

## WEIGHTED OPTIMIZATION THEORY FOR NONLINEAR SYSTEMS\*

CIPRIAN FOIAS† AND ALLEN TANNENBAUM‡

**Abstract.** In this paper, the solution of a nonlinear version of the weighted sensitivity  $H^\infty$ -optimization problem is discussed. It is shown that the natural object to be considered in this context is a certain "sensitivity operator," which will be optimized locally in a given "energy ball" (see § 5 for the details). In the linear case, the authors are reduced again to the classical sensitivity minimization technique of Zames [21]. The methods were very strongly influenced by the complex analytic power series ideas of [3], [4], [5]. See also the recent results of Ball and Helton [6] for another approach to this subject.

**Key words.** sensitivity operator, nonlinear control, dilation theory, skew Toeplitz operator

**AMS(MOS) subject classifications.** 93B35, 93C05

**1. Introduction.** Recently, there has been a great deal of research devoted to the weighted  $H^\infty$ -optimization of linear systems. See [13] for a rather extensive list of references. Much of the underlying theory for this work has been based on the ideas of Adamjan, Arov, and Krein [1], generalized interpolation theory in  $H^\infty$  due to Sarason [17], and, most generally, on the Sz.-Nagy-Foias commutant lifting theorem [19].

In the papers [3], [4] an extension of the commutant lifting theorem to a local nonlinear setting was given, together with a discussion of how this result could be used to develop a design procedure for nonlinear systems. In the present paper, we continue this line of research with a constructive extension of the linear  $H^\infty$  theory to nonlinear systems. We should note that our colleagues Ball and Helton [6] have developed a completely different, novel approach to this problem based on a nonlinear version of Ball-Helton theory.

In the theory presented below, we will consider majorizable input/output operators (see § 3 for the precise definition). In particular, these operators are analytic in a ball around the origin in a complex Hilbert space, and it turns out that it is possible to express each  $n$ -linear term of the Taylor expansion of such an operator as a linear operator on a certain tensor space. (Our class of operators also includes Volterra series of fading memory [8].) This allows us to iteratively apply the classical commutant lifting theorem in designing a compensator. (The general technique we call the *iterative commutant lifting procedure*. See § 6 for the details.) For single input/single output (SISO) systems, this leads to the construction of a compensator which is optimal relative to a certain sensitivity function that will be defined in § 5. Moreover, in complete generality (i.e., for multiple input/multiple output (MIMO) systems), our procedure will ameliorate (in the sense of our nonlinear weighted sensitivity criterion) any given design. We note that for linear systems, our method reduces to the standard  $H^\infty$  design technique as discussed, for example, in [13] and initiated in [21].

In developing the present theory, we have had to extend some of the skew Toeplitz techniques of [7] and [11] to linear operators defined on certain tensor spaces. This

---

\* Received by the editors May 16, 1988; accepted for publication (in revised form) November 20, 1988. This research was supported in part by grants from the Research Fund of Indiana University, the Department of Energy (DE-FG02-86ER25020), the National Science Foundation (ECS-8704047) and (DMS-8811084) and the Air Force Office of Scientific Research (AFOSR-88-0020).

† Department of Mathematics, Indiana University, Bloomington, Indiana 47405.

‡ Department of Electrical Engineering, University of Minnesota, 123 Church Street SE, Minneapolis, Minnesota 55455.

has led to several novel results in computational operator theory, and, for example, provides a way of iteratively constructing the nonlinear intertwining dilation of the nonlinear commutant lifting theorem considered in [3] and [4]. Moreover, we provide a generalization of a formula due to Sarason [17] for the optimal interpolant in terms of a maximal vector. See § 8 for the details.

An important point is that many of our results are constructive and lead to physically implementable compensators. In fact, we reduce a nonlinear optimization problem to an iterative linear procedure, each step of which we know how to solve. This is illustrated by an example in § 9.

**2. Analytic mappings on Hilbert space.** We would like to discuss here a few standard results about analytic mappings on Hilbert spaces. We are essentially following the treatments of [3]–[5] and [8] to which the reader may refer for all of the details. In particular, input/output operators that admit Volterra expansions are special cases of the operators which we study here. See [8], [16], [20].

Let  $G$  and  $H$  denote complex Hilbert spaces. Set

$$B_{r_0}(G) := \{g \in G : \|g\| < r_0\}$$

(the open ball of radius  $r_0$  in  $G$  about the origin). Then we say that a mapping  $\phi: B_{r_0}(G) \rightarrow H$  is *analytic* if the complex function  $(z_1, \dots, z_n) \rightarrow \langle \phi(z_1 g_1 + \dots + z_n g_n), h \rangle$  is analytic in a neighborhood of  $(1, 1, \dots, 1) \in \mathbb{C}^n$  as a function of the complex variables  $z_1, \dots, z_n$  for all  $g_1, \dots, g_n \in G$  such that  $\|g_1 + \dots + g_n\| < r_0$ , for all  $h \in H$ , and for all  $n > 0$ . (Note that we denote the Hilbert space norms in  $G$  and  $H$  by  $\|\cdot\|$  and the inner products by  $\langle \cdot, \cdot \rangle$ .)

We will now assume that  $\phi(0) = 0$ . It is easy to see that if  $\phi: B_{r_0}(G) \rightarrow H$  is analytic, then  $\phi$  admits a convergent Taylor series expansion, i.e.,

$$\phi(g) = \phi_1(g) + \phi_2(g, g) + \dots + \phi_n(g, \dots, g) + \dots,$$

where  $\phi_n: G \times \dots \times G \rightarrow H$  is an  $n$ -linear map. Clearly, without loss of generality we may assume that the  $n$ -linear map  $(g_1, \dots, g_n) \rightarrow \phi_n(g_1, \dots, g_n)$  is symmetric in the arguments  $g_1, \dots, g_n$ . This assumption will be made throughout this paper for the various analytic maps that we consider. For  $\phi$  a Volterra series,  $\phi_n$  is basically the  $n$ th-Volterra kernel.

Now set

$$\hat{\phi}_n(g_1 \otimes \dots \otimes g_n) := \phi_n(g_1, \dots, g_n).$$

Then  $\hat{\phi}_n$  extends in a unique manner to a dense subset of  $G^{\otimes n} := G \otimes \dots \otimes G$  (tensor product taken  $n$  times). Note by  $G^{\otimes n}$  we mean the Hilbert space completion of the algebraic tensor product of the  $G$ 's. Clearly if  $\hat{\phi}_n$  has finite norm on this dense subset, then  $\hat{\phi}_n$  extends by continuity to a bounded linear operator  $\hat{\phi}_n: G^{\otimes n} \rightarrow H$ . By abuse of notation, we will set  $\phi_n := \hat{\phi}_n$ , and  $\phi_n(g) := \phi_n(g \otimes \dots \otimes g)$  (the tensor product taken  $n$  times).

It is important to note that in principle we can determine  $\phi_n$  quite easily from the input/output operator  $\phi$ . Indeed, we have the following elementary lemma.

**LEMMA 2.1.** *Let  $\phi: B_{r_0}(G) \rightarrow H$  be analytic,  $\phi(0) = 0$ . Suppose, moreover, that if*

$$\phi(g) = \phi_1(g) + \dots + \phi_n(g) + \dots,$$

*then each of the  $\phi_n$  defines a bounded linear operator  $G^{\otimes n} \rightarrow H$  as above (and is symmetric in its arguments). Then for  $g_j \in G$  ( $j = 1, \dots, n$ ) with  $\|g_1\| = \dots = \|g_n\| < r_0$ , we have*

$$\begin{aligned} n! \phi_n(g_1 \otimes \dots \otimes g_n) &= \frac{1}{(2\pi)^n} \int_0^{2\pi} \dots \int_0^{2\pi} \phi(\exp(i\theta_1)g_1 + \dots + \exp(i\theta_n)g_n) \\ &\quad \times \exp(-i(\theta_1 + \dots + \theta_n)) d\theta_1 \dots d\theta_n. \end{aligned}$$

*Proof.* Expand  $\phi(z_1 g_1 + \cdots + z_n g_n)$  in powers of  $z_1, \dots, z_n$ . Then it is easy to see that the coefficient of  $z_1 \cdots z_n$  is precisely  $n! \phi_n(g_1 \otimes \cdots \otimes g_n)$ . The required result then follows immediately from the Cauchy formula.  $\square$

*Remark 2.2.* We should note that if  $\phi$  is analytic, then each  $\phi_n$  is continuous (as an  $n$ -multilinear map); hence, the associated linear map extends to the  $n$ th projective power of  $G$ . Lemma 2.1 is valid in this more general situation as well.

We now conclude this section with two key definitions.

