Coordinate optimization for bi-convex matrix inequalities*

J. William Helton
Department of Mathematics
University of California, San Diego
La Jolla, California, 92093
whelton@.ucsd.edu

Orlando Merino
Department of Mathematics,
University of Rhode Island
Kingston, RI 02881
merino@math.uri.edu

August 23, 1998

Abstract

We consider optimization of the largest eigenvalue of a smooth selfadjoint matrix valued function $\Gamma(X,Y)$ of two vector or matrix variables X and Y. We shall assume that Γ is concave or convex in Y and separately in X, but possibly has bad joint behavior. A typical problem one faces in control design are matrix versions of minimizing in Y and maximizing in X. Also minimizing in X and Y is an important problem. When joint behavior in X and Y is bad existing commercial software must be applied to each coordinate separately, and so can be used only to give a coordinate optimization algorithm . We give strong evidence in this article that on "well behaved Γ" coordinate optimization always gives a local optimum for the $\min_Y \max_X$ problem and that it almost never gives a local solution to the $\min_Y \min_X$ problem.

Also, this article treats second order optimality conditions for optimization of matrix functions. Second order tests are important because optimization of matrix valued Γ based on linearization or coordinate descent will produce critical points which are not solutions to the problem. Sufficient conditions, especially if they include second order information, become very valuable in these cases, as they can be used to tell which among the critical points correspond to true local optimal points, and to provide good update directions.

The matrix problems studied here are of course certain non-smooth optimization problems. Another class of non-smooth problems is functions space (frequency domain) optimization with an L^∞ worst case performance criterion. Such problems exhibit similar behavior to that presented here for matrix optimization problems, since in face this paper gives analogs of H^∞ optimization results derived earlier by the authors.

*partly supported by the AFOSR, the NSF, the Ford Motor Co., and the CDRF.

Contents

1 Problem statement

1.1 Coordinate Optimization

There are many problems in engineering which center on finding matrices or vectors X and Y which optimize a selfadjoint matrix valued function (or "performance measure") $\Gamma(X,Y)$. Here we treat two types of problems which are matrix versions of $\min_Y \max_X$ and of $\min_Y \min_X$ problems. A formal statement of these problems is this.

Given a smooth function

$$\Gamma : \mathbf{C}^K \times \mathbf{C}^L \to S_{n\times n}$$

from two complex vector spaces $\mathbf{C}^K$ and $\mathbf{C}^L$ into the set of self-adjoint $n \times n$ matrices denoted $S_{n\times n}$, *find the smallest real number* γ^* *so that for* $\gamma > \gamma^*$,

(MINMAX) there exists $Y \in C^L$ such that $\gamma I \geq \Gamma(X,Y)$ for all $X \in C^K$

(MINMIN) there exists $Y \in C^L$ and $X \in C^K$ such that $\gamma I \geq \Gamma(X,Y)$

Also find (X^*, Y^*) such that $\gamma^* I \geq \Gamma(X^*, Y^*)$, and such that (for the MINMAX problem case only) $\gamma^* I \geq \Gamma(X, Y^*)$ for all X, and $\gamma^* I \not> \Gamma(X^*, Y^*)$. Such pair (X^*, Y^*) is called a *solution.*

This paper pertains to what might be called *coordinate optimization* numerical algorithms for solving these problems.

The problem MINMIN is a particular case of the following more general problem.

(MIN) Given a smooth function $\Gamma : \mathbf{C}^K \to S_{n\times n}$,
find the smallest real number γ^* so that for $\gamma > \gamma^*$,
there is a$V \in \mathbf{C}^L$ such that $\gamma I \geq \Gamma(V)$
Also find V^* such that $\gamma^* I \geq \Gamma(V^*)$.

Our emphasis in this article is on the case V of the form (X,Y), but we shall give results for MIN as well.

