H^∞ Optimization with Plant Uncertainty and Semidefinite Programming*

J. William Helton
Dept of Mathematics
U. Calif. San Diego
La Jolla, CA 92093

Orlando Merino
Dept of Mathematics
U. of Rhode Island
Kingston, RI 02881

Trent E. Walker
Dept of Mathematics
U. Calif. Berkeley
Berkeley, CA 94720

Abstract

The fundamental H^∞ problem of control is that of finding the stable frequency response function that best fits worst case frequency domain specifications. This is a non smooth optimization problem that underlies the frequency domain formulation of the H^∞ problem of control; it is the main optimization problem in Qualitative Feedback Theory for example. It is shown in this article how the fundamental H^∞ optimization problem of control can be naturally treated with modern primal-dual interior point (PDIP) methods. The theory introduced here generalizes and unifies approaches to solving large classes of optimization problems involving matrix valued functions, a subclass of which are commonly treated with Linear Matrix Inequalities techniques. Also, in this article new optimality conditions for H^∞ optimization problems over matrix valued functions are proved, and numerical experience on natural (PDIP) algorithms for these problems is reported. In experiments we find the algorithms exhibit (local) quadratic convergence rate in many instances. Finally, H^∞ optimization problems with an uncertainty parameter are considered. It is shown how to apply the theory developed here to obtain optimality conditions and derive algorithms. Numerical tests on simple examples are reported.

*Supported in part by AFOSR, Ford Motor Co. and NSF

Contents

0 Introduction

The fundamental H^∞ problem of control is that of finding the stable frequency response function that best fits worst case frequency domain specifications. This is a non smooth optimization problem that underlies the frequency domain formulation of the H^∞ problem of control; it is the main optimization problem in Qualitative Feedback Theory for example. We show in this article how the fundamental H^∞ optimization problem of control can be naturally treated with modern primal-dual interior point (PDIP) methods. The theory introduced in §1 and §2 generalizes and unifies approaches to solving large classes of optimization problems involving matrix valued functions, a subclass of which are commonly treated with Linear Matrix Inequalities techniques.

The method that we suggest to be very effective on H^∞ optimization problems might be thought of as a hybrid between primal-dual methods and those for traditional optimization of smooth objective functions. For example, a simple H^∞ optimization problem is, find

$$\inf_{f\in H^\infty} \sup_{\theta} \left(\mathrm{re}[\frac{1}{e^{i\theta}} - f(e^{i\theta})]\right)^2 + 4\left(\mathrm{im}[\frac{1}{e^{i\theta}} - f(e^{i\theta})]\right)^2 \tag{1}$$

The non-differentiable part of the problem comes from the supnorm and this is best handled by putting a primal-dual feature into the algorithm. On the other hand, the performance function in (1) is smooth in θ and f and while it could be handled via duality, a very natural approach resembles a smooth type of optimization method (purely primal). Over the last 5 years, we have found such primal-dual-smooth hybrids to be quite effective. As we describe them here they generalize certain aspects of theory and methods now widely popular for solving Linear Matrix Inequalities (LMI).

The core of our H^∞ optimization theory is a set of equations which locally characterize optimal functions. Indeed there are many such characterizations most of which tend to underdetermine the answer. This prevents Newton's method from being effective. In a previous article [HMW] we described 2 ways to cut down a characterization when the performance function is scalar valued to produce a set of optimality conditions on which Newton's method is effective. Unfortunately this does not extend to the more general situations treated in this article.

In §3 we generalize the optimality conditions from scalar valued to matrix valued performance functions. This adds the complication of yet further types of nonsmoothness to the H^∞ optimization problem and the difficulty that local optima typically are not isolated. This latter occurrance is a traditional bane to the second order convergence of Newton's method, however an accomplishment of this article is that we present methods which appear to overcome this second difficulty in many cases. The added generality of the matrix problems is physically important since this contains engineering problems with competing constraints and plant uncertainty. Among the numerical schemes we propose and examine (in §4 and §5) are the so called "interior point methods". These of course iterate Newton steps. We find on several tests of local (rather than global behavior) that interior point methods do not offer great advantages. This is discussed in §6. An important conclusion of our tests in §6 is that the order of convergence seems to be intimately related to the characteristics of the optimal dual function. Roughly, numerical experiments indicate that quadratic convergence rate holds when the optimal dual function is smooth, linear convergence rate holds when the optimal dual function is continuous but not smooth, and very slow convergence rate holds when the optimal dual function is not continuous.

Applications of our theory to H^∞ optimization with an uncertainty parameter are considered in §7 and §8. There, we apply the theory developed in previous sections to obtain optimality conditions and derive algorithms. Numerical tests on simple examples are reported.

Part I

Optimization Over H^∞

1 Recollections about Semidefinite Programming

We operate at a high level of abstraction since this brings out the basic issues and actually makes understanding easier. Also since H^∞ optimization is infinite dimensional we are forced to something beyond finite dimensional vector spaces.

1.1 Basic Set Up

Let X be a real Banach space. A subset X_+ of X is called a *cone* if it is closed under addition and multiplication by nonnegative scalars. If x and y are elements of X, we say that $x \geq y$ provided $x - y$ is an element of X_+.

Let X' be the dual of X. The *dual cone* of a closed cone X_+ is

$$X'_+ = \{\lambda : \langle \lambda, p \rangle \geq 0 \text{ for all } p \in X_+\} \tag{2}$$

Note that X'_+ is closed.

Definition 1.1 *Let X_+ be a closed cone in a Banach space X such that $X_+ \neq X$. Let I be a fixed element of X and set*

$$\eta^*(x) := \min\{\eta : \eta I - x \geq 0\} \tag{3}$$

The function η^* is called a size *function (with respect to I) if*

$$|\, \eta^*(x) \,| \leq \|\, x \,\|_X \qquad (\forall x \in X) \tag{4}$$

Note that the definition of η^* requires that $\{0\} \neq X_+ \neq X$. Some properties of size functions are given in Proposition 2 in the Appendix.

Definition 1.2 *Let X be a real Banach space with cone X_+ and size function η^*. We say that η^* satisfies the* Hahn-Banach Condition *if*

$$\text{(HBC)} \qquad \eta^*(x) \;=\; \sup_{\lambda \in X'_+} \frac{\langle \lambda, x \rangle}{\|\lambda\|_{X'}} \qquad (\forall x \in X)$$

The supremum in (HBC) is always attained (See Corollary 3 in the Appendix.)

1.2 The Optimization Problem

Let X be a Banach space, X_+ be a closed cone in X, η^* be a size function on X with respect to $I \in X$, and F be a real Banach vector space with dual F'. Let $\mathcal{F} : F \to X$ be a C' (continuously differentiable) map, and k be a linear functional on F. A basic problem that we consider in this paper is:

$$\text{Find } \gamma = \gamma_* \in R,\ f = f_* \in F \text{ solving}$$

$$\textbf{(PO)} \qquad \begin{array}{ll} \textbf{minimize} & \gamma + k(f) \\[1em] \textbf{subject to} & \gamma I - \mathcal{F}(f) \geq 0 \end{array}$$

We shall refer to (PO) as the *Primal Optimization Problem* and refer to γ and f as the *primal variables.*

A pair (γ_*, f_*) is a *local solution* to (PO) if there is a neighborhood V of (γ_*, f_*) such that the latter point is a solution to the problem obtained from (PO) by adding the constraint $(\lambda, f) \in V$. Note that if $\mathcal{F}$ is affine in f, then local solutions are also global. This is not true for general $\mathcal{F}$.

Now we list optimality conditions and an associated primal-dual problem. The latter is given in terms of the differential $\mathcal{DF}_f[\cdot]$ of $\mathcal{F}$ at a given $f \in F$, and its adjoint[1] $(\mathcal{DF}_f)^\dagger : X' \to F'$. Recall that the adjoint operator satisfies and is defined by

$$\langle (\mathcal{DF}_f)^\dagger[\lambda], h\rangle = \langle \lambda, \mathcal{DF}_f[h]\rangle \ \ \forall \lambda \in X',\ h \in F\ .$$

The *Primal-Dual Optimization problem* is,

$$\textbf{(PDO)}\qquad \begin{array}{l} \text{Find } \gamma_* \in R,\ f_* \in F \text{and } \lambda_* \in X'_+ \text{ solving} \\[1ex] \textbf{minimize} \quad \gamma - \langle \lambda, \mathcal{F}(f)\rangle \\[1ex] \textbf{subject to} \\ \qquad\qquad \begin{array}{rcl} \lambda & \in & X'_+ \\ \|\lambda\|_{X'} & = & 1 \\ \gamma I - \mathcal{F}(f) & \geq & 0 \\ \mathcal{DF}^\dagger_{f_*}[\lambda] + k^\dagger[\lambda] & = & 0 \end{array} \end{array}$$

We refer to λ as *dual variable.* A triple $(\gamma_*, f_*, \lambda_*)$ is a *local solution* to (PDO) if there is a neighborhood V of $(\gamma_*, f_*, \lambda_*)$ such that the latter point is a solution to the problem obtained from (PDO) by adding the constraint $(\gamma, f, \lambda) \in V$.

The quantity

$$\Delta(\gamma, f, \lambda) \ = \ \gamma - \langle \lambda, \mathcal{F}(f)\rangle \tag{5}$$

is called the *duality gap.* The duality gap is zero at local solutions to (PDO).

Proposition 1 *If $(\gamma_*, f_*, \lambda_*)$ is a local solution to (PDO), then* $\Delta(\gamma_*, f_*, \lambda_*) = 0$

Proof. Since $\gamma_* - \mathcal{F}(f_*) \geq 0$, we have that

$$\gamma_* \geq \eta^*(\mathcal{F}(f_*)) \tag{6}$$

Now strict inequality in (6) does not hold, since otherwise γ_* may be replaced by a smaller γ_1 which still satisfies the inequality constraint, while reducing the value of Δ, which is not possible. Therefore

$$\gamma_* = \eta^*(\mathcal{F}(f_*)) \tag{7}$$

We also have that, by the definition of η^*,

$$\eta^*(\mathcal{F}(f_*)) \geq \langle \lambda_*, \mathcal{F}(f_*)\rangle \tag{8}$$

Strict inequality in (8) does not hold, since otherwise one may choose $\lambda_1 \in X_+$ with unit norm such that

$$\eta^*(\mathcal{F}(f_*)) = \langle \lambda_1, \mathcal{F}(f_*)\rangle$$

Such λ_1 exists by Corollary 3. Replacing λ_* by λ_1 (or if necessary by a suitable scaling of a convex combination of λ_* and λ_1) yields a smaller value of Δ, thus contradicting the fact that $(\gamma_*, f_*, \lambda_*)$ is a local solution. Therefore we arrive at the relation

$$\eta^*(\mathcal{F}(f_*)) = \langle \lambda_*, \mathcal{F}(f_*)\rangle \tag{9}$$

[1] We use $L^\dagger$ to denote the adjoint of a linear operator L.

Combine (7) with (9) to obtain the conclusion of the proposition. □

Proposition 1 leads to the following system of equations in λ, f and γ which must be satisfied by local solutions to (PDO):

$$\textbf{(PDEQ)} \qquad \begin{aligned} \gamma - \langle \lambda, \mathcal{F}(f) \rangle &= 0 && \text{(i)} \\ \| \lambda \|_{X'} &= 1 && \text{(ii)} \\ \mathcal{DF}^{\dagger}_{f_*}[\lambda] + k^{\dagger}[\lambda] &= 0 && \text{(iii)} \end{aligned}$$

Adding the inequality constraints

$$\begin{aligned} \lambda &\geq 0 && \text{(iv)} \\ \gamma I - \mathcal{F}(f) &\geq 0 && \text{(v)} \end{aligned}$$

to (PDEQ) gives a set of relations we refer to as (**PDEQ**$^+$).

1.3 Main High Level Theorem

Our next result states that (PO) and (PDO) are closely related.

Theorem 1 *Suppose that η^* satisfies (HBC). If (γ_*, f_*) is a local solution to (PO) such that $\gamma_* > 0$, then there exists $\lambda_* \in X'_+$ such that $(\gamma_*, f_*, \lambda_*)$ satisfies (PDEQ$^+$).*

Theorem 1 implies that solving (PDEQ$^+$) leads to excellent candidates for solutions to (PO). Our algorithms are based on this.

Proof. First we prove the theorem with $k(\cdot) = 0$, since a simple transformation gives the result involving k. Note that the optimal γ_* in (PO), by definition of size η^*, is equal to $\eta^*(\mathcal{F}(f_*))$. Therefore there exists $\epsilon > 0$ such that f_* and γ_* satisfy

$$\gamma_* = \eta^*(\mathcal{F}(f_*)) = \inf_{\{h \in F : \|h\| < \epsilon\}} \eta^*(\mathcal{F}(f + h)) \tag{10}$$

The first step in the proof is to show that that the following relation holds:

$$\gamma_* = \eta^*(\mathcal{F}(f_*)) = \inf_{h \in F} \eta^* \left(\mathcal{F}(f_*) + \mathcal{DF}[h]\right) \tag{11}$$

In order to prove (11), it suffices to show that

$$\gamma_* = \eta^*(\,\mathcal{F}(f_*)\,) \leq \eta^*(\mathcal{F}(f_*) + \mathcal{DF}_{f_*}[h]), \quad \forall\, h \in F \tag{12}$$

To prove (12), fix $h \in F$. The function $\phi(t) = \eta^*(\mathcal{F}(f_*) + t\, \mathcal{DF}_{f_*}[h])$ is a convex function of t. If $\phi(t) < \phi(0)$, then it follows that $\phi'_+(0)$, the right derivative of ϕ at $t = 0$, is negative. (The right derivative exists because ϕ is convex.) Since by Proposition 2 in the Appendix η^* is Lipschitz continuous, we have that

$$\eta^*(\mathcal{F}(f + th)) = \eta^*(\mathcal{F}(f_*) + t\mathcal{DF}[h]) + o_1(t).$$

Then, for all $t > 0$ small enough,

$$\begin{aligned} \eta^*(\mathcal{F}(f + th)) &= \phi(0) + \phi'_+(0)t + o_2(t) + o_1(t) \\ &= \eta^*(\mathcal{F}(f_*)) + \phi'_+(0)t + t\,(\,\phi'_+(0) \;+\; o_2(t)/t\,) \end{aligned} \tag{13}$$

The functions o_1 and o_2 in (13) satisfy

$$\frac{1}{t}(o_1(t)+o_2(t)) \longrightarrow 0 \text{ as } t \longrightarrow 0,$$

hence there exists $t > 0$ such that

$$\eta^*(\mathcal{F}(f_* + th)) \; < \; \eta^*(\mathcal{F}(f_*))$$

which is impossible. Thus the right derivative of ϕ is non-negative. Similar reasoning tells us that the left derivative of ϕ is non-negative. The proof of (12) is now complete. Relation (11) is readily obtained from (12) by taking infimum of both members of the equality.

Since $\mathcal{F}(f_*)$ is not an element of the closure of $\mathcal{DF}_{f_*}[F]$, by Theorem 5 in the Appendix there exists a functional $\lambda_* \in X_+^*$ with unit norm such that

$$\langle \lambda^*, \mathcal{F}(f_*) \rangle = \eta^*(\mathcal{F}(f_*)) \tag{14}$$

and

$$\langle \lambda^*, \mathcal{DF}_{f_*})[h] \rangle = 0, \quad \forall h \in F \tag{15}$$

Thus (i), (ii) and (iv) of PDEQ$^+$ are satisfied. Also, (iii) holds since we are assuming for now that $k = 0$, while a local version of (v) is automatically satisfied.

We have finished the proof except for the k wrinkle. To include $k(f)$ in the (PDO) formulas, rewrite (PO) as

$$\begin{array}{ll} \textbf{minimize} & \gamma + k(f) \\ \textbf{subject to} & (\gamma + k(f))I - (\mathcal{F}(f) + k(f))I \geq 0 \end{array} \tag{16}$$

Replace γ with $\tilde{\gamma} := \gamma + k(f)$ to get the problem we have been treating. This immediately yields the conclusion of our theorem. □

1.4 Examples

Example 1. Consider $X = R^2$ with norm $\|(x,y)\| = \max\{|x|,|y|\}$. Set $X_+ = \{(x,y) : x \geq 0, \ y \geq 0\}$, and let $I = (1,1)$. Note that $X'_+ = X_+$ and $\|(w,v)\|_{X'} = |w| + |u|$. Here $\eta^*(x,y) = max\{x,y\}$ is a size function.

