. Then H is in (4.2) if and only if

$$(4.4) \quad \rho(\mathfrak{G}(H), \mathfrak{G}(F)) \leq \operatorname{arctanh} r.$$

Consequently, an H in $H^\infty(M_n)$ lies in (4.2) if and only if there is a K in $\mathfrak{G}(\mathfrak{B}H^\infty(M_n))$ which lies in the disk

$$(4.5) \quad \{K \in \mathfrak{B}L^\infty(M_n) : \rho(K, \mathfrak{G}(F)) \leq \operatorname{arctanh} r\}.$$

Thus we have a variation on question (4.1). Instead of determining if $H^\infty(M_n)$ intersects a certain disk we must determine if $\mathfrak{G}(\mathfrak{B}H^\infty(M_n))$ intersects a certain disk (given by (4.5)). Fortunately (for the rational case), the machinery necessary for doing this was developed in §3(a)(e)(f).

Theorem 3.1 tells us that the range of \mathfrak{G} is

$$(4.6) \quad \mathfrak{G}(\mathfrak{R}\mathfrak{B}H^\infty) \cong (G + \Phi\mathfrak{R}H_l^\infty) \cap \mathfrak{B}L^\infty$$

for a G and Φ in $\mathfrak{R}H^\infty$ which are explicitly computed in the course of proving the theorem, and l given by Theorem 3.3. In a moment we write out the recipe for obtaining l , G and Φ from F , but now we proceed with our computation of the distance from F to $\mathfrak{B}H^\infty$. A function in the set (4.6) is in the disk (4.5) if and only if there is an L in H_l^∞ satisfying

$$\rho(G + \Phi L, \mathfrak{G}(F)) \leq \operatorname{arctanh} r.$$

That is, by footnote 10, we must find an L in H_l^∞ and in the disk

$$(4.7) \quad (\Phi^{-1}C_{\mathfrak{G}(F),r} - \Phi^{-1}G - L)P_{\mathfrak{G}(F),r}^2(\Phi^{-1}C_{\mathfrak{G}(F),r} - \Phi^{-1}G - L)^* \leq R_{\mathfrak{G}(F),r}^2.$$

Theorem 4.2 tells precisely for which F (and subsequently derived G , Φ , and l) such an L exists. This gives many tests (one for each \mathfrak{G}) to determine if F is within distance r of $\mathfrak{B}H^\infty(M_n)$. This class of tests will be stated formally as a theorem (Theorem 4.4) but first let us look at some examples.

Our first example is the proof of Theorem 4.1. This is a rather vague example in that Theorem 4.1 involved operators A and B which weren't explicitly computed. Subsequent examples will compute several choices for A and B .

PROOF OF THEOREMS 4.1 AND 4.1'. The function F has Poincaré distance $\Delta_F \leq \operatorname{arctanh} \sqrt{\lambda}$ if and only if there is an H in $\mathfrak{B}H^\infty(M_n)$ satisfying

$$(4.8) \quad \rho(F, H) \leq \operatorname{arctanh} \sqrt{\lambda}.$$

Suppose \mathcal{G} is a map from $\mathcal{R}U(n, n)$ satisfying $\mathcal{G}(F) = 0$. For example,

$$(4.9) \quad \mathcal{G}_0^F(S) = -F + (1 - FF^*)^{1/2}S(1 - F^*S)^{-1}(1 - F^*F)^{1/2}$$

is one such map and all others are ‘equivalent’ to it through multiplication by functions with unitary boundary values. With this choice of \mathcal{G} formula (4.7) is simple, because as one instantly checks

$$C_{0,\sqrt{\lambda}} = 0, \quad P_{0,\sqrt{\lambda}} = I, \quad \text{and} \quad R_{0,\sqrt{\lambda}} = \sqrt{\lambda}I.$$

One computes Φ, G, l for \mathcal{G} and obtains from Theorem 4.2 that F satisfies (4.8) if and only if $\lambda \geq \lambda_l$, the l th from largest eigenvalue for $Ax = \lambda Bx$, where

$$(4.10) \quad A = \mathcal{H}_{\Phi^{-1}G} \mathcal{H}_{\Phi^{-1}G}^* \quad \text{and} \quad B = I.$$

Thus Theorems 4.1 and 4.1’ hold with this choice of A, B and an l which is computed from G and Φ . \square

The geometric interpretation of the coordinate change \mathcal{G} is very informative. As λ increases the H ’s satisfying (4.8) form a family of increasing disks containing F ; while F is at the ‘Poincaré center’ of these disks it is not at the Euclidean center unless $F = 0$. The map \mathcal{G} sends F to 0 and since it is a ρ -isometry \mathcal{G} takes the family (4.8) of circles to a family of circles with (both Poincaré and Euclidean) center at 0 and radius $\operatorname{arctanh} \sqrt{\lambda}$. As a result the radius of the circles appears linearly in the matrix test of Theorem 4.2 and determination of the disk of minimum radius which intersects H^∞ becomes an eigenvalue problem (namely, the one in Theorem 4.1).

Further examples and details and the coordinate changes appear in the next subsection.

(c) *Explicit formulas.* In this subsection we want to demonstrate that the operators A and B which appear in the solution to the Poincaré distance problem can be computed explicitly and we want to give a sample computation. We make no attempt to be encyclopedic since very general formulas have been listed and proved elsewhere [H1, H3].

Our main achievement here will be to derive a whole class of formulas for determining if the Poincaré distance δ_F of an F in $\mathcal{R}\mathcal{B}L^\infty$ is $\leq r$. These formulas will be stated as Theorem 4.4. Each of these formulas results from a particular ‘coordinate change’ \mathcal{G} and all basic ideas behind what we shall do appear in 4(b). Missing in §4(b) were explicit formulas. Our first step is to list a few definitions; then we state one corollary of Theorem 4.4; then we state Theorem 4.4; then we show how one specialization of it gives the corollary and Theorem 4.1’’ for $n = 1$. After that we prove Theorem 4.4; then we change to the topic of interpolation and formulas based on interpolation.

Our explicit formulas require that we recall some definitions from Lemma 3.5. Given a function K in $\mathcal{R}L^\infty$, let $K = RV_K^{-1}$ be the coprime factorization of K with V_K inner. Recall that K_I is a function in $\mathcal{R}H^\infty$ with the property that $[K_I - K^{-1}]/V_K^2$ is analytic near the zeroes of V_K in Δ . Define $\langle K \rangle$ to be

$$\langle K \rangle = \frac{-\bar{K} - (1 - |K|^2)K_I}{(1 - |K|^2)V_K^2}.$$

One consequence of the main theorem of this section is

COROLLARY 4.3. *If F is in $\mathfrak{R}\mathfrak{B}L^\infty$, then Theorems 4.1 and 4.1' hold with*

$$A = \mathfrak{H}_{\langle \bar{F} \rangle} \mathfrak{H}_{\langle \bar{F} \rangle}^* \quad \text{and} \quad B = \mathfrak{X}_{(1-|F|^2)^{-2}}.$$

This corollary is one instance ($D = F$) of the following general result.

THEOREM 4.4. *To determine if the 'Poincaré distance' δ_F of a rational function F to $\mathfrak{B}H^\infty$ is less than or equal to r , select any function D in $\mathfrak{R}\mathfrak{B}L^\infty$. Let $\bar{D} = RV^{-1}$ be the coprime factorization of \bar{D} , form the operator*

$$M = r\mathfrak{X}(1 - |\check{\Gamma}|^2) / (1 - |\check{D}|^2)^2 (1 - r|\check{\Gamma}|^2) - \mathfrak{H}_C \mathfrak{H}_C^*$$

where

$$C = \frac{(1-r)\Gamma(1-r|\Gamma|^2)^{-1} + D + (1-|D|^2)[\bar{D}]_I}{(1-|D|^2)V^2}$$

and

$$\Gamma = -D + (1-|D|^2)F(1-\bar{D}F)^{-1}.$$

Then $\delta_F \leq r$ if and only if M has at most l negative eigenvalues where $l = \#$ poles of \bar{D} inside Δ .

An excellent illustration of how one uses the rather complicated looking Theorem 4.4 is to derive Corollary 4.3 and Theorem 4.1'' when $n = 1$. All that one does is take $D = F$ in Theorem 4.4. We get $\Gamma = \mathfrak{G}_0^F(F) = 0$ and

$$C = \frac{F + (1-|F|^2)[\bar{F}]_I}{(1-|F|^2)V^2} = \langle \bar{F} \rangle$$

where $\bar{F} = RV^{-1}$ is the coprime factorization of \bar{F} . This gives Theorem 4.3 immediately.