We now restate the problems just introduced in terms of optimization of eigenvalues. Let $\lambda(M)$ denote the largest eigenvalue of a selfadjoint matrix M. Then

$$\lambda(M) = \min_{\eta I \geq M} \eta$$

and the problem MINMAX can now be restated as that of finding solutions to

$$\min_Y \max_X \ \lambda(\Gamma(X,Y))$$

A similar restatement applies to MINMIN and to MIN, e.g., MINMIN can be restated as the problem of finding solutions to

$$\min_Y \min_X \ \lambda(\Gamma(X,Y))$$

By a *coordinate optimization algorithm for* MINMAX we mean:

Coordinate Optimization (MINMAX)
Given (X_k, Y_k) *a guess at the solution, to update to* (X_{k+1}, Y_{k+1})

(MAX) *find* X_{k+1} *such that* $\lambda(\Gamma(X_{k+1}, Y_k) = \max_X \lambda(\Gamma(X, Y_k))$

Then,

$$\textbf{(MAX)} \qquad \textit{find } Y_{k+1} \textit{ such that } \quad \lambda(\Gamma(X_{k+1}, Y_{k+1}) = \max_{Y} \lambda(\Gamma(X_{k+1}, Y))$$

Coordinate optimization for MINMIN is defined analogously. Here we do not say how one solves MIN and MAX, but we assume

($XokYok$) Solutions to MIN (resp. MAX) always exist, and there is an algorithm that finds one.

Note that solutions to MIN or MAX need not be unique. We say a bit more about "bi-convexity", since in practice this yields ($XokYok$). For the MINMAX problem we use bi-convexity to mean that Γ is matrix convex in Y and matrix concave in X (see Section 7 for a definition). For the MINMIN problem bi-convexity means that Γ is matrix convex in X and Y. Using this terminology bi-convex functions Γ meeting some compactness conditions on sub and super level sets of Γ satisfy assumption ($XokYok$).

This is a fairly practical set up, since there are numerous bi-convex, indeed bi-linear problems see [SI] [VB] [GSL] of this type that arise in engineering and current conventional software solves only LMI's or convex matrix inequalities. In many cases this software applies only to each coordinate separately, so what people do in practice when confronted with hard problems like MINMIN or MINMAX is coordinate optimization. Another value of coordinate optimization is that in some cases it is faster than other algorithms. An example of this is D, K-iteration which is a popular approach to the μ-synthesis problem of H^∞ control [DGZ].

In this article we shall point out that for a large class of Γ's, coordinate optimization almost never reaches a local optimum in the case of MINMIN (Section 4), while in the case of MINMAX it typically does (Section 2.) Also we give a test to determine if a "solution" obtained via coordinate optimization actually is not a true local solution to the MINMIN problem. The test is easy to implement on a computer. The serious limitations of MINMIN are well understood in the case of H^∞ optimization, see [HMer], but here we see that they apply in many other circumstances.

Constrained optimization fits the set up we have described, see Section 5.

In Section 6 we present second order optimality conditions for MINMIN. Second order tests are important because optimization of matrix valued Γ based on linearization or coordinate descent will produce critical points which are not solutions to the problem. Sufficient conditions, especially if they include second order information, become very valuable in these cases, as they can be used to tell which among the critical points correspond to true optimal points, and to provide good update directions.

In Section 7 we discuss the concept of "matrix convexity" and its connection to second order optimality results. In Section 8 we give a brief description of of H^∞ optimization problems, which is a class of nonsmooth optimization problems closely related to the problems treated in this article. Indeed, some of the issues that arise with coordinate optimization with matrix valued functions have appeared in the context of H^∞ optimization.

2 Optima for MINMAX

Suppose that we run a coordinate optimization algorithm to try to solve MINMIN or MINMAX and that it makes progress and finally stops at (X^*, Y^*). The question is, is (X^*, Y^*) a solution to the original problem?