Example 2. If A is an $n \times m$ matrix, denote with

$$s_1(A) \geq s_2(A) \geq \dots \tag{17}$$

the singular values of A. We define functions σ_1 and σ_{max} on the space of $n \times m$ matrices as follows:

$$\begin{aligned} \sigma_1(A) \; &= \; \textstyle\sum_{\ell=1}^{m} s_\ell(A) \\ \sigma_{max}(A) \; &= \; \max_\ell s_\ell(A) \end{aligned} \tag{18}$$

Let $X = M =$ the self adjoint $n \times n$ matrices with entries in C equipped with the norm σ_∞. One can easily show that $M' = M$ (as a set) and $X'_+ = X_+$, via

$$\langle x, y \rangle = re\, tr\, (y^t x) \tag{19}$$

Here y^t is the transpose of y, and tr denotes the trace function. Also one can prove that

$$\|B\|_{M'} = \sigma_1(B) \tag{20}$$

Let M_+ be the cone of the nonnegative definite matrices, and let I be the identity matrix. Clearly, $M'_+ = M_+$. Then

$$\eta^*(B) = \max\{s : det(sI - B) = 0\}$$

That is, $\eta^*(B)$ is the largest eigenvalue of B. One can show that η^* satisfies (HBC).

Example 3. Let $X = \mathcal{C}_{n\times n}$ the set of continuous functions on the unit circle ∂D of the complex plane C, with values in M, the set of $n \times n$ self adjoint matrices. (see previous example.) Equip the space $\mathcal{C}_{n\times n}$ with the norm

$$\|g\|_{\mathcal{C}_{n\times n}} = \sup_{\theta} \sigma_{max}(g(e^{i\theta})) \tag{21}$$

Let I be the constant function equal to the identity matrix, and let X_+ be the cone of continuous functions on ∂D that take nonnegative definite matrix values. Then,

$$\eta^*(g) = \max_{\theta}\ \max\left\{\mu : det(\mu I_{n\times n} - g(e^{i\theta})) = 0\right\} \tag{22}$$

It can be shown that η^* satisfies (HBC). The dual space of $\mathcal{C}_{n\times n}$ consists of functionals which can be represented as $n \times n$ self adjoint matrix valued bounded Borel measures. More details about this example are given in section 3.1.

2 Numerical Algorithms

In this section we give a conceptual picture of algorithms that can be used to solve PDEQ$^+$. Later in Section 4 we specialize the algorithms to H^∞ optimization.

2.1 The Interior Point Algorithm IPA

Interior methods take a key constraint and perturb it in hopes of producing a more robust algorithm. Viewed geometrically, when using these algorithms one is finding a point belonging to an ε-level set of certain function. The ε level set is a subset of the 0-level set, which actually contains all the solutions to the original problem. As computation progresses, one obtains a sequence of "interior points" that approach the 0-level set ("the boundary") and if convergent, approach a solution.

We now modify the duality gap relation 1. of PDEQ$^+$ by replacing the r.h.s. of condition (i) of $(PDEQ^+)$ by a positive constant ε, to obtain

$$\mathbf{PDEQ}_\varepsilon^+ \qquad \begin{array}{lrcl} 1. & \langle \lambda, \gamma I - \mathcal{F}(f)\rangle & = & \varepsilon \\ 2. & \mathcal{D}\mathcal{F}_f^\dagger[\lambda] + k^\dagger[\lambda] & = & 0 \\ 3. & \|\lambda\|_{X'} & = & 1 \\ 4. & \lambda & \geq & 0 \\ 5. & \gamma I - \mathcal{F}(f) & \geq & 0 \end{array}$$

The equalities 1.–3. in (PDEQ$_\varepsilon^+$) can be written in the form

$$(\mathbf{OCI}_\varepsilon) \qquad \mathcal{T}_\varepsilon[x] = 0.$$

where $\mathcal{T}_\epsilon$ is a nonlinear operator defined on a suitable space V containing elements $x = (\gamma, f, \lambda)$. Depending on the problem and its setting, the inequalities 4. and 5. of (PDEQ$_\varepsilon^+$) may be incorporated into the operator equation ($\mathbf{OCI}_\varepsilon$), for example via a barrier function. Thus to solve PDEQ$^+$, one chooses an initial ε and performs the following iteration repeatedly until a prescribed stopping criterion is satisfied:

Iteration of IPA. Given $\varepsilon > 0$,

I. Solve ($\mathbf{OCI}_\varepsilon$) approximately by Newton's method or by a modification of it.

II. Decrease ε.

On occasions there is enough structure to solve in

$$\langle \lambda, \gamma - \mathcal{F}(f) \rangle = \varepsilon$$

for the dual variable λ:

$$\lambda = \mathcal{G}(\gamma, \mathcal{F}, f, \varepsilon) \tag{23}$$

While this might seem like a rare event, it happens for a large class of H^∞ problems, as we shall see in §4. Obviously, (23) can be used to eliminate λ from (PDEQ$^\infty$).

These simplifications and others, while unimportant from a theoretical perspective, have a direct impact in practical optimization since they translate into savings of computation time and storage.

A final remark is that in the context of H^∞-optimization the ε perturbation of equation (i) of PDEQ$^+$ can be thought of as a regularization of the original problem. This suggests that IPA type algorithms might be a good choice for numerically difficult problems.

2.2 Boundary Algorithms BCIPA

In some situation instead of having ε iterate towards zero, it can be set to 0 right at the onset, thus effectively reducing the Interior Algorithm to exactly one iteration. We call this the *Boundary Case of the Interior Point Algorithm* BCIPA.

The BCIPA is actually useful in H^∞-optimization. For example, the algorithms discussed in [HMW] are of this type. Convergence estimates and report of numerical experience in [HMW] performed on a class of (scalar) H^∞ optimization problems indicate that BCIPA is superior to IPA. Later in §5 we give results of further comparison tests performed on other optimization problems.

Tricks of the trade. In H^∞ Optimization computation can be accelerated because an H^∞ optimum produces flat performance. For many H^∞ problems this is expressed in the high level notation as

$$\gamma I = \mathcal{F}(f)$$

This is a strengthend version of $\langle \lambda, \gamma I - \mathcal{F}(f) \rangle = 0$. Even further reduction is obtained by "projecting out" the variable γ (see [HMW]). In this way one obtains a problem with no unknown γ, and with no constraint $\gamma I - \mathcal{F}(f) \geq 0$.

2.3 Optimization Problems

2.3.1 LMI's and related stories.

Let $A_0, A_1, \ldots, A_K$ be $n \times n$ real symmetric matrices, and let $c \in R^K$. For $f = (f_1, \ldots, f_K) \in R^K$ let $\mathcal{F}(f) = A_0 + A_1 f_1 + \ldots + A_K f_K$. The optimization problem

$$\text{Find } f = (f_1, \ldots, f_K) \in R^K \text{ solving}$$

$$\textbf{minimize} \quad c^t f \tag{24}$$

$$\textbf{subject to} \quad \mathcal{F}(f) \geq 0$$

has been studied by several authors, see [BV] for a beautiful and thorough discussion, extensive bibliography and applications. It is the central problem treated with Linear Matrix Inequalities techniques. The problem (24) is very close to a problem (PO) when X is the space of real symmetric matrices, X_+ is the space of positive semidefinite matrices in X, and $\mathcal{F}$ is an affine map from R^K to X. Indeed, one can show that with the additional condition of at least one of the A_ℓ's being positive definite, then problem (24) is a special case of (PO).

The Interior Point Algorithm (§2.1) is in the spirit of algorithms discussed in [BV], although barrier functions are used heavily there. The algorithm in Haeberly and Overton [HO] is of the type BCIPA.

2.3.2 Smooth optimization

Traditional unconstrained optimization of a smooth function $\mathcal{F}$ on R^n fits into the §1 setting. To see this take $X = R$, $X_+ = R^+$, $\mathcal{F} : R^n \to R$, and $k(f) = 0$. Then in (PO) the minimization of γ gives $\gamma = \mathcal{F}(f)$, so (PO) becomes

$$\textbf{minimize } \mathcal{F}(f). \tag{25}$$

Now we give the optimality conditions ($PDEQ^+$) for (25). First note that λ is a positive real number, which the normalization forces to be 1. Next note that $0 = \mathcal{D}\mathcal{F}^\dagger_{f^*}[1] = \nabla\mathcal{F}(f_*)$. This is the classical gradient equal to zero condition.

2.3.3 H^∞ optimization

We shall see that H^∞ optimization combines both nonsmooth and smooth structure.

Definition 1 *A performance function is a smooth function Γ of $e^{i\theta} \in \partial D$ and $z \in C^N$, that takes $n \times n$ nonnegative definite matrix values, and such that*

i. Γ *is* real symmetric*:* $\Gamma(e^{i\theta}, z) = \Gamma(e^{-i\theta}, \overline{z})$.

ii. For each $e^{i\theta}$, the function $\Gamma(e^{i\theta}, \cdot)$ is strictly plurisubharmonic.

Remark 2.1 The condition of being real symmetric is dictated by the fact that we are dealing with functions on the unit circle ∂D whose Fourier coefficients are real. The condition of being strictly plurisubharmonic is technical and we believe that most performance functions of practical importance are plurisubharmonic. Also, in previous articles ([HMer1], [HMer2], [HMW]), performance functions were defined to be (possibly nonsmooth) positive valued functions. The class of performance functions is extended in this article, to include matrix valued positive definite functions. However, non-smoothness is no longer allowed in a performance function. The price to pay is not high, as we will see later that many optimization problems with certain nonsmooth real valued performance functions can be restated as optimization problems with a (smooth) matrix valued performance.

If D is a $n \times n$ matrix, denote with $\|D\|_{n\times n}$ the operator norm of D, that is, using our previous notation we have $\|D\|_{n\times n} = \sigma_{max}(D)$. Let $\mathcal{A}_N$ be the C^N valued functions on the unit circle ∂D which are continuous and extend analytically to the unit disk D. If Γ is a performance function and $f \in \mathcal{A}_N$, the *performance at* f is

$$\gamma(f) = \sup_\theta \|\Gamma(e^{i\theta}, f(e^{i\theta}))\|_{n\times n}$$

The minimization of $\gamma(f)$ is the *fundamental* H^∞ *problem of control*[2]:

$$\text{(OPT)}\qquad \begin{array}{l} \text{Given a performance function } \Gamma, \\ \\ \text{Find } f_* \in A_N \text{ and } \gamma_* \in R \text{ such that} \\ \\ \gamma_* = \sup_\theta \|\Gamma(e^{i\theta}, f_*(e^{i\theta}))\|_{n\times n} = \inf_{f\in A_N} \sup_\theta \|\Gamma(e^{i\theta}, f(e^{i\theta}))\|_{n\times n} \end{array} \tag{26}$$

A pair (γ_*, f_*) is a local solution to (OPT) if it solves the problem obtained from (OPT) by restricting the set A_N of feasible functions to a neighborhood of f_*. If f_* is a function in $\mathcal{A}_N$, we say that f_* is a local solution to (OPT) whenever the pair $(\gamma_*, f_*) = (\gamma(f_*), f_*)$ is a local solution to (OPT).

We now state the problem (OPT) as a problem of the type (PO).

$$\textbf{(PO-}H^\infty\textbf{)}\qquad \begin{array}{l} \text{Find } \gamma = \gamma_* \in R,\ f = f_* \in \mathcal{A}_N \text{ solving} \\ \\ \textbf{minimize}\quad \gamma \\ \\ \textbf{subject to}\quad \gamma I - \Gamma(\cdot, f) \geq 0 \end{array}$$

With the set up of Example 3 in §1.4 we have that Theorem 1 applies to local solutions to (OPT). That is, *local solutions to* (OPT) *must satisfy relations* (PDEQ+).

Many optimization problems can be rephrased as problems of the type (OPT). Below we illustrate this by presenting two examples, one involving rectangular matrix valued functions, and the other one involving the maximum of two performance functions.

Rectangular matrix valued functions. If $Q(e^{i\theta}, z)$ is a function that takes $m \times n$ complex matrix values, then the optimization problem of finding minimizers $f_* \in \mathcal{A}_N$ to

$$\inf_{f\in\mathcal{A}_N} \sup_\theta \|Q(e^{i\theta}, f(e^{i\theta}))\|_{m\times n} \tag{27}$$

can be stated as a problem (OPT). To do this[3], recall that for $m \times n$ matrices A, the matrix A^TA is an $n \times n$ nonnegative definite matrix, and that $\|A\|^2_{m\times n} = \|A^TA\|_{n\times n}$. Hence, the set of solutions to (27) is the the same as the set of solutions to

$$\inf_{f\in\mathcal{A}_N} \sup_\theta \|Q(e^{i\theta}, f(e^{i\theta}))^T Q(e^{i\theta}, f(e^{i\theta}))\|_{n\times n} \tag{28}$$

which is a problem of the type (OPT) with performance $\Gamma(e^{i\theta}, z) = Q(e^{i\theta}, z)^T Q(e^{i\theta}, z)$.

Competing Constraints. Consider the following problem: Given $n_\ell \times n_\ell$ matrix valued performance functions Γ_ℓ, $1 \leq \ell \leq 2$, find minimizers to

$$\inf_{f\in\mathcal{A}_N} \max \left\{ \sup_\theta \| \Gamma_1(\cdot, f) \|_{n_1\times n_1} \ ,\ \sup_\theta \| \Gamma_2(\cdot, f) \|_{n_2\times n_2} \right\} \tag{29}$$

[2]See [HMar], [HMer1], [HMer2] , [HMW]. For issues like existence of f_*, uniqueness etc. see [HMer3].
[3]A^T is the conjugate transpose of A.

or equivalently,

$$\inf_{f\in\mathcal{A}_N} \sup_\theta \max \ \{ \parallel \Gamma_1(\cdot,f) \parallel_{n_1\times n_1} \ , \ \parallel \Gamma_2(\cdot,f) \parallel_{n_2\times n_2} \} \tag{30}$$

We now derive an OPT problem. Set

$$\Gamma(e^{i\theta},z) = \begin{pmatrix} \Gamma_1(\cdot,z) & 0 \\ 0 & \Gamma_2(\cdot,z) \end{pmatrix} \tag{31}$$

Since

$$\parallel \Gamma(e^{i\theta},z) \parallel_{(n_1+n_2)\times(n_1+n_2)} \ = \max\{ \parallel \Gamma_1(e^{i\theta},z) \parallel_{n_1\times n_1} , \parallel \Gamma_2(e^{i\theta},z) \parallel_{n_2\times n_2} \} \tag{32}$$

holds the problem (30) can be rewritten as a problem (OPT). It should be clear that the same idea applies to situations involving any finite number of functions Γ_ℓ.

3 Characterizing H^∞ Optima

3.1 Definitions and Notation

For $p \in [1,\infty)$, let $L^p = L^p_1$ be the the space of complex valued Lebesgue measurable functions f on the unit circle ∂D having finite p-th norm $\|f\|_p = \{\int |f(e^{i\theta})|^p\}^{1/p}$. For $p=\infty$, L^∞ is the space of measurable functions f on ∂D such that $\|f\|_\infty = \text{ess sup}|f(e^{i\theta})|$ is finite. If g is a function on ∂D that takes matrix values, we say that g is in (matrix valued) $L^p_{n\times m}$ if the entries of g belong to (scalar) L^p. Functions g in $L^1_{n\times n}$ have an associated Fourier series expansion

$$g \sim \sum_{\ell=-\infty}^{\infty} \hat{g}_\ell$$

where $\hat{g}_\ell$ is an $n\times m$ matrix, $\ell \in Z$. We make that assumption (standard in engineering but not in pure mathematics) that elements of $L^p_{n\times m}$ have Fourier coefficients $\hat{g}_\ell$ which are matrices that have real entries only. The Hardy space $H^p_{n\times m}$ consists of those g in $L^p_{n\times m}$ for which $0 = \hat{g}_{-1} = \hat{g}_{-2} = \ldots$. Thus H^p, L^p are *real* vector spaces.

