H^∞ Optimization 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

This article shows how the fundamental H^∞ optimization problem of control can be naturally treated with modern primal-dual interior point (PDIP) methods. The fundamental H^∞ problem of control is that of finding the stable frequency response function which best fits worst case frequency domain specifications. This is a non smooth optimization problem which underlies the frequency domain formulation of the H^∞ problem of control; it is the main optimization problem in QFT for example. Also, in this article we present new optimality conditions for matrix valued H^∞ problems, and compare natural (PDIP) algorithms for these problems, as well as fit them into the context of classical H^∞ theory.

The method which we expect will 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 objectives functions. For example, a simple H^∞ optimization problem is, given g, find

$$\inf_{f \in H^\infty} \sup_\theta Re(g(e^{i\theta}) - f(e^{i\theta}))^2 + 4Im(g(e^{i\theta}) - f(e^{i\theta}))^2 \tag{1}$$

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 years, the authors 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).

1. Recollections about Semidefinite Programming

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

1.1 Basic Set Up

Let X be a real Banach space, and let X_+ be a closed cone in X called the *cone of nonnegative elements.* If x and y are elements of X, we say that $x \geq y$ if $x - y$ is an element of X_+. Let X' be the dual of X. The closed *dual cone* of X_+ is

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

Throughout this paper, we assume that there exists a distinguished element $I \in X_+$ with the property that $\eta^*(x) = \inf\{\eta : \eta I - x \geq 0\}$ is finite for every $x \in X$, and that $\eta^*(x) = \|x\|$ for every $x \in X_+$. The existence of I and η^* must be verified in specific cases.

1.2 The Optimization Problem

Let F be a normed real vector space with dual F', $\mathcal{F} : F \to X$ be a C' (continuously differentiable) map, and let k be a linear functional on F. A basic problem that we consider in this paper is:[1]

$$\textbf{(PO)} \qquad \begin{array}{ll} \multicolumn{2}{l}{\text{Find } \gamma = \gamma_* \in R,\ f = f_* \in F \text{ solving}} \\ \textbf{minimize} & \gamma + k(f) \\ \textbf{subject to} & \gamma I - \mathcal{F}(f) \geq 0 \end{array}$$

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 (γ_*, f_*) is a solution to the problem obtained from (PO) by adding the constraint $(\gamma, f) \in V$. 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. This is based on the differential $\mathcal{DF}_f[\cdot]$ of $\mathcal{F}$ at a given $f \in F$, and on its adjoint[2] $(\mathcal{DF}_f)^\dagger : X' \to F'$.

Define the *primal-dual optimization problem* to be

*Supported in part by the AFOSR and the NSF

[1] The basic LMI problem [BV] is similar, but without the γ terms.

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

$$
\textbf{(PDO)} \quad \begin{array}{ll} \multicolumn{2}{l}{\text{Find } \gamma=\gamma_* \in R,\ f=f_* \in F} \\ \multicolumn{2}{l}{\text{and } \lambda_* \in X'_+ \text{ solving}} \\ \textbf{minimize} & \gamma - \langle \lambda, \mathcal{F}(f)\rangle \\ \textbf{subject to} & \lambda \in X'_+ \\ & \|\lambda\|_{X'} \ = \ 1 \\ & \gamma I - \mathcal{F}(f) \geq 0 \\ & \mathcal{DF}^\dagger_{f_*}[\lambda] + k^\dagger[\lambda] = 0 \end{array}
$$

The dual variable is λ. The quantity

$$
\Delta(\gamma, f, \lambda) \ = \ \gamma - \langle \lambda, \mathcal{F}(f)\rangle \qquad (3)
$$

is called the *duality gap*. We shall see in Theorem 1 that the duality gap is zero at solutions to (PDO). This leads to the following system of equations in λ, f and γ which must hold for solutions to (PDO):

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

Adding the inequality constraints $\lambda \geq 0$ and $\gamma I - \mathcal{F}(f) \geq 0$ to (PDEQ) gives a set of relations we refer to as (**PDEQ**$^+$).

1.3 Main High Level Theorem

Theorem 1 shows that the problems are closely related.

Theorem 1 *[HMW3] Under technical hypotheses, if (γ_*, f_*) is a local solution to (PO) then there exists $\lambda_* \in X'_+$ such that $(\gamma_*, f_*, \lambda_*)$ solves (PDEQ$^+$). Moreover, $\Delta(\gamma_*, f_*, \lambda_*) = 0$.*

Section 3 contains examples where the theorem applies. Theorem 1 has the implication that solving (PDEQ$^+$) leads to excellent candidates for solutions to (PO).