We say that coordinate optimization initialized at (X_0, Y_0) *stops at* (X^*, Y^*) if the sequence (X_0, Y_0), $(X_1, Y_1), \ldots$ generated by coordinate optimization converges to (X^*, Y^*). The point (X^*, Y^*) is called *stopping point* for the sequence. It is not hard to show that

Lemma 1 *Let (X^*, Y^*) be a stopping point of coordinate optimization. If coordinate optimization is re-started with initial point (X^*, Y^*) then a constant sequence is produced.*

Theorem 1 *Suppose Γ is smooth and satisfies assumption $(XokYok)$. If coordinate optimization stops at (X^*, Y^*), and if MINMAX has a solution, then (X^*, Y^*) is a solution to MINMAX.*

Proof. Suppose (X^{**}, Y^{**}) is a true solution to MINMAX, and let $\gamma^{**} := \Gamma(X^{**}, Y^{**})$. Then

$$\Gamma(X, Y^{**}) \leq \gamma^{**} I \qquad \text{for all } X$$

In particular $\Gamma(X^*, Y^{**}) \leq \gamma^{**} I$. However, by Lemma 1 we have that

$$\gamma^* = \min_Y \lambda(\Gamma(X^*, Y^*))$$

This implies that $\gamma^* \leq \gamma^{**}$.

Next we consider $\Gamma(X, Y^*)$. Since coordinate optimization on X yields value γ^*, we have

$$\Gamma(X, Y^*) \leq \gamma^* I \leq \gamma^{**} I \qquad \text{for all } X$$

Thus Y^*, γ^* satisfy the conditions in MINMAX, but γ^{**} is a minimal such number and so $\gamma^{**} = \gamma^*$, hence (X^*, Y^*) solves MINMAX. $\bullet$

3 A common class of Γ

Local optimality for MINMIN is tenuous and in this section we introduce techniques used in our study of it.

3.1 Rational matrix functions

What occurs in practice are functions Γ which are polynomial or rational in matrix variables. Examples are better than a ponderous formal definition. If M is an $n \times n$ matrix, the conjugate transpose of M is denoted by M^T, and we define

$$sym[M] := M + M^T$$

Examples of polynomials are

$$\begin{aligned} \Gamma(X) &= sym[AXBXCXD + EXFXG + Q] && (1) \\ \Gamma(X) &= sym[A_0X^TB_0XC_0 + A_1XB_1XC_1 + A_2XB_2XC_2 + A_3XB_3] && (2) \\ \Gamma(X, Y) &= sym[AXBYCXD + EXF + GYH] && (3) \end{aligned}$$

Rational functions of X are polynomials in X and in inverses of polynomials of X. For example,

$$\Gamma(X, Y) = sym[A(I + DX)^{-1}CY + E(I + Y)^{-1}F] \qquad (4)$$

3.2 Derivatives

Later we study optimality conditions and these are based on derivatives of Γ and their adjoints.

To begin, consider a function $\Gamma : C^N \to S_{n\times n}$. Now $\Gamma(Z)$ may be thought of as depending on two independent variables, Z and Z^T. We shall write $\Gamma(Z, Z^T)$ to emphasize this dependence. We define derivatives of Γ as follows.

$$\begin{aligned} D_Z\Gamma(Z)[H] &= \lim_{t\to 0} \tfrac{1}{t} \left\{\Gamma(Z+tH, Z^T) - \Gamma(Z, Z^T)\right\} \\ &= \tfrac{d}{dt}\, \Gamma(Z+tH, Z^T)\Big|_{t=0} \end{aligned}$$

$$\begin{aligned} D_{Z^T}\Gamma(Z)[H] &= \lim_{t\to 0} \tfrac{1}{t} \left\{\Gamma(Z, Z^T+tH^T) - \Gamma(Z)\right\} \\ &= \tfrac{d}{dt}\, \Gamma(Z, Z^T+tH^T)\Big|_{t=0} \end{aligned}$$

When Γ is a map from $C^L \times C^K$ to $S_{n\times n}$ write $Z = (X, Y)$, and we define partial derivatives of Γ as a function of four independent variables: X, X^T, Y, and Y^T. Just as before, we write $\Gamma(X, X^T, Y, Y^T)$ instead of just $\Gamma(X, Y)$ whenever we wish to emphasize this dependence.

$$\begin{aligned} D_X\Gamma(X,Y)[H] &:= \lim_{t\to 0} \frac{1}{t} \left(\Gamma(X+tH, X^T, Y, Y^T) - \Gamma(X, X^T, Y, Y^T)\right) \\ &= \frac{d}{dt}\, \Gamma(X+tH, X^T, Y, Y^T)\Big|_{t=0} \end{aligned}$$