An elegant special case of this is Theorem 4.1'' for $n = 1$. The hypothesis of Theorem 4.1'' is that $F^* \in \mathfrak{B}H^\infty$, so $\bar{D} \in \mathfrak{B}H^\infty$ and $l = \#$ poles of $\bar{D} = 0$; consequently $V = 1$ and we may take $[\bar{D}]_I = 0$. So in this case $M = r\mathfrak{X}_{(1-|F|^2)^{-2}} - \mathfrak{H}_{F\bar{D}} \mathfrak{H}_{F\bar{D}}^*$. The number of negative eigenvalues of M is $\leq l = 0$ if and only if $M \geq 0$. Thus we have proved that $\delta_F \leq r$ if and only if

$$(4.11) \quad \mathfrak{H}_{F\bar{D}} \mathfrak{H}_{F\bar{D}}^* \leq r\mathfrak{X}_{(1-|F|^2)^{-2}}.$$

This actually is a nicer result than Theorems 4.1' + 4.1'' since (4.11) is simpler than the form of A and B given there. The reason is that now we are treating only scalar valued F and for matrix valued F the answer is a little more complicated. To obtain the A and B of Theorem 4.1'' from (4.11) just take the outer Wiener-Hopf factorization $Q\bar{Q}$ of $1 - |F|^2$, substitute it into the right side of (4.11), and use the Toeplitz and Hankel tricks of the type which follow (4.3) to convert inequality (4.11) to the inequality $A \leq B$ where A and B are given by Theorem 4.1''.

We remark that the flexibility of Theorem 4.4 appears to be very helpful for engineering problems. In practical situations one is not given the function F

completely; one only has (imperfect) measurements of its values on some interval. The article [H3] is devoted to such problems and how different choices of D in Theorem 4.4 have different numerics and advantages.

PROOF OF THEOREM 4.4. This theorem formalizes the tests we were deriving in §4(b) based on a "change of coordinates" \mathcal{G} . The derivation in §4(b) concluded with equation (4.7) and the application of Theorem 4.2 to it. This actually constituted a proof of Theorem 4.4 except for the fact that various components of (4.7), namely the functions G , Φ and l , had not been explicitly computed. Thus the only remaining work for our proof is explicit computation of these functions.

Go back now to §4(b) where we are in the process of analyzing the effect of changing coordinates via a map \mathcal{G} . The first thing we claim is that if we multiply \mathcal{G} by a function ψ of modulus one on Π to obtain a new map $\mathcal{G}^1(S) = \psi\mathcal{G}(S)$, then the new coordinate change \mathcal{G}^1 gives the same answer as the old one \mathcal{G} . To be precise the disk consisting of L 's satisfying inequality (4.7) based on \mathcal{G} and on \mathcal{G}^1 is the same. There are several ways to verify this; one is to go to footnote 6 which contains the definitions of $C_{\mathcal{G}(F),r}$, etc., go to the recipe in §3(e) which contains the definition of Φ and G and see (easily) that $\Phi^{-1}C_{\mathcal{G}(F),r}$ for \mathcal{G} and $\Phi^{-1}C_{\mathcal{G}^1(F),r}$ for \mathcal{G}^1 are equal, that $\Phi^{-1}G$ and $\Phi^{-1}G_1$, etc. are equal, etc. In other words, ψ effectively cancels out of (4.7).

The preceding paragraph allows us to take the \mathcal{G} we are studying and to convert it to another \mathcal{G}' of simpler form. We do this by first denoting by D the function which \mathcal{G} maps to 0; that is $\mathcal{G}(0) = 0$. It is easy to check that $\mathcal{G} = \psi_1\mathcal{G}_0^D$ where \mathcal{G}_0^D was defined in (4.9) for a unimodular ψ_1 . Thus it suffices to analyze \mathcal{G}_0^D . Actually an even easier map to study is \mathcal{G}^1 defined by

$$\mathcal{G}^1(S) = \frac{Q_1^2}{(1 - |D|^2)V_B^2} \mathcal{G}_0^D(S) = Q_1^2 \left[\frac{-D}{(1 - |D|^2)V_B^2} + \frac{1}{V_B^2} S(1 - \bar{D}S)^{-1} \right]$$

where Q_1 is the outer Wiener-Hopf factor of $1 - |D|^2$. Since $Q_1^2/(1 - |D|^2)V_B^2$ is unimodular \mathcal{G}^1 comes from $\mathcal{R}U(1,1)$. Henceforth abbreviate V_B^2 to V . The recipe in §3(e) or even more simply Lemma 3.5 tells us that range \mathcal{G}^1 equals

$$\mathcal{G}^1(\mathcal{R}\mathcal{B}H^\infty) = [G_1 + H_l^\infty] \cap \mathcal{R}\mathcal{B}L^\infty$$

where $l = \#$ poles of \bar{D} , and

$$G_1 = Q_1^2 \left[\frac{-D}{(1 - |D|^2)V^2} - \frac{[\bar{D}]_l}{V^2} \right] = Q_1^2 \langle \bar{D} \rangle.$$

Now all that we must do is substitute $G = Q_1^2 \langle D \rangle$, $\Phi = 1$, and $\mathcal{G}^1(F)$ into (4.7). Abbreviate $\mathcal{G}_0^D(F)$ by Γ , note that $\mathcal{G}^1(F) = Q_1^2 \Gamma / (1 - |D|^2)V^2$, note that $|\mathcal{G}^1(F)| = |\Gamma|$ and use footnote 10 to evaluate the coefficients of (4.7):

$$\Phi C_{\Gamma,r} = (1 - r) \frac{Q_1^2 \Gamma}{(1 - |D|^2)V^2} (1 - r|\Gamma|^2)^{-1},$$

$$R_{\Gamma,r}^2 = r(1 - |\Gamma|^2)(1 - r|\Gamma|^2)^{-1}, \quad P_{\Gamma,r}^2 = (1 - r|\Gamma|^2)(1 - |\Gamma|^2)^{-1}.$$

With these values (4.7) becomes

(4.12)

$$\left| Q_1^2(1-r) \frac{\Gamma(1-r|\Gamma|^2)^{-1}}{(1-|D|^2)V^2} - Q_1^2\langle \bar{D} \rangle - L \right|^2 \leq r(1-|\Gamma|^2)^2(1-r|\Gamma|^2)^{-2}$$

after dividing by (the scalar function) $P_{\Gamma,r}^2$. Next divide by the outer function Q_1^2 to get

$$(4.13) \quad |C - L_1|^2 \leq \frac{r(1-|\Gamma|^2)}{(1-|D|^2)^2(1-r|\Gamma|^2)}$$

where L_1 is an arbitrary function in H^∞ . Theorem 4.4 follows directly by applying Theorem 4.2 to the disk determined by (4.13). \square

Throughout this entire section we have never mentioned interpolation. Instead we have used the equivalent sup norm approximation view toward the subject. Some readers are probably puzzled by this because §3 was done almost entirely from the interpolation viewpoint. The reason is that if one works on problems like these but with added constraints, the constraints fit better into the interpolation framework. In §6 we shall present a variety of problems; the easier ones lead to the problems we just treated, the harder ones absolutely require interpolation. Thus the only unified way to cover a large territory is through interpolation. We conclude this section by pursuing the interpolation approach to Theorem 4.1 and writing down the explicit matrices A, B which it gives. As usual we stick to a special case.

THEOREM 4.5. *Suppose F is in $\mathcal{R}\mathcal{B}L^\infty$. Let Φ be an inner function in $\mathcal{R}H^\infty$ minimal with respect to the property that $\Phi\langle \bar{F} \rangle$ lies in $\mathcal{R}H^\infty$, and let z_1, \dots, z_N denote the zeroes of Φ which we assume to have multiplicity one. Let Q be the outer Wiener-Hopf factor of $1 - |F|^2$. Let w_j be the function $\Phi Q^2\langle \bar{F} \rangle$ evaluated at z_j . Then Theorem 4.1 holds with*

$$A = \left\{ \frac{(w_i, w_j)}{1 - z_i \bar{z}_j} \right\}_1^N, \quad B = \left\{ \frac{1}{1 - z_i \bar{z}_j} \right\}_1^N.$$

The proof is straightforward and dull so we do not inflict it upon the reader.

5. Orbits and Poincaré contractions. So far we have presented two disjoint topics. First there was an exposition of old fashion filtering theory, then came a solution to a mathematical problem on Poincaré distance. At the moment these appear to be disjoint topics, but they aren't and we now begin a campaign to derive the connection. This is done in two sections, §§5 and 6. §5 presents two purely mathematical results; one classifies the orbits of the group of linear fractional maps \mathcal{G}_g with g in $\mathcal{R}U(n, n)$, the other asserts that many $U(n, n)$ invariant metrics on $\mathcal{B}M_n$ are contracted by a \mathcal{G}_g with g in $C(n, n)$. §6 lists a half dozen electronic power transfer problems from §2 and uses everything developed in the paper to either solve or simplify them.

a. *Orbit structure.* As one would expect in studying the transformation group $\mathfrak{R}U(n, n)$ acting on $\mathfrak{R}M_n$, its orbits play an important role. For our purposes it suffices to classify the orbits of the constant functions. To describe them define $M(j, k, l)$ to be those matrices in M_{j+k+l} with j singular values less than 1, k singular values equal to 1, and l singular values greater than 1. Recall the singular values of a matrix m are the eigenvalues of $|m| = (mm^*)^{1/2}$. As usual $\mathfrak{R}M(j, k, l)$ denotes matrices with rational entries whose values (on the circle) lie in $M(j, k, l)$.

THEOREM 5.1 (THEOREM 1.1 IN [B-H1]). *Each set $\mathfrak{R}M(j, k, l)$ is an orbit of $\mathfrak{R}U(n, n)$ with $n = j + k + l$ which intersects the constants. In fact that is all of them.*

Theorem 5.1 has an old engineering result called Darlington's theorem as an immediate consequence.