By $\mathcal{C}_{n\times m}$ we denote the subset of continuous functions in $L^\infty_{n\times m}$. We equip $\mathcal{C}_{n\times m}$ with the norm

$$\|g\|_\infty = \sup_\theta \sigma_{max}\left(g(e^{i\theta})\right) \qquad (g\in\mathcal{C}_{n\times m})$$

while the norm we assume on $L^1_{n\times m}$ is

$$\|g\|_{L^1} = \int \sigma_1\left(g(e^{i\theta})\right)\frac{d\theta}{2\pi} \qquad (g\in L^1_{n\times m})$$

The space $\mathcal{C}'_{n\times m}$ (the dual of $\mathcal{C}_{n\times m}$) consists of bounded Borel measures λ. These measures may be represented in the form

$$\lambda = \begin{pmatrix} \lambda_{1,1} & \ldots & \lambda_{1,m} \\ \vdots & & \vdots \\ \lambda_{n,1} & \ldots & \lambda_{n,m} \end{pmatrix} \tag{33}$$

where it is understood that the action of λ on an element $g \in \mathcal{C}_{n\times m}$ is given by

$$\langle\lambda, g\rangle = \sum_{\ell,k} \text{re} \int g_{\ell,k}\, d\lambda_{\ell,k} = \int \text{re}\; tr(g^t d\lambda) \tag{34}$$

If λ is absolutely continuous with respect to Lebesgue measure, then there exists a function Ψ in $L^1_{n\times m}$ such that

$$d\lambda = \Psi \tfrac{d\theta}{2\pi} \tag{35}$$

The function Ψ is the *density* of λ (with respect to Lebesgue measure.) With (35), relation (34) can be rewritten as

$$\langle \lambda, g\rangle = \int \mathrm{re}\, tr\ (g^t\ \Psi) \tfrac{d\theta}{2\pi} \tag{36}$$

The norms of λ and Ψ are related by

$$\|\lambda\|_{\mathcal{C}'_{n\times n}} = \|\Psi\|_{L^1} = \int \sigma_1(\Psi(e^{i\theta})) \tfrac{d\theta}{2\pi} \tag{37}$$

Moreover, if λ is in the dual cone to the continuous functions with $n\times n$ nonnegative definite values and if λ is absolutely continuous with respect to Lebesgue measure with density Ψ, then Ψ takes nonnegative definite values, and in this case,

$$\|\Psi\|_{L^1} = \int \ tr\ \Psi\ \tfrac{d\theta}{2\pi} \tag{38}$$

3.2 Performance, Jacobians and adjoints

To simplify notation, in this paper the convention of writing $\Gamma(e^{i\theta}, z)$ instead of $\Gamma(e^{i\theta}, z, \overline{z})$ will be used. If Γ is a performance function, then it (k,ℓ)-th entry $\Gamma_{k,\ell}$ is a scalar valued function of $e^{i\theta}$ and of $z = (z_1, \dots, z_N)^t$, which for $e^{i\theta}$ fixed has as differential in z given by

$$\mathcal{D}_z(\Gamma_{\ell,k})[w] = \mathcal{D}_z(\Gamma_{\ell,k})[w_1,\dots,w_N] = \sum_{j=1}^{N} \frac{\partial \Gamma_{\ell,k}}{\partial z_j}(e^{i\theta}, z) w_j + \frac{\partial \Gamma_{\ell,k}}{\partial \overline{z_j}}(e^{i\theta}, z)\overline{w_j}$$

Then the differential of Γ for $e^{i\theta}$ fixed is

$$\mathcal{D}_z\Gamma_z[w] = \mathcal{D}_z\Gamma_z[w_1,\dots,w_N] = \sum_{j=1}^{N} \frac{\partial \Gamma}{\partial z_j}(e^{i\theta}, z) w_j + \frac{\partial \Gamma}{\partial \overline{z_j}}(e^{i\theta}, z)\overline{w_j} \tag{39}$$

where $\frac{\partial\Gamma}{\partial z_j}(e^{i\theta}, z)$ and $\frac{\partial\Gamma}{\partial \overline{z_j}}(e^{i\theta}, z)$ are $n\times n$ matrices with entries $\frac{\partial\Gamma_{\ell,k}}{\partial z_j}(e^{i\theta}, z)$ and $\frac{\partial\Gamma_{\ell,k}}{\partial \overline{z_j}}(e^{i\theta}, z)$ respectively.

One can consider (with abuse of notation) the function Γ as an operator mapping continuous functions to continuous functions:

$$\begin{aligned} \Gamma:\ A_N &\longrightarrow \quad \mathcal{C}_{n\times n} \\ f \quad &\longrightarrow \Gamma(\cdot, f(\cdot)) \end{aligned} \tag{40}$$

The differential of the operator in (40) at a given f is a linear map

$$\begin{aligned} \mathcal{D}\Gamma_f:\ A_N &\longrightarrow \quad \mathcal{C}_{n\times n} \\ h \quad &\longrightarrow \mathcal{D}\Gamma_f[h] \end{aligned} \tag{41}$$

where

$$\mathcal{D}\Gamma_f[h] = \mathcal{D}\Gamma_f[h_1,\dots,h_N] = \sum_{j=1}^{N} \frac{\partial \Gamma}{\partial z_j}(\cdot, f) h_j + \frac{\partial \Gamma}{\partial \overline{z_j}}(\cdot, f)\overline{h_j} \tag{42}$$

We now state a result that will be needed later.

Lemma 3.1 *Let λ be in $\mathcal{C}'_{n\times n}$, let Γ be a performance function and let $f \in \mathcal{A}_N$. If $\lambda \geq 0$, then*

$$\langle \mathcal{D}\Gamma_f^\dagger[\lambda], h\rangle = \sum_{j=1}^{N} \int 2\, re\, tr\, h_j(e^{i\theta}) \left(\frac{\partial \Gamma}{\partial z_j}(e^{i\theta}, f(e^{i\theta}))\, d\lambda \right),\ \forall h \in \mathcal{A}_N \tag{43}$$

PROOF. For $h \in \mathcal{A}_N$ we have that

$$\begin{aligned} \langle \mathcal{D}\Gamma_f^\dagger[\lambda], h\rangle &= \langle \lambda, \mathcal{D}\Gamma_f[h]\rangle \\ &= \int re\, tr \left\{ \left(\sum_{j=1}^N \tfrac{\partial \Gamma}{\partial z_j}(e^{i\theta}, f(e^{i\theta}))h_j(e^{i\theta}) + \tfrac{\partial \Gamma}{\partial \bar{z}_j}(e^{i\theta}, f(e^{i\theta}))\overline{h_j(e^{i\theta})} \right)^t d\lambda \right\} \end{aligned} \tag{44}$$

But for $1 \leq j \leq N$ we also have that

$$\begin{aligned} \int re \left[\overline{h_j(e^{i\theta})}\, tr \left\{ \left(\tfrac{\partial \Gamma}{\partial \bar{z}_j}(e^{i\theta}, f(e^{i\theta}))^t \right) d\lambda \right\} \right] &= \int re \left[h_j(e^{i\theta})\, tr \left\{ \overline{\left(\tfrac{\partial \Gamma}{\partial \bar{z}_j}(e^{i\theta}, f(e^{i\theta}))^t \right) \overline{d\lambda}} \right\} \right] \\ &= \sum_{k,\ell} \int re \left[h_j(e^{i\theta}) \overline{\left(\tfrac{\partial \Gamma_{k,\ell}}{\partial \bar{z}_j}(e^{i\theta}, f(e^{i\theta})) \right) \overline{d\lambda_{k,\ell}}} \right] \\ &= \sum_{k,\ell} \int re \left[h_j(e^{i\theta}) \left(\tfrac{\partial \overline{\Gamma_{k,\ell}}}{\partial z_j}(e^{i\theta}, f(e^{i\theta})) \right) \overline{d\lambda_{k,\ell}} \right] \\ &= \sum_{k,\ell} \int re \left[h_j(e^{i\theta}) \left(\tfrac{\partial \Gamma_{\ell,k}}{\partial z_j}(e^{i\theta}, f(e^{i\theta})) \right) d\lambda_{\ell,k} \right] \\ &= \int re \left[h_j(e^{i\theta})\, tr \left\{ \left(\tfrac{\partial \Gamma}{\partial z}(e^{i\theta}, f(e^{i\theta}))^t \right) d\lambda \right\} \right] \end{aligned} \tag{45}$$

The statement of the Lemma follows from combining (44) and (45). □

3.3 Main H^∞ Optimization result

Theorem 2 *Let $f_* \in \mathcal{A}_N$ be a local solution to (OPT). Set*

$$\begin{aligned} \gamma_* &= \sup_\theta \|\Gamma(e^{i\theta}, f_*(e^{i\theta}))\|_{n\times n} \\ \mathcal{U} &= \begin{pmatrix} \frac{\partial \Gamma_{1,1}}{\partial z_1}(\cdot, f_*) & \cdots & \frac{\partial \Gamma_{\ell,k}}{\partial z_1}(\cdot, f_*) & \cdots & \frac{\partial \Gamma_{n,n}}{\partial z_1}(\cdot, f_*) \\ \vdots & & & & \vdots \\ \frac{\partial \Gamma_{1,1}}{\partial z_N}(\cdot, f_*) & \cdots & \frac{\partial \Gamma_{\ell,k}}{\partial z_N}(\cdot, f_*) & \cdots & \frac{\partial \Gamma_{n,n}}{\partial z_N}(\cdot, f_*) \end{pmatrix} \end{aligned} \tag{46}$$

If

$$\mathcal{U}^T \mathcal{U} \quad \text{is pointwise invertible on } \partial D \tag{47}$$

then there exists $\Psi_* \in L^1_{n\times n}$ such that the triple (γ_*, f_*, Ψ_*) satisfies

$$\mathbf{PDEQ^+H^\infty} \qquad \begin{array}{ll} 1. & (\gamma_* I - \Gamma(\cdot, f_*))\Psi_* = 0 \\ 2. & tr(\frac{\partial \Gamma}{\partial z_\ell}(\cdot, f_*)^t \Psi_*) \in zH^1, \qquad 1 \leq \ell \leq N \\ 3. & \int tr \Psi_* \frac{d\theta}{2\pi} = 1 \\ 4. & \Psi_* \geq 0 \\ 5. & \gamma_* I - \Gamma(\cdot, f_*) \geq 0 \end{array}$$

We refer to the first three conditions of $\mathbf{PDEQ^+H^\infty}$ as $\mathbf{PDEQH^\infty}$.

Remark 3.2 It can be easily shown that if Γ has diagonal or block diagonal structure, then the function Ψ in Theorem 2 can be chosen to have the same diagonal structure.

PROOF. Let $\lambda_* \in \mathcal{C}'_{n\times n}$ be given by Theorem 1, that is, $(\gamma_*, f_*, \lambda_*)$ is a local solution to (PDEQ$^+$).

We now claim that λ_* is absolutely continuous with respect to Lebesgue measure. To see this, start from condition 3 of PDEQ$^+$ and apply Proposition 3.1 to conclude that every $h \in \mathcal{A}_N$ satisfies

$$\begin{aligned} 0 &= \langle\, \lambda\,,\, \textstyle\sum_{j=1}^N \frac{\partial\Gamma}{\partial z_j}(\cdot,f)h_j + \frac{\partial\Gamma}{\partial \overline{z_j}}(\cdot,f)\overline{h_j}\,\rangle \\ &= 2\,\langle\, \lambda\,,\, \textstyle\sum_{j=1}^N \frac{\partial\Gamma}{\partial z_j}(\cdot,f)h_j\,\rangle \\ &= 2\,\mathrm{re}\, \textstyle\int \mathrm{tr}\, \left[\left(\sum_{j=1}^N \frac{\partial\Gamma}{\partial z_j}(\cdot,f)\,h_j \right)^t d\lambda \right] \\ &= 2\, \textstyle\sum_{j=1}^N \mathrm{re} \int \mathrm{tr}\, \left[h_j \left(\frac{\partial\Gamma}{\partial z_j}(\cdot,f)^t\, d\lambda \right)\right] \end{aligned} \tag{48}$$

The last member of relation (48) defines a real linear functional, which by the chain of equalities is the zero functional:

$$\sum_j 2\,\mathrm{re} \int \left[h_j^t\, \mathrm{tr}\left(\tfrac{\partial\Gamma}{\partial z_j}(\cdot,f)^t\, d\lambda \right)\right] = 0, \qquad \forall h = (h_1,\ldots,h_N) \in \mathcal{A}_N \tag{49}$$

For $j = 1,\ldots,N$, the C^N-valued Borel measure tr $\left(\frac{\partial\Gamma}{\partial z_j}(\cdot,f)\,d\lambda\right)$ annihilates $\mathcal{A}_1$. Hence it is absolutely continuous with respect to Lebesgue measure, and its density ϕ_j belongs to zH^1_N. Therefore, we may write

$$\sum_{1\le k_1,k_2\le n} \frac{\partial\Gamma_{k_1,k_2}}{\partial z_\ell}(\cdot,f)\, d\lambda_{k_1,k_2} = \phi_\ell\, \tfrac{d\theta}{2\pi} \qquad {\scriptstyle (1\le \ell\le N)} \tag{50}$$

Relation (50) can be rewritten in the form

$$\mathcal{U}\begin{pmatrix} d\lambda_{1,1} \\ \vdots \\ d\lambda_{n,n}\end{pmatrix} = \Phi\, \tfrac{d\theta}{2\pi} \tag{51}$$

where $\Phi = (\phi_1,\ldots,\phi_N)^t$. Finally, by (47) we may solve for $d\lambda$ in (51) to get

$$\begin{pmatrix} d\lambda_{1,1} \\ \vdots \\ d\lambda_{n,n}\end{pmatrix} = (\,\mathcal{U}^T\mathcal{U}\,)^{-1}\,\mathcal{U}^T\,\Phi\tfrac{d\theta}{2\pi} \tag{52}$$

Hence λ is absolutely continuous with respect to Lebesgue measure. Let Ψ be the density function of λ. Note that substituting $d\lambda = \Psi_* \frac{d\theta}{2\pi}$ into (50) gives relations (2.) of $\mathbf{PDEQ^+H^\infty}$.

Since λ_* is in the dual cone, we have that by the discussion in section 3.1 that Ψ_* takes nonnegative definite values, that is, 4. of PDEQH$^\infty$ holds. Also by (38), and since λ_* has unit norm, we have that 3. of PDEQH$^\infty$ holds.

Now we may rewrite 1. of PDEQ$^+$ as

$$\int tr\; \{\gamma_* I - \Gamma(\cdot,f_*)\}\; \Psi \tfrac{d\theta}{2\pi}{=}0 \tag{53}$$

Since $\gamma_* I - \Gamma(\cdot, f_*)$ and Ψ take nonnegative definite values, the trace of their product is a nonnegative real valued function, and since the integral of the latter is 0, we have that

$$tr\ (\gamma_* I - \Gamma(\cdot, f_*))\Psi = 0 \tag{54}$$

It is an exercise in linear algebra to show that if A and B are $n \times n$ nonnegative definite matrices such that $tr(AB) = 0$, then $AB = 0$. Applying this result to equation (54) implies 1. of PDEQH$^\infty$. $\square$

Corollary 1 FLATNESS OF PERFORMANCE. *At local optima $f_* \in \mathcal{A}_N$ the largest singular value of the performance is a constant function of θ, i.e.,*

$$\sigma_{max}\left(\Gamma(e^{i\theta}, f_*(e^{i\theta}))\right) = \gamma_*, \qquad \forall \theta \tag{55}$$

We will use this fact to define a key optimality diagnostic used in computer runs.