The derivatives $D_{X^T}\Gamma(X,Y)[\cdot]$, $D_Y\Gamma(X,Y)[\cdot]$, and $D_{Y^T}\Gamma(X,Y)[\cdot]$ are defined similarly. For example, the derivative of Γ in (3) is

$$D_X\Gamma(X,Y)[H] = AHBYCXD + AXBYCHD + EHF$$

and the derivative of Γ in (4) is

$$D_Y\Gamma(X,Y)[H] = A(I+DX)^{-1}CH - E(I+Y)^{-1}H(I+Y)^{-1}F$$

we have illustrated two things.

Lemma 2 *If $\Gamma(Z)$ is rational, then $D_Z(Z)[H]$ has the form*

$$L[H] = \sum_{\ell} A_\ell H B_\ell$$

The same holds when $Z = (X, Y)$ for $D_X\Gamma(X,Y)$ and $D_Y\Gamma(X,Y)$.

Remark. Directional derivatives of matrix rational functions are extremely straightforward to compute. For example, the noncommutative algebra program NCAlgebra (http://math.ucsd.edu/∼ ncalg) has a command which does this automatically.

Remark. For $Z = (X, Y)$ we have

$$D_Z\Gamma(X,Y)[(H,G)] = D_X\Gamma(X,Y)[H] + D_Y\Gamma(X,Y)[G] \tag{5}$$

To obtain sufficient conditions for optimization we must use the second order terms of a Taylor expansion of $\Gamma(V+tH)$ about $t = 0 \in R$.

$$\Gamma(V+tH) = \Gamma(V) + D\Gamma(V)[H]\,t + \mathcal{H}\Gamma(V)[H]\,t^2 + \dots \tag{6}$$

Here $\mathcal{H}\Gamma$ denotes the Hessian of Γ and is defined by

$$\mathcal{H}\Gamma(V)[H] := \frac{d^2}{dt^2}\Gamma(V+tH)\Big|_{t=0}$$

One can easily show that the second derivative of a matrix rational function has the form

$$\mathcal{H}\Gamma(V)[H] = sym\left[\sum_{\ell=1}^{k} A_\ell H^T B_\ell H C_\ell \;+\; \sum_{\ell=1}^{k} D_\ell H E_\ell H F_\ell\right]$$

3.3 Adjoints of Derivatives

The proof of our main theorem on MINMIN uses certain adjoint maps, however, a reader who is interested only in the statement of results can skip this subsection.

The matrix rational function Γ is a map from $\mathbf{C}^K \times \mathbf{C}^L$ to $M_{n\times n}$. For fixed X, Y a directional derivative is a linear map, e.g.

$$D_X\Gamma(X,Y) : \mathbf{C}^K \longrightarrow M_{n\times n}$$

One must specify inner products on $\mathbf{C}^K$ and on $M_{n\times n}$ for adjoints to make sense. We use the trace inner product on $M_{n\times n}$ and the usual Euclidean inner product on $\mathbf{C}^K$.

Lemma 3 *Suppose $L : M_{s\times t} \longrightarrow M_{n\times n}$ is given by the function*

$$L(H) = \sum_j A_j H B_j$$

where $A_j : \mathbf{C}^s \to \mathbf{C}^n$ and $B_j : \mathbf{C}^n \to \mathbf{C}^t$. Then its adjoint is the map $L^\dagger : M_{n\times n} \to M_{st}$ defined by

$$L^\dagger(\psi) := \sum_j B_j \psi A_j$$

Proof. $tr\ \psi L(H) = \sum_j tr\ B_j\psi A_j H = tr\ L^\dagger(\psi)H$ •

4 First Order Optimality Conditions

4.1 First Order Conditions for MIN

The first result gives the first order necessary conditions for optimality. This result is due to J.-P.A. Haeverly and M. L. Overton. Our notation is different.