2. Numerical Algorithms

Here are two algorithms for solving (PDEQ$^+$). We give a high level conceptual presentation. In [HMW3] we give more details.

2.1 Interior Point Algorithm (IPA)

Replace 0 in condition (i) of (PDEQ) by ε to obtain

$$
(\textbf{PDEQ}^+_\varepsilon) \quad \begin{array}{l} \langle \lambda, \gamma I - \mathcal{F}(f)\rangle = \varepsilon \\ \|\lambda\|_{X'} \ = \ 1 \\ \mathcal{DF}^\dagger_{f_*}[\lambda] + k^\dagger[\lambda] = 0 \\ \gamma I - \mathcal{F}(f) \geq 0, \quad \lambda \geq 0. \end{array}
$$

We can write all of the above except for $\gamma I - \mathcal{F}(f) \geq 0$ and $\lambda \geq 0$ in the following form, for some function $\mathcal{T}_\epsilon$,

$$
(\textbf{OCI}_\varepsilon) \qquad \mathcal{T}_\varepsilon(\lambda, f, \gamma) = 0.
$$

To solve these equations, fix $\varepsilon > 0$, do Newton steps to solve (OCI$_\varepsilon$), then make ε smaller and repeat. One generally needs to keep an eye on the terms $\gamma I - \mathcal{F}(f)$ and λ in order to insure their positivity, but in general this does not turn out to be difficult.

2.2 Boundary Algorithms

Many engineering problems have special structure which allow one to apply Newtons method directly to (PDEQ). Sometimes setting $\varepsilon = 0$ in §2.1 is effective, we call this the *Boundary Case of the Interior Point Algorithm* (BCIPA).

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

$$
(1') \qquad \gamma I = \mathcal{F}(f).
$$

This is a strengthend version of $\langle \lambda, \gamma I - \mathcal{F}(f)\rangle = 0$.

For our algorithm we must introduce the projection $\mathcal{P}_{H^{2\perp}}$. If f is a L^2 function and $f(e^{i\theta}) = \sum_{\ell=-\infty}^{\infty} a_\ell e^{i\ell\theta}$, then $\mathcal{P}_{H^{2\perp}} f(e^{i\theta}) = \sum_{\ell=-\infty}^{-1} a_\ell e^{i\ell\theta}$. Since γ is a constant $(1')$ can be rewritten as

$$
(1'') \quad \mathcal{P}_{H^{2\perp}}[\mathcal{F}(f)] = 0.
$$

This leads to an equivalent way of writing PDEQ$^+$:

$$
\begin{array}{l} \mathcal{P}_{H^{2\perp}} \mathcal{F}(f) = 0 \\ \|\lambda\|_{X'} \ = \ 1, \quad \lambda \geq 0 \\ \mathcal{DF}^\dagger_{f_*}[\lambda] + k^\dagger[\lambda] = 0 \end{array}
$$

Note that *we have eliminated the unknown γ and the constraint $\gamma I - \mathcal{F}(f)$ from PDEQ$^+$.* Conceptually it is best to think of these equations in condensed notation $\mathcal{T}(\lambda, f) = 0$. One can apply Newton's method to solve this system of equations. We call this the *Boundary Algorithm* (BA).

3. Examples

The following examples illustrate the theory and also introduce notation that will be used later.

3.1 Spaces

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 $\|(w,v)\|_{X'} = |w| + |u|$.

2. Let $X = M =$ the self adjoint $n \times n$ matrices, with norm $\sigma_{max}(A)$ the maximum singular value of the matrix A. Consider the cone $M_+ =$ the nonnegative definite matrices, and let I be the identity matrix. The dual is M via $\langle x, y\rangle = tr(y^T x)$ and $X'_+ = X_+$. Here y^T is the conjugate transpose of y, and tr denotes the trace function.

3. Let $X = \mathcal{C}_{n\times n}$ the space of continuous functions on the circle with values in M, the space of self adjoint matrices, (see previous example). The space $\mathcal{C}_{n\times n}$ is equipped with the norm

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

Take X_+ to be the set of elements g of $\mathcal{C}_{n\times n}$ which take nonnegative definite values, and let $I =$ the constant function equal to the $n \times n$ identity matrix at each θ. The space X' contains the matrices with entries in L^1 of the circle, and $\langle x, y\rangle = \int tr(y^T x)\frac{d\theta}{2\pi}$.