**DEFINITION 2.3.** (i) Notation as above. By a *majorizing sequence* for the holomorphic map  $\phi$ , we mean a sequence of positive numbers  $\alpha_n, n = 1, 2, \dots$  such that  $\|\phi_n\| \leq \alpha_n$  for  $n \geq 1$ . Suppose that  $\rho := \limsup \alpha_n^{1/n} < \infty$ . Then it is completely standard ([8]) that the Taylor series expansion of  $\phi$  converges at least on the ball  $B_r(G)$  of radius  $r = 1/\rho$ .

(ii) If  $\phi$  admits a majorizing sequence as in (i), then we will say that  $\phi$  is *majorizable*.

We will see in the next section that a very important class of input/output operators from systems and control theory are in point of fact majorizable.

**3. Operators with fading memory.** In this section, we will show that perhaps the most natural class of input/output operators from the systems standpoint are majorizable. Moreover, for this class of operators we will even derive an a priori majorizing sequence. We begin with the following key definition:

**DEFINITION 3.1.** An analytic map  $\phi: B_{r_0}(G) \rightarrow H$ ,  $\phi(0) = 0$  has *fading memory* if its nonlinear part  $\phi - \phi'(0)$  admits a factorization

$$\phi - \phi'(0) = \hat{\phi} \circ W,$$

where  $\hat{\phi}$  is an analytic map defined in some neighborhood of  $0 \in G$ , and  $W$  is a linear Hilbert-Schmidt operator. (In this case, we can assume that there exists an orthonormal basis of eigenvectors for  $W$  in  $G$ ,  $\{e_k\}$ ,  $k = 1, 2, \dots$  such that  $W e_k = \lambda_k e_k$  with

$$\|W\|_2^2 := \sum_{k=1}^{\infty} |\lambda_k|^2 < \infty.$$

$\|W\|_2$  is called the *Hilbert-Schmidt norm* of  $W$ .)

*Remark 3.2.* System-theoretically fading memory input/output operators have the property that any two input signals, which are close in the recent past but not necessarily close in the remote past, will yield present outputs which are close. For more details about this important class of operators, see [8].

For fading memory operators, we can construct an explicit majorizing sequence.

**LEMMA 3.3.** Let  $\phi: B_{r_0}(G) \rightarrow H$ ,  $\phi(0) = 0$ , have fading memory. Suppose, moreover, that if we write

$$\phi - \phi'(0) = \hat{\phi} \circ W$$

as in (3.1), then  $\hat{\phi}: B_{r_1}(G) \rightarrow B_{r_2}(H)$ . Then the sequence

$$\alpha_1 := \|\phi'(0)\|$$

$$\alpha_n := \frac{r_2 e^n \|W\|_2^n}{r_1^n}$$

for  $n \geq 2$ , is a majorizing sequence for  $\phi$ .

*Proof.* For complete details see [4, Lemma (3.5)]. However, since we will use some estimates from the proof for Proposition (3.5) below, we will give an outline of the proof.

First, without loss of generality we may assume that  $W$  is positive. Since  $\hat{\phi}: B_{r_1}(G) \rightarrow B_{r_2}(H)$ , from (2.1) we obtain

$$\|\phi_n(g_1 \otimes \cdots \otimes g_n)\| \leq \frac{1}{n!} r_2,$$

for  $\|g_1\| = \cdots = \|g_n\| \leq r_1/n$ .

Now, since  $\{e_{i_1} \otimes \cdots \otimes e_{i_n}; 1 \leq i_1, \dots, i_n\}$  is an orthonormal basis of  $G^{\otimes n}$ , we can write  $g \in G^{\otimes n}$  as

$$\sum_{1 \leq i_1, \dots, i_n} \alpha_{i_1, \dots, i_n} e_{i_1} \otimes \cdots \otimes e_{i_n}$$

and

$$\|\hat{g}\|^2 = \sum_{1 \leq i_1, \dots, i_n} |\alpha_{i_1, \dots, i_n}|^2 < \infty.$$

Now, from the above we can easily compute (see [4] for the details) that

$$\begin{aligned} & \left\| \phi_n \left( \sum_{1 \leq i_1, \dots, i_n} \alpha_{i_1, \dots, i_n} e_{i_1} \otimes \cdots \otimes e_{i_n} \right) \right\| \\ & \leq \frac{n^n}{r_1^n} \sum |\lambda_{i_1} \cdots \lambda_{i_n} \alpha_{i_1, \dots, i_n}| \left\| \hat{\phi}_n \left( \frac{r_1}{n} e_{i_1} \otimes \cdots \otimes \frac{r_1}{n} e_{i_n} \right) \right\| \\ & \leq \frac{n^n}{r_1^n} \frac{r_2}{n!} \sum |\lambda_{i_1} \cdots \lambda_{i_n} \alpha_{i_1, \dots, i_n}| \\ & \leq \frac{n^n}{r_1^n} \frac{r_2}{n!} \|W\|_2^n \sum \alpha_{i_1, \dots, i_n} e_{i_1} \otimes \cdots \otimes e_{i_n}. \end{aligned}$$

This implies that

$$\|\phi_n\| \leq \frac{n^n}{r_1^n} \frac{r_2}{n!} \|W\|_2^n \leq r_2 \frac{e^n}{r_1^n} \|W\|_2^n$$

for  $n \geq 2$  as required.  $\square$

**Remark 3.4.** (i) From the above proof it follows that  $\hat{\alpha}_n$ , where

$$\hat{\alpha}_1 := \|\phi'(0)\|$$

$$\hat{\alpha}_n := \frac{n^n r_2 \|W\|_2^n}{n! r_1^n} \quad \text{for } n \geq 2$$

is a majorizing sequence for  $\phi$ . In computations it turns out that it is easier to work with the majorizing sequence  $\alpha_n$  given in the formulation of Lemma 3.3.

(ii) Note, moreover, we have that

$$\rho := \limsup (\alpha_n)^{1/n} = \frac{e \|W\|_2}{r_1}.$$

(iii) In what follows, we will assume that all of the input/output operators we consider are causal and are majorizable.

An interesting and useful property of fading memory operators is the following proposition.

PROPOSITION 3.5. *The notation and hypotheses are as in Lemma 3.3. Then each  $\phi_n$  (regarded as a linear operator on  $G^{\otimes n}$ ) is compact for  $n \geq 2$ .*

*Proof.* Let the sequence in  $G^{\otimes n}$

$$x^{(k)} := \sum_{i_1, \dots, i_n=1}^{\infty} \alpha_{i_1 \dots i_n}^{(k)} e_{i_1} \otimes \dots \otimes e_{i_n} \rightarrow 0$$

weakly. Define a projection in  $G$  for each natural number  $N > 0$  by

$$P_N e_j = \begin{cases} 0 & j \leq N \\ e_j & j \geq N+1. \end{cases}$$

Then from the above proof of Lemma 3.3, for fixed  $n$ , we have that there exist constants  $C$  and  $\hat{C}$  such that

$$\begin{aligned} \|\phi_n(x^{(k)})\| &\leq C \sum_{i_1, i_2, \dots, i_n \geq 1} |\lambda_{i_1} \dots \lambda_{i_n} \alpha_{i_1 \dots i_n}^{(k)}| \\ &\leq C \sum_{i_1, i_2, \dots, i_n \leq N} |\lambda_{i_1} \dots \lambda_{i_n} \alpha_{i_1 \dots i_n}^{(k)}| + \hat{C} \|WP_N\|_2 \|W\|_2^{n-1}. \end{aligned}$$

Thus,

$$\limsup \|\phi_n(x^{(k)})\| \leq \hat{C} \|WP_N\|_2 \|W\|_2^{n-1}.$$

Hence as  $N \rightarrow \infty$ , we see that

$$\limsup \|\phi_n(x^{(k)})\| = 0,$$

which shows that  $\phi_n$  is compact.  $\square$

**4. Control theoretic preliminaries.** We start here with the control problem definition. First, we will need to consider the precise kind of input/output operator we will be considering. See [3], [4] for closely related discussions. As mentioned above we are assuming that all of the operators we consider are causal and are majorizable. For a discussion of causality in the nonlinear context, see [3]-[6]. Throughout this paper,  $H^2(\mathbf{C}^k)$  will denote the standard Hardy space of  $\mathbf{C}^k$ -valued functions on the unit circle ( $k$  may be infinite, i.e., in this case  $\mathbf{C}^k$  is replaced by  $h^2$ , the space of one-sided square-summable sequences). We now have the following definition.

DEFINITION 4.1. Let  $S: H^2(\mathbf{C}^k) \rightarrow H^2(\mathbf{C}^k)$  denote the canonical unilateral right shift. Then we say an input/output operator  $\phi$  is *locally stable* if it is causal and majorizable,  $\phi(0) = 0$ , and if there exists an  $r > 0$  such that  $\phi: B_r(H^2(\mathbf{C}^k)) \rightarrow H^2(\mathbf{C}^k)$  with  $S\phi = \phi \circ S$  on  $B_r(H^2(\mathbf{C}^k))$ . We set

$$C_l := \{\text{space of locally stable operators}\}.$$

Since the theory we are considering is local, the notion of local stability is sufficient for all of the applications we have in mind. The interested reader can compare this notion with the more global notions of stability as, for example, discussed in [6].