COROLLARY 5.2. *The orbit of 0 under $\mathfrak{R}U^+(n, n)$ is $\mathfrak{R}M(n, 0, 0) \cap H^\infty(M_n)$, that is, it equals $\mathfrak{B}\mathfrak{R}H^\infty(M_n)$.*

This corollary follows immediately from the fact (in Theorem 5.1) that the orbit of 0 under $\mathfrak{R}U(n, n)$ equals $\mathfrak{R}M(n, 0, 0)$ and a trivial computation which associates to any g in $\mathfrak{R}U(n, n)$ for which $\mathcal{G}_g(0)$ is in $\mathfrak{B}H^\infty(M_n)$ a Blaschke product ϕ so that

$$g' \stackrel{\Delta}{=} g \begin{pmatrix} \varphi I & 0 \\ 0 & I \end{pmatrix}$$

is in $\mathfrak{R}U^+(n, n)$. Now $\mathcal{G}_{g'}(s) = \mathcal{G}_g(\varphi s)$, so $\mathcal{G}_{g'}(0) = \mathcal{G}_g(0)$.

For $n = 1$ this is due to S. Darlington; for arbitrary n it is attributed to Ono-Yasuro, Belevitch, or in Russia to Potopov. It says that any passive n -port can be made by terminating n ports of a certain $2n$ -port lossless circuit in n unit resistors.

There is a further useful property of linear fractional maps. It describes how they behave with respect to 'Schwartz reflection' $\tilde{s} = s^{*-1}$ of matrices.

LEMMA 5.3. *If $g \in \text{GL}(2n)$ then $\widetilde{\mathcal{G}}_g(s) = \mathcal{G}_{g_0}(\tilde{s})$ where $g_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \tilde{g} \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}$. Consequently if $g \in U(n, n)$, then $g_0 = g$.*

The proof is straightforward algebra and can be found (in greater generality) in [B-H1, Lemma 1.2].

Hopefully, Theorem 5.1 is sufficiently intuitive that we can omit its proof without disorienting the reader. The cleanest proof [B-H5] uses the symplectic Beurling-Lax-Halmos theorem of §3(f). Straightforward algebra would prove the theorem except for the fact that the square root of a rational function may not be rational. This causes considerable difficulties with the brute force approach.

c. *Invariants.* It was demonstrated by C. L. Siegel that the 'cross ratio' of two matrices is preserved (up to similarity) under maps \mathcal{G}_g with g in $U(n, n)$. A more general cross ratio of four matrices z_1, z_2, z_3, z_4 was introduced by Youla; it is

$$\mathcal{C}(z_1, z_2, z_3, z_4) = (z_1 - z_2)(z_2 - z_3)^{-1}(z_3 - z_4)(z_4 - z_1)^{-1}$$

whenever the necessary inverses exist. It is invariant under a linear fractional mapping up to similarity. That is, if $z' = \mathcal{F}(z)$ then $\exists m$ invertible so that

$$\mathcal{C}(z'_1, z'_2, z'_3, z'_4) = m\mathcal{C}(z_1, z_2, z_3, z_4)m^{-1}.$$

J. H. C. van Heuven and T. Z. Rozzi [H-R] described a very general procedure for using \mathcal{C} to generate invariants. Suppose τ_1 and τ_2 are transformations on M_n and that \mathcal{F} is a linear fractional map on M_n (that is, $\mathcal{F} = \mathcal{G}_g$ for a $g \in GL(2n)$) which intertwines τ_1 and τ_2 ; that is $\tau_1\mathcal{F} = \mathcal{F}\tau_2$. Define

$$\mathcal{E}_j(z_1, z_2) = \mathcal{C}(z_1, z_2, \tau_j(z_1), \tau_j(z_2)).$$

Then clearly

$$\mathcal{E}_1(\mathcal{F}(z_1), \mathcal{F}(z_2)) \quad \text{and} \quad \mathcal{E}_2(z_1, z_2)$$

are similar. In particular if $\tau_1 = \tau_2$, then \mathcal{E}_1 is a similarity invariant for \mathcal{F} .

This invariance applies readily to linear fractional \mathcal{G}_g with g in $U(n, n)$. This is because the \sim operation defined by $\tilde{m} = m^{*-1}$ commutes with \mathcal{G}_g ; that is $\mathcal{G}_g(\tilde{m}) = \widetilde{\mathcal{G}_g(m)}$ (see Lemma 5.3). Thus

$$\mathcal{E}(z_1, z_2) \stackrel{\Delta}{=} \mathcal{C}(z_1, z_2, \tilde{z}_1, \tilde{z}_2)$$

is invariant under \mathcal{G}_g . In fact \mathcal{E} is essentially the C. L. Siegel cross ratio. All eigenvalues of it are positive and

$$p(s, h) = \text{arctanh}[\text{b.e.v. } \mathcal{E}(s, h)]^{1/2}$$

is the Carathéodory metric defined at the beginning of §4. Thus *the mappings \mathcal{G}_g with g in $U(n, n)$ are isometries in the metric p .*

There is a generalization of these invariance properties of physical importance. If \mathcal{F} is any analytic map of $\mathfrak{B}M_n$ into $\mathfrak{B}M_n$ then \mathcal{F} is a contraction in the metric p , that is,

$$p(\mathcal{F}(s), \mathcal{F}(h)) \leq p(s, h)$$

for all s, h in $\mathfrak{B}M_n$. The generalizations of it to domains other than $\mathfrak{B}M_n$ are treated by Earle and Hamilton [E-H] and Kobayashi [K]. This result on the Carathéodory metric concerns only the biggest eigenvalue of $\mathcal{E}(s, h)$, but there are circumstances when we need to deal with all of the eigenvalues of $\mathcal{E}(s, h)$. In particular let $\lambda_l[m]$ denote the l th from largest eigenvalue of the selfadjoint matrix m and call an analytic mapping $\mathcal{F}: \mathfrak{B}M_n \rightarrow \mathfrak{B}M_n$ *totally contractive* provided

$$\lambda_l[\mathcal{E}(\mathcal{F}(s), \mathcal{F}(h))] \leq \lambda_l[\mathcal{E}(s, h)]$$

for all l . The paper [B-H3] describes exactly which \mathcal{F} are totally contractive. A consequence is

THEOREM 5.5 (COROLLARY 3.1 IN [B-H3]). *If g is in $\mathcal{C}(n, n)$, then the linear fractional map \mathcal{G}_g is totally contractive.*

6. Physical problems and how our machinery applies to them. In this section we shall state a variety of mathematics problems which we wish to solve. Each one arises from a power gain problem in §2 and requires the optimization of a

particular functional on an orbit of $\mathfrak{R}U^+(n, n)$ or $\mathfrak{R}C^+(n, n)$ possibly with some constraints. Most can be reduced to a more appealing form, namely, that of finding the Poincaré distance of a given function F in $\mathfrak{B}L^\infty \mathfrak{B}M_n$ to a certain subset $E \subset \mathfrak{B}H^\infty(M_n)$. Some of the Poincaré distance problems have been solved explicitly and all of them can be analyzed to produce qualitative results. This section is broken into six parts, a, b, c, d, e, and f. Each part presents a problem about orbits and then reduces it to a Poincaré distance problem. Also, since each problem comes directly from circuit theory, we indicate physical motivation in each case so an interested reader can refer to §2 and see how the problem fits in. We begin with the simplest and most basic problem.

a. *The broadband impedance matching problem for n -ports* (see Introduction (ii) and §2(d), problem (a) with $Q = 0$) gives rise directly to the mathematics problem:

Given S in $\mathfrak{R} \mathfrak{B}H^\infty(M_n)$, find

$$\mu \stackrel{\Delta}{=} \inf_{g \in \mathfrak{R}U^+(n, n)} \|\mathfrak{G}_g(S)\|$$

and the optimizing g .

A few principles from §5 easily reduce this to computing a Poincaré distance as we now see. Observe

$$\operatorname{arctanh} \mu = \inf_{g \in \mathfrak{R}U^+} \rho(0, \mathfrak{G}_g(S)) = \inf_{g \in \mathfrak{R}U^+} \rho(\mathfrak{G}_{g^{-1}}(0), S).$$

By Corollary 5.2 and the easily checked fact that g^{-1} sweeps $\mathfrak{R}U^-$ as g sweeps $\mathfrak{R}U^+$ we get

$$\operatorname{arctanh} \mu = \inf_{H \in \mathfrak{R} \mathfrak{B}H^\infty(M_n)} \rho(H, S^*).$$

Also from the optimizing H one can construct the optimizing g in the original problem (using the construction behind Theorem 5.2). Thus the problem is

Find the ρ -closest point in $\mathfrak{R} \mathfrak{B}H^\infty(M_n)$ to the function S^ and the ρ -distance.*

This is solved by Theorems 4.1, 4.1', 4.1'' and 4.4 and settles the broadband matching problem in several different ways.

b. *Optimizing the gain of a reflection type amplifier using lossless circuitry* (§2(d) Figure 2.13 with g lossless) hands one the mathematical problem:

Given S in $\mathfrak{R}H^\infty(M_n)$ with

$$\inf_{\theta} \inf_{\|x\|=1} \|S(e^{i\theta})x\|_{\mathbb{C}^n} > 1,$$

find

$$\gamma_S \stackrel{\Delta}{=} \max_{g \in \mathfrak{R}U^+(n, n)} \inf_{\theta} \inf_{\|x\|=1} \|\mathfrak{G}_g(S)(e^{i\theta})x\|_{\mathbb{C}^n}.$$

That is find the amplifier whose minimum gain (over all modes of operation is the largest).