PROOF. Suppose that (55) does not hold at some θ. Since $\sigma_{max}(\Gamma(\cdot, f_*))$ is a continuous function of $e^{i\theta}$, there is an arc $J \subset \partial D$ such that

$$\sigma_{max}\left(\Gamma(e^{i\theta}, f_*(e^{i\theta}))\right) < \gamma_*, \qquad \forall e^{i\theta} \in J \tag{56}$$

By 1. of PDEQ$^+H^\infty$, $\Psi(e^{i\theta}) = 0$ for $e^{i\theta} \in J$. Then the function Φ in (51) is necessarily the function 0. But then (52) implies that $\Psi = 0$, which contradicts 3. of PDEQ$^+H^\infty$. $\square$

Theorem 2 has as a corollary the optimality condition result that is known for scalar valued performance Γ, see [HM]:

Corollary 2 *Let Γ be a scalar valued performance ($n = 1$), and let $f_* \in \mathcal{A}_N$ be a local solution to (OPT). If $\frac{\partial \Gamma}{\partial z}(\cdot, f_*)$ is never 0 on ∂D, then*

$$\text{Flatness:} \qquad \Gamma(\cdot, f_*) = \gamma_*$$

$$\text{Gradient Alignment:} \qquad \exists \psi > 0 \quad \text{measurable such that} \quad \textstyle\int \psi \frac{d\theta}{2\pi} = 1 \quad \text{and}$$

$$\psi\, \frac{\partial \Gamma}{\partial z}(\cdot, f_*) \ \in\ z H^1_N$$

PROOF. The flatness relation is the statement of Corollary 1. Note that the Gradient Alignment condition is just 2. of **PDEQ$^+$H$^\infty$**. $\square$

3.4 Distance of a continuous matrix valued function to analytic functions

To illustrate Theorem 2 we give the optimality conditions PDEQ$^+$ for the problem of minimizing the distance of a 2×2 -matrix valued continuous function to the space of analytic functions $\mathcal{A}_{2\times 2}$. More concretely, let the function $g \in \mathcal{C}_{2\times 2}$ be given, and let $f_* \in \mathcal{A}_{2\times 2}$ be such that

$$\sup_\theta \| g(e^{i\theta}) \ - \ f_*(e^{i\theta}) \|_{2\times 2} = \inf_{f \in \mathcal{A}_{2\times 2}} \sup_\theta \| g(e^{i\theta}) \ - \ f(e^{i\theta}) \|_{2\times 2} \tag{57}$$

In order to apply Theorem 2, rewrite (57) as

$$\sup_\theta \|(g(e^{i\theta}) - f_*(e^{i\theta}))^T (g(e^{i\theta}) - f_*(e^{i\theta}))\|_{2\times 2} = \inf_{f \in \mathcal{A}_{2\times 2}} \sup_\theta \|(g(e^{i\theta}) - f_*(e^{i\theta}))^T (g(e^{i\theta}) - f(e^{i\theta}))\|_{2\times 2} \tag{58}$$

and let γ_* be the real number equal to either term of equation (58). With

$$\Gamma(e^{i\theta}, z_1, z_2, z_3, z_4) = \begin{pmatrix} g_1(e^{i\theta}) - z_1 & g_2(e^{i\theta}) - z_2 \\ g_3(e^{i\theta}) - z_3 & g_4(e^{i\theta}) - z_4 \end{pmatrix}^T \begin{pmatrix} g_1(e^{i\theta}) - z_1 & g_2(e^{i\theta}) - z_2 \\ g_3(e^{i\theta}) - z_3 & g_4(e^{i\theta}) - z_4 \end{pmatrix}$$

the matrix $\mathcal{U}$ of Theorem 2 becomes

$$\mathcal{U} = \begin{pmatrix} -(g_1 - f_1) & 0 & 0 & 0 \\ 0 & -(g_2 - f_2) & 0 & 0 \\ 0 & 0 & -(g_3 - f_3) & 0 \\ 0 & 0 & 0 & -(g_4 - f_4) \end{pmatrix} \tag{59}$$

That is, the hypothesis of $\mathcal{U}^T\mathcal{U}$ being invertible reduces to

$$g_\ell - f_\ell \text{ is never } 0, \text{ for } 1 \le \ell \le 4 \tag{60}$$

Under condition (60), Theorem 2 guarantees that there exists a 2×2 selfadjoint matrix valued function

$$\Psi = \begin{pmatrix} \psi_1 & \psi_2 \\ \psi_3 & \psi_4 \end{pmatrix}$$

with entries $\psi_\ell \in L^1$ with ψ_1, ψ_4 real valued, $\psi_3 = \overline{\psi_2}$, and such that the triple (γ_*, f_*, Ψ_*) satisfies

$$\begin{array}{lrcl} 1. & (\gamma_* I - (g-f)^T(g-f))\Psi & = & 0 \\ 2. & \overline{(g_\ell - f_\ell)}\, \psi_\ell & \in & zH^1, \qquad \ell = 1, \ldots, 4 \\ 3. & \int (\psi_1 + \psi_4) \frac{d\theta}{2\pi} & = & 1 \\ 4. & \Psi & \ge & 0 \\ 5. & \gamma_* I - (g-f)^T(g-f) & \ge & 0 \end{array} \tag{61}$$

The existence of the function Ψ is guaranteed by hypothesis (60).

The case when hypothesis (60) does not necessarily hold may be analyzed by proceeding as in the proof of Theorem 2. In this case we have that for $1 \le \ell \le 4$ there exist Borel measures λ_ℓ, and there exist functions ϕ_ℓ in zH^1 such that

$$-\,(g_\ell - f_\ell)\, d\lambda_\ell = \phi_\ell \tfrac{d\theta}{2\pi} \tag{62}$$

For each ℓ one can distinguish three cases:

i. $g_\ell - f_\ell$ is bounded away from 0. Then one can solve for λ in (62), so λ is absolutely continuous with respect to Lebesgue measure:
$$d\lambda = \frac{1}{g_\ell - f_\ell} \tfrac{d\theta}{2\pi}$$

ii. $g_\ell - f_\ell$ is zero on a subset of the circle with positive Lebesgue measure. This forces the function ϕ_ℓ in (62) to be the function 0. This implies that the complement of the support of λ is a subset of $\{e^{i\theta}\,;\, g(e^{i\theta}) - f(e^{i\theta}) = 0\}$.

iii. $g_\ell - f_\ell$ is zero on a subset of the circle with null Lebesgue measure. In this case the function ϕ_ℓ may not be the function 0 -although we do not know of any examples. In any case it does not follow from relation (62) that λ is absolutely continuous with respect to Lebesgue measure.

4 Numerical Algorithms for H^∞ optimization.

We specialize the high level computer algorithms in §2 to H^∞ optimization problems by using the equations in PDEQH$^\infty$.

Functions in L^1 appear naturally in H^∞ optimization as *dual variables.* In many cases, these dual variables are actually bounded or at least L^2 integrable. This is always the case, for example, in the classical scalar case of approximating a continuous function g on the circle by functions in A, provided g is differentiable and a mild technical condition is satisfied.

The statement of the algorithms that we consider in this paper require that *dual variables (such as Ψ in Theorem 2) be elements of L^2.* A stronger assumption may be necessary in order to prove convergence of the algorithms, see for example [HMW], where convergence is discussed in the context of C^∞ primal and dual variables.

For example, if Ψ is a function in H^2, the statement

$$\frac{\partial\Gamma}{\partial z}(\cdot, f)\Psi \in zH^1$$

is equivalent to

$$P_{H_0^{2\perp}}[\frac{\partial\Gamma}{\partial z}(\cdot, f)\Psi] = 0$$

Implementation of the algorithms require more than just the optimal dual variables belong to L^2, as approximate or discrete representation in the computer of unbounded functions is higly inaccurate. Therefore we state the main assumption on the class of problems to be solved approximately by the algorithms in this article: **Every optimal dual function belongs to L^2.**

4.1 Interior Point Algorithm

The duality gap equation (1.) of PDEQ becomes in the current setting

$$((\gamma - \epsilon)I - \Gamma(\cdot, f))\Psi = 0 \tag{63}$$

which we rewrite as

$$(\gamma I - \Gamma(\cdot, f))\Psi = \epsilon\Psi \tag{64}$$

To allow more flexibility in computer algorithms, we replace the right hand side of (64) by a small multiple of another function, say Φ, which takes nonnegative definite values and where Φ may be either known in advance, or may depend on the unknowns of the problem. We obtain,

$$(\gamma I - \Gamma(\cdot, f))\Psi = \epsilon\Phi \tag{65}$$

Obviously (65) includes (64) as a special case. Another possible choice is $\Phi = I$. Combining (65) with 2. and 3. of $\mathbf{PDEQ^+H^\infty}$ we get

$$\mathbf{PDEQ^+H^\infty_\epsilon} \qquad \begin{array}{lrcl} 1. & (\gamma_* I - \Gamma(\cdot, f_*))\Psi_* & = & \epsilon\Phi \\ 2. & \mathcal{P}_{zH^{2\perp}}[tr(\frac{\partial\Gamma}{\partial z}(\cdot, f)\Psi_*)] & = & 0 \\ 3. & \int tr\Psi_* \frac{d\theta}{2\pi} & = & 1 \end{array}$$

Solve for f, Ψ and γ using Newton's method, with barriers or restrictions on the line search imposing positivity conditions $\Psi \geq 0$ and $\gamma I - \Gamma(\cdot, f) \geq 0$.

An example of the equations one solves is:

Optimality conditions for competing performance functions. Let $\Gamma_1(e^{i\theta}, z)$ and $\Gamma_2(e^{i\theta}, z)$ be scalar valued performance functions of $e^{i\theta}$ and $z \in C$, and set

$$\Gamma(e^{i\theta}, z) = \begin{pmatrix} \Gamma_2(e^{i\theta}, z) & 0 \\ 0 & \Gamma_1(e^{i\theta}, z) \end{pmatrix}$$

It is easy to verify that if Ψ_* is a dual optimal function, then so is the function $\Psi_1 = diag(\psi_1, \psi_2)$ obtained from Ψ_* by replacing all the off-diagonal terms by 0. The equations corresponding to $\mathbf{PDEQ^+H^\infty_\epsilon}$ with $\Phi = I$ can be written as follows:

$$\begin{array}{lcl}
(i.a) & (\gamma - \Gamma_1)\psi_1 & = \varepsilon \\
(i.b) & (\gamma - \Gamma_2)\psi_2 & = \varepsilon \\
(ii) & \mathcal{P}_{zH^{2\perp}}\,[\,\psi_1 \frac{\partial \Gamma_1}{\partial z}(\cdot, f) + \psi_2 \frac{\partial \Gamma_2}{\partial z}(\cdot, f)\,] & = 0 \\
(iii) & \int \psi_1 + \psi_2 & = 1.
\end{array}$$

The article [HV] gives conditions under which at optimum we have $\Gamma_1(\cdot, f_*) = \gamma_*$ and $\Gamma_2(\cdot, f_*) = \gamma_*$.

4.2 Boundary Algorithms

We use the name Boundary Algorithms to refer to Newton type methods applied directly to the optimality conditions or a modification of them that does not involve the real parameter ϵ.

4.2.1 Boundary Case of the Interior Point Algorithm

This consists in setting $\Phi = 0$ in $\mathbf{PDEQH^\infty_\epsilon}$ (so the equations no longer depend on ϵ), and using a Newton type method to solve the resulting problem.

4.2.2 Boundary Algorithm for scalar valued performance

This algorithm applies when the performance is scalar valued ($m = 1$), and consists in solving $\mathbf{PDEQH^\infty_\epsilon}$ with $\Phi = 0$ and with the elimination of the primal variable γ from the problem. To see how this is done, note that by Corollary 2 Ψ_* is scalar positive valued function. Then Ψ can be eliminated from (1.) of $\mathbf{PDEQH^\infty_\epsilon}$ when $\Phi = 0$, to obtain

$$\Gamma(\cdot, f_*(\cdot)) = \gamma$$

Now apply a projection to get

$$\mathcal{P}_{H^{2\perp}}[\Gamma(\cdot, f)] = 0$$

Note that by doing this the inequality constraint

$$\gamma - \Gamma(\cdot, f) \geq 0$$

is no longer part of the optimization problem. See [HMW] for details.

4.3 Interior Point Algorithm 2

The distinctive feature of this algorithm is that the dual variable Ψ is eliminated from the problem. We now describe some simple algebra which will do this.

Take $\Phi = I$ in $\mathrm{PDEQH}^\infty_\epsilon$. We may solve 1. of $\mathrm{PDEQH}^\infty_\epsilon$ for Ψ and get

$$\Psi_\varepsilon = \varepsilon(\gamma - \Gamma(\cdot, f))^{-1}.$$

which is an equation of type 23. Substituting this into (2) of PDEQH^∞ gives

$$(1.L_\varepsilon) \qquad \mathcal{P}_{zH^2}[(\gamma - \Gamma(\cdot, f))^{-1}\frac{\partial \Gamma}{\partial z}(\cdot, f)] \;= 0.$$

Also, the normalization requirement in (3) of PDEQH$^\infty$ becomes the normalization

$$(2.L_\varepsilon) \qquad \int tr\,(\gamma - \Gamma(\cdot, f))^{-1}\,\frac{d\theta}{2\pi} \;=\; 1/\varepsilon.$$

Finally one solves $((1.L_\varepsilon))$ and $((2.L_\varepsilon))$ for γ, f.

There is an elegant way to reformulate equation $(2L_\varepsilon)$. The left term $(2L_\varepsilon)$ can be written as

$$(\gamma_* - \Gamma)^{-1}\frac{\partial(\gamma_* - \Gamma)}{\partial z}$$

which is a logarithmic derivative. Since (1_ε) implies Ψ is a uniformly positive definite matrix function, there exists a selfadjoint matrix L satisfying $\Psi = e^{-L}$. Thus $(2L_\varepsilon)$ becomes

$$\mathcal{P}_{zH^2}[\frac{\partial L}{\partial z}] \;= 0$$

Now we describe some connections with traditional H^∞ optimization. If γ is bigger than the optimal value γ^* for (OPT), then there are many $f \in A_N$ which satisfy $\gamma \geq \Gamma(e^{i\theta}, f(e^{i\theta}))$. One way to pinpoint a unique function is by choosing the f_*^γ which maximizes the entropy[4]

$$E(f) := \int_0^{2\pi} \ln\det[\gamma I - \Gamma(e^{i\theta}, f(e^{i\theta}))]\tfrac{d\theta}{2\pi}$$

That is,

$$f_*^\gamma \;:=\; \underset{\{f\in A_N : \Gamma(e^{i\theta}, f(e^{i\theta}))\leq\gamma\ \forall\theta\}}{\operatorname{argmax}} \;\; E(f)$$

The connection with IPA2 comes from

Lemma 4.1 *If the maximum entropy solution f_*^γ is continuous, then it satisfies $(1.L_\epsilon)$ as well as $PDEQ^+H_\epsilon^\infty$ for $\epsilon = \gamma - \gamma_*$.*

Proof: We shall use a basic fact about an invertible matrix valued function $A(x)$ of x. This is,

$$\frac{d}{dx}\ln\det A(x) = tr[A(x)^{-1}\frac{dA(x)}{dx}] \tag{66}$$

Since f_*^γ maximizes the entropy and also is in the interior of the set [BV] $\{f \in A_N : \Gamma(e^{i\theta}, f(e^{i\theta})) \leq \gamma\,\forall\theta\}$, we have that

$$\frac{d}{dt}\;E(f_*^\gamma + th)|_{t=0} = 0 \;\;\forall h \in \mathcal{A}_N \tag{67}$$

Equation (67) combined with (66) gives

$$\sum_{\ell=1}^{N}\int_0^{2\pi} tr((\gamma I - \Gamma)^{-1}\frac{\partial\Gamma}{\partial z_j})h_j\;\tfrac{d\theta}{2\pi} = 0 \tag{68}$$

this is equivalent to $(1.L_\epsilon)$.

[4]See[DG] for the special case Nehari Problem of approximating a continuous function by analytic functions in the sup-norm.