The theory we are about to give holds for all plants which admit coprime locally stable factorizations. However, for simplicity we will assume that our plant is also locally stable. Accordingly, let  $P, W$  denote locally stable operators, with  $W$  invertible. Referring to Fig. 1,  $P$  represents the plant, and  $W$  the weight or filter. Now we see that the feedback compensator  $C$  locally stabilizes the closed loop if the operator

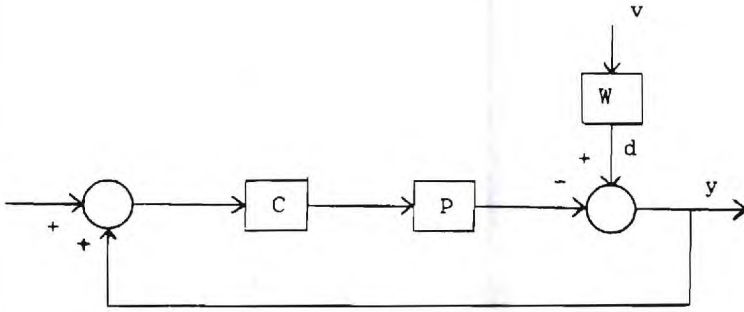


FIG. 1

$(I + P \circ C)^{-1}$  and  $C \circ (I + P \circ C)^{-1}$  are well defined and locally stable. By a result of [2],  $C$  locally stabilizes the closed loop if and only if

$$(1) \quad C = \hat{q} \circ (I - P \circ \hat{q})^{-1}$$

for some  $\hat{q} \in C_l$ . Note then that the weighted sensitivity  $(I + P \circ C)^{-1} \circ W$  can be written as  $W - P \circ q$ , where  $q := \hat{q} \circ W$ . (Since  $W$  is invertible, the data  $q$  and  $\hat{q}$  are equivalent.) In this context, we will call such a  $q$  a *compensating parameter*. From the compensating parameter  $q$ , we get a locally stabilizing compensator  $C$  via the formula (1).

The problem we would like to solve here is a version of the classical disturbance attenuation problem associated to the feedback loop in Fig. 1 (see [7], [21]). This, of course, corresponds to the "minimization" of the "sensitivity"  $W - P \circ q$  taken over all locally stable  $q$ . In order to formulate a precise mathematical problem, we need to say in what sense we want to minimize  $W - P \circ q$ . This we will do in the next section where we will propose a notion of "sensitivity minimization" which seems quite natural to analytic input/output operators.

**5. Sensitivity function.** In this section we define a fundamental object, namely a nonlinear version of *sensitivity*. We will see that while the optimal  $H^\infty$  sensitivity is a real number in the linear case, the measure of performance which seems to be more natural in this nonlinear setting is a certain function defined in a real interval.

In order to define our notion of sensitivity, we will first have to partially order the space of analytic mappings defined in a ball about the origin. All of the input/output operators here will be locally stable. We also follow here our convention that for given  $\phi \in C_l$ ,  $\phi_n$  will denote the bounded linear map on the tensor space  $(H^2(C^k))^{\otimes n}$  associated with the  $n$ -linear part of  $\phi$ , which we also denote by  $\phi_n$  (and which we always assume without loss of generality is symmetric in its arguments). The context will always make the meaning of  $\phi_n$  clear.

We can now state the following key definitions.

**DEFINITION 5.1.** (i) For  $W, P, q \in C_l$  ( $W$  is the weight,  $P$  the plant, and  $q$  the compensating parameter), we define the *sensitivity functions*  $S(q)$ ,

$$S(q)(\rho) := \sum_{n=1} \rho^n \|(W - P \circ q)_n\|$$

for all  $\rho > 0$  such that the sum converges. Note that for fixed  $P$  and  $W$ , for each  $q \in C_l$ , we get an associated sensitivity function.

(ii) We write  $S(q) \leq S(\tilde{q})$ , if there exists a  $\rho_0 > 0$  such that  $S(q)(\rho) \leq S(\tilde{q})(\rho)$  for all  $\rho \in [0, \rho_0]$ . If  $S(q) \leq S(\tilde{q})$  and  $S(\tilde{q}) \leq S(q)$ , we write  $S(q) \equiv S(\tilde{q})$ . This means that  $S(q)(\rho) = S(\tilde{q})(\rho)$  for all  $\rho > 0$  sufficiently small, i.e.,  $S(q)$  and  $S(\tilde{q})$  are equal as germs of functions.

(iii) If  $S(q) \leq S(\tilde{q})$ , but  $S(\tilde{q}) \not\leq S(q)$ , we will say that  $q$  *ameliorates*  $\tilde{q}$ . Note that this means  $S(q)(\rho) < S(\tilde{q})(\rho)$  for all  $\rho > 0$  sufficiently small.

Now with Definition 5.1, we can define a notion of "optimality" relative to the sensitivity function.

DEFINITION 5.2. (i)  $q_o \in C_l$  is called *optimal* if  $S(q_o) \leq S(q)$  for all  $q \in C_l$ .

(ii) We say  $q \in C_l$  is *optimal with respect to its  $n$ th term*  $q_n$ , if for every  $n$ -linear  $\hat{q}_n \in C_l$ , we have

$$S(q_1 + \cdots + q_{n-1} + q_n + q_{n+1} + \cdots) \leq S(q_1 + \cdots + q_{n-1} + \hat{q}_n + q_{n+1} + \cdots).$$

If  $q \in C_l$  is optimal with respect to all of its terms, then we say that it is *partially optimal*.

Clearly, if  $q$  is optimal, then it is partially optimal; however, the converse may not hold. Note, moreover, that if  $\phi$  is a Volterra series, then our definition of sensitivity measures in a precise sense the amplification of energy of each Volterra kernel on signals whose energy is bounded by a given  $\rho$ . For this reason, it appears that in this context, Definition 5.1 of the sensitivity function  $S(q)$  seems physically natural. In the next section, we will discuss a procedure for constructing partially optimal compensating parameters, and then in § 7 we will show how this procedure leads to the construction of optimal compensating parameters for SISO systems. Of course, from formula (1) above, we can derive the corresponding partially optimal (respectively, optimal) compensator from the partially optimal (respectively, optimal) compensating parameters.

**6. Iterative commutant lifting method.** In this section, we discuss the main construction of this paper from which we will derive both partially optimal and optimal compensators relative to the sensitivity function given in Definition 5.1 above. As before,  $P$  will denote the plant and  $W$  the weighting operator, both of which we assume are locally stable. As in the linear case, we always suppose that  $P_1$  is an isometry, i.e.  $P_1$  is *inner*. In order to state our results, we will need to make a few preliminary remarks and set up some notation.

We begin by noting the following key relationship:

$$(W - P \circ q)_k = W_k - \sum_{1 \leq j \leq k} \sum_{i_1 + \cdots + i_j = k} P_j(q_{i_1} \otimes \cdots \otimes q_{i_j}).$$

Note that once again for  $\phi$  majorizable,  $\phi_n$  denotes the  $n$ -linear part of  $\phi$ , as well as the associated linear operator on the appropriate tensor space.

We are now ready to formulate the *iterative commutant lifting procedure*. Let  $\Pi: H^2(\mathbb{C}^k) \rightarrow H^2(\mathbb{C}^k) \ominus P_1 H^2(\mathbb{C}^k)$  denote orthogonal projection. Using the linear commutant lifting theorem (CLT) (see [19] for the details), we may choose  $q_1$  such that

$$\|W_1 - P_1 q_1\| = \|\Pi W_1\|.$$

Now given this  $q_1$ , we choose (using the CLT)  $q_2$  such that

$$\|W_2 - P_2(q_1 \otimes q_1) - P_1 q_2\| = \|\Pi(W_2 - P_2(q_1 \otimes q_1))\|.$$

Inductively, given  $q_1, \dots, q_{n-1}$ , set

$$(2) \quad A_n := (W_n - \sum_{2 \leq j \leq n} \sum_{i_1 + \cdots + i_j = n} P_j(q_{i_1} \otimes \cdots \otimes q_{i_j}))$$

for  $n \geq 2$ . Then from the CLT, we may choose  $q_n$  such that

$$(2a) \quad \|A_n - P_1 q_n\| = \|\Pi A_n\|.$$

We now come to the key point on the convergence of the iterative commutant lifting method.