Theorem 2 *Suppose $\Gamma(Z)$ is a matrix rational function. Let $Z^* \in C^N$ be given, and let $\gamma^* := \lambda(\Gamma(Z^*))$. Let the directional derivative of Γ at Z^* be given by*

$$D_Z\Gamma(Z^*)[H] = \sum_{i=1}^{r_1} E_i H\ F_i$$

If Z^* is a local solution for MIN, then there exists a non zero $n\times n$ self adjoint matrix ψ, so that

$$\psi(\gamma^* I - \Gamma(Z^*) = 0 \tag{7}$$

$$\sum_{i=1}^{r_1} F_i\psi E_i = 0 \tag{8}$$

Call a ψ satisfying (7) (8) *maximal rank* if it has the property that no other such $\widetilde{\psi}$ has range properly containing Range ψ.

4.2 First Order Conditions for MINMIN

The MINMIN problem has much worse behavior (in terms of coordinate optimization) than the MINMAX problem even if Γ is convex in X and convex in Y. This separate convexity does not imply that Γ is jointly convex in X and Y, thus in many cases to check that (X^*, Y^*) is a local optimizer for MINMIN one needs both first and second variation optimality conditions.

Theorem 3 *Suppose Γ is a matrix rational function satisfying $(XokYok)$, and let $(X^*, Y^*) \in C^K \times C^L$ be given, and let $\gamma^* := \lambda(\Gamma(X^*, Y^*))$. Let the directional derivatives of Γ at (X^*, Y^*) be given by*

$$D_X\Gamma(X^*, Y^*)[H] = \sum_{i=1}^{r_1} E_i H \ F_i$$

$$D_Y\Gamma(X^*, Y^*)[H] = \sum_{i=1}^{r_2} P_i H \ Q_i$$

If (X^*, Y^*) is a stopping point for coordinate optimization applied to MINMIN, then there exist non zero $n \times n$ self adjoint matrices ψ_1, ψ_2, so that

$$\psi_\ell(\gamma^* I - \Gamma(X^*, Y^*)) = 0 \quad \ell = 1, 2 \tag{9}$$

$$\sum_{i=1}^{r_1} F_i \psi_1 E_i = 0 \tag{10}$$

$$\sum_{i=1}^{r_2} Q_i \psi_2 P_i = 0 \tag{11}$$

$$\psi_1 \geq 0 \quad \psi_2 \geq 0. \tag{12}$$

If (X^*, Y^*) also solves MINMIN, then

(**S**) *there exists a nonnegative matrix $\psi \neq 0$ which solves (9), (10), (11) simultaneously.*

Call (S) the **simultaneity** *condition.*

If Γ is scalar valued (i.e., $n = 1$), then the simultaneity condition (S) always holds.

The theorem holds for non rational Γ too, but in such case one obtains formulas more complicated than (10) and (11). Intuitively, one thinks of the linear in X, linear in Y case where for example Matlab's LMI toolbox is based on algorithms which manipulate primal and dual variables. In principle this makes ψ_1 and ψ_2 accessible. The theorem does not say that for (X^*, Y^*) to solve MINMIN that ψ_1 must equal ψ_2. Certainly if this happened, then the simultaneity condition is satisfied. However, the practical issue is, suppose an LMI solver yields X^*, ψ_1^* and Y^*, ψ_2^*, determine if a ψ exists meeting the simultaneity condition. The following corollary converts this to an LMI and so gives a practical test.

Corollary 1 *Let $\mathcal{N} \subset \mathbf{C}^n$ denote the null space of $\gamma^* I - \Gamma(X^*, Y^*)$. Then*

$$\mathcal{N} \supset [\text{Range } \psi_1 + \text{Range } \psi_2],$$

and usually[1] inclusion in the reverse direction holds. Let ρ denote the orthogonal projection of $\mathbf{C}^n$ onto $\mathcal{N}$. The simultaneity condition (S) has a solution if and only if there is a non-negative matrix ψ satisfying

$$\sum_{i=1}^{r_1} F_i \rho \psi \rho E_i = 0 \tag{13}$$

[1] In practice $\mathcal{N}$ would be produced as [Range ψ_1 + Range ψ_2]. There is nothing in the coordinate descent algorithm that would dispose $\mathcal{N}$ to being larger than [Range ψ_1 + Range ψ_2].

$$\sum_{i=1}^{r_2} Q_i \rho \psi \rho P_i = 0 \tag{14}$$

Note that (13) (14) and $\psi \geq 0$ constitute a conventional LMI , thus solving these numerically is straight forward and practical.