3.2 Optimization Problems

3.2.1 LMI's and related stories.

Given $f = (f_1, \ldots, f_m)$, X the space of $n \times n$ real symmetric matrices, F_i in X, and

$$\mathcal{F}(f) = F_0 + F_1 f_1 + \ldots + F_m f_m.$$

Then, (PO) is similar to the situation discussed in equations (1) and (2) of the Introduction of [BV]. There the authors give a beautiful and thorough discussion of affine $\mathcal{F}$ acting on matrix spaces, which includes LMI. The Interior Point Algorithm (§2.1) is in the spirit of algorithms discussed in [BV]. IPA is, in the matrix context, a special case of the algorithm in Haeberly and Overton [HO]. Barrier functions are also used heavily in [BV].

3.2.2 Smooth optimization

Observe that old fashion unconstrained optimization of a smooth function $\mathcal{F}$ on R^n fits into the §1 setting. Here we 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

$$\textbf{minimize } \mathcal{F}(f) \tag{4}$$

Now we give the optimality conditions $(PDEQ^+)$ for (4). First note that λ is a positive real number, which the normalization forces to be 1. Next note that $0 = \mathcal{D}\mathcal{F}^{\dagger}_{f^*}[1] = \nabla\mathcal{F}(f_*)$. These actually constitute $(PDEQ^+)$. So for the smooth case it gives exactly the classical gradient equal zero condition.

3.2.3 H^∞ optimization

We shall see that this combines both nonsmooth and smooth structure. Let $\mathcal{A}$ be the C^N valued functions on the unit circle ∂D which are continuous and extend analytically to the unit disk D, and let $\Gamma(e^{i\theta}, z)$ be a C' function of $e^{i\theta} \in \partial D$ and $z \in C^N$ with values in M.

The *fundamental H^∞ problem of control* is[3]

$$\textbf{(OPT)} \quad \begin{aligned} \gamma_* &= \|\Gamma(\cdot, f_*)\|_{\mathcal{C}_{n\times n}} \\ &= \inf_{f\in\mathcal{A}} \sup_\theta \|\Gamma(e^{i\theta}, f(e^{i\theta}))\|_{n\times n} \end{aligned}$$

Observe that (OPT) is of type (PO) and so Theorem 1 applies to characterize solutions by equations (PDEQ+), provided Γ takes positive definite matrix values [4]. Also from now on we assume that $\gamma_* > 0$.

4. Characterizing Optima

Next we show that Theorem 1 implies a new basic characterization of optima for H^∞ problems. This establishes that Theorem 1 as well as algorithms §2.1 contain basic LMI (see [BV] and [HO]) as well as H^∞ numerical methods.

[3] For issues like existence of f_*, uniqueness etc. see [HM book].

[4] If Γ is not positive definite or even not square, then there are ways to treat the optimization problem [HMW3].

Theorem 2 *Suppose that $\Gamma(e^{i\theta}, z)$ takes $n \times n$ positive definite values. Let $f_* \in \mathcal{A}$ and set $\gamma_* = \|\Gamma(\cdot, f_*)\|_{\mathcal{C}_{n\times n}}$, $\Gamma_* = \Gamma(\cdot, f_*(\cdot))$, $\Gamma_{*z} = \frac{\partial\Gamma}{\partial z}(\cdot, f_*(\cdot))$.*

Suppose $(\gamma_*, f_*, \lambda_*)$ is a solution to (PDO) for which $(\Gamma_{*z})^T\ \Gamma_{*z}$ is pointwise invertible and that λ_* satisfies a technical condition. Then there is a matrix valued function Ψ_* (it's effectively the density of λ) such that

$$\textbf{PDEQH}^\infty \qquad \begin{array}{ll} 1. & (\gamma_* I - \Gamma_*)\Psi_* = 0 \\ 2. & \mathcal{P}_{zH^{2\perp}}[tr\ \{\Gamma^t_{*z}\Psi_*\}] = 0 \\ 3. & \gamma_* I - \Gamma_* \geq 0 \\ 4. & \Psi_* \geq 0 \\ 5. & \int tr\ \Psi_* \frac{d\theta}{2\pi} = 1 \end{array}$$

Idea of Proof: This is gotten by evaluating the PDEQ+ conditions in this special case.