PROPOSITION 6.1. *With the above notation, let  $q^{(1)} := q_1 + q_2 + \dots$ . Then  $q^{(1)} \in C_1$ .*

*Proof.* It suffices to show that  $\sum \|q_n\| \rho^n$  converges for all  $0 \leq \rho$  sufficiently small. Then from (2a)

$$\|A_n - P_1 q_n\| = \|\Pi A_n\| \leq \|A_n\|$$

and so (using the fact that  $P_1$  is an isometry)

$$(3) \quad \|q_n\| \leq 2\|A_n\| \leq 2\|W_n\| + 2 \sum_{2 \leq j \leq n} \sum_{i_1 + \dots + i_j = n} \|P_j(q_{i_1} \otimes \dots \otimes q_{i_j})\|.$$

Clearly from the majorizability hypothesis, we can find positive constants  $M_o, R_o, M, R$  such that

$$(4) \quad \|W_i\| \leq \frac{1}{2} M_o R_o^i$$

$$(5) \quad \|P_j\| \leq \frac{1}{2} M R^j$$

for  $i \geq 1$ , and for  $j \geq 2$ . Thus,  $\|q_1\| \leq M_o R_o$  and

$$(6) \quad \|q_n\| \leq M_o R_o^n + \sum_{2 \leq j \leq n} M R^j \sum_{i_1 + \dots + i_j = n} \|q_{i_1}\| \dots \|q_{i_j}\|$$

for  $n \geq 2$ . Let  $f(z) = \sum_{n=0}^{\infty} f_n z^n$ , and  $g(z) = \sum_{n=0}^{\infty} g_n z^n$  be formal power series. Then we write  $f \ll g$  if  $|f_n| \leq |g_n|$  for all  $n \geq 0$ .

We introduce the notation

$$\tilde{q}(z) := \sum_{n=1}^{\infty} \|q_n\| z^n$$

$$a(z) := \sum_{n=1}^{\infty} M_o R_o^n z^n$$

$$b(z) := \sum_{n=2}^{\infty} M R^n z^n.$$

With this notation, (3) may be equivalently written as

$$(7) \quad \tilde{q}(z) \ll a(z) + b(\tilde{q}(z)).$$

Now (formally) define

$$\mu(z) \equiv a(z) + b(\mu(z)).$$

Then we claim the following:

- (i)  $\mu(z) \gg 0$ ;
- (ii)  $\tilde{q}(z) \ll \mu(z)$ ;
- (iii)  $\mu$  is analytic in some sufficiently small neighborhood of zero.

Clearly, the verification of this claim would complete the proof of the proposition.

In order to do this, let  $f$  be analytic in some ball of radius  $r_o$  centered at the origin. Then we set

$$\|f\|_{(r)} := \sup \{|f(z)| : |z| \leq r\}$$

for  $r < r_o$ . Next, we define an operator on the set of analytic functions defined in some neighborhood of the origin by  $F(f) := a + b(f)$  whenever  $F(f)$  is well defined as an analytic function near zero. Then for given  $\delta > 0$ , and  $r \leq 1/2 R_o < 1/R_o$  (this choice for  $r$  will be made clear below), we let

$$B := \{f \text{ analytic near } 0 : \|f - a\|_{(r)} \leq \delta\}.$$

We want to choose  $\delta$ , such that  $F$  is well defined in  $B$ ,  $F: B \rightarrow B$ , and such that  $F$  is contractive in  $B$ .

Now it is easy to see that

$$\|f\|_{(r)} \leq \delta + \|a\|_{(r)} \leq \delta + \frac{M_o R_o r}{1 - R_o r} \leq \delta + 2M_o R_o r.$$

Clearly, we can choose  $r, \delta$  such that

$$(8a) \quad 0 < \delta + 2M_o R_o r \leq \frac{1}{2R} < \frac{1}{R}.$$

However,

$$(8b) \quad \|F(f) - a\|_{(r)} \leq \|b(f)\|_{(r)} \leq \frac{MR^2(\delta + 2M_o R_o r)^2}{1 - (\delta + 2M_o R_o r)R} \leq 2MR^2(\delta + 2M_o R_o r)^2.$$

We require then that  $\delta$  and  $r$  satisfy

$$(9) \quad 2MR^2(\delta + 2M_o R_o r)^2 \leq \delta.$$

With these choices we clearly have that  $F: B \rightarrow B$ . Now

$$\begin{aligned} \|F(f) - F(g)\|_{(r)} &\leq \|b(f) - b(g)\|_{(r)} \leq \left\| \frac{M^2 f^2}{1 - Mf} - \frac{M^2 g^2}{1 - Mg} \right\|_{(r)} \\ &\leq \left\| \frac{M^2(f^2 - g^2)}{1 - Mf} \right\|_{(r)} + M^2 \|g^2\|_{(r)} \left\| \frac{1}{1 - Mf} - \frac{1}{1 - Mg} \right\|_{(r)} \\ &\leq 2M^2 \|f + g\|_{(r)} \|f - g\|_{(r)} + 4M^3 \|g\|_{(r)}^2 \|f - g\|_{(r)} \\ &\leq (4M^2(\delta + 2M_o R_o r) + 4M^3(\delta + 2M_o R_o r)^2) \|f - g\|_{(r)}. \end{aligned}$$

If we choose  $\delta$  and  $r$  such that

$$\theta := (4M^2(\delta + 2M_o R_o r) + 4M^3(\delta + 2M_o R_o r)^2) < 1,$$

we see that

$$\|F(f) - F(g)\|_{(r)} \leq \theta \|f - g\|_{(r)}.$$

Hence by the contraction mapping theorem, we get (iii). Moreover, (i) now follows immediately by definition of  $\mu$  and the fact that  $a(z) \gg 0$  and  $b(z) \gg 0$ . Finally, we can prove (ii) by induction. Indeed, let

$$\tilde{q}_k(z) := \sum_{n=1}^k \|q_n\| z^n$$

$$\mu_k(z) := \sum_{n=1}^k \mu_n z^n.$$

Clearly  $\tilde{q}_1(z) \ll \mu_1(z)$ , and suppose by induction that  $\tilde{q}_n(z) \ll \mu_n(z)$  for  $1 \leq n \leq N$ . Then note that there exists a polynomial  $p$  with positive coefficients depending on  $a$  and  $b$  such that  $\tilde{q}_{N+1}(z) \ll p(\tilde{q}_1, \dots, \tilde{q}_N)$  and  $\mu_{N+1} = p(\mu_1, \dots, \mu_N)$ , from which (ii) follows immediately. This completes the proof of Proposition 6.1.  $\square$

Note that given any  $q \in C_l$ , we can apply the iterative commutant lifting procedure to  $W - P \circ q$ . Now set

$$S_{\Pi}(q)(\rho) := \sum_{n=1}^{\infty} \rho^n \|\Pi(W - P \circ q)_n\|.$$

Clearly,  $S_{\Pi}(q) \equiv S(q)$  (as functions). We can now state the following result whose proof is immediate from the above discussion.

**PROPOSITION 6.2.** *Given  $q \in C_1$ , there exists  $\tilde{q} \in C_1$ , such that  $S(\tilde{q}) \equiv S_{\Pi}(q)$ . Moreover,  $\tilde{q}$  may be constructed from the iterated commutant lifting procedure.*

Moreover, we easily have the following result.

**PROPOSITION 6.3.**  *$q$  is partially optimal if and only if  $S(q) \equiv S_{\Pi}(q)$  (i.e.,  $S(q)(\rho) = S_{\Pi}(q)(\rho)$  for all  $\rho > 0$  sufficiently small; see § 5).*

*Proof.* Assume that  $q$  is partially optimal. Then,  $q$  must be optimal with respect to its first term  $q_1$ . However, we have seen that there exists  $\hat{q}_1$  such that  $\|W_1 - P_1 \hat{q}_1\| = \|\Pi W_1\|$ . If  $\|W_1 - P_1 q_1\| > \|\Pi W_1\|$ , then since we are considering germs of functions, we would have  $S(q) \not\equiv S(\hat{q}_1 + q_2 + \dots)$ , contradicting the partial optimality of  $q$ .

By induction, assume that we have proven

$$\|(W - P \circ q)_j\| = \|\Pi(W - P \circ q)_j\|$$

for  $1 \leq j \leq n$ . Then again if

$$\|(W - P \circ q)_{n+1}\| > \|\Pi(W - P \circ q)_{n+1}\|,$$

by the above construction, using the commutant lifting theorem, we can find a  $\hat{q}_{n+1}$  such that

$$\|\Pi(W - P \circ q)_{n+1}\| = \|(W - P \circ (q_1 + q_2 + \dots + q_n + \hat{q}_{n+1} + \dots))_{n+1}\|.$$

So once more,  $S(q) \not\equiv S(q_1 + \dots + q_n + \hat{q}_{n+1} + q_{n+2} + \dots)$ , contradicting the partial optimality of  $q$ . Hence, we get that  $S(q) \equiv S_{\Pi}(q)$ . The proof of the converse direction is similar.  $\square$

We can now summarize the above discussion with the following theorem.