5 Numerical Results

5.1 Finding local solutions to OPT

Each one of the algorithms described in this paper at some point requires us to solve an equation of the form

$$\mathcal{T}[x] = 0 \tag{69}$$

where x is an element of a suitable space. When solving nonlinear equations such as (69) Newton-type methods generate sequences $x_1, x_2, \ldots, x_\ell, \ldots$ from solutions to the Newton step equation

$$\mathcal{D}\mathcal{T}_{x_\ell}[h_\ell] = -\mathcal{T}[x_\ell] \tag{70}$$

or a modification of it, followed by a search step to determine $t_\ell > 0$ that gives the update

$$x_{\ell+1} = x_\ell + t_\ell h_\ell$$

Our implementation of algorithms used in the tests reported in this section is based on representing functions on the circle by using a grid of equally spaced points on the circle, or equivalently, by a finite sequence of (discrete) Fourier coefficients. The discrete representation of the Newton step equation (70) is just an ordinary finite dimensional system of equations where the unknowns are the Fourier coefficients of the unknown functions of the original problem. Such a system of equations is solved on the computer by using standard linear algebra techniques.

We remind the reader that the methods for solving OPT discussed in this article are based on first order conditions that must be satisfied by solutions. Clearly, some convergent sequences generated by the algorithms in this paper may have a limit which is a local solution and not a global one, or the limit might be a "saddle point".

5.2 Numerical Experiments

For the numerical experiments $n \times n$ matrix valued performance functions $\Gamma(e^{i\theta}, z)$ were chosen with $n = 1$ or $n = 2$, with various N values (recall that $z \in C^N$), and with sublevel sets

$$\mathcal{S}_\theta(c) = \{z \in C^N : \ \|\Gamma(e^{i\theta}, z)\|_{n\times n} \leq c\} \tag{71}$$

having different characteristics (smoothness, convexity, etc.) The list of cases is given in Table 1.

When the performance Γ has diagonal structure (e.g., cases III − V in Table 1) one can view the corresponding OPT problem as an optimization problem with *competing performance functions.* See §2.3.3 and §4.1.

Table 1

Case	Performance $\Gamma(e^{i\theta}, z)$	N, n, and sublevel sets
I	$Re[g_1(e^{i\theta}) + z_1]^2 + 4Im[g_1(e^{i\theta}) + z_1]^2$ $+Re[g_2(e^{i\theta}) + z_2]^2 + 9Im[g_2(e^{i\theta}) + z_2]^2$	$N = 2,\ n = 1,$ smooth convex
	Results of tests: Local quadratic convergence rate to the solution. *Properties of the solutions:* The optimal primal and dual functions are smooth and unique.	
II	$\|z_1\|^2 + \|100 + e^{i\theta} z_1 + 0.1(z_1 + z_2 + z_1 z_2)\|^2$ $+\|z_2\|^2 + \|100 + e^{i\theta} z_2 + 0.1(z_1 + z_2 + z_1 z_2)\|^2$	$N = 2,\ n = 1,$ smooth non-convex
	Results of tests: Local quadratic convergence rate to local solutions or to "saddle points". See [HMW]. *Properties of the solutions:* The optimal primal and dual functions are smooth. Optimal primal functions are locally unique. The dual function associated with each optimal primal function is unique.	
III	$\begin{pmatrix} re[\frac{1}{e^{i\theta}} + z]^2 & 0 \\ 0 & im[\frac{1}{e^{i\theta}} + z]^2 \end{pmatrix}$	$N = 1,\ n = 2$ non-smooth convex
	Results of tests: Convergence to solution is apparent, but the numerics become bad once the current iterate is within 0.01 from the solution. Increasing the grid size produces very little improvement in the numerics, in contrast to other better behaved cases. *Properties of the solutions:* The optimal primal function f_* and the dual optimal function Φ_* are unique. Numerical evidence indicates that f_* is continuous and non-differentiable on ∂D, and that Ψ_* is not bounded. Moreover, $\Psi_* = diag(\psi_1, \psi_2)$ where for $\ell = 1, 2$, ψ_ℓ is 0 on a subset J_ℓ of the circle, and $J_1 \cap J_2$ consists of just two points.	
IV	$\begin{pmatrix} re[\frac{1}{e^{i\theta}} + z_1]^2 + im[\frac{1}{e^{i\theta}} + z_2]^2 & 0 \\ 0 & im[\frac{1}{e^{i\theta}} + z_1]^2 + re[\frac{1}{e^{i\theta}} + 12 + z_2]^2 \end{pmatrix}$	$N = 2,\ n = 2$ non-smooth convex
	Results of tests: Linear convergence rate was obtained in all tests. Increasing the grid size produces very little improvement in the numerics, in contrast to other better behaved cases. *Properties of the solutions:* Numerical evidence suggests that the (primal) solution f_* is smooth, and that dual functions are continuous but not smooth. Moreover, $\Psi_* = diag(\psi_1, \psi_2)$ where for $\ell = 1, 2$, ψ_ℓ is 0 on a subset J_ℓ of the circle, and $J_1 \cap J_2$ contains nontrivial arcs of the circle.	
V	$\begin{pmatrix} \|\frac{1}{e^{i\theta}} + z_1\|^2 & 0 \\ 0 & \|\frac{0.2}{e^{i\theta}} + z_2\|^2 \end{pmatrix}$	$N = 2,\ n = 2$ non-smooth convex
	Results of tests: Local quadratic convergence rate. *Properties of the solutions:* There exists an open set of optimal primal functions f. The dual function associated with each optimal primal function is unique.	
VI	$\begin{pmatrix} \frac{(-1+e^{i\theta})^2}{4e^{i\theta}} - z_1 & 4 - z_2 \\ \frac{(-1+e^{i\theta})^2}{1+3(e^{i\theta})^2} - z_3 & \frac{1}{e^{i\theta}} - z_4 \end{pmatrix}^T \begin{pmatrix} \frac{(-1+e^{i\theta})^2}{4e^{i\theta}} - z_1 & 4 - z_2 \\ \frac{(-1+e^{i\theta})^2}{1+3(e^{i\theta})^2} - z_3 & \frac{1}{e^{i\theta}} - z_4 \end{pmatrix}$	$N = 4,\ n = 2$ non-smooth convex
	Results of tests: Local quadratic convergence in all tests. *Properties of the solutions:* There exists an open ball of optimal primal functions f. At the solution the dual variable Ψ is a pointwise rank 1 square 2×2 matrix, with eigenvalues uniformly separated by a nonzero gap, thus this is a "smooth" situation.	

5.2.1 Conclusions

All the conclusions in this subsection concern local convergence only. In all examples tested the optimal dual function turned out to be unique. In the computer runs performed (with BCIPA) the order of convergence is related to the characteristics of the optimal dual function Ψ.

a. **Quadratic order of convergence**. This occurs in all cases where the optimal dual function Ψ is smooth. (Cases I, II, V, and VI). Note that Cases I and II have scalar valued performance, case V has diagonal-matrix valued performance, and case VI has (full) 2×2 matrix valued performance.

b. **Linear order of convergence**. Happens in case IV, where the optimal dual function Ψ is continuous but not differentiable.

c. **Bad numerical behavior**. This occurred in case III, where the optimal dual function is not continuous. Once the iterate is in certain neighborhood, the update gives marginal improvement, or no improvement at all. Increasing the number of gridpoints on the circle used in the discretization does not help.

A remarkable situation occurs in cases V and VI. In each case, there is an infinite number of primal solutions f (although the optimal dual function is unique). Consequently, the jacobian of the nonlinear operator $\mathcal{T}$ at a primal-dual solution has a nontrivial null space. It is very surprising that the BCIPA algorithm delivers quadratic order of convergence. This is not an isolated behavior, as we performed additional tests whose results are consistent and that rule out this possibility. This is in agreement with the results tests reported in [HO], where a BCIPA type algorithm was applied to the (considerably simpler) case of minimizing the largest singular value of a nonlinear pencil of matrices over R^K (In [HO], R^K replaces the analytic functions.) In this article we do not do analysis of convergence, but only report on numerical tests. Numerical analysis issues will be discussed in a future article.

As we saw six different performance functions in Table 1 were used to test BCIPA. Further tests were performed to test IPA, with results that seemed to indicate that, except for the fact that IPA has an additional outer loop so it is less efficient, both algorithms exhibit similar behavior. We do not report details of those tests here since for local convergence BCIPA seems superior.

To this point our attention has focused on local convergence and what modifications of these methods are required for good global convergence has not been addressed. The algorithms of section 6 gave reasonable global results in informal tests when $n = 1$, as in cases I as in II.

5.2.2 Set up of numerical experiments

At least 10 computer runs were done for each performance function listed in Table 1. The initial guess was produced by a random perturbation of size ≤ 0.1 of the first five Fourier coefficients of an approximate local solution. The size of grid on the circle for sampling functions in all tests was 256. The number of Fourier coefficients used in building discrete approximations to the nonlinear operator $\mathcal{T}$ in (69) were 32. The runs were conducted on a Sun SparcStation 20 with 48MB RAM. The algorithm was implemented in Mathematica version 2.2. We report only the tests on BCIPA.

The implementation of the algorithms requires a reparametrization. For example, in Case VI of Table 1 we used the following expression for the dual variable:

$$\Psi = \begin{pmatrix} re(\beta_1) & \beta_2 + i\beta_3 \\ \beta_2 - i\beta_3 & \beta_4 \end{pmatrix}$$

Here β_ℓ , $\ell = 1, \ldots, 4$ are scalar valued analytic functions, where β_3 satisfies the following constraint:

$$\int \beta_3 \tfrac{d\theta}{2\pi} = 0$$

Also, equation 1 of PDEQH$^+$ is redundant. To cut down unnecessary equations, note that at optimum the error function $\gamma I - \Gamma(\cdot, f)$ and the dual function Ψ share a basis of eigenvectors (at each θ). Therefore they commute, which can be used to show that at optimum, equation 1 of PDEQH$^+$ is equivalent to the symmetization of it, i.e.,

$$(\gamma I - \Gamma(\cdot, f)\Psi + ((\gamma I - \Gamma(\cdot, f)\Psi)^T = 0$$

This equation is what is actually used in calculations. Clearly, a similar thing can be done with PDEQH$^+_\epsilon$.

5.2.3 Computer output

To give a more precise idea of the experiments we give in Table 2 a description of the computer output diagnostics, and in Table 3 we present sample output from computer runs with algorithm BCIPA.

Table 2

Diagnostic	Meaning
Iter	j, the current iteration number.
Perf. Value	$\sup_\theta \|\Gamma(\cdot, f_j)\|_{n\times n}$, norm of current value of performance.
Oper. Value	$\|\mathcal{T}[x_j]\|_{L^2}$, norm of current value of nonlinear operator in (69). Here x_j contains both primal and dual variables.
Flatness	$(\max_\theta \sigma_{max}(\Gamma(\cdot, f)) - \min_\theta \sigma_{max}(\Gamma(\cdot, f)))/\max_\theta \sigma_{max}(\Gamma(\cdot, f))$. By Corollary 1, Flatness diagnostic $= 0$ at local solutions f.
P.D. Step	$\|x_j - x_{j-1}\|_{L^2}$, Primal-Dual step, L^2 norm of change in variable x.
P. Error	$\sup_\theta \|f(e^{i\theta}) - f_*(e^{i\theta})\|_{C^N}$, the error in the primal variable, when the solution f_* is known.

Table 3

Iter. j	Perf. Value $\sup_\theta \|\Gamma(\cdot, f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness	P.D. Step $\|x_j - x_{j-1}\|_{L^2}$	P. Error $\|f_j - f_*\|_\infty$
0	2.24089067735161	9.68942	0.464695	*None*	0.397374
1	1.966525096922556	0.726568	0.354695	3.64753	0.293018
2	1.774815446305852	0.247886	0.218135	5.01624	0.145684
3	1.643104279677681	0.047166	0.040159	3.41706	$2.69\,10^{-2}$
4	1.612600392051017	$2.35\,10^{-3}$	$6.54\,10^{-4}$	0.71695	$4.10\,10^{-3}$
5	1.612022018047374	$4.18\,10^{-5}$	$7.28\,10^{-6}$	0.03752	$1.62\,10^{-5}$
6	1.612017921876376	$1.38\,10^{-6}$	$9.73\,10^{-7}$	$5.14\,10^{-4}$	$8.83\,10^{-7}$

An run of BCIPA with the performance function in case VI of Table 1. Gridsize: 256. Number of Fourier coefficients: 32. The initial guess was generated by perturbing each one of the first five Fourier coefficients of the solution with a random perturbation with magnitude < 0.1.

6 More Algorithms for H^∞ Optimization

6.1 The class of Algorithms $\mathcal{H}$

In order to improve global behavior of algorithms we consider an analog of steepest descent in R^K optimization. The basic problem here is that our Newton type methods converge to saddle points in function space, analogously to classical optimization of smooth functions in R^K.

We now consider a variant of a method introduced in [HMer1]. one replaces the original performance Γ with a modified performance

$$\Gamma_H(\cdot, z) = \Gamma(\cdot, f) + \mathcal{D}\Gamma_f[z] + H[e^{i\theta}, z]$$

where $H[e^{i\theta}, z]$ is a positive definite self adjoint valued continuous function of $e^{i\theta} \in \partial D$ and $z \in C^N$ that is quadratic and that one chooses at convenience. Note that for θ fixed, the terms of order up to 1 in the Taylor expansions of $\Gamma_z(\cdot, z)$ about $z = 0$ and of $\Gamma(\cdot, z)$ about $z = f$ coincide. We now state the step.

Step of Algorithm Given a performance function $\Gamma(e^{i\theta}, z)$, a positive definite valued bilinear form $H[e^{i\theta}, z]$ and a guess f at the solution to (OPT), to update f

i. *Subproblem:*Find $h \in \mathcal{A}_N$ that minimizes $\sup_\theta \|\Gamma_H(\cdot, h)\|_{n\times n}$.

ii. *Line search:* Find $t \geq 0$ that minimizes $\sup_\theta \|\Gamma(\cdot, f + th)\|_{n\times n}$.

iii. *Update:*$f \leftarrow f + th$.

One can show that if the subproblem (i) produces an update direction $h \in \mathcal{A}_N$ which is not in the null space of $\mathcal{D}\Gamma$, then the norm of the performance decreases along this direction (see [HMer2]).

Note that, in contrast to the scalar valued performance case (n=1), in the matrix valued case ($n > 1$) the subproblem (i) is not a "disk problem", that is, the sublevel sets $\mathcal{S}_\theta(c) = \{z \in C^N : \|\Gamma_H(e^{i\theta}, z)\|_{n\times n} \leq c\}$ are not disks in C^N. Furthermore, $\mathcal{S}_\theta(c)$ has corners in its boundary.

6.2 Saddle points vs. minima

When the function H in Algorithm $\mathcal{H}$ is not near or equal to the true hessian, one expects no better than linear convergence rate. In R^m minimization this type of algorithms with generic initialization produce sequences convergent to local minimum, while Newton type methods produce sequences that often converge to a saddle point. We have observed a similar phenomenon with H^∞ optimization. In [HMW] an example of convergence of iterates produced with Newton's method to a saddle point was reported (scalar valued performance). Here we report our experience with the same problem but treated with an algorithm $\mathcal{H}$.

The (scalar) performance function is

$$\Gamma = |z_1|^2 + |z_2|^2 + |100 + e^{i\theta} z_1 + .1(z_1 + z_2 + z_1 z_2)|^2 + |100 + e^{i\theta} z_2 + .1(z_1 + z_2 + z_1 z_2)|^2$$

The performance function has a saddle point at $f_0 = (0, 0)$ with value $20,000$. The optimality conditions PDEQH$^\infty$ are satisfied at f_0.

We tested two algorithms with the initial guess $f_1(e^{i\theta}) = (.009 + .001e^{i\theta}, .002 - .003e^{i\theta})$ and a gridsize of 256. As reported in [HMW], BCIPA produces iterates that converge to the saddlepoint f_0. However, the iterates produced by the $\mathcal{H}$ algorithm with $H[z,z] = \|z\|^2$ roll off the saddle point and approach a local minimizer (See Table 4).