Also note that for most functions Γ it is reasonable to expect that $\gamma^* I - \Gamma(X^*, Y^*) \neq 0$. In all such cases one must have that $\mathcal{N}$ is not all of C^n, so the projection matrix ρ has rank $< n$.

Proof of Theorem 3. The optimality conditions (9) (10) (11) (12)are an immediate specialization of the general Theorem [HMW2] to this case. For example, the derivative conditions in the general case are

$$D_X \Gamma(X, Y)^\dagger [\psi_1] = 0$$

and similarly for D_Y. The computations in the previous section yield that this is (10) and similarly one gets (11).

The simultaneity condition is actually independent of the coordinate optimization condition and is a direct specialization of Theorem [HMW2] to the MINMIN problem. •

Proof of Corollary 1. Suppose coordinate optimization has produced X^*, Y^*, ψ_1, ψ_2 and γ^* . Since $\gamma^* I - \Gamma(X^*, Y^*) \geq 0$ has *Range* (ψ_1) and *Range* (ψ_2) in its null space, it also has their span in its null space $\mathcal{N}$. Now any non-negative matrix $\hat{\psi}$ supported on $\mathcal{N}$ satisfies (9) and can be written as $\hat{\psi} = \rho \psi \rho$ for some matrix $\psi \geq 0$. The only thing that remains is the additional requirements (10) (11) on derivatives and those are expressed by (13) (14). •

4.3 Genericity

The previous result leads us to suspect that the generic behavior of MINMIN is simple. For the generic MINMIN problems with $n > 1$ coordinate optimization does not reach a local optimum.

The reasoning goes as follows. In practice $\dim \mathcal{N}$ will be substantially less than n. The simultaneity condition will have $n(n+1)$ equations on $\frac{\dim \mathcal{N}}{2}(\dim \mathcal{N} + 1)$ unknowns (after reparametrization). Thus if these equations were randomly generated the probability of there existing a solution is 0. This is not a proof. The process of optimization might *in principle* produce equations (10) and (11) with extremely special properties, in addition to the ones we have written down. In practice this seems unlikely but only numerical experience will tell.

As we shall mention in Section 8 an analog situation occurs in H^∞ *optimization* when one optimizes over two functions f_1 and f_2, see [HMer]. In the H^∞ situation there is theoretical evidence based on optimality conditions very similar to what we have in this paper. Empirical evidence shows that H^∞ *coordinate optimization almost never hits a true local minimum*, for example in the few computer tests reported in [HMer] it misses the optimal value by about 30%.

The above considerations and caveats suggest that in order to solve MINMIN one has to solve (9)—(12) directly. One possibility for solving the relations is quasi-Newton algorithms applied to (9)—(11), with additional safeguards on linesearch etc. for preserving positivity constraints (12). See [HMW2], [HO] for an analogous set up.

5 Constraints on X and Y

The objective of this section is to point out that constrained problems can be treated by the theory in this paper

Suppose X, Y are self adjoint matrices and we wish to solve the problem

$$(MINMIN_C) \qquad \begin{array}{l} \textit{Find the smallest } \gamma^* \textit{ so that} \\ \textit{for all } \gamma > \gamma^*, \textit{ there exist } X, Y: \\ \gamma I \geq \Gamma(X,Y) \\ \textit{Denote optimizers by } \gamma^*, X^*, Y^* \end{array}$$

This is close to a MINMIN problem in that if we fix a number α and set

$$\widetilde{\Gamma}(X,Y) = \begin{pmatrix} \Gamma(X,Y) & 0 & 0 \\ 0 & \alpha - X & 0 \\ 0 & & \alpha - Y \end{pmatrix}$$

then the inequality

$$\gamma I \geq \widetilde{\Gamma}(X,Y) \qquad \forall X, Y$$

with $\alpha = \gamma^*$ has exactly the same solutions (X^*, Y^*) as the original problem. Thus optimality conditions for MINMIN$_C$ are easily gotten from those for MINMIN. Consequently we can and have restricted our studies to MINMIN. A similar reduction holds for MINMAX.