**THEOREM 6.4.** *For given  $P$  and  $W$  as above, any  $q \in C_1$  is either partially optimal or can be ameliorated by a partially optimal compensating parameter.*

*Proof.* The proof follows immediately from Propositions 6.1–6.3.  $\square$

It is important to emphasize that a partially optimal compensating parameter need not be optimal in the sense of Definition 5.1(i). Basically, what we have shown here is that using the iterated commutant lifting procedure, we can ameliorate any given design. The question of optimality will be considered in the next section.

**7. Optimal compensators.** In this section we will derive our main results about optimal compensators. Basically, we will show that in the single input/single output setting, the iterated commutant lifting procedure leads to an optimal design. We begin with the following theorem.

**THEOREM 7.1.** *There exist optimal compensators.*

*Proof.* We will only sketch the proof. Note that our proof is not constructive and makes use of the weak compactness property of weakly closed, bounded, convex sets of operators on Hilbert space.

First of all, set

$$O^{(1)} := \{q_1: q_1 \text{ is optimal relative to } W_1 \text{ and } P_1\} = \{q_1: \|\Pi W_1\| = \|W_1 - P_1 q_1\|\}.$$

It follows from the classical theory [1] that  $O^{(1)}$  is a bounded, weakly closed, convex set of operators. Now set

$$\begin{aligned} O_{q_1}^{(2)} &:= \{q_2: q_2 \text{ is optimal relative to } W_2 - P_2(q_1 \otimes q_1) \text{ and } P_1\} \\ &= \{q_2: \|W_2 - P_2(q_1 \otimes q_1) - P_1 q_2\| = \|\Pi(W_2 - P_2(q_1 \otimes q_1))\|\}. \end{aligned}$$

Next let

$$\hat{W}_2(q) := W_2 - P_2(q \otimes q).$$

Further, we write

$$O^{(2)} := \bigcup_{q_1 \in O^{(1)}} O_{q_1}^{(2)}.$$

Then we can find a sequence  $q_{2j} \in O_{q_{1j}}^{(2)}$  such that

$$\|\hat{W}_2(q_{1j}) - P_1 q_{2j}\| \rightarrow \inf \{\|\hat{W}_2(q_1) - P_1 q_2\| : q_1 \in O^{(1)}, q_2 \in O^{(2)}\} =: \sigma_2.$$

Without loss of generality, we can assume that  $q_{1j} \rightarrow q_1$  weakly. Obviously  $q_1 \in O^{(1)}$ . Moreover, since  $\{q_{2j}\}$  is a bounded sequence, we can also assume without loss of generality that  $q_{2j} \rightarrow q_2$  weakly. Thus,

$$\|\hat{W}_2(q_1) - P_1 q_2\| \leq \liminf \|\hat{W}_2(q_{1j}) - P_1 q_{2j}\|,$$

and hence  $\|\hat{W}_2(q_1) - P_1 q_2\| = \sigma_2$ .

Clearly the above procedure can be iterated step by step. Convergence follows by the same argument as that used in Proposition 6.1.  $\square$

For the construction of the optimal compensator in Theorem 7.3 below, we will need one more technical result. Accordingly, we will need to set up a bit more notation. First set  $H^2 := H^2(\mathbb{C})$ , and  $H^\infty := H^\infty(\mathbb{C})$  (the space of bounded analytic complex-valued functions on the unit disc). Let  $m \in H^\infty$  be a nonconstant inner function, let  $\Pi_1 : H^2 \rightarrow H^2 \ominus mH^2 =: H(m)$  denote orthogonal projection, and set  $T := \Pi_1 S|_{H(m)}$ , where  $S$  is the canonical unilateral shift on  $H^2$ . ( $T$  is the compressed shift.) For  $H$  a complex separable Hilbert space, let  $S_\infty : H \rightarrow H$  denote a unilateral shift, i.e., an isometric operator with no unitary part. This means that  $S_\infty^{*n} h \rightarrow 0$  for all  $h \in H$  as  $n \rightarrow \infty$ . (See [15] and [19].) We can now state the following generalization of a nice result which appears in [18].

LEMMA 7.2. *Notation as above. Let  $A : H \rightarrow H^2 \ominus mH^2$  be a bounded linear operator which attains its norm, i.e., such that there exists  $h_0 \in H$  with  $\|Ah_0\| = \|A\| \|h_0\| \neq 0$ . Suppose moreover that*

$$AS_\infty = TA.$$

*Then there exists a unique minimal intertwining dilation  $B$  of  $A$ , i.e., an operator  $B : H \rightarrow H^2$  such that  $BS_\infty = SB$ ,  $\|A\| = \|B\|$ , and  $\Pi_1 B = A$ .*

*Proof.* First of all, without loss of generality, we can assume that  $\|A\| = 1$ . The existence of  $B$  follows from the commutant lifting theorem [19]. For the uniqueness, we use the results of [10]. Indeed, let

$$F := \{D_T Ah \oplus D_A h : h \in H\}^-$$

where for a contraction  $K$ , we set  $D_K^2 := (I - K^*K)$ ,  $D_K \geq 0$ . Then by [10],  $B$  is unique if  $F = D_T \oplus D_A$ , where  $D_T = \overline{D_T H(m)}$ , and  $D_A = \overline{D_A H}$ . Now it is well known that  $D_T f = \langle f, \mu \rangle \hat{\mu}$  where  $\mu := \bar{z}(m(z) - m(0))$ , and  $\hat{\mu} := \mu / \|\mu\|$ . Thus  $D_T Ah = \langle Ah, \mu \rangle \hat{\mu}$ , and so

$$F = \{\langle Ah, \mu \rangle \hat{\mu} \oplus D_A h : h \in H\}^-.$$

Since  $h_0 \in H$  is such that  $\|Ah_0\| = \|h_0\| \neq 0$ , we have

$$D_T Ah_0 \oplus D_A h_0 = \langle Ah_0, \mu \rangle \hat{\mu} \oplus 0.$$

We consider the following two cases.

Case (i). Suppose  $\langle Ah_0, \mu \rangle \neq 0$ . Then  $C\hat{\mu} \oplus 0 \subset F$ , which implies that  $F \supset 0 \oplus D_A$ , from which we get that  $F = D_T \oplus D_A$ .

Case (ii). Suppose  $\langle Ah_0, \mu \rangle = 0$ . We claim that there exists  $j \geq 1$  such that  $\langle T^j Ah_0, \mu \rangle \neq 0$ . Indeed, suppose not. Then  $\langle T^j Ah_0, \mu \rangle = 0$  for all  $j \geq 1$ ; hence,  $\|T^j Ah_0\| = \|Ah_0\| = \|h_0\|$ . Let  $M$  be the Hilbert space generated by the elements  $T^j Ah_0$  for  $j \geq 0$ . Then  $M$  is  $T$ -invariant, and  $T|_M$  is an isometry. Since  $T$  is of class  $C_0$  (see [19]), this is impossible. Thus, we can find a minimal  $j$  such that  $\langle T^{j+1} Ah_0, \mu \rangle \neq 0$ . However,

$$\|AS_\infty^j h_0\| = \|T^j Ah_0\| = \|Ah_0\| = \|h_0\|.$$

Hence replacing  $h_0$  by  $S_\infty^j h_0$ , we are back to the first case, from which we can complete the proof.  $\square$

We now come to the main result of this section.

**THEOREM 7.3.** *Let  $W$  and  $P$  be SISO locally stable operators, with  $W$  the weight and  $P$  the plant. Suppose that  $\Pi W_j$  is compact for  $j \geq 1$  and  $\Pi P_k$  is compact for  $k \geq 2$ . ( $\Pi: H^2 \rightarrow H^2 \ominus P_1 H^2$  denotes orthogonal projection.) Let  $q_{opt}$  be a partially optimal compensating parameter as constructed by the iterated commutant lifting procedure. Then  $q_{opt}$  is optimal.*

*Proof.* First of all, since  $\Pi W_1$  attains its norm, from Lemma 7.2 we have that the optimal  $q_1$  constructed relative to  $W_1$  and  $P_1$  is unique. (Actually, in this special case, since we are working in  $H^2$ , this follows from [18].) Now from our above hypotheses, each  $\Pi A_k$  is compact for  $k \geq 2$ ; hence, each  $\Pi A_k$  attains its norm. Therefore, by Lemma 7.2 each optimal  $q_k$  constructed by the iterated commutant lifting procedure is unique. Theorem 7.3 now follows immediately from Theorem 6.4.  $\square$

**COROLLARY 7.4.** *Let  $P$  be locally stable and SISO, with linear part  $P_1$  rational. Then the partially optimal compensating parameter  $q_{opt}$  constructed by the iterated commutant lifting procedure is optimal.*

*Proof.* Indeed, since  $P_1$  is SISO rational (recall that we also always assume that  $P_1$  is inner),  $H^2 \ominus P_1 H^2$  is finite-dimensional, and so we are done by Theorem 7.3.  $\square$

**Remark 7.5.** Corollary 7.4 gives a constructive procedure for finding the optimal compensator under the given hypotheses. Indeed, when  $P_1$  is SISO rational, the iterative commutant lifting procedure can be reduced to *finite dimensional matrix calculations*. We will illustrate this important point via an example in § 9. In a subsequent paper, we will show that when the hypotheses of Theorem 7.3 are satisfied, the skew Toeplitz theory of [7] provides an algorithmic design procedure for distributed nonlinear systems as well.