The reduction starts with the observation that for an invertible matrix m

$$\inf_{\|x\|=1} \|mx\|_{\mathbb{C}^n} = \frac{1}{\|m^{-1}\|_{M_n}} = \frac{1}{\|\tilde{m}\|_{M_n}}.$$

Use this trick and then Lemma 5.3 to get

$$\gamma_S = \max_g \frac{1}{\|\widetilde{\mathcal{G}}_g(\tilde{S})\|_{L^\infty(M_n)}} = \max_g \frac{1}{\|\mathcal{G}_g(\tilde{S})\|_{L^\infty(M_n)}}$$

so

$$(6.1) \quad \frac{1}{\gamma_S} = \inf_{g \in \mathfrak{R}U^+} \|\mathcal{G}_g(\tilde{S})\|_{L^\infty(M_n)}.$$

This is exactly the same problem as in part (a) except that \tilde{S} need not belong to H^∞ . Instead \tilde{S} is in $\mathfrak{B}H_l^\infty(M_n)$ where l is the number of zeroes of S outside the unit disk. The reduction from here on is the same as in part (a) only we obtain not \mathcal{G}^{S^*} but $\mathcal{G}^{S^{-1}}$ which corresponds to a g in $\mathfrak{R}U(n, n)$ but not in $\mathfrak{R}U^+(n, n)$. Again we get that the crux of the matter is a Poincaré distance problem. Namely,

Given S^{-1} rational, find

$$\operatorname{arctanh} \frac{1}{\gamma_S} = \inf_{H \in \mathfrak{B}H^\infty(M_n)} \rho(H, S^{-1})$$

and find the optimizing H .

This is solved by Theorems 4.1, 4.1', 4.3 and 4.4 when $n = 1$; for the general case see Theorem 5.1 in [H1].

c. *Passive equalization.* A common variation on (a) and (b) is to let g sweep through $\mathfrak{R}C^+(n, n)$ rather than $\mathfrak{R}U^+(n, n)$ as we did before (actually this is the problem presented in §2(d)). Problem (a) with this variation proves to be vacuous so we concentrate on problem (b). The precise problem now is to find

$$\gamma'_S \stackrel{\Delta}{=} \max_{g \in \mathfrak{R}C^+(n, n)} \inf_{\theta} \inf_{\|x\|=1} \|\mathcal{G}_g(S)(e^{i\theta})x\|_{C^n}.$$

The physical circumstance is the same as before but now in the basic configuration shown in Figure 2.13 we allow g to be any passive (rather than energy conserving) circuit.

The reduction goes just as before to give

$$\frac{1}{\gamma'_S} = \inf_{g \in \mathfrak{R}C^+(n, n)} \left\| \mathcal{G}_{\binom{l}{0} \binom{0}{-l}} \tilde{g}_{\binom{l}{0} \binom{0}{-l}}(\tilde{S}) \right\|_{L^\infty(M_n)}.$$

Now $g_0 \stackrel{\Delta}{=} \left[\binom{l}{0} \binom{0}{-l} \tilde{g}_{\binom{l}{0} \binom{0}{-l}} \right]^{-1}$ is pointwise in $C(n, n)$ so Theorem 5.5 says that \mathcal{G}_{g_0} is a Poincaré contraction; thus

$$(6.2) \quad \operatorname{arctanh} \frac{1}{\gamma'_S} \geq \inf_{g \in \mathfrak{R}C^+(n, n)} \rho\left(\mathcal{G}_{\binom{l}{0} \binom{0}{-l}} \tilde{g}_{\binom{l}{0} \binom{0}{-l}}(0), \tilde{S}\right).$$

For g in $\mathfrak{R}C^+(n, n)$ one can readily check that

$$(6.3) \quad H \stackrel{\Delta}{=} \mathcal{G}_{\binom{l}{0} \binom{0}{-l}} \tilde{g}_{\binom{l}{0} \binom{0}{-l}}(0)^*$$

is in $\mathfrak{B}H^\infty(M_n)$. So by Corollary 3.2 we have

$$(6.4) \quad \operatorname{arctanh} \frac{1}{\gamma'_S} \geq \inf_{H \in \mathfrak{B}H^\infty(M_n)} \rho(H, S^{-1}).$$

However, the right side equals $\operatorname{arctanh}(1/\gamma_S)$ from part (b). Therefore we conclude $\gamma_S = \gamma'_S$.

Thus we have proved the intuitively plausible fact that one cannot get more gain by using passive equalizing circuits than by using only lossless equalizing circuits. This phenomenon persists through many types of amplifier design and the reason always is that the \mathcal{G}_g from $C(n, n)$ are totally Poincaré contractive as in Theorem 5.5 or in a stronger sense [B-H3]. For more on this see [B-H1, §8].

d. *Stability: There are very important restrictions which one can add in problems (b) and (c).* Ultimately problems (b) and (c) amounted to computing the Poincaré distance from a given function F in $\mathfrak{R} \mathfrak{B}L^\infty(M_n)$ to $\mathfrak{R} \mathfrak{B}H^\infty$. The problem we shall now discuss ultimately reduces to computing the distance from F to the subset

$$E_F = \{H \in \mathfrak{B}H^\infty(M_n) : H \text{ and } F \text{ never agree}\}$$

of $\mathfrak{B}H^\infty(M_n)$. Here the term H and F never agree means that $H(z) = F(z)$ is an invertible matrix for each z in $|z| < 1$. This turns out to be quite a nice mathematical problem which is entirely open when $n > 1$. There is strong motivation for studying it as we now see.

In problems (a)–(c) we were given a function $S \in H^\infty(M_n)$; then we acted on it by $\mathcal{G}_g(S)$ where g was in $\mathfrak{R}U^+(n, n)$ or $\mathfrak{R}C^+(n, n)$. When $S \in \mathfrak{B}H^\infty(M_n)$ then $\mathcal{G}_g(S)$ is automatically in $H^\infty(M_n)$ (even in $\mathfrak{B}H^\infty(M_n)$). When $\|S\|_{L^\infty} \nless 1$, the function $\mathcal{G}_g(S)$ may have poles inside the disk. This is anathema in amplifier design because such a pole corresponds to an instability in the circuit as described in §2(d).

Thus the *all important stability constraint* on the amplifier in Figure 2.13 with coupler g_0 and amplifying element S_0 is that $\mathcal{G}_g(S)$ have no poles in $|z| < 1$ for any g and S which are near to g_0 and S_0 . Consequently the variant of problem (c) of true practical interest is

Given S , find

$$\gamma'_S \triangleq \max_{\substack{g \in \mathfrak{R}C^+(n, n) \\ g, S \text{ stable}}} \inf_{\theta} \inf_{\|x\|=1} \|\mathcal{G}_g(S)(e^{i\theta})x\|_{C^n}$$

and a g which (approximately) produces the maximum.

Now we sketch how one again uses the principles of §5 to convert this problem to the E_F -Poincaré distance problem just cited.

The first step is to analyze the stability constraint and phrase it in convenient terms. To do this write $\mathcal{G}_g(S) = A + BS(1 - DS)^{-1}C$ as in §3(d). If g is in $\mathfrak{R}C^+(n, n)$, then one can take A, B, C, D to be in $H^\infty(M_n)$. Thus if $(1 - DS)^{-1}$ is in H^∞ , the function $\mathcal{G}_g(S)$ will be too. The converse is not true because B, S or C could cancel singularities of $(1 - DS)^{-1}$. However, this cancellation is destroyed by perturbation of A, B, C, D and S , so in practice the pair g, S is stable if and only if $(1 - DS)^{-1}$ lies in $H^\infty(M_n)$. It is straightforward to verify that $(1 - DS)^{-1}$ is in $H^\infty(M_n)$ if and only if D and S^{-1} do not agree (see [B-H2]).

The function D has a neat characterization in terms of \mathcal{G}_g , namely,

$$(6.5) \quad \mathcal{G}_{\binom{1}{0} \binom{0}{-1}} \binom{1}{0} \binom{0}{-1}(0) = D^*.$$

To check this note that the left side of (6.5) is

$$\mathfrak{G}_{(-\beta^* \gamma^*)}(0) = -\kappa^* \gamma^{*-1}.$$

The conversion rule (3.5) from \mathfrak{G}_g to \mathfrak{F}_u simply says that $-\kappa^* \gamma^{*-1} = D^*$.