Table 4

Iter. j	Perf. Value $\sup_\theta \|\Gamma(\cdot, f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness	P.D. Step $\|x_j - x_{j-1}\|_{L^2}$
0	20004	N/A	0.000435	42.4215
1	20000	9.47	0.000063	42.3667
2	18097	2724.74	0.020397	31.4222
3	8756	1315.86	0.230767	17.2136
4	7123	2521.01	0.255536	16.2617

One can see in Table 4 that the iterates move away from the saddle point. When the computer run was terminated, f was somewhat close to the local minimum $(30,-30)$ and is making slow progress. Due to the fact that no smoothing[5] was used, numerical error began to creep into the system. Mild smoothing will cause the code to make even better progress.

A mixed strategy in which the $\mathcal{H}$ algorithm is used in initial stages, and then followed with pure Newton iterations produces good results in practice. We illustrate this by restarting the run shown in Table 4. The last iterate in the run shown in Table 4 is used to initialize a run with Pure Newton (after a small amount of smoothing). Results are shown in Table 5.

Table 5

Iter. j	Perf. Value $\sup_\theta \|\Gamma(\cdot, f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness	P.D. Step $\|x_j - x_{j-1}\|_{L^2}$
0	7180	N/A	0.263999	N/A
1	4833	486.43	0.121588	8.93833
2	4490	358.75	0.100381	7.65556
3	3960	137.90	0.050897	3.04164
4	3861	63.77	0.022891	1.25022
5	3816	19.61	0.009958	0.61894
6	3802	2.65	0.001730	0.09880
7	3800	0.08	0.000034	0.00215
8	3800	310^{-5}	1.210^{-8}	610^{-7}

Note the quadratic convergence rate showing in iterations 7, 8 of Table 4. We have found that this function in particular has proved difficult for certain methods which rely only on the first order terms of Γ. See [HMW] for an example of this.

And one final remark on these runs. Due to the small grid in the run of Table 4 a small amount of numerical error creeps in. In our run above, the aggregate affect of this is noticeable by the last few iterations. In practice it is convenient to smooth the current guess as the run progresses in order to improve performance. Doing this would improve dramatically the numbers in Table 4. However, we presented a more conservative, strict interpretation of the algorithm $\mathcal{H}$ above without smoothing.

We have not performed a theoretical analysis of the methods discussed in this section but experiments behave very predictably and give us confidence that these methods are effective for treating OPT.

[5] By smoothing we mean tapering the Fourier coefficients.

Part II

PLANT UNCERTAINTY

7 The Problem

7.1 Engineering Motivation

Suppose our objective is to design a linear, time invariant, system in the frequency domain description, with stability a major consideration. Typically part of the system is given and we are forced to use it (in control this is called the plant), and another part f we get to choose (or designable part) (see Fig. 1). The stipulation that the designable part of the circuit be stable amounts to requiring that f has no poles in the right half plane (R.H.P.)

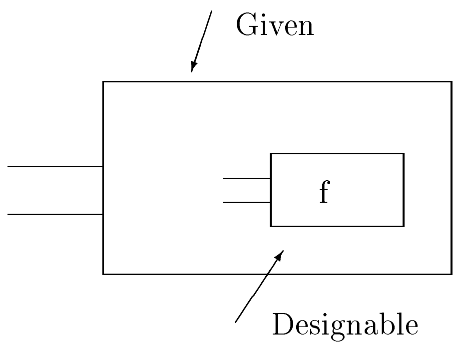

Figure 1.

Suppose also that to carry out the design (i.e., choose f) we use a given performance function Γ. In many cases the performance takes the form of a positive function[6] of frequency ω and of the designable part f at that frequency. Thus we write $\Gamma(\omega, f(i\omega))$ for the performance at frequency ω and for designable function f.

The performance is a cost function: big is bad. The "worst case performance" occurs at the frequency ω at which

$$\sup_{\omega \in R} \Gamma(\omega, f(i\omega)) \tag{72}$$

is attained. We wish to minimize (72) over the set $\mathcal{A}_{RHP}$ of R.H.P. stable functions f. Hence we have the following (scalar) OPT problem

OPT-RHP Given a (nonnegative, scalar valued) performance $\Gamma(\omega, z)$, find

$\inf_{f \in \mathcal{A}_{RHP}} \ \sup_{\omega \in R} \ \Gamma(\omega, f(i\omega))$

and find a minimizer $f_* \in \mathcal{A}_{RHP}$ if it exists.

A minimizer f_* is also called solution. A local solution is a solution to OPT over some open subset of $\mathcal{A}_{RHP}$.

[6]More general performance functions are possible but not considered now.

In physical systems there is always uncertainty. A way to treat the uncertainty is to write the performance as a function of an uncertainty parameter $\alpha \in R^K$. We denote a performance function that is also a function of an uncertainty parameter α by

$$\tilde{\Gamma}(\ \omega\ ,\ \alpha\ ,\ f(\ i\omega\ )\ )$$

The range $\mathcal{P}_\omega \subset R^K$ of the uncertainty parameter α may change with the frequency ω and is often known before choosing f.

If we fix the designable function f and the frequency ω, the worst case in terms of uncertainty is

$$\Gamma(\omega, f((i\omega))) := \sup_{\alpha \in \mathcal{P}_\omega} \tilde{\Gamma}(\omega, \alpha, f(i\omega)) \tag{73}$$

Thus we have the following optimization problem (with scalar performance):

UNCOPT-RHP Given a (nonnegative, scalar valued) performance $\tilde{\Gamma}(\omega, \alpha, z)$, find

$\inf_{f \in A_{RHP}}\ \sup_{\omega \in R}\ \sup_{\alpha \in \mathcal{P}_\omega}\ \tilde{\Gamma}(\omega, \alpha, f(i\omega))$

and find a minimizer $f_* \in A_{RHP}$ if it exists.

7.2 The Problem UNCOPT

We now change coordinates from the imaginary axis to the unit circle, and from the RHP to the unit disk.

Definition 2 *A* family of uncertainty sets *is a one parameter family $\mathcal{P} = \{\mathcal{P}_\theta :\ 0 \le \theta < 2\pi\ \}$ of subsets of R^k that are given by a real valued function $g(e^{i\theta}, \alpha)$ as follows:*

$$\begin{aligned} \mathcal{P}_\theta &= \{\alpha \in R^K : g(e^{i\theta}, \alpha) < 0\} \\ \partial\mathcal{P}_\theta &= \{\alpha \in R^K : g(e^{i\theta}, \alpha) = 0\} \end{aligned} \tag{74}$$

Furthermore, the function g satisfies $g(e^{i\theta}, \alpha) = g(e^{-i\theta}, \alpha)$ for all θ and all $\alpha \in \mathcal{P}_\theta$, so in particular, $\mathcal{P}_\theta = \mathcal{P}_{-\theta}$. The function $g(e^{i\theta}, \cdot)$ is referred to as defining function *for the set $\mathcal{P}_\theta$. The family is* smooth *if the function g is smooth (that is, continuously differentiable).*

Definition 3 *A* (scalar) uncertainty performance function *is a nonnegative valued function $\tilde{\Gamma}$ of $(e^{i\theta}, \alpha, z) \in \partial D \times R^K \times C^N$ such that*

1. *$\tilde{\Gamma}$ is real symmetric: $\tilde{\Gamma}(e^{i\theta}, \alpha, z) = \tilde{\Gamma}(e^{-i\theta}, \alpha, \overline{z})$.*
2. *$\tilde{\Gamma}$ is strictly plurisubharmonic.*

The H^∞ optimization problem with an uncertainty parameter is:

UNCOPT Given an uncertainty performance $\tilde{\Gamma}(e^{i\theta}, \alpha, z)$ and a family $\mathcal{P}$ of uncertainty sets, find

$\inf_{f \in A}\ \sup_\theta\ \sup_{\alpha \in \mathcal{P}_\theta}\ \tilde{\Gamma}(e^{i\theta}, \alpha, f(e^{i\theta}))$

and find a minimizer $f_* \in A$ if it exists.

A didactic example. The problem is to find minimizers f for

$$\inf_{f\in\mathcal{A}_\mathcal{N}} \sup_{\theta} \sup_{\alpha_1^2+\alpha_2^2<0.25} \left| \frac{1}{e^{i\theta}-\alpha_1-\alpha_2 i} - f(e^{i\theta}) \right|^2 . \tag{75}$$

Note that this is an UNCOPT problem with uncertainty performance function

$$\tilde{\Gamma}(e^{i\theta},\alpha_1,\alpha_2,z) = \left| \frac{1}{e-\alpha_1-\alpha_2 i} - z \right| \tag{76}$$

Our discusion below foreshadows some ideas that will be used later. For fixed $z \in C$ and $e^{i\theta} \in \partial D$, the function $\tilde{\Gamma}$ gives the distance from z to a point of the disk

$$B(\frac{4}{3}e^{-i\theta}, \frac{2}{3}) = \{\beta : \beta = \frac{1}{e-\alpha_1-\alpha_2 i} \text{ for some } \alpha_1, \alpha_2\}$$

By geometric considerations we see that whenever

$$z \neq \frac{4}{3}e^{-i\theta} \tag{77}$$

that is, z is not the center of $B(\frac{4}{3}e^{-i\theta}, \frac{2}{3})$, the function $\tilde{\Gamma}(e^{i\theta},\alpha_1,\alpha_2,z)$ has exactly two smoothly varying critical points $(\alpha_1,\alpha_2) = a_1(e^{i\theta},z)$ and $(\alpha_1,\alpha_2) = a_2(e^{i\theta},z)$, one giving a minimum and the other giving a maximum.

If values of the solution $f_* \in \mathcal{A}_1$ to the problem (75) do not come close to violating (77), then solutions to (75) may be written as

$$\inf_{f\in A} \sup_{\theta} \max\left\{ \tilde{\Gamma}(\cdot,a_1(\cdot,f),f) \ , \ \tilde{\Gamma}(\cdot,a_2(\cdot,f),f) \right\} \tag{78}$$

Note that dependence on $e^{i\theta}$ has been supressed in several terms of (78).

Expression (78) suggests that UNCOPT may be transformed to an OPT problem with diagonal matrix valued performance. Indeed,

$$\max\{\Gamma_1(e^{i\theta},z),\Gamma_2(e^{i\theta},z)\} = \left\| \begin{pmatrix} \Gamma_1(e^{i\theta},z) & 0 \\ 0 & \Gamma_2(e^{i\theta},z) \end{pmatrix} \right\|_{2\times 2}$$

7.3 Optimality conditions for UNCOPT

We now give the optimality conditions for local solutions to UNCOPT. See [HMW2] for the statement of an earlier result.

Definition 4 *Let $\tilde{\Gamma}(e^{i\theta},\alpha,z)$ be an uncertainty performance function and let $\mathcal{P}$ be a family of uncertainty sets*

$$\mathcal{P}_\theta = \left\{\alpha \in R^k : g(e^{i\theta},\alpha) \le 0\right\} \tag{79}$$

For $e^{i\theta_0} \in \partial D$ and $z_0 \in C^N$, we say that $\alpha_0 \in R^K$ is an interior critical point *of $\tilde{\Gamma}(e^{i\theta_0},\cdot,z_0)$ if*

$$g(e^{i\theta_0},\alpha_0) < 0 \quad and \quad \frac{\partial\tilde{\Gamma}}{\partial\alpha}(e^{i\theta_0},\alpha_0,z_0) = 0, \tag{80}$$

and we say that $\alpha_0 \in R^K$ is a boundary critical point *of $\tilde{\Gamma}(e^{i\theta_0},\cdot,z_0)$ if*

$$g(e^{i\theta_0},\alpha_0) = 0 \quad and \quad \exists\eta \in R \quad such\ that \quad \frac{\partial\tilde{\Gamma}}{\partial\alpha}(e^{i\theta_0},\alpha_0,z_0) - \eta\,\frac{\partial g}{\partial\alpha}(e^{i\theta_0},\alpha_0) = 0. \tag{81}$$

Theorem 3 *Let $\tilde{\Gamma}(e^{i\theta}, \alpha, z)$ be a uncertainty performance function and let $\mathcal{P}$ be a family of uncertainty sets*

$$\mathcal{P}_\theta = \left\{\alpha \in R^k : g(e^{i\theta}, \alpha) \leq 0\right\} \tag{82}$$

Suppose that there exist continuously differentiable functions $a_1(e^{i\theta}, z), \ldots, a_m(e^{i\theta}, z)$ such that for every fixed $z \in C$ and every fixed $e^{i\theta} \in \partial D$,

$$\alpha \text{ is a critical point of } \tilde{\Gamma}(e^{i\theta}, \cdot, z) \iff \alpha \in \left\{a_1(e^{i\theta}, z), \ldots, a_m(e^{i\theta}, z)\right\} \tag{83}$$

Let $f_* \in \mathcal{A}_N$ be a local solution to (UNCOPT). For each $r \in \{1, \ldots, m\}$ let

$$\alpha_r = a_r(\cdot, f_*), \tag{84}$$

and set

$$\gamma_* = \sup_\theta \max_{1 \leq r \leq m} \tilde{\Gamma}(e^{i\theta}, \alpha_r(e^{i\theta}), f_*(e^{i\theta})) \tag{85}$$

If

$$\sum_{r=1}^m \sum_{\ell=1}^N \left|\frac{\partial \tilde{\Gamma}}{\partial z_\ell}(\cdot, \alpha_r, f_*)\right|^2 \quad \text{is pointwise invertible on } \partial D, \tag{86}$$

then there exist nonnegative valued functions $\Psi_r \in L^1$, $1 \leq r \leq m$, such that

$$\begin{array}{lrcll}
1. & \left(\gamma_* - \tilde{\Gamma}(\cdot, \alpha_r, f_*)\right) \Psi_r & = & 0 & 1 \leq r \leq m \\
2. & \sum_{r=1}^m \frac{\partial \tilde{\Gamma}}{\partial z_\ell}(\cdot, \alpha_r, f_*)\Psi_r & \in & zH^1, & 1 \leq \ell \leq N \\
3. & \sum_{r=1}^m \int \Psi_r \frac{d\theta}{2\pi} & = & 1 & \\
4. & \Psi_r & \geq & 0 & 1 \leq r \leq m \\
5. & \gamma_* - \Gamma(\cdot, \alpha_r, f_*) & \geq & 0 & 1 \leq r \leq m
\end{array} \tag{87}$$

Proof. Set

$$\Gamma_r(e^{i\theta}, z) := \tilde{\Gamma}(e^{i\theta}, \alpha_r(e^{i\theta}, z), z)$$

and

$$\Gamma(e^{i\theta}, z) := diag\left(\Gamma_1(e^{i\theta}, z), \ldots, \Gamma_m(e^{i\theta}, z)\right)$$

Then $\Gamma(e^{i\theta}, z)$ is a smooth performance function (with no uncertainty parameter). We have,

$$\begin{aligned}
\|\Gamma(e^{i\theta}, z)\|_{m \times m} &= \|diag\left(\Gamma_1(e^{i\theta}, z), \ldots, \Gamma_m(e^{i\theta}, z)\right)\|_{m \times m} \\
&= \max_{1 \leq k \leq m} \Gamma_k(e^{i\theta}, z) \\
&= \max_{1 \leq k \leq m} \tilde{\Gamma}(e^{i\theta}, \alpha_r(e^{i\theta}, z), z) \\
&= \max_{\alpha \in \mathcal{P}_\theta} \tilde{\Gamma}(e^{i\theta}, \alpha, z)
\end{aligned} \tag{88}$$

so it is clear from (88) that solving UNCOPT for $\tilde{\Gamma}$ is equivalent to solving OPT for Γ, and that both problems have the same local solutions in $\mathcal{A}_N$.