**8. Maximal vectors and optimal interpolants.** In order to apply the iterative commutant lifting procedure to an actual example, we will need a generalization of a result due to Sarason [17] on the optimal interpolant. More precisely, for  $K$  a bounded linear operator on a Hilbert space,  $k_0$  is a *maximal vector*, if  $\|Kk_0\| = \|K\| \|k_0\| \neq 0$ . Then for SISO systems, Sarason [17] derives a formula for the optimal interpolant in terms of a maximal vector of the associated Hankel operator (see [1] for a similar result).

In order to state our result, we will first need a few preliminary remarks. Let  $H := H^2(\mathbb{C}^k)$ . As above, we let  $m \in H^\infty$  be nonconstant inner, and let  $\Pi_1: H^2 \rightarrow H^2 \ominus mH^2 =: H(m)$  denote orthogonal projection, with  $T$  the compression of the canonical shift on  $H^2$  to  $H(m)$ . Moreover,  $S_\infty$  will denote the canonical shift on  $H$ , defined by multiplication by  $e^{\prime\prime}$ . Now given  $h \in H$ , we can write  $h$  as a column vector (perhaps infinite)

$$h = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \end{bmatrix}.$$

We then set

$$h^* := [\bar{h}_1 \bar{h}_2 \cdots].$$

Moreover, given any bounded linear operator  $B : H \rightarrow H^2$  such that  $BS_\infty = SB$ , we have that for  $z \in D$  (the unit disc),

$$(Bh)(z) = \sum_{j \geq 1} b_j(z)h_j(z).$$

That is, we can express  $B$  as the row matrix

$$[b_1 \ b_2 \ \cdots]$$

with  $b_j \in H^\infty$  for  $j \geq 1$ . We will identify  $B$  with this row matrix. With this notation, we can now state the following resulting proposition.

**PROPOSITION 8.1.** *Notation as above. Let  $A : H \rightarrow H^2 \ominus mH^2$  be a bounded linear operator such that  $AS_\infty = TA$ . Suppose, moreover, that  $A$  has a maximal vector  $h_0$ . Let  $B : H \rightarrow H^2$  be the minimal intertwining dilation of  $A$ , i.e.,  $\Pi_1 B = A$ ,  $BS_\infty = SB$ , and  $\|A\| = \|B\|$ . Then if we let  $\lambda := \|A\|^2$ , we have that*

$$B = \frac{\lambda h_0^*}{Ah_0}.$$

*Proof.* First of all, given  $h_0 \in H$ , we represent  $h_0$  as a column vector with components  $h_j$ ,  $j \geq 1$  as above. Then, as we have seen, we have that  $(Bh)(z) = \sum_{j \geq 1} b_j(z)h_j(z)$  (for  $z \in D$ ), and

$$\|B\| = \sup \left\{ \left( \sum_{j=1}^{\infty} |b_j(z)|^2 \right)^{1/2} : |z| < 1 \right\} = \text{ess sup} \left\{ \left( \sum_{j=1}^{\infty} |b_j(\zeta)|^2 \right)^{1/2} : |\zeta| = 1 \right\}.$$

However,

$$\|A\|^2 \|h_0\|^2 = \|Ah_0\|^2 \leq \|Bh_0\|^2 \leq \|B\|^2 \|h_0\|^2 = \|A\|^2 \|h_0\|^2.$$

Thus  $\|Ah_0\|^2 = \|Bh_0\|^2$ , and since  $\Pi Bh_0 = Ah_0$ , we have that  $Ah_0 = Bh_0$ . Next note that  $\sum_{j \geq 1} |b_j(e^{it})|^2 \leq \lambda$  almost everywhere, and

$$\frac{1}{2\pi} \int_0^{2\pi} \left( \lambda \sum_{j=1}^{\infty} |h_j(e^{it})|^2 - \left| \sum_{j=1}^{\infty} b_j(e^{it})h_j(e^{it}) \right|^2 \right) dt = 0.$$

(This follows from the fact that  $\lambda \|h_0\|^2 = \|Bh_0\|^2$ .) But using the Cauchy-Schwarz inequality, the expression under the integral sign is nonnegative. Thus,

$$\lambda \sum_{j \geq 1} |h_j(e^{it})|^2 = \left| \sum_{j \geq 1} b_j(e^{it})h_j(e^{it}) \right|^2 \leq \left( \sum_{j \geq 1} |b_j(e^{it})|^2 \right) \left( \sum_{j \geq 1} |h_j(e^{it})|^2 \right) \leq \lambda \sum_{j \geq 1} |h_j(e^{it})|^2$$

almost everywhere, which implies that

$$\sum_{j \geq 1} |b_j(e^{it})|^2 = \lambda$$

almost everywhere, and

$$h_j = \phi(e^{it}) \overline{b_j(e^{it})}$$

almost everywhere for all  $j \geq 1$ , and for some function  $\phi \in H^2$  satisfying

$$Ah_0 = Bh_0 = \lambda \phi.$$

Thus, for

$$B(e^{it}) = [b_1(e^{it}) \ b_2(e^{it}) \ \cdots]$$

we have

$$B(e^{i\theta}) \overline{Ah_0(e^{i\theta})} = \lambda h_0(e^{i\theta})^*$$

almost everywhere, as required.  $\square$

We will apply Proposition 8.1 in our computation of an optimal compensator in the next section.

**9. Example.** In this section, we will give an example of our nonlinear design procedure. Since we have been working in the disc, we will here take discrete-time systems, even though our techniques obviously go through in a similar manner for continuous-time systems as well. In what follows below,  $H_{D^2}$  will denote the space of  $\mathbb{C}$ -valued analytic functions on the bidisc  $D^2$  with square integrable boundary values.

We let

$$W(z) = \frac{1-z}{2}$$

and  $P = P_1 + P_2$ , where  $P_1 = z^2$  (in the discrete Fourier domain), and

$$P_2(F) = \frac{1}{2\pi i} \int_{|\zeta|=1} F(z\zeta^{-1}, \zeta) \frac{d\zeta}{\zeta}$$

for  $F \in H_{D^2} \cong H^2 \otimes H^2$ . More precisely, as we explained above, we can regard a bilinear map  $P_2$  on  $H^2 \times H^2$  as a linear map on  $H^2 \otimes H^2$ , and then it is easy to see that  $H^2 \otimes H^2$  can be naturally identified with  $H_{D^2}$ . (The identification is given by  $z \otimes 1 \rightarrow z_1$  and  $1 \otimes z \rightarrow z_2$ .) Note that in the discrete-time domain,  $P_2$  is just a discrete Fourier transform of the "squaring" map, i.e., given the square integrable sequence  $\{a_n\}$ , we have that  $P_2$  is the Fourier transform of the mapping  $\{a_n\} \rightarrow \{a_n^2\}$ .

We now apply our procedure to the weight  $W$  and the plant  $P$ . Accordingly, if we let  $M_W: H^2 \rightarrow H^2$  denote multiplication by  $W$ , and let  $\Pi: H^2 \rightarrow H^2 \ominus P_1 H^2 =: H_1$  be orthogonal projection, we set  $A_0 := \Pi M_W|_{H_1}$ . Notice that  $H_1 \cong \mathbb{C}^2$ , and that via this isomorphism, we have the identification

$$A_0 = \begin{bmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

However,

$$A_0^* A_0 = \begin{bmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

from which we get that  $\|A_0\| = (\sqrt{5}+1)/2$ , and that a maximal vector  $h_0$  (i.e., a vector such that  $\|A_0 h_0\| = \|A_0\| \|h_0\| \neq 0$ ) is given by

$$h_0 := \begin{bmatrix} 1 \\ -\beta \end{bmatrix}$$

where  $\beta := (\sqrt{5}-1)/2$ . Using then the Sarason formula [17] mentioned in the previous section, we can then compute that the optimal compensating parameter is

$$q_1 := \frac{\beta}{2(1-\beta z)}.$$

Of course, the above computation was based on standard linear  $H^\infty$ -optimization theory. We now want to show how to get the optimal second-order compensating parameter. Accordingly, following the iterative commutant lifting procedure, we note that

$$\begin{aligned} P_2(q_1 \otimes q_1)(F) &= \frac{1}{2\pi i} \int_{|\zeta|=1} q_1(z\zeta^{-1})q_1(\zeta)F(z\zeta^{-1}, \zeta) \frac{d\zeta}{\zeta} \\ &= \frac{\beta^2}{8\pi i} \int_{|\zeta|=1} \frac{1}{1-\beta z\zeta^{-1}} \frac{1}{1-\beta\zeta} F(z\zeta^{-1}, \zeta) \frac{d\zeta}{\zeta} \end{aligned}$$

for  $F \in H_{D^2}$ .  $P_2(q_1 \otimes q_1)$  will be the "weight" for which we will apply the commutant lifting procedure relative to the "plant"  $P_1$ .