Now we convert the stable amplifier problem (d) to the Poincaré distance problem. One simply tracks through the reduction in (c) and observes the effect of the stability constraint. The crucial place is in passing from inequality (6.1) to inequality (6.4). The statement (6.5) quite miraculously is the same as the statement (6.3) and says that H in (6.3) equals D . Thus (6.4) for the stable g , S which we are considering is

$$\operatorname{arctanh} \frac{1}{\dot{\gamma}'_S} \geq \inf_{H \in \mathfrak{R} E_{S^{-1}}} \rho(H, S^{-1}).$$

Likewise, problem (b) with the stability condition reduces to the same Poincaré function space approximation problem; namely, the optimum for (b) with stability equals $\dot{\gamma}'_S$ rather than $\dot{\gamma}_S$.

The crucial problem is

Find

$$\operatorname{arctanh} \frac{1}{\dot{\gamma}_S} = \inf_{H \in \mathfrak{R} E_{S^{-1}}} \rho(H, S^{-1})$$

and the optimizing H .

Now we have finally come to a problem which is not solved in §4 and which is in fact not fully solved. The attack on Poincaré distance problems presented in §4 still applies here but the interpolation view is now more appropriate than the sup norm approximation view emphasized in §4. To convert the crucial problem to a constrained interpolation problem we again track through the basic reduction described in §4(b). We now review a special case of that reduction quickly.

Set $F = S^{-1}$ and consider the issue of computing the ρ -distance δ_F of F to $\mathfrak{B}H^\infty$. The mapping \mathfrak{G}_0^F given by (4.9) takes F to 0, consequently

$$\rho(H, F) = \rho(\mathfrak{G}_0^F(H), 0) = \operatorname{arctanh} \|\mathfrak{G}_0^F(H)\|_{L^\infty(M_n)}.$$

Theorem 3.1 tells us that $\mathfrak{G}_0^F(H)$ sweeps through a set of interpolating functions

$$\mathfrak{T}_{\Phi, \psi, l}(G) \cap \mathfrak{R} \mathfrak{B}L^\infty(M_n)$$

as H sweeps through $\mathfrak{R} \mathfrak{B}H^\infty(M_n)$; abbreviate this set by Ξ . Thus computing δ_F becomes the problem of finding the sup norm smallest element of Ξ and this is just (a very general version of) the classical Nevanlinna-Pick interpolation problem settled in §3.

The same argument together with the *fact*¹² that

$$\mathfrak{G}_0^F(\mathfrak{R} E_F) \text{ contains } \Phi_1 \{K \in \mathfrak{G}_0^F(\mathfrak{R} \mathfrak{B}H^\infty(M_n)) : K \text{ never agrees with } 0\} \Phi_2$$

¹²This seems plausible because $\mathfrak{G}_0^F(F) = 0$, so we would expect $\mathfrak{G}_0^F(H)$ not to agree with 0 for any H which does not agree with F . This is basically true up to some fluky cancellations: that's why the statement reads 'contains' rather than 'equals'. A complete proof is in [B-H2].

densely for some functions Φ_j with $\Phi_j(e^{i\theta})$ unitary says that computing the ρ -distance of F to E_F is equivalent to computing the sup norm of the smallest function in $\mathfrak{R}\hat{\Omega}_F$ where $\hat{\Omega}_F$ is the set

$$\hat{\Omega}_F = \Xi \cap \{K \in \mathfrak{R}\mathfrak{B}H^\infty(M_n): K(z) \neq 0 \text{ for } z \in \Delta\}.$$

Thus the issue is not just Ξ as before but the set of functions in Ξ which do not vanish. This is exactly the problem addressed in §3(g) and Theorem 3.8 applies directly.

It implies that no smallest K in $\mathfrak{R}\hat{\Omega}_F$ exists. However, there is a minimizing sequence K_j and a normal families argument applies to give $K_j \rightarrow K_\infty \in \hat{\Omega}_F$ in a weak sense. One can show that K actually is the smallest function in $\hat{\Omega}_F$. The reduction we gave reverses to show that K_∞ yields a solution H_∞ to the problem of finding the ρ -closest point in E_F to F . So for practical analytical purposes E_F (rather than $\mathfrak{R}E_F$) seems to be the best set to use. The limit function K_∞ is always discontinuous which leads to some physically bizarre consequences which will be discussed in §7.

e. *The common amplifier.* Finding an equalizing circuit to give the amplifier having the configuration in Figure 2.12 with optimal gain immediately gives the following mathematics problem:

Given S in $H^\infty(M_2)$, find

$$G_S = \max_{g \in \mathfrak{R}U^+(2,2)} \inf_{\theta} \left| \left[\mathfrak{G}_g(S)(e^{i\theta}) \right]_{12} \right|.$$

Here m_{12} means the $(1, 2)$ entry of the matrix m . The function S typically satisfies b.e.v. $|S(e^{i\theta})| \gg 1$ and s.e.v. $|S(e^{i\theta})| < 1$ where b.e.v. and s.e.v. denote biggest and smallest eigenvalues.

Our reduction begins by rewriting G_S as

$$G_S = \max_{g \in \mathfrak{R}U^+(2,2)} \max_{U, V \in \mathfrak{R}U(2)} \inf_{\theta} \left\| \left(U(e^{i\theta}) \mathfrak{G}_g(S)(e^{i\theta}) V(e^{i\theta}) \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \right\|_{\mathbb{C}^2}.$$

It is intuitively plausible that the U, V sweep a large enough set to give

$$G_S = \max_{g \in \mathfrak{R}U^+(2,2)} \inf_{\theta} \|\mathfrak{G}_g(S)\|_{M_2}.$$

This indeed is the case (see §6 [H1]). So

$$G_S^{-1} = \inf_{g \in \mathfrak{R}U^+(2,2)} \sup_{\theta} (\text{s.e.v. } |\mathfrak{G}_g(S)|^{-1}).$$

We therefore obtain the exact analog of (6.1)

$$G_S^{-1} = \inf_{g \in \mathfrak{R}U^+(2,2)} \|\text{s.e.v. } |\mathfrak{G}_g(\tilde{S})|\|_{L^\infty(\mathbb{C}^n)}$$

which is a more canonical statement of the problem. Further reduction of the problem proceeds exactly as it did in §6(a), (b) and yields

$$G_S^{-1} = \inf_g \|\text{s.e.v. } \mathfrak{G}(\mathfrak{G}_g^{-1}(0)^*, S^{-1})\|_{L^\infty(\mathbb{C}^1)}.$$

Thus the ‘Poincaré approximation’ analog here is to find

$$G_S^{-1} = \inf_{H \in \mathfrak{R}\mathfrak{B}H^\infty(M_2)} \|\text{s.e.v. } \mathfrak{G}(H, S^{-1})\|_{L^\infty}.$$

Again one can reduce this to minimizing something over a set of interpolating functions. To convert this to an interpolation problem requires more machinery than we have developed. The reason is that one must generalize most results presented to a signed bilinear form. So now we give only the beginnings of the reduction.

Theorem 5.1 gives a $\mathcal{G}_1^{S^{-1}}$ which maps S^{-1} to $\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \end{pmatrix}$ and one can use Corollary 5.2 and the invariance of \mathfrak{E} to obtain

$$G_S^{-1} = \inf_{H \in \mathfrak{B} \mathfrak{R} H^\infty(M_2)} \left\| \text{s.e.v. } \mathfrak{E} \left(\mathcal{G}_1^{S^{-1}}(H), \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \end{pmatrix} \right) \right\|_{L^\infty(\mathbb{C})}.$$

In the earlier Poincaré distance problem we had $\mathcal{G}_0^{S^{-1}}$ instead of $\mathcal{G}_1^{S^{-1}}$ and consequently the matrix $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ instead of $\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \end{pmatrix}$; so the expression involving \mathfrak{E} simplified immediately. To simplify the expression here we apply the constant coefficient L.F.T. denoted \mathfrak{K} which maps $\begin{pmatrix} 0 & 0 \\ 0 & \sqrt{2} \end{pmatrix}$ to $\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ and obtain $\mathfrak{K}(\mathcal{G}_1^{S^{-1}}(\mathfrak{B} H^\infty(M_2)))$ as the crucial set of functions. The map \mathfrak{K} is not in $\mathfrak{R}U(2, 2)$ and the machinery needed to analyze its range space, its relationship to \mathfrak{E} , etc. is developed in [B-H1]. That machinery reduces this problem to a rather advanced interpolation problem.

f. *The stable amplifier.* It is no surprise that the stability constraint should also be applied to the common amplifier in §6(d). Analyzing its effect is a straightforward combination of procedures in §§6(d) and 6(e). So we don't drag the reader through the boring details, but simply state that optimization problems corresponding to the design of a stable (common) amplifier with maximum gain is

Given $S \in \mathfrak{R} H^\infty M(1, 0, 1)$, find

$$(6.6) \quad \dot{G}_S^{-1} = \inf_{H \in \mathfrak{R} E_{S^{-1}}} \|\text{s.e.v. } \mathfrak{E}(H, S^{-1})\|_{L^\infty}.$$

Most linearized amplifiers must be stable and have the form in Figure 2.12 which we are studying but with some passive rather than lossless g . Combination of the procedures in §6(c)–(e) apply easily to prove that the maximum gain obtainable from an amplifier of this type again equals (6.6). As in parts (c) and (d) the machine in [B-H1] gives us the same interpolation problem in part (c) but with the added constraint that the interpolating matrix functions have invertible values everywhere on the disk.