The bounded linear functionals λ on $\mathcal{C}_{n\times n}$ have the form

$$\langle\, \lambda\, ,\, g\, \rangle = \int tr(\, g^t d\lambda)$$

The technical condition is that λ_* is absolutely continuous with respect to Lebesgue measure with L^2 density. Indeed this produces Ψ_* via $d\lambda_* = \Psi_* \frac{d\theta}{2\pi}$

Condition 2 follows from

$$\mathcal{D}\Gamma_z[w] = \frac{\partial\Gamma^t}{\partial z}(\cdot, z)w + \frac{\partial\Gamma^t}{\partial \overline{z}}(\cdot, z)\overline{w}$$

for each $w \in C^N$ which is a concrete formula for abstract $\mathcal{D}\Gamma$. Here B^t means the transpose of B. ∎

Another remark is that Theorem 2 gives exactly the optimality condition results that are known for scalar valued performance Γ, see [HM]:

Corollary 1 *Suppose the hypotheses of Theorem 2 hold, in addition to $\Gamma(e^{i\theta}, z)$ being scalar valued. Then the following conditions on f_* hold:*

Flatness: *for all θ, $\Gamma(e^{i\theta}, f_*(e^{i\theta})) = \gamma_*$.*

Gradient alignment: $\exists\Psi_* > 0$ *measurable such that*

$$\mathcal{P}_{zH^{2\perp}}[\Psi_*(e^{i\theta})\frac{\partial\Gamma}{\partial z}(e^{i\theta}, f_*(e^{i\theta}))] = 0$$

5. Plant Uncertainty and Optimization

Now we turn to frequency domain optimization when there is plant uncertainty. A simple example sets the stage.

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

$$\inf_{f\in A} \sup_\theta \sup_{[\alpha^x]^2+[\alpha^y]^2<0.25} \left| \frac{1}{e^{i\theta} - \alpha^x - \alpha^y i} - f(e^{i\theta}) \right|^2. \tag{5}$$

Here $e^{i\theta}$ is the frequency variable, $(\alpha^x, \alpha^y) \in R^2$ is the uncertain plant parameter, and f is the designable H^∞ parameter. The general problem is find

$$\inf_{f\in A} \sup_\theta \sup_{g(e^{i\theta},\alpha)\leq 0} \tilde{\Gamma}(e^{i\theta}, \alpha, f(e^{i\theta})). \tag{6}$$

Here $\tilde{\Gamma}$ and g are smooth and $f, \alpha = (\alpha^x, \alpha^y)$ are like those above.

There is an appropriate setting in which this is a problem of type (PO). Our goal now is simply to indicate what (PDEQ) becomes. One gets: *At local solutions to (6) there exist real functions $\mu_1 \geq 0$ and $\mu_2 \geq 0$, with $\mu_1 + \mu_2 > 0$ and $\int \mu_1 + \mu_2 = 1$, R^2-valued functions α_1, α_2, and R-valued functions $\eta_1 \geq 0, \eta_2 \geq 0$, such that*

$$\begin{array}{rcl} (\ell = 1,2) \qquad \mu_\ell \; (\; \gamma^* - \tilde{\Gamma}(\cdot, \alpha_\ell, f_*) \;) & = & 0 \\ \mathcal{P}_{zH^{2\perp}}[\; \mu_1 \; \frac{\partial \tilde{\Gamma}}{\partial z}(\cdot, \alpha_1, f_*) + \mu_2 \; \frac{\partial \tilde{\Gamma}}{\partial z}(\cdot, \alpha_2, f_*) \;] & = & 0 \\ (\ell = 1,2) \qquad \frac{\partial \tilde{\Gamma}}{\partial \alpha}(\cdot, \alpha_\ell, f_*) - \eta_\ell \frac{\partial g}{\partial \alpha}(e^{i\theta}, \alpha_\ell) & = & 0 \end{array}$$

See [HMWcdc94] for the actual theorem.

6. Numerical Algorithms for H^∞ optimization.

We can specialize the high level computer algorithms in §2 to H^∞ optimization problems by using the equations in PDEQH$^\infty$ [HMW3]. Similar algorithms may be used to solve the equations in §5.