For  $F \in H_{D^2}$ , let

$$F(z_1, z_2) = \sum_{j,k=0}^{\infty} F_{jk} z_1^j z_2^k.$$

Then,

$$\begin{aligned} \frac{4}{\beta^2} P_2(q_1 \otimes q_1)(F) &= \sum_{j,k=0}^{\infty} \left( \frac{1}{2\pi i} \int_{|\zeta|=1} \frac{\zeta}{\zeta-\beta z} \frac{1}{1-\beta\zeta} z^j \zeta^{k-j} \frac{d\zeta}{\zeta} \right) F_{jk} \\ &= \sum_{j,k=0}^{\infty} z^j F_{jk} \frac{(\beta z)^{k-j}}{1-\beta^2 z} + \sum_{j,k=0}^{\infty} z^j F_{jk} \frac{1}{2\pi i} \int_{0 < |\zeta| = \epsilon \ll 1} \frac{\zeta^{k-j}}{(\zeta-\beta z)(1-\beta\zeta)} d\zeta \\ &= \sum_{j,k=0}^{\infty} F_{jk} \frac{\beta^{k-j} z^k}{1-\beta^2 z} + \sum_{j>k} \frac{z^j F_{jk}}{-\beta z} \frac{1}{2\pi i} \\ &\quad \times \int_{0 < |\zeta| = \epsilon \ll 1} \frac{(\sum_{l=0}^{\infty} \zeta^l / (\beta z)^l) (\sum_{l=0}^{\infty} \beta^l \zeta^l)}{\zeta^{j-k}} d\zeta \\ &= \sum_{j,k=0}^{\infty} \frac{\beta^{k-j} z^k}{1-\beta^2 z} + \sum_{j>k} z^{j-1} F_{jk} \frac{-1}{\beta} \frac{1}{2\pi i} \int_{0 < |\zeta| = \epsilon \ll 1} \frac{\sum_{h,l} \beta^h (\zeta^{h+1} / (\beta z)^l)}{\zeta^{j-k}} d\zeta \\ &= \sum_{j,k=0}^{\infty} F_{jk} \frac{\beta^{k-j} z^k}{1-\beta^2 z} + \sum_{j>k} \frac{-1}{\beta} F_{jk} \sum_{l+h=j-k-1} \beta^{h-l} z^{h+k} \\ &= \sum_{j \leq k} F_{jk} \frac{\beta^{k-j} z^k}{1-\beta^2 z} + \sum_{j>k} \beta^{k-j} F_{jk} z^k \frac{(z\beta^2)^{j-k}}{1-\beta^2 z} \\ &= \sum_{j,k} F_{jk} \frac{\beta^{|k-j|} z^{\max\{j,k\}}}{1-\beta^2 z}. \end{aligned}$$

Set  $A := -\Pi P_2(q_1 \otimes q_1)$ . Then from the above computations, we have that

$$-\frac{4}{\beta^2} AF = F_{00} + (\beta F_{10} + \beta F_{01})z + \beta^2 F_{00}z + F_{11}z.$$

Moreover, if we let

$$A_1 := -\frac{4}{\beta^2} A|(\ker A)^{\perp}$$

we clearly have that

$$A_1 \begin{bmatrix} F_{00} \\ F_{10} \\ F_{01} \\ F_{11} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \beta^2 & \beta & \beta & 1 \end{bmatrix} \begin{bmatrix} F_{00} \\ F_{10} \\ F_{01} \\ F_{11} \end{bmatrix}$$

where we identify  $(\ker A)^{\perp}$  with  $\mathbf{C}^4$  in the natural way. Now

$$A_1 A_1^* = \begin{bmatrix} 1 & \\ \beta^2 & (\beta^2 + 1)^2 \end{bmatrix},$$

and then it is easy to compute that  $\|A_1\|^2 =: \lambda \cong 2.048924$ ,  $\|A_1\| \cong 1.431406$ , and that a maximal vector for  $A_1$  is given by

$$h_1 := \begin{bmatrix} \lambda \\ (\lambda - 1)/\beta \\ (\lambda - 1)/\beta \\ (\lambda - 1)/\beta^2 \end{bmatrix}.$$

Now we must write the Fourier representation of  $h_1$  in order to apply Proposition 8.1, and so we must express  $H_{D^2}$  as some  $H^2(\mathbf{C}^k)$ . Accordingly, we apply the techniques of [19], to which we refer the reader for all the details about Fourier representations. More precisely, given  $F = \sum_{j,k=0}^{\infty} F_{jk} z_1^j z_2^k$ , we have that the Fourier representation of  $F$ , denoted by  $F(\zeta)$ , is given by

$$(10) \quad F(\zeta) := \sum_{n=0}^{\infty} \zeta^n \begin{bmatrix} F_{n,n} \\ F_{n+1,n} \\ F_{n,n+1} \\ F_{n+2,n} \\ F_{n,n+2} \\ \vdots \end{bmatrix}$$

for  $\zeta \in \partial D$ . Thus via the above identifications, the Fourier representation of  $h_1$ , denoted by  $h_1(\zeta)$ , is

$$h_1(\zeta) = \begin{bmatrix} \lambda \\ (\lambda - 1)/\beta \\ (\lambda - 1)/\beta \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix} + \zeta \begin{bmatrix} (\lambda - 1)/\beta^2 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}.$$

Applying Proposition 8.1 (and using the same notation), we get that the minimal intertwining dilation of  $A_1$ ,  $B_1$ , is given (in the Fourier space) by

$$B_1(\zeta) = \frac{\zeta \begin{bmatrix} \lambda & \lambda - 1 & \lambda - 1 & 0 & 0 & \cdots \end{bmatrix} + \begin{bmatrix} \lambda - 1 \\ \beta^2 & 0 & 0 & \cdots \end{bmatrix}}{\zeta + (\lambda - 1)/\beta^2}.$$

(Note that  $(\lambda - 1)/\beta^2$  is about  $2.74 > 1$ , and hence  $1/(z + (\lambda - 1)/\beta^2)$  is analytic and bounded in  $\bar{D}$ .) Using the Fourier representation (10) of  $F$ , we have that in the Fourier space

$$(B_1 F)(\zeta) = B_1(\zeta) \sum_{n=0}^{\infty} \zeta^n \begin{bmatrix} F_{n,n} \\ F_{n+1,n} \\ F_{n,n+1} \\ F_{n+2,n} \\ F_{n,n+2} \\ \vdots \end{bmatrix} = \sum_{n=0}^{\infty} \zeta^n \frac{\zeta(\lambda F_{n,n} + (\lambda - 1)\beta^{-1}F_{n+1,n} + (\lambda - 1)\beta^{-1}F_{n,n+1}) + (\lambda - 1)\beta^{-2}F_{n,n}}{\zeta + (\lambda - 1)\beta^{-2}}.$$

We are almost done! Indeed, still working with the Fourier representations, the optimal  $q_2$  may be derived from the equality (for  $z \in D$ )

$$-(4/\beta^2)P_2(q_1 \otimes q_1)F - z^2 q_2 F = -B_1 F.$$

Thus, we see that

$$(q_2 F)(z) = \frac{1}{z^2} \frac{(\sum_{n=0}^{\infty} z^n \{z(\lambda F_{n,n} + (\lambda - 1)\beta^{-1}F_{n+1,n} + (\lambda - 1)\beta^{-1}F_{n,n+1}) + (\lambda - 1)\beta^{-2}F_{n,n}\})}{z + (\lambda - 1)\beta^{-2}} - \frac{1}{z^2} \sum_{j,k} F_{jk} \frac{\beta^{|k-j|} z^{\max(j,k)}}{1 - \beta^2 z}.$$