Claim: $\frac{\partial \Gamma_r}{\partial z_\ell}(e^{i\theta}, f_*(e^{i\theta})) = \frac{\partial \tilde{\Gamma}}{\partial z_\ell}(e^{i\theta}, \alpha_r(e^{i\theta}), f_*(e^{i\theta}))$
To prove the claim note that the chain rule gives

$$\frac{\partial \Gamma_k}{\partial z_\ell}(e^{i\theta}, f_*) = \frac{\partial \tilde{\Gamma}}{\partial \alpha}(\cdot, \alpha_r, f_*)\frac{\partial a_r}{\partial z}(\cdot, f_*) + \frac{\partial \tilde{\Gamma}}{\partial z_\ell}(\cdot, \alpha_r, f_*) \tag{89}$$

By hypothesis (83) we have that for each fixed $e^{i\theta}$ there are two possibilities:

1.) α_r is an interior critical point, hence

$$\frac{\partial \tilde{\Gamma}}{\partial \alpha}\left(e^{i\theta}, \alpha_r(e^{i\theta}), f_*(e^{i\theta})\right) = 0 \tag{90}$$

2.) α_r is a boundary critical point. In this case one may combine equation (81) with the fact that $\frac{\partial a}{\partial z_\ell}(e^{i\theta}, \alpha_r(e^{i\theta}))$ is tangent to the boundary of the uncertainty set at $\alpha = f_*(e^{i\theta})$, to conclude that

$$\frac{\partial \tilde{\Gamma}}{\partial \alpha}\left(e^{i\theta}, \alpha_r(e^{i\theta}), f_*(e^{i\theta})\right)\frac{\partial a_r}{\partial z}\left(e^{i\theta}, f_*(e^{i\theta})\right) = 0 \tag{91}$$

Therefore we see from (90) and (91) that the right hand side of (89) reduces to $\frac{\partial \tilde{\Gamma}}{\partial z_\ell}(\cdot, \alpha_r, f_*)$. This concludes the proof of the claim.

By the claim we have that hypothesis (86) is just hypothesis (47) when Theorem 2 is applied to Γ. By the latter result there exists a self adjoint matrix valued function Ψ with entries in L^1 such that the optimality relations PDEQ$^+H^\infty$ hold. By remark (3.2), the function Ψ may be chosen so that it has block diagonal structure identical to that of Γ, that is,

$$\Psi = diag(\Psi_1, \ldots, \Psi_m) \tag{92}$$

where Ψ_k is a nonnegative valued function, $k = 1, \ldots, m$. Combine (92) with relations ($\mathbf{PDEQ^+H^\infty}$) to obtain the conclusion of the theorem. $\square$

Obviously the critical point functions $a_k(e^{i\theta}, z)$ in Theorem 3 need only be defined in a neighborhood of $z = f_*(e^{i\theta})$.

In many cases uncertainty performance functions will have, at solutions f_* to UNCOPT, critical points $\alpha_r(e^{i\theta}) = a_k(e^{i\theta}, f_*(e^{i\theta}))$ which for fixed k are always (in θ) interior or always boundary points. One can show that under a nondegeneracy condition (this is just the implicit function theorem) the functions $a_k(e^{i\theta}, z)$ are continuously differentiable near $z = f_*(e^{i\theta})$, so Theorem 3 may be applied.

If one or more of the critical points $\alpha_r(e^{i\theta})$ migrate from the interior of the uncertainty set to the boundary as θ varies, one may prove that α_r is continuous, but differentiability is likely to no longer be a property of α_r. The conclusion of Theorem 3 may still hold in this case, and the algorithms for finding approximate solutions to UNCOPT may still be viable. An example is given in §8.2 which follows.

8 Solving UNCOPT numerically

8.1 Primal Dual relations for UNCOPT

We now rewrite the optimality conditions in a way that leads to numerical algorithms.

$$\textbf{PD-UNCOPT}\qquad \begin{array}{lrcll}
1. & \left(\gamma-\tilde{\Gamma}(\cdot,\alpha_r,f)\right)\Psi_r & = & 0 & 1\le r\le m \\[2mm]
2. & P_{zH^{2\perp}}\left[\sum_{r=1}^{m}\frac{\partial\tilde{\Gamma}}{\partial z_\ell}(\cdot,\alpha_r,f)\Psi_r\right] & = & 0, & 1\le \ell\le N \\[2mm]
3. & \sum_{r=1}^{m}\int\Psi_r\frac{d\theta}{2\pi} & = & 1 & \\[2mm]
4. & \Psi_r & \ge & 0 & 1\le r\le m \\[2mm]
5. & \gamma-\Gamma(\cdot,\alpha_r,f) & \ge & 0 & 1\le r\le m \\[2mm]
6. & \frac{\partial\tilde{\Gamma}}{\partial\alpha}(\cdot,\alpha_r,f)-\eta_r(\cdot)\frac{\partial g}{\partial\alpha}(\cdot,\alpha_r) & = & 0 & 1\le r\le m \\[2mm]
7. & \eta_r(\cdot)\,g(\cdot,\alpha_r) & = & 0 & 1\le r\le m \\[2mm]
8. & g(\cdot,\alpha_r) & \le & 0 & 1\le r\le m \\[2mm]
9. & \eta_r\Psi_r & \ge & 0 & 1\le r\le m
\end{array}$$

The unknowns in (PD-UNCOPT) are

$$\gamma,\ \Psi_r,\ f,\ \alpha_r,\ \eta_r \qquad (1\le r\le m).$$

Remark 8.1 As mentioned in §4, our algorithms require that the function Ψ be an element of L^2, so that projections can be applied to obtain equation 2 of PD-UNCOPT. Equation 6 of PD-UNCOPT is just the statement that α_r is a critical point for each θ. Equation 7 of PD-UNCOPT is a complementary condition: if α_r is an interior critical point, then necessarily $\eta = 0$. Inequality 9 of PD-UNCOPT specifies that there is a maximum at points $\alpha_r(e^{i\theta})$ in $\partial\mathcal{P}_\theta$ where $\tilde{\Gamma}(e^{i\theta},\alpha_r(e^{i\theta}),f(e^{i\theta}))=\gamma$.

Remark 8.2 Relations 1—8 of (PD-UNCOPT) characterize stationary points to UNCOPT. Therefore the same set of conditions applies to stationary points of the following optimization problem:

$$\inf_{f\in A}\ \sup_{\theta}\ \inf_{\alpha\in\mathcal{P}_\theta}\ \tilde{\Gamma}(e^{i\theta},\alpha,f(e^{i\theta})) \tag{93}$$

This problem appears in applications, for example, the celebrated μ synthesis problem [ZDK].

Computer algorithms for solving UNCOPT are based on the fact that solutions to PD-UNCOPT are good candidates to solutions to UNCOPT, just like with the problem OPT. Algorithms of type IPA or BCIPA discussed in §4 are directly applicable.

A sketch of the main feature of each algorithm is given below:

$$\begin{array}{c}\text{Algorithms}\\ \text{for UNCOPT}\end{array}\left\{\begin{array}{l}
\text{IPA}\left\{\begin{array}{ll}
\text{IPA.1} & \text{Set r.h.s. of equation 1 in PD-UNCOPT to } \epsilon I \\[2mm]
\text{IPA.2} & \text{Set r.h.s. of equation 1 in PD-UNCOPT to } \epsilon I \text{ and} \\
 & \text{eliminate dual variable } \Psi \text{ via substitution}
\end{array}\right. \\[8mm]
\text{BCIPA Apply Newton's method to PD-UNCOPT}
\end{array}\right.$$

To avoid duplication we do not discuss algorithms further but only just remark that the interior point algorithm IPA.2 yields important memory savings when doing H^∞ optimization with uncertainty parameters, as the dual variable Ψ_r is eliminated through substitution. The set of equations one solves in this case, in addition to 5. — 8. of PD-UNCOPT, consists of

$$\begin{array}{lrcl} 1'. & P_{H^{2\perp}}\left[\sum_{r=1}^m \frac{\partial}{\partial z_\ell}\log(\gamma^* - \tilde{\Gamma}(\cdot,\alpha_r,f_*))\right] & = & 0.\ \ 1\le \ell \le N \\ 2'. & \sum_{r=1}^m \int \frac{1}{\gamma^*-\tilde{\Gamma}(\cdot,\alpha_r,f_*)}\frac{d\theta}{2\pi} - \frac{1}{\epsilon} & = & 0 \\ 3'. & \frac{\partial\tilde{\Gamma}}{\partial\alpha}(\cdot,\alpha_r,f_*) - \eta_r\,\frac{\partial g}{\partial\alpha}(e^{i\theta},\alpha_r) & = & 0,\ \ 1\le r\le m \end{array} \tag{94}$$

Condition $3'$ of (94) arises from the constraint $\int \sum_{r=1}^m tr\Psi_r = 1$.

8.2 Numerical examples

In this section we present results of numerical tests. Tables 6 – 8 below show output of computer runs of an implementation of the BCIPA algorithm for solving UNCOPT.

We need the following auxiliary complex valued function:

$$\beta(\alpha_1,\alpha_2) = \frac{\alpha_1 + i\alpha_2}{1+\alpha_1^2+\alpha_2^2}, \qquad (\alpha_1,\alpha_2)\in R^2 \tag{95}$$

Note that the range of $\beta(\cdot)$ is the closed disk centered at 0 and radius 0.5.

The uncertainty performance functions used in the examples of this section that are listed below illustrate different situations arising in connection with the uncertainty parameter α.

$$\tilde{\Gamma}_1(e^{i\theta},\alpha,z) = \left|\frac{1}{e^{i\theta}-\beta(\alpha_1,\alpha_2)} + z\right|^2 \qquad \forall z\in C,\ \forall\alpha\in R^2$$

$$\tilde{\Gamma}_2(e^{i\theta},\alpha,z) = \left|\frac{1}{e^{i\theta}-\beta(\alpha_1,\alpha_2)} + z\right|^2 + \left|\frac{1}{e^{i\theta}+\beta(\alpha_1,\alpha_2)} + z\right|^2 \qquad \forall z\in C,\ \forall\alpha\in R^2$$

$$\tilde{\Gamma}_3(e^{i\theta},\alpha,z) = \left|\frac{1}{e^{i\theta}-\beta(\alpha_1,\alpha_2)} + z\right|^2 + \left|\frac{1}{e^{i\theta}+\beta(\alpha_1,\alpha_2)} + z\right|^2 + 0.02\left|\frac{1}{e^{i\theta}}+\beta(\alpha_1,\alpha_2)\right| \qquad \forall z\in C,\ \forall\alpha\in R^2$$

The problem UNCOPT for these performance functions is unconstrained in the uncertainty variable α, thus stationary points are obtained by solving PD-UNCOPT with $\eta_r = 0$. Tables 6, 7 and 8 below present output from computer runs that illustrate the performance of one of the algorithms.

Table 6

Iter. j	Perf. Value $\sup_\theta \max_k \|\Gamma(\cdot,\alpha_{k,j},f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness	P.D. Step $\|x_j - x_{j-1}\|_{L^2}$	P. Error $\|f_j - f_*\|_\infty$
0	4.150898227485347	$1.1\ 10^{-1}$	$8.3\ 10^{-2}$	N/A	$4.0\ 10^{-2}$
1	4.011006981438838	$2.7\ 10^{-2}$	$5.5\ 10^{-2}$	$1.2\ 10^{0}$	$2.3\ 10^{-3}$
2	4.000069588581157	$4.4\ 10^{-4}$	$5.9\ 10^{-5}$	$1.7\ 10^{-1}$	$2.3\ 10^{-5}$
3	4.000000003877265	$1.0\ 10^{-4}$	$5.1\ 10^{-9}$	$5.9\ 10^{-2}$	$8.0\ 10^{-9}$

A run of BCIPA with the performance function $\tilde{\Gamma}_1$. At the theoretical solution $f_* = 0$ there are exactly two critical points in the uncertainty parameter: a unique maximizer given by $\alpha_1 + i\,\alpha_2 = e^{i\theta}$, and a unique minimizer given by $\alpha_1 + i\,\alpha_2 = -e^{i\theta}$. The gridsize is 128. The number of Fourier coefficients used in building the approximation to the operator $\mathcal{T}$ is 32. The run was initialized with $f_0 = .04$, $\Psi_0 = .98$ and .04 on the diagonals, and $\alpha_1 = (.93e^{i\theta} + .02(e^{i\theta})^2, .97e^{i\theta})$ and $\alpha_2 = (-.95e^{i\theta}, -.96e^{i\theta} + .02(e^{i\theta})^2)$.

Table 7

Iter. j	Perf. Value $\sup_\theta \max_k \|\tilde{\Gamma}(\cdot, \alpha_{k,j}, f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness $\|x_j - x_{j-1}\|_{L^2}$	Step $\|f_j - f_*\|_\infty$
0	4.976688616559667	N/A	$2.1\ 10^{-1}$	N/A
1	4.629519306750855	$5.8\ 10^{-1}$	$8.3\ 10^{-2}$	$7.9\ 10^{0}$
2	4.523949273177587	$4.1\ 10^{-1}$	$6.1\ 10^{-2}$	$3.1\ 10^{0}$
3	4.460544063227239	$4.0\ 10^{-2}$	$1.9\ 10^{-2}$	$9.7\ 10^{-1}$
4	4.444787426990420	$6.4\ 10^{-4}$	$4.5\ 10^{-4}$	$1.3\ 10^{-1}$
5	4.444444577581560	$1.7\ 10^{-7}$	$1.4\ 10^{-7}$	$2.2\ 10^{-3}$

An run of BCIPA with the performance function $\tilde{\Gamma}_2$. There is no theoretical solution that we have been able to find; however, the solution does not appear to be trivial. The gridsize is 128. The number of Fourier coefficients used in building the approximation to the operator $\mathcal{T}$ is 32. The run was initialized with $f_0 = .1$, $\Psi_0 = .7I_{2\times 2}$, and $\alpha_1 = (.93e^{i\theta}, .97e^{i\theta})$ and $\alpha_2 = (-.8e^{i\theta}, -.6e^{i\theta})$. We note that this solution is somewhat far from the theoretical solution (observe the step size above). At solution, both α_1 and α_2 are the maximizers.

Table 8

Iter. j	Perf. Value $\sup_\theta \max_k \|\tilde{\Gamma}(\cdot, \alpha_{k,j}, f_j)\|_{n\times n}$	Oper. Value $\|\mathcal{T}[x_j]\|_{L^2}$	Flatness $\|x_j - x_{j-1}\|_{L^2}$	Step $\|f_j - f_*\|_\infty$
0	6.269444444444444	N/A	$4.0\ 10^{-1}$	N/A
1	4.517727459024402	$4.8\ 10^{-2}$	$1.5\ 10^{-2}$	$2.7\ 10^{0}$
2	4.489406850177974	$9.3\ 10^{-2}$	$3.5\ 10^{-3}$	$3.0\ 10^{0}$
3	4.484260091239294	$1.0\ 10^{-1}$	$1.0\ 10^{-3}$	$3.9\ 10^{-0}$
4	4.483291826928294	$5.0\ 10^{-2}$	$6.4\ 10^{-4}$	$2.8\ 10^{-0}$

An run of BCIPA with the performance function $\tilde{\Gamma}_3$. There is no theoretical solution that we have been able to find; however, the solution does not appear to be trivial. The gridsize is 128. The number of Fourier coefficients used in building the approximation to the operator $\mathcal{T}$ is 32. The run was initialized with $f_0 = .2 + .1e^{i\theta}$, $\Psi_0 = .7I_{2\times 2}$, and $\alpha_1 = (e^{i\theta}, e^{i\theta})$ and $\alpha_2 = (-e^{i\theta}, -e^{i\theta})$. This solution is somewhat far from the theoretical solution (observe the step size above). Here, neither α_1 nor α_2 is the maximizer or minimizer. For certain θ, α_1 is the maximizer while for other θ, α_2 is the maximizer. We illustrate this with the figures below.