7. Physical conclusions. We begin this section with two details of physical importance which mathematically are not enlightening. After discussing them we indicate physical results. It should be clear from §6 that theoretical problems in designing circuits with maximum gain are mathematics problems of a surprisingly canonical type. §4 gave explicit (and numerically efficient) solutions to several of these problems. These methods also give some simple qualitative principles of gain equalization (even for problems too complicated to be explicitly solvable). In this section we sketch the physical consequences of our mathematical studies. We shall not list detailed formulas and tedious recipes, since the solutions of the physical problems are much in the spirit of §4, and they can be found in [H3]. Instead we indicate which problems have

complete explicit solutions, we give references, and then we devote most of our attention to qualitative principles.

(a) *Added realism.* The first item is that all frequency response functions for circuits are real on the real axis. After transforming to the unit disk this gives functions in $\mathfrak{B}H_1^\infty(M_n)$ which are real on the real axis, that is, have real Fourier coefficients. The constraint this produces changes the structure we have successfully analyzed very little; if you can solve the problem without the constraint, then you can solve it with the constraint. Thus we have given this issue little attention. Quite possibly as the more complicated situations are fully analyzed insuring that the solutions are real on the real axis will become a more serious problem.

The second item is that instead of computing distances in the ρ -metric which is a supremum type metric one could use a weighted supremum Poincaré metric. Theorem 4.2 actually settles this question and all of the theorems in §4 (to include Theorems 4.1, 4.1', 4.1'' and 4.4) with slight modification deal with weighted Poincaré metrics as effectively as we dealt with the unweighted ones. This allows us to treat the physical problem of building a circuit with prescribed gain function rather than just maximum gain function; this is problem (b), §2. It is a bit surprising at first, but it turns out that one can get results for this problem exactly as strong as the ones we have obtained for the gain maximization problem.

The mathematics in §§3–6 modified in the ways just described produces the following physical results.

(b) *Power transfer in passive circuits.* The problem of transferring a prescribed amount of power $\Gamma(\omega)$ to a strictly passive load S from a source with $Q = 0$ (see Figure 2.11, problem (b), §2) is completely solved. It corresponds to the $l = 0$ case of results in §4 for a weighted Poincaré metric.

There is a matrix test as in Theorems 4.1, 4.2, 4.3, 4.4, to determine if a given power transfer ratio G can be delivered to S . If it can, then several algorithms are in principle capable of finding the specifications (frequency response function) for the optimal 'equalizing' circuit g in Figure 2.11. For explicit formulas, see §5, [H1] and [H3].

These quantitative results produce quickly several general properties: Fix a load circuit with rational S .

MONOTONICITY PRINCIPLE [H2]. *Suppose G is a rational function, $0 \leq G < 1$. If there is an energy conserving coupling circuit g with power transfer ratio to S greater than (or equal to) G for all frequencies, then there is an energy conserving g_1 which gives power transfer ratio to S identically equal to G .*

EXTREME MODULUS PRINCIPLE (FOR ONE PORTS). *Fix rational $G \geq 0$. Let κ_0 be the largest number κ for which κG is a physically realizable gain (with energy conserving g). Then there is NO physically realizable gain G_1 strictly bigger than $\kappa_0 G$. For example, if one chooses g_0 to maximize $\max_g \inf_\omega G_g(\omega)$, then the resulting gain G_{g_0} is constant.*

To prove the Monotonicity Principle simply consult Theorems 4.1 and 4.1''; the gain G is physically realizable if and only if $A \leq \mathfrak{X}_{(1-G)}$. An elementary

property of positive Toeplitz operators— $\mathfrak{X}_{G_1} \geq \mathfrak{X}_G$ if $G_1(\omega) \geq G(\omega)$ for all ω —implies the principle. The Extreme Modulus Principle follows from the fact that the $\|\cdot\|_{L^\infty}$ smallest interpolating function has constant modulus.

The engineering theory of impedance matching is mostly due to Fano (1950) [Fn] and Youla (1964) [Y1]. It effectively reduces the 1-port matching problem to interpolation. This is what appears in the textbook accounts [K-R, C, W]. The texts attack the interpolation problem by converting it to a highly nonlinear system of inequalities in many variables which they do not attempt to solve. In fact, a beginning step in the standard recipe is simply to make a guess at the maximum gain possible in Figure 2.11. If one were to apply Nevanlinna-Pick theory to the interpolation problem obtained in [F, Y1] one would ultimately obtain the $n = 1$ case of Theorem 4.5 which would compute the gain. Theorems 4.2, 4.2' and 4.4 which also compute the gain are a radical departure from the classical approach.

Nevanlinna-Pick theory entered the engineering literature in a 1967 article by Saito and Youla [S-Y]. They showed how interpolation theoretic constructions correspond to circuit theoretic ones. They also mention that Pick's theory applies to broadband matching, but unfortunately this and Pick's theory were ignored in the electronics community. Chen's 1976 text [C] on impedance matching uses Pick's theorem to prove the Extreme Modulus Principle for constant gains provided $n = 1$ and all poles of $\bar{s}(1 - |s|^2)^{-1}$ are order one, a fact essentially due to Youla [Y1]; but to compute the optimum gains or an optimizing complex g one is directed to the nonlinear inequalities. Computation of optimum gains from eigenvalues first appeared in [H4, H5].

(c) *Amplifiers*. The situation we analyze is always this. You are given a linear amplifying device and you want to add on lossless circuitry so that the result is a good amplifier for use with a simple source (unit internal resistance) and a simple load (a resistor); see Figures 2.12 and 2.13. The aspect we analyze is that of designing lossless circuitry, called equalizing circuitry, to maximize the gain of the amplifier over all frequencies. What is the best gain possible and what are the specifications (frequency response functions) for the equalizing circuit?

(i) *Reflection type amplifiers*. The reflection type amplifier (see Figure 2.13 and §6(b)) without the stability constraint is well understood and the Monotonicity Principle as well as the Extreme Modulus Principle hold for the gain of such amplifiers. Since the stability constraint has overriding importance we don't belabor these results, but move quickly to a discussion of stable amplifiers.

(ii) *Stable reflection-type amplifiers*. Now we turn to the most rudimentary considerations in the design of a stable amplifier. We showed in §6(d) that the key issue is computation of the Poincaré distance from a function F to the set E_F of functions in $\mathfrak{B}H_l^\infty(M_n)$ which don't agree with F . This was reduced to the interpolation problem addressed in §3(e).

This problem is completely open when $n > 1$. When $n = 1$ and $l = 0$ one can obviously use Theorem 3.8 to obtain a solution; when $n = 1$ there is a partial solution. There is little reason to faithfully list this solution here; the reader can deduce it immediately from Theorem 3.8 or actually look it up in

[B-H2]. However, there is an interesting point to be made: while one cannot actually compute the distance from F to E_F or find the closest H_0 in E_F one can make qualitative statements about the closest point. In particular, the error function $F - H_0$ has constant modulus on Π and is continuous except at a finite number of points $e^{i\theta_1}, e^{i\theta_2}, \dots, e^{i\theta_n}$ on Π .

This is very peculiar in light of the underlying physical motivation. After all we started with the problem of designing a very simple type of stable amplifier. We found that the gain function for the 'optimum' amplifier is constant as hoped, but it is not actually achievable, because the existence of discontinuities forces the optimal equalizing circuit g in Figure 2.13 to have a discontinuous frequency response function g . Since a physically realizable g must be a rational function we see that in practice we are forced to approximate g by rational functions g_m . All of this sounds very plausible so far, but the strange thing is that it is *impossible* to approximate g well. In fact, one can prove (by studying the approximation of a singular inner function with rational outer functions) the

SPIKED GAIN PRINCIPLE. *If the gain function $G(\omega)$ for a particular (linear one port strictly active reflection type) amplifier cannot be significantly increased ($G'(\omega) > G(\omega)$) by some judicious modification of the passive circuitry in that amplifier, then there are frequencies $\omega_1, \omega_2, \dots, \omega_n$ at which G is very, very large. The only optimal amplifiers for which no spikes ω_j exist in the gain are those based on amplifying devices whose behavior is completely independent of frequency. For example a (reflection type) amplifier for which $\inf_{\omega} G(\omega)$ is near maximal has nearly constant gain except for a finite number of spikes at points $\omega_1, \omega_2, \dots, \omega_n$ which are determined by the amplifying devices within the amplifier.*

This is actually a stable amplifier version of the Extreme Modulus Principle. In fact for a fixed S the gains and 'ideal' gains obtainable taken together will satisfy something very close to the Monotonicity and Extreme Modulus Principles. The rub is that when one eliminates 'ideal' gains from consideration the principles change because near optimal actual gains have spikes.

Whether this phenomenon has serious practical implications I do not know. In part it derives from the fact that we were maximizing gain over all frequencies. The true physical problem is to maximize gain over an interval $[a, b]$ of frequencies. While we have not discussed such problems in this article such a generalization of the theory here is straightforward. Upon studying optimization of gain over $[a, b]$ one finds that for a particular problem singularities might or might not occur. Since most design problems call for an amplifier with flat gain, one would usually be forced to add the constraint that no singularity θ_j occurs in $[a, b]$, the operating band. Fortunately, one can reduce this constraint to one on the interpolation problem by tracking through the procedure in §6. Miraculously, one obtains a classical interpolation constraint which has been digested in the literature. This permits one to analyze the new problem to the extent that the original one was analyzed. For example, one can prove that nearly maximum gains with no singularities are nearly constant.