6.1 Interior Point Algorithm

In this method we replace (1) in (PDEQH$^\infty$) by

$$(1_\varepsilon) \qquad (\gamma I - \Gamma)\Psi = \varepsilon I.$$

The other conditions remain the same. Fix ε and solve (1_ε) and (2), (5) of (PDEQH$^\infty$) for f, Ψ and γ using Newton's method, with barriers or restrictions on the line search imposing positivity conditions (3), (4). An example of the equations one solves is:

Multiple performances. Take Γ to be diagonal, $\Gamma = \mathrm{diag}(\Gamma_1, \Gamma_2)$, where Γ_ℓ is scalar valued, $\ell = 1, 2$. One can take $\Psi = \mathrm{diag}(\Psi_1, \Psi_2)$, with Ψ_ℓ nonnegative scalar valued. We rewrite equations in PDEQH$^\infty$ as

$$\begin{array}{rcl} \mathcal{P}_{zH^{2\perp}}\,[\,(\gamma - \Gamma_1)\Psi_1\,] = \mathcal{P}_{zH^{2\perp}}\,[\,(\gamma - \Gamma_2)\Psi_2\,] & = & \varepsilon \\ \mathcal{P}_{zH^{2\perp}}\,[\,\Psi_1 \frac{\partial \Gamma_1}{\partial z}(\cdot, f) + \Psi_2 \frac{\partial \Gamma_2}{\partial z}(\cdot, f)\,] & = & 0 \\ \int \Psi_1 + \Psi_2 & = & 1. \end{array}$$

6.2 Boundary Algorithms

There are two cases. BCIPA is clear; set $\varepsilon = 0$ in §6.1.

The Boundary Algorithm (BA) for scalar valued Γ solves PDEQH$^\infty$ with (1) replaced by

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

and (3) removed. Use Newton's method to solve for the unknowns f, Ψ with barriers or restrictions on the line search imposing the positivity condition (4).

7. Numerical Results

We tested on (PDEQH$^\infty$) the following examples with about 10 randomly generated initial guesses.

a. $\Gamma_1(e^{i\theta}, z_1, z_2) = Re[g_1(e^{i\theta}) + z_1]^2 + 4Im[g_1(e^{i\theta}) + z_1]^2 + Re[g_2(e^{i\theta}) + z_2]^2 + 9Im[g_2(e^{i\theta}) + z_2]^2$. Here g_1, g_2 are certain complex valued rational fuctions. Note the sublevel sets of Γ in z are ellipses.

b. $\Gamma_2(e^{i\theta}, z_1, z_2) = |z_1|^2 + |z_2|^2 + |100 + e^{i\theta} z_1 .1(z_1 + z_2 + z_1 z_2)|^2 + |100 + e^{i\theta} z_2 + .1(z_1 + z_2 + z_1 z_2)|^2$ (sublevel sets are not convex)

c. $\Gamma_3 = \max\{\Gamma_{3,1}, \Gamma_{3,2}\}$, with $\Gamma_{3,j}$ various smooth scalar functions dependent on one z. (sublevel sets have corners)

We have tested BA on smooth scalar valued Γ extensively and found it to be very effective. The Interior Point Algorithm (IPA) an our Boundary Case Interior Point Algorithm (BCIPA) tests are recent and are not comprehensive.

For problems where Γ has a corner (like c), we tried (BCIPA). We expected the (IPA) to be more accurate than the BCIPA, since an interior method does not ever actually hit a corner. However, in our (very limited) numerical testing of problems of type (c) we did not see a distinct difference in accuracy.

The result in all examples tested indicate that the accuracy of the two methods is almost identical. The IPA requires more iterations than BCIPA since it contains an inner loop. We tested this using the usual rule of thumb for decreasing ε in proportion to the optimality diagnostic for the inner loop and found the IPA takes about 3 times as many iterations as the BCIPA method.

Further tests (to include the problems found in section 5) are in progress and will be reported in the longer version of this paper [HMW3].

We conclude with a remark on classical state-space H^∞ control. While formulas, etc. are all done in state-space, the key ones are based on flatness of the optimal error function. Thus boundary rather than interior methods are more in the spirit of classical H^∞ control Riccati equations. If one solves these with "LMI numerics" that brings in an interior method.

References

[BV] S. Boyd and L. Vandenberghe, Semidefinite Programing, to appear SIAM Review.

[HMar] J. W. Helton and D. Marshall. U. of Indiana Math. J. 1991.

[HM] J. W. Helton and O. Merino. SIAM J. Cont. Opt., 1993.

[HMbook] ibid., Preprint

[HMW1] J. W. Helton, O. Merino and T. E. Walker. CDC 1992, Tucson, Arizona.

[HMW2] ibid. , U. of Indiana Math. J. (42) No. 3, pp 839–874.

[HMW3] ibid., In Preparation.

[HO] J. P. Haeberly and M. Overton. ACC94

The authors are grateful to M. Overton and Steve Boyd for very helpful conversations.