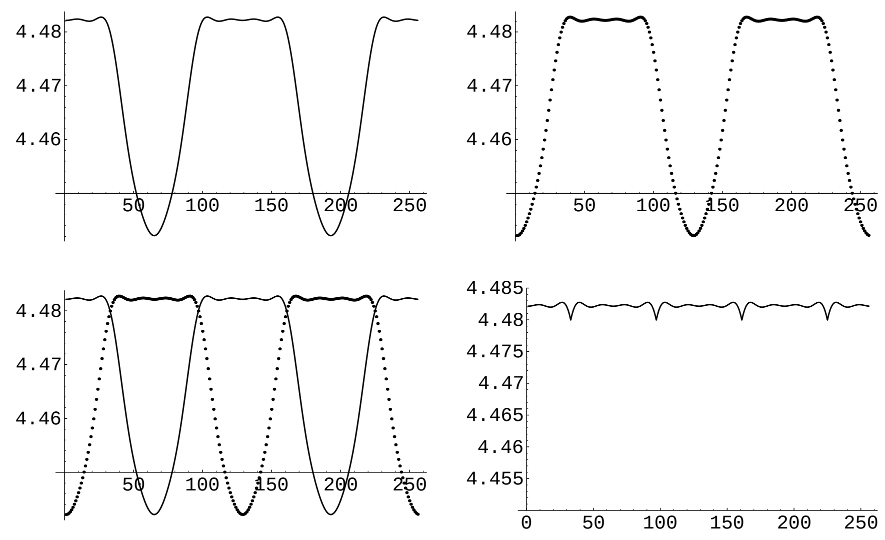

Fig. 2. The x-axis corresponds to θ from 0 to 2π and has 256 increments. We have plots of (*top left*) $\tilde{\Gamma}(\cdot, \alpha_1(\cdot), f^*(\cdot))$ (*top right*) $\tilde{\Gamma}(\cdot, \alpha_2(\cdot), f^*(\cdot))$ (*bottom left*) the overlay of (1) and (2) (*bottom right*) the maximum, $\Gamma(\cdot, , f^*(\cdot))$, of the two graphs (1) and (2). The graph in the lower right indicates that the flatness diagnostic is approximately $10^{-}4$.

Remark 8.3 The problems UNCOPT for the performance functions $\tilde{\Gamma}_1$, $\tilde{\Gamma}_2$ and $\tilde{\Gamma}_3$ correspond to a constrained UNCOPT problem over $\beta \in B(0, \frac{1}{2})$, which may be solved either by applying Newton's method to PD-UNCOPT directly, or by first introducing a barrier function depending on the uncertainty parameter and then using PD-UNCOPT for interior critical points (that is, with $\eta_k = 0$). For example,

$$-\varepsilon \log(0.25 - \alpha_1^2 - \alpha_2^2)$$

may be incorporated in the performance to insure that $\alpha_1^2 + \alpha_2^2 < 0.25$. If the uncertainty sets are closed convex bounded neighborhoods of the origin with corresponding Minkowski functionals ω_θ, one can use as barrier function the function

$$-\epsilon \; \log(1 - \omega_\theta(\alpha))$$

or a variation of it.

Another possibility when dealing with convex uncertainty sets is to use the maps defined by

$$\phi_\theta(\alpha) = \frac{2}{1 + \omega_\theta(\alpha)^2} \, \alpha \tag{96}$$

Clearly ϕ_θ maps $\partial\Omega_\theta$ onto $\partial\Omega_\theta$, and ϕ_θ maps Ω_θ onto Ω_θ. Also, ϕ_θ is smooth if $\partial\Omega_\theta$ is smooth. Thus the function ϕ_θ can be used to change coordinates, so that the transformed UNCOPT problem is unconstrained.

Conclusions of the numerical tests. A few tests were performed to illustrate the use of algorithm BCIPA for solving UNCOPT. The examples include a case with unique maximizer in the uncertainty variable α, a case with two pointwise global maximizers, and a case with two maximizers which are global over strict subsets of the unit circle. The output in the first two cases suggests quadratic convergence rate (local), while the convergence rate in the third example seems to be linear.

The tests aim to illustrate rather than compare different methods. More tests are needed to obtain conclusions with regards to performance of algorithms.

Acknowledgement. The authors wish to thank Michael Overton for discussions which presented his views and approach to numerical LMI's. Also discussions with Steve Boyd were informative.

9 References

[BV] S. Boyd and L. Vandenberghe, Semidefinite Programming, to appear SIAM Reviews.
[DG1] H. Dym and I. Gohberg. A maximum entropy principle for contractive interpolants, J. Funct. Anal, (65) pp. 83-125, 1986
[DG2] H. Dym and I. Gohberg. A new class of contractive interpolants and maximum entropy principles. In I. Gohberg, ed. Topics in Operator Theory and Interpolation. OT vol 29, pp. 117-150, Birkhauser Verlag, 1988.
[HH] J. W. Helton and R. Howe. A bang-bang principle for the frequency domain. J. Approx. Th. **47** (1986), 101-121.
[HMar] J. W. Helton and D. Marshall. Optimization in the frequency domain and analytic selections. U. of Indiana Math. J. 1991.
[HMer1] J. W. Helton and O. Merino. A Novel Approach for Accelerating Newton's method in SupNorm Optimization Problems in H^∞ Control. J. Optimization Theory and Applications. 1993. **[HMer2]** J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., 1993.
[HMer3] J. W. Helton and O. Merino. In preparation.
[HMW] J. W. Helton, O. Merino and T. E. Walker. Algorithms for Optimizing over Analytic Functions. U. of Indiana Math. J. (42) No. 3, pp 839–874.
[HMW2] J. W. Helton, O. Merino and T. E. Walker. Proceedings of the IEEE's 33th Conference on Decision and Control, Orlando, Florida., Dec. 1994.
[HMW3] J. W. Helton, O. Merino and T. E. Walker. H^∞ Optimization and Semidefinite Programming. Proceedings of the IEEE's 34th Conference on Decision and Control, New Orleans, LA, Dec. 13-15
[HO] J. P. Haeverly and M. Overton. Optimizing eigenvalues of symmetric definite pencils. Proceedings of American Control Conference, 1994.
[HV1] J. W. Helton and A. Vityaev. On optimization with competing performance criteria. Proceedings of the IEEE's 34th Conference on Decision and Control, New Orleans, LA, Dec. 13-15
[HV2] J. W. Helton and A. Vityaev. Analytic functions Optimizing Competing Constraints. To appear, SIAM J. of Mathematical Analysis.
[LO] A. Lewis and M. Overton. Optimizing Eigenvalues. To appear in Acta Numerica.
[ZDG] K. Zhow, with J. C. Doyle and K. Glover. Robust and Optimal Control. Prentice Hall. 1995.

Part III

APPENDIX

10 Appendix: Proofs

Proposition 2 *Let η^* be a size function with respect to I. Then,*

1. *η^* is convex and positive-homogeneous.*
2. *η^* is subadditive: for every x, y in X, $\eta^*(x+y) \leq \eta^*(x) + \eta^*(y)$.*
3. *η^* is Lipschitz: for every x, y in X, $\mid \eta^*(x) - \eta^*(y) \mid \leq \| x - y \|_X$*
4. *$\eta^*(0) = 0$, and $\eta^*(I) = 1$.*
5. *If both x and $-x$ are elements of X_+ then $x = 0$.*
6. *If x is in X_+, then $\eta^*(-x) \leq 0$ and $\eta^*(x) = \|x\|_X$.*
7. *$\eta^*(tI + x) = t + \eta^*(x)$, for every $t \in R$ and $x \in X$.*

PROOF. The proof of Proposition 1 is straightforward. We only present the proof of the convexity of η^*. For $t \in [0,1]$ we have,

$$\begin{aligned} \eta^*(tx + (1-t)y) &= \inf\{\eta : \eta I - (tx + (1-t)y) \geq 0\} \\ &= \inf\{\eta : \eta I - (t\eta^*(x)I \ + \ (1-t)\eta^*(y)I) + t(\eta^*(x)I - x) + (1-t)(\eta^*(y)I - y) \geq 0\} \\ &\leq \inf\{\eta : \eta I - (t\eta^*(x)I \ + \ (1-t)\eta^*(y)I) \geq 0\} \\ &= t\eta^*(x) + (1-t)\eta^*(y) \end{aligned} \tag{97}$$

□

Proposition 3 *Suppose η^* satisfies (HBC). If $\lambda \in X'_+$ is such that there exists $x \in X_+$ satisfying*

$$\eta^*(x) = \frac{\langle \lambda, x \rangle}{\|\lambda\|_{X'}}$$

then $\langle \lambda, I \rangle = \|\lambda\|_{X'}$.

PROOF. From $\langle \lambda, \eta^*(x)I - x \rangle \geq 0$ we have that

$$\langle \lambda, I \rangle \eta^*(x) \geq \langle \lambda, x \rangle = \eta^*(x)$$

Hence $\langle \lambda, I \rangle \geq 1$. We also have that

$$1 = \eta^*(I) = \sup_{\lambda \in X'_+} \frac{\langle \lambda, I \rangle}{\|\lambda\|} \geq \langle \lambda, I \rangle$$

□

Theorem 4 *Let p be a sublinear functional on a Banach space X, that is, p is real valued and*

$$\begin{aligned} p(x+p(y) &\leq p(x+y) \quad \forall x,y \in X \qquad (i) \\ p(tx) &= tp(x), \qquad \forall t \geq 0, \ \forall x \in X \ \ (ii) \end{aligned} \tag{98}$$

Let X_+ be a closed cone in X such that

$$p(x) \geq 0 \quad \textit{and} \quad p(-x) \leq 0, \qquad \forall x \in X_+ \tag{99}$$

Then for every $x_0 \in X_+$, $x_0 \neq 0$, there exists $\lambda_0 \in X'_+$ such that

$$\begin{aligned} \langle \lambda_0, x \rangle &\leq p(x), \ \forall x \in X \\ \langle \lambda_0, x_0 \rangle &= p(x_0) \end{aligned} \tag{100}$$

Proof. The existence of $\lambda_0 \in X'$ such that equations (100) hold is a consequence of the Hahn-Banach theorem. The only thing left to verify is that λ_0 is a nonnegative functional. From

$$-\langle \lambda_0, x \rangle = \langle \lambda_0, -x \rangle \leq p(-x), \ \forall x \in X$$

we conclude that

$$-p(-x) \leq \langle \lambda_0, x \rangle \leq p(x), \ \forall x \in X \tag{101}$$

If in particular $x \in X_+$, then it follows from (101) and (99) that

$$\langle \lambda_0, x \rangle \geq 0$$

□

Corollary 3 *If η^* is a size function that satisfies* (HBC)*, then for every $x \in X$ there exists $\lambda \in X'_+$ such that*

$$\eta^*(x) \, \|\lambda\|_{X'} = \langle \lambda, x \rangle$$

That is, the supremum in (HBC) *is attained.*

Proof. First assume that x_0 is a nonzero element in X_+. Take $p = \eta^*$ in Theorem 4 to obtain $\lambda_0 \in X'_+$ such that (100) holds. By (101),

$$|\langle \lambda_0, x \rangle| \leq \max\{|\eta^*(x)|, |\eta^*(-x)|\} \leq \|x\|_X$$

hence $\|\lambda_0\| \leq 1$. We also have that

$$\langle \lambda_0, x_0 \rangle = \|x_0\|_X$$

by 6. of Proposition 2, so $\|\lambda_0\|_{X'} = 1$. This proves the corollary in the case when x_0 is a nonnegative element in X.

Now suppose that x_0 is any element of $X \setminus X_+$. Since

$$\eta^*(-x_0) = \inf\{\eta \, ; \, \eta I + x_0 \in X_+\}$$

we have that $x_1 = \eta^*(-x_0)I + x_0$ is in X_+. By the first part of the proof there exists $\lambda_0 \in X'_+$ with unit norm such that $\langle \lambda_0, x \rangle \leq \eta^*(x)$ for all $x \in X$, and such that

$$\eta^*(x_0 + \eta^*(-x_0)I) = \langle \lambda_0, x_0 + \eta^*(-x_0)I \rangle \tag{102}$$

By (102) and 7. of Proposition 2 we have that

$$\eta^*(x_0) + \eta^*(-x_0) = \langle \lambda_0, x_0 \rangle + \eta^*(-x_0)\langle \lambda_0, I \rangle \tag{103}$$

Therefore,

$$\eta^*(x_0) - \langle \lambda_0, x_0 \rangle = \eta^*(-x_0)\left(\langle \lambda_0, I \rangle - 1\right) \tag{104}$$

Since the left hand side of (104) is nonnegative, we have that

$$\eta^*(-x_0)\left(\langle \lambda_0, I \rangle - 1\right) \geq 0 \tag{105}$$

Proposition 2 implies that Since $\eta^*(I) = 1$ and λ_0 has unit norm, we must have that $\langle \lambda_0, I \rangle \leq 1$; note that $\langle \lambda_0, I \rangle < 1$ is not possible, since otherwise $\eta^*(-x_0) \leq 0$ by (105), thus yielding $x_0 \in X_+$, a contradiction. Therefore we have that

$$\langle \lambda_0, I \rangle = 1 \tag{106}$$

Finally substitute (106) into (103) to obtain

$$\eta^*(x_0) = \langle \lambda_0, x_0 \rangle$$

□

Theorem 5 *Let η^* be a size function such that (HBC) holds. Let x_0 be an nonzero element of X and M be a closed subspace of X such that $x_0 \in X_+ \setminus M$ and*

$$\eta^*(x_0) = \inf_{y \in M} \eta^*(x_0 + y) \tag{107}$$

Then, there exists $\lambda \in X_+$ satisfying

$$\begin{aligned} \|\lambda\| &= 1 \\ \langle \lambda, x_0 \rangle &= \rho^*(x_0) \\ \langle \lambda, y \rangle &= 0, \ \forall y, \in M \end{aligned} \tag{108}$$

Proof. Consider the function $\hat{\eta} : X/M \to R$, where

$$\hat{\eta}(\hat{x}) = \inf_{y \in M} \eta^*(x + y)$$

The function $\hat{\eta}$ is convex. It is also finite valued, since by convexity, for $\hat{\eta}$ to be bounded below it is sufficient to notice that $\hat{\eta}(x_0)$ is finite. Note that $X_+ + M$ is a nontrivial closed cone in X/M. By theorem 4, there exists bounded linear functional $\hat{\lambda}$ on X/M such that relations (100) hold with $p = \hat{\eta}$. Also, note that $\hat{\lambda}$ is nonnegative on the (nontrivial) cone $X_+ + M$. The linear functional $\hat{\lambda}$ induces in a natural way a nonnegative functional λ on X via

$$\langle \lambda, x \rangle = \langle \hat{\lambda}, \hat{x} \rangle, \ \forall x \in X$$

Note that $\langle \lambda, y \rangle = 0, \ \forall y \in M$. Clearly relations (100) hold with $p = \eta$ and $\lambda_0 = \lambda$, since by hypothesis $\eta^*(x_0) = \hat{\eta}^*(\hat{x}_0)$. Therefore,

$$|\langle \lambda, x \rangle| \leq \max\{|\eta^*(x)|, |\eta^*(-x)|\} \leq \|x\|_X$$

so $\|\lambda\|_{x'} \leq 1$ holds. But since $x_0 \in X_+$,

$$\|x_0\|_X = \eta^*(x_0) = \langle \lambda, x_0 \rangle$$

so we conclude that $\|\lambda\|_{x'} = 1$. □

TODO

TO DO LIST !!

TEST NUMERICAL UNCOPT: BCIPA VS IPA,

TRENTS TRICK VS PURE BOUNDARY VS LOG BARRIER FUNCTION

spelling:

c.f contionuous occuring ricatti

$Hessian = I$ —¿ True Hesian

which Hessian ??

Gamma Hessian -get a nehari type problem (but not a Nehari type soln). Has nothingto do with winding num probs. Helps numerics sometimes. use is rolls off saddle point.

Proof: As the hessian is the identity, then the equations solved in a Newton step are equal to the optimality equations solved for some Nehari problem which locally approximates the original gamma at our current position. Hence we have reduced the problem to showing that Newton always "goes down" on a Nehari type problem.

Show this and done.

Question: Does Newton always solve the vector Nehari Case????? If so then done. ENDPROOF

ALSO: Want to compare two methods for finding uncertainty parameters.

(1) Simultaneous update of alpha and f (via Newton) versus

(2) Co-ordinate descent type methods.

(3) Anything else???

Existing methods are μ-synthesis which is a co-ordinate DEscent method, Packard-Gahnet Computation of Simultaneous solution to many ricatti inequalities depending on alpha. ALSO see Megretski's or Rancer paper.

Proj onto H^1 is fishy. Do we want to change it.