The main theoretical engineering work on stable reflection amplifiers is [Y3] and [C-K]. It converts the problem to that of finding a function which meets a set of constraints several typed pages long.

(iii) *The common amplifier.* The study of the common amplifier led us to the problem: Find

$$\inf_{H \in \mathfrak{R} \mathfrak{B} H^\infty(M_2)} \|\text{s.e.v. } \mathfrak{G}(H, S^{-1})\|_{L^\infty}$$

where $S^{-1} \in \mathfrak{R} M(1, 0, 1)$. This actually can be reduced to an interpolation problem using a method which is an exact generalization of the one used in §4(a)–(d) to spaces with a signed bilinear form. That is the subject of much of [B-H1]. It is best accomplished by working in the Grassmannian setting with subspace valued analytic functions and while we could go through the procedure here we haven't since it is too specialized. The interpolation problem which one gets is a natural generalization of the ones solved in §3, but it is well beyond the present state of the art. Thus explicit solution of the common amplifier problem awaits a considerable advance in interpolation theory. However, existing interpolation theory gives qualitative information.

PRINCIPLE. The gain of a nearly optimal (common linear amplifier) is constant. Also the Monotonicity Principle holds: namely, if it is possible to build an amplifier whose gain is $\geq G(\omega)$ for all ω , then one can build an amplifier whose gain nearly equals G . Also front and back end reflection of this amplifier are nearly 0.

Sadly, lacking from the principle is the constraint that the amplifier is stable. That makes it useless. The author expects that a variant of the spiked gain conjecture holds for the common amplifier. While the spikes may not be in the gain they should show up somewhere. Hopefully, much of the work for understanding this is already done [B-H1]. What is needed are some qualitative results on matrix interpolation and it is too early to see how difficult these will be to obtain.

(iv) *Passive equalization.* Everything stated so far holds whether one is optimizing over all passive equalizing circuits g or only optimizing over lossless ones. Either optimization problem results in the same answer. This was derived in §6(c) for reflection type amplifiers. It is true also for common amplifiers and stable amplifiers and we summarize all this as

DISSIPATION PRINCIPLE. *If any (stable) amplifier which is built with passive equalizing circuitry has a certain gain function G , then there is a (stable) amplifier built with the same amplifying device S and of the same type, but built with lossless circuitry g which has gain greater than or equal to G . In other words, you can't get better gain with passive circuitry than with lossless circuitry.*

It is amusing to note that for 1-port reflection amplifiers, this is just a physical interpretation of the old conundrum that every analytic map f of the unit disk onto itself decreases the Poincaré distance between points. Actually the physical principle only requires the mathematical one for the special case of linear fractional f . This is fortunate because the mathematical principle breaks down in higher dimensions for many analytic f but holds for linear fractional

maps in a very strong form: Theorem 5.5 is just one example of this and a stronger theorem is needed to obtain the Dissipation Principle; see [B-H3].

8. Epilogue. Each of the sciences has its own character and this influences the type of mathematical support it requires. Electrical engineering (and mathematics for that matter) have the distinctive feature that they are far less determined by nature than, say, physics or biology. The works of nature which are the subject of modern physics are more orderly and in a way more limited than those works of man which constitute engineering. Whereas physics is broken into a few large parts each trying to find the fundamental principles which underlie (and always will underlie) a basic phenomenon of nature, engineering is by comparison chaotic.

For the sake of discussion let us distinguish between two types of engineering. We will call one applied physics; the other black box engineering. Applied physics brings to mind hull design, wing design, various types of electromagnetic theory and things of that sort. By black box engineering I have in mind electronics and communications equipment—the sort of high technology that most of us buy in the store. It is this second field which is something of a circus. The reason for its wild nature is that every few years the technology changes radically. Just imagine what the field of aerodynamics and wing design would be like if the Navier-Stokes equations completely changed every five years. In the electronics field there are¹³ distinguished people who feel that journals over 10 years old are useless and might as well be burned.

With everything changing all the time can there be any theory of enduring importance? Probably most people think so and are prepared to argue violently about what and how much there is of it. Much of the leadership in academic electronics theory is engaged in a valiant struggle to keep theoretical efforts pointed at modern electronics. One might argue that electronics is going through one of those golden periods as physics did in the 1920's when producing quantum mechanics. Several powerful devices with radically new behavior have been developed within the last few years and quantitative improvements have been so profound as to produce qualitative revolutions in practice. In my singular view the implication for theoreticians is simple—there aren't enough of them. Most electronics theoreticians are rushing to the front to figure out how one takes advantage of the marvelous new charge coupled devices, the new digital filter technology, how to sensibly put a million transistors on a chip, and how to cope with the nonlinear behavior of many modern circuits. The gap this leaves in the study of basics presents a good opportunity for mathematicians.

This exaggerated description of the backdrop in front of which black box engineering plays was designed to highlight distinctions with subjects more familiar to mathematicians. In fact the electronics community consists of various camps, some with a reverence for basics, a few with a fondness for great abstraction, many who think that only the newest work will apply to new circuits and many who don't like theory much at all. I have a vested interest in

¹³According to a survey performed by R. Brockett during coffee breaks at MIT.

basics, so as you might guess I believe they are worth doing. Historically, they have fared well.

The evolution of the power transfer theory treated in this paper may contain a lesson. Interest developed in the late 1940's after the advent of radar. There was more good work in the early 1960's prompted by the discovery of the tunnel diode. Hopes for the tunnel diode faded and the basic problems frequently were said to be solved by influential theoreticians even though no recipe existed and the proposed solutions required guessing. So work stopped. Then in the middle 1970's came a revolution in broadband microwave devices which produced transistors badly in need of gain equalization and produced a serious need for some theory. Unfortunately the theory from the 1960's hadn't been developed or polished to the point of being widely useful beyond compromised versions of the defunct tunnel diode. Consequently the theory of broadband gain equalization had little industrial impact. One might speculate that if work in the 1960's had been encouraged beyond the problem at hand procedures might have been available when needed in the 1970's. Who knows. Anyway one could argue that there is practical value in supplementing the goal oriented impatience of down to earth engineering with a study of structure for its own sake.

Another point is that while technologies come and go some of the mathematics underlying them has an eerie permanence. For example, while modern digital filters and classical analog filters have completely different components much of the theory of digital design is a transform of classical analog theory (this business is in the spirit of problem (iv) in the Introduction).

We conclude this discourse by observing a peculiarity in the relationship between mathematics and black boxes. While nature does not imitate mathematics, black boxes frequently do. While a physicist spends his days finding the mathematics which best describes nature, much of engineering is devoted to making devices behave in simple manageable ways so that they fit easily into a mathematical design procedure. For example, engineers like devices which correspond to a simple function into which more complicated functions can be easily decomposed. Thus the mathematics of the decomposition theory to some extent dictates the type of devices which are built. While this seems to be a distinction between electronics and most other branches of science, I'm not sure what implications it has. An optimist like myself thinks it means that mathematics has a significant practical role to play. Another possible implication is that mathematicians should be a little more open minded about some of the rather formal and seemingly contrived applications of mathematics. I can recall dismissing out of hand some engineering talks on "hemidemi groupoids of rings over..." and later discovering that the fellow had some pretty decent physical examples in mind. On the other hand, most things along that line are indeed as they appear—flakey. It's best to always press for a naive example and a little history of the topic. These ups and downs remind us of the first point in this diatribe—the black box business compared to a gently flowing subject like mathematics is a bit wild. While parts of it are of little mathematical interest other parts are fascinating and shimmer with the allure of high rewards.