Despite its seemingly complicated form, we will now see that  $q_2$  has an integral expression in the Fourier domain, which translates into a rather simple two-linear function in the time-domain. Explicitly, we may write (11) equivalently as

$$q_2 = S_1 - S_2 + \frac{\lambda F_{1,1} + (\lambda - 1)\beta^{-1}(F_{2,1} + F_{1,2})}{z + (\lambda - 1)\beta^{-2}} - \frac{\lambda F_{1,1}}{(1 - \beta^2 z)(z + (\lambda - 1)\beta^{-2})} - \frac{\beta^2 \lambda F_{0,0} + \beta \lambda (F_{1,0} + F_{0,1})}{(1 - \beta^2 z)(z + (\lambda - 1)\beta^{-2})},$$

where

$$S_1 := \frac{(\sum_{n=2}^{\infty} z^{n-2} \{z(\lambda F_{n,n} + (\lambda - 1)\beta F_{n+1,n} + (\lambda - 1)\beta F_{n,n+1}) + (\lambda - 1)\beta^2 F_{n,n}\})}{z + (\lambda - 1)\beta^2}$$

and

$$S_2 := \sum_{n=2}^{\infty} \frac{z^{n-2}}{1 - \beta^2 z} \left( \sum_{0 \leq j \leq n} (F_{jn} + F_{nj}) \beta^{n-j} \right).$$

Clearly in order to find a computable expression for  $q_2$ , we must first find such an expression for the map  $M_{m,n} : H_{D^2} \rightarrow \mathbb{C}$ , defined by  $M_{m,n}(F) := F_{m,n}$  where  $m, n = 0, 1, \dots$  are fixed. Let  $a = \{a_j\}$  and  $b = \{b_j\}$  ( $j \geq 0$ ) be sequences in the "discrete time-domain"  $h^2$ . By slight abuse of notation, we also let  $a = a(\zeta) = \sum_{j=0}^{\infty} a_j \zeta^j$ , and  $b = b(\zeta) = \sum_{j=0}^{\infty} b_j \zeta^j$  denote their discrete Fourier transforms. Then it is easy to see that

$$M_{m,n}(a \otimes b) = M_{m,n}(a, b) = \left( \frac{1}{2\pi i} \right)^2 \int_{|\zeta_1|=1} \int_{|\zeta_2|=1} \zeta_1^{-m} \zeta_2^{-n} a(\zeta_1) b(\zeta_2) \frac{d\zeta_1}{\zeta_1} \frac{d\zeta_2}{\zeta_2} = a_m b_n.$$

In this way, we get that

$$S_1 = \frac{1}{z + (\lambda - 1)/\beta^2} T_1(F)$$

where

$$T_1(F) := \sum_{n=2}^{\infty} z^{n-2} \{z(\lambda F_{n,n} + (\lambda - 1)/\beta F_{n+1,n} + (\lambda - 1)/\beta F_{n,n+1}) + (\lambda - 1)/\beta^2 F_{n,n}\}.$$

Hence, we see that

$$T_1(a \otimes a) = \lambda \sum_{n=2}^{\infty} z^{n-1} a_n^2 + 2(\lambda - 1)\beta^{-1} \sum_{n=2}^{\infty} z^{n-1} a_n a_{n+1} + (\lambda - 1)\beta^{-2} \sum_{n=2}^{\infty} z^{n-2} a_n^2$$

which, of course, is the transform of a very simple quadratic map in the time domain. In the exact same way, we can write down explicit expressions for all the terms of  $q_2$  appearing in formula (12).

Note that our above computations essentially amount to finite-dimensional matrix manipulations. We have then that  $q_1 + q_2$  is the optimal compensating parameter up to order two. A similar computation allows us to find the optimal compensating parameter up to any order, and by Proposition 6.1, our procedure is guaranteed to converge.

**10. Conclusions.** In this paper we have introduced a novel notion of "sensitivity minimization," and have given a method for constructing optimal compensators for SISO systems, and partially optimal compensators for MIMO systems. This generalizes the standard  $H^\infty$  linear theory in a rather natural way. However, in contrast to the linear case, the measure of performance is now given by (the germ of) a certain sensitivity function instead of a real number. The key idea is the utilization of an iterative commutant lifting procedure which can also be employed to ameliorate any given design in the sense of § 5.

The techniques we have used here are local and very much inspired by the previous work in [3]-[5]. The interested reader can contrast this approach with the nonlinear Ball-Helton method as given in [6]. An intriguing problem would be to compare nonlinear designs derived from these two approaches (which, of course, coincide in the linear case). This we would like to consider in some future work as well as attempt to derive a more global theory. There are, of course, a number of open questions still remaining even in our local setting. A key problem is to design optimal controllers for nonlinear MIMO plants. Indeed, even though we can ameliorate any design, because of nonuniqueness in the choice of the various minimal intertwining dilations in the iterative commutant lifting procedure, for MIMO systems we cannot guarantee optimality but only partial optimality. In a subsequent paper, we plan to show how the skew Toeplitz techniques of [7] provide a design methodology for distributed nonlinear systems as well.

At the Systems Research Center of Honeywell in Minneapolis, an interesting partial dynamic inversion technique due to Elgersma and Morton [9] has recently been employed to obtain some nonlinear designs related to a sixth degree of freedom aircraft model. A project on which we are now embarked is the utilization of the iterative commutant lifting procedure in order to ameliorate this kind of design. Finally, in the SISO case (in which there is a rather complete theory), our procedure is algorithmic, and we are presently working on software for its digital implementation with our colleagues at Honeywell along the lines of the work already done in the linear framework based on [11] and [12].

## REFERENCES

- [1] V. M. ADAMJAN, D. Z. AROV, AND M. G. KREIN, *Infinite Hankel matrices and generalized problems of Caratheodory-Fejer and F. Riesz*, *Functional Anal. Appl.*, 2 (1968), pp. 1-18.
- [2] V. ANANTHARAM AND C. DESOER, *On the stabilization of nonlinear systems*, *IEEE Trans. Automat. Control*, AC-29 (1984), pp. 569-573.
- [3] J. BALL, C. FOIAS, J. W. HELTON, AND A. TANNENBAUM, *On a local nonlinear commutant lifting theorem*, *Indiana J. Mathematics*, 36 (1987), pp. 693-709.
- [4] ———, *Nonlinear interpolation theory in  $H^\infty$* , in *Modelling, Robustness, and Sensitivity in Control Systems*, R. Curtin, ed., NATO-ASI Series, Springer-Verlag, New York, 1987.
- [5] ———, *A Poincare-Dulac approach to a nonlinear Beurling-Lax-Halmos theorem*, *J. Math. Anal. Appl.*, to appear.
- [6] J. BALL AND J. W. HELTON, *Sensitivity bandwidth optimization for nonlinear feedback systems*, Technical Report, Department of Mathematics, University of California at San Diego, 1988.
- [7] H. BERCOVICI, C. FOIAS, AND A. TANNENBAUM, *On skew Toeplitz operators*, I, *Operator Theory: Adv. Appl.*, 29 (1988), pp. 21-44.
- [8] S. BOYD AND L. CHUA, *Fading memory and the problem of approximating nonlinear operators with Volterra series*, *IEEE Trans. Circuits and Systems*, CAS-32 (1985), pp. 1150-1161.
- [9] M. ELGERSMA AND B. MORTON, *Nonlinear flying quality parameters based on dynamic inversion*, Technical Report AFWAL-TR-87-3079, 1987.
- [10] C. FOIAS AND A. FRAZHO, *On the Schur representation in the commutant lifting theorem*, I, *Operator Theory: Adv. Appl.*, 18 (1986), pp. 207-217.
- [11] C. FOIAS AND A. TANNENBAUM, *On the four block problem*, II: *the singular system*, *Integral Equations Operator Theory*, 11 (1988), pp. 726-767.
- [12] C. FOIAS, A. TANNENBAUM, AND G. ZAMES, *Some explicit formulae for the singular values of certain Hankel operators with factorizable symbol*, *SIAM J. Math. Anal.*, 19 (1988), pp. 1081-1089.
- [13] B. FRANCIS, *A Course in  $H^\infty$  Control Theory*, McGraw-Hill, New York, 1987.
- [14] E. HILLE AND R. S. PHILLIPS, *Functional Analysis and Semigroups*, AMS Colloquium Publications XXIII, Providence, 1957.
- [15] M. ROSENBLUM AND J. ROVNYAK, *Hardy Classes and Operator Theory*, Oxford University Press, New York, 1985.
- [16] W. J. RUGH, *Nonlinear System Theory: the Volterra/Wiener Approach*, Johns Hopkins Univ. Press, Baltimore, 1981.
- [17] D. SARASON, *Generalized interpolation in  $H^\infty$* , *Trans. Amer. Math. Soc.*, 127 (1967), pp. 179-203.
- [18] ———, *Function Theory on the Unit Circle*, Lecture Notes, Virginia Polytechnic Institute, Blacksburg, VA, 1978.
- [19] B. SZ.-NAGY AND C. FOIAS, *Harmonic Analysis of Operators on Hilbert Space*, North-Holland, Amsterdam, 1970.
- [20] N. WIENER, *Nonlinear Problems in Random Theory*, Technology Press of M.I.T., Cambridge, MA, 1958.
- [21] G. ZAMES, *Feedback and optimal sensitivity: model reference transformations, multiplicative seminorms, and approximate inverses*, *IEEE Trans. Automat. Control*, AC-26 (1981), pp. 301-320.