BIBLIOGRAPHY

- [A] D. Z. Arov, *Darlington realization of matrix-valued functions*, *Izv. Akad. Nauk SSSR Ser Mat.* **37** (1973), 1299–1331; English transl. in *Math. USSR Izv.* **7** (1973).
- [A-A-K1] V. M. Adamajan, D. Z. Arov and M. G. Krein, *Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem*, *Math. USSR-Sb.* **15** (1971), 31–73.
- [A-A-K2] _____, *Infinite block Hankel matrices and their connection with the interpolation problem*, *Akad. Nauk Armenia SSR Ivestia Mat.* **6** (1971); *Amer. Math. Soc. Transl. (2)* **111** (1978), 133–156.
- [A-C-F] Gr. Arsene, Z. Ceausescu and C. Foiaş, *On intertwining dilations*. VIII, *J. Operator Theory* **4** (1980), 55–91.
- [An] T. Ando, *On a pair of commutative contractions*, *Acta Sci. Math. (Szeged)* **24** (1963), 88–90.
- [B] J. Ball, *Interpolation problems of Pick-Nevanlinna and Löwner types for meromorphic matrix functions* (preprint available).
- [B-H1] J. A. Ball and J. W. Helton, *Lie groups over the field of rational functions, signed spectral factorization, signed interpolation, and amplifier designs*, *J. Operator Theory* (to appear).
- [B-H2] _____, *Interpolation with outer functions and gain equalization amplifiers*, *Proc. of the Internat. Conf. on Math. Theory of Circuits and Systems (Delft, July 1979)*.
- [B-H3] _____, *Subinvariants for analytic mappings on matrix balls*, *J. Anal. and Appl.* (to appear).
- [B-H4] _____, *A Beurling-Lax theorem for the Lie group $U(m, n)$ which contains most classical interpolation theory*, *J. Operator Theory* (to appear).
- [B-H5] _____, *Factorization results related to shifts in an indefinite metric* (manuscript).
- [Bdr] J. Border, *Nonlinear Hardy spaces and electrical power transfer*, U.C.S.D. Thesis, 1979.
- [C] W. Chen, *Theory and design of broadband matching networks*, Pergamon, Oxford, 1976.
- [C-K] Y. T. Chan and E. Kuh, *A general matching theory and its applications to tunnel a diode amplifier*, *IEEE Trans. C.T.* **13** (1966), 6–18.
- [C-L] L. Chua and Pen-Min-Lin, *Computer aided analysis of electronic circuits: Algorithms and computational techniques*, Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- [C] D. Clark, *On the spectra of bounded hermitian Hankel matrices*, *Amer. J. Math.* **90** (1968), 627–656.
- [D] W. F. Donoghue, *Monotone matrix functions and analytic continuation*, Springer, New York, 1974.
- [D-F] K. Davis and I. Foiaş, *Operators with bounded characteristic function and their J -unitary dilation*, *Acta Sci. Math. (Szeged)* **32** (1971), 127–139.
- [D-H] R. G. Douglas and J. W. Helton, *Inner dilations of analytic matrix functions and Darlington synthesis*, *Acta Sci. Math. (Szeged)* **34** (1973), 61–67.
- [DeW] P. DeWilde, *Roomy scattering matrix synthesis*, Univ. of California, Berkeley Tech. Report, 1971.
- [D-K] C. A. Desoer and E. S. Kuh, *Basic circuit theory*, McGraw-Hill, New York, 1969.
- [D-M-P] R. G. Douglas, P. Muhly and C. Pearcy, *Lifting commuting operators*, *Michigan Math. J.* **13** (1968), 385–393.
- [D-V] P. DeWilde and T. Vandewalle, *On the factorization of a nonsingular rational matrix*, *IEEE Trans. Circuits and Systems*, CAS-22 (1972), 637–645.
- [E-H] C. Earle and R. S. Hamilton, *A fixed point theorem for holomorphic mappings*, *Global Analysis*, *Proc. Sympos. Pure Math.*, vol. 15, Amer. Math. Soc., Providence, R. I., 1970.
- [E-P] A. V. Efimov and V. P. Potapov, *J -expanding functions and their role in the analytic theory of electrical circuits*, *Russian Math. Survey* **28** (1973).
- [E-H] J. Evans and J. W. Helton, *Applications of Pick-Nevanlinna interpolation theory to retention-solubility studies in the lungs* (preprint available).
- [Fa] R. M. Fano, *Theoretical limitations on broadband matching of arbitrary impedances*, *J. Franklin Inst.* **249** (1960), 57–83; 139–155.
- [F] P. Fuhrman, *On Hankel operator ranges meromorphic pseudo-continuations and factorization of operator analytic functions*, *J. London Math. Soc. (2)* **13** (1975), 323–327.
- [Fd1] I. I. Fedchin, *Description of solutions of the tangential Nevanlinna-Pick problem*, *Doklady Akad. Nauk Armenian SSR* **60** (1975), 37–42.

- [Fd2] _____, *Tangential Nevanlinna-Pick problem with multiple points*, Doklady Akad. Nauk Armenian SSR **61** (1975), 214–218.
- [Fo] C. Foiaş, *Contractive intertwining dilations and waves in layered media*, Proc. Internat. Congr. Math. (Helsinki, 1978), Vol. 1, Acad. Sci. Fenn., Helsinki, 1980, pp. 605–613.
- [G-K] I. C. Gohberg and M. G. Krein, *Systems of integral equations on a half-line with kernel depending on the difference of the arguments*, Amer. Math. Soc., Transl. (2) **14** (1960).
- [H1] J. William Helton, *The distance from a function to H^∞ in the Poincaré metric; electrical power transfer*, J. Funct. Anal. **38** (1980), 273–314.
- [H2] _____, *Orbit structure of the Möbius transformation semigroup action on H^∞ (broadband matching)*, Advances in Math. Suppl. Studies, vol. 3, Academic Press, New York, 1978, pp. 129–197.
- [H3] _____, *Broadbanding: gain equalization directly from data*, IEEE Trans. on Circuits and Systems **28** (1981).
- [H3E] _____, *Broadbanding: gain equalization directly from data*, An expanded version of [H3], available upon request.
- [H4] _____, *Operator theory and broadband matching*, Proc. of the Allerton Conference on Circuits and Systems Theory (1976).
- [H5] _____, *A simple test to determine gain-bandwidth limitations*, IEEE International Conference on Circuit Systems (Phoenix, 1977), pp. 628–631.
- [H6] _____, *Systems with infinite dimensional state space: the Hilbert space approach*, Proc. IEEE **64** (1976), 145–160.
- [Hs] H. Helson, *Invariant subspaces*, Academic Press, New York, 1964.
- [HPPE] Hewlett-Packard, E. E. Pac I programs for the HP-67 and 97 calculators.
- [HP2] _____, *S-parameter design*, Application Note 154, April 1972.
- [H-R] J. H. C. van Heuven and T. Z. Rozzi, *The invariant properties of a multivalued n -port in a linear embedding*, IEEE Trans. Circuit Theory **19** (1972).
- [Htz] W. Hintzman, *Best uniform approximation via annihilating measures*, Bull. Amer. Math. Soc. **76** (1970), 1062–1066.
- [Iv1] T. S. Ivanchenko, *The Nevanlinna-Pick problem in the case of indefinite metrics*, Doklady Akad. Nauk Ukrainian SSR (1980), 8–13.
- [Iv2] _____, *The Schur problem in the case of indefinite metrics*, submitted to VINITI, July 31st, 1979, No. 2868-79.
- [Iv3] _____, *Nevanlinna-Pick problem with simple poles*, submitted to VINITI, June 24, 1980, No. 2558-80.
- [K] T. Kailath, *Linear systems*, Princeton-Hall, Englewood Cliffs, New Jersey, 1980.
- [K-P] I. V. Kovalishyn and V. P. Potapov, *Indefinite metric in the Pick-Nevanlinna problem*, Doklady Akad. Nauk Armenian SSR **59** (1974), 3–9.
- [K] S. Kobayashi, *Hyperbolic manifolds and holomorphic mappings*, Dekker, New York, 1970.
- [K-R] E. Kuh and R. A. Roher, *Theory of linear active networks*, Holden-Day, San Francisco, Calif., 1967.
- [N-F] B. Sz-Nagy and C. Foiaş, *Harmonic analysis of operators on Hilbert space*, North-Holland, Amsterdam, 1970.
- [Nu1] A. A. Nudelman, *On a new problem of the type of moment problem*, Dokl. Akad. Nauk SSSR **233** (1977), 792–795.
- [Nu2] _____, *On a generalization of classical interpolation problems*, Dokl. Akad. Nauk SSSR **256**, (1981), 790–793.
- [P] V. P. Potapov, *The multiplicative structure of J -contractive matrix functions*, Trudy Mos. Mat., Obsc. **4** (1955), 125–136. (Russian) Amer. Math. Soc. Transl. **13** (1960), 131–243.
- [R] R. Redheffer, *On a certain linear fractional transformation*, J. Math. Phys. **39** (1960), 269–286.
- [R-R] M. Rosenbloom and J. Rovnyak, *An operator theoretic approach to theorems of the Pick-Nevanlinna and Löwner types. I* (to appear).
- [S] D. Sarason, *Generalized interpolation in H^∞* , Amer. Math. Soc. Transl. **127** (2) (1967), 180–203.
- [S-Y] M. Saito and D. C. Youla, *Interpolation with positive real functions*, J. Franklin Inst. **284** (2) (1967), 77–108.

[V-D] J. Vandewalle and P. DeWilde, *On the irreducible cascade synthesis of a system with real rational transfer matrix*, Trans. IEEE Circuit Systems, CAS **24** (1977), 481-484.

[W] R. Wohlers, *Lumped and distributed passive networks*, Academic Press, New York, 1969.

[W] R. Wohlers, *Linear distributed and passive networks*, Academic Press, New York and London, 1969.

[Y1] D. C. Youla, *A new theory of broadband matchings*, IEEE Trans. CT-**11** (1964), 30-50.

[Y2] D. C. Youla, L. Kaplan and D. Stock, *The analytic foundation of linear time invariant n-ports*, Polytechnic Inst. of New York, Brooklyn, New York, PIB-MRI Report 1028-62.

[Z-F] G. Zames and B. A. Francis, *Feedback and minimax sensitivity*, Adv. Group for Aerospace Research and Development NATO Lecture Notes, 117, Multivariable Analysis and Design Techniques.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT SAN DIEGO, LA JOLLA, CALIFORNIA 92093