6 Second Order Optimality Conditions for MIN and MINMIN

6.1 Main Result

Now we give a second order test for optimality for the MINMIN problem. Of course this contains the special case of minimizing a performance function $\Gamma(V)$ over V. Conversely, the "special case" but with $V = (X, Y)$ gives a second order test for MINMIN.

Theorem 4 *If $V^* = (X^*, Y^*)$ is a local solution for MINMIN and if ψ is a maximal rank solution with ρ the projection onto Range ψ, then*

(NONNEG)
$$\begin{array}{c} for\ each\ H\ satisfying\ \rho D\Gamma(V^*)[H]\rho \leq 0\ the\ matrix \\ \rho\mathcal{H}\Gamma(V^*)[H]\rho \\ has\ a\ non-negative\ eigenvalue, \end{array}$$

Conversely, if $\psi \geq 0$ is a solution to (7) (8), with maximal rank, and if

(POS)
$$\begin{array}{c} for\ each\ H\ satisfying\ \rho D\Gamma(V^*)[H]\rho \leq 0\ the\ matrix \\ \rho\mathcal{H}\Gamma(V^*)[H]\rho \text{ has a positive eigenvalue,} \end{array}$$

then V^* is a local optimum for MIN.

We can rewrite the key condition NONNEG as

(NONNEG)
$$\lambda\left(\rho D\Gamma(V^*)[H]\rho\right) \leq 0 \quad \Rightarrow \quad \lambda\left(\rho\mathcal{H}\gamma(V^*)[H]\rho\right) > 0$$

Similarly for **(POS)** . In this notation we obtain the following corollary.

Corollary 2 *Suppose $\mathcal{N} = \mathcal{N}_*$, and that conditions (7) (8) (??) on $\gamma^*, V^*, \psi \geq 0$ hold. If*

$$\lambda\left(\rho\mathcal{H}\Gamma(V^*)[H]\rho\right) > 0\ for\ each\ H\ that\ satisfies\ \rho D\Gamma(V^*)[H]\rho = 0$$

then V^* is a local minimizer for MIN.

The converse is true with non-negative eigenvalue $\lambda \geq 0$ replacing $\lambda > 0$. The condition in the corollary can be implemented numerically in a conventional way, since it requires testing if a quadratic form on a subspace is non-negative definite.

6.2 Proofs without duality

Proof. Consider the Taylor expansion (6) compressed to $\mathcal{N}$. Then we can improve the performance of V^* with a small update in a direction H satisfying

$$\rho D\Gamma(V^*)[H]\rho < 0. \tag{15}$$

Thus at an optimum V^* such an H had better not exist. Conversely, any H making (15) have a positive eigenvalue can not give an update which improves $\lambda(\Gamma)$.

At the next level of subtlety when some H makes (15) a negative semidefinite operator, analysing the effect of an update in that direction requires second order analysis. If H also satisfies that

($\mathbf{NEG_1}$)
$$\mathcal{H}\Gamma(V^*)[H] is\ negative\ on\ \mathcal{N} \ominus Range(\rho D\Gamma(V^*)[H]\rho),$$

then taking a small update in direction H will improve performance, so V^* is not an optimum. Conversely,

$$(\mathbf{POS_1}) \qquad \begin{array}{l} \textit{if every } H \textit{ making } \rho D\Gamma(V^*)[H]\rho \leq 0 \\ \textit{produces } \mathcal{H}\Gamma(V^*)[H] \textit{ which compressed to} \\ \mathcal{N}_H := \mathcal{N} \ominus (\textit{Range } \rho D\Gamma(V^*)[H]\rho) \textit{ has a positive eigenvalue,} \end{array}$$

then γ^*, V^* is a local optimum.

This condition appears unwieldy, but can be improved. This is based on using a subspace $\mathcal{N}_*$ of $\mathcal{N}$, which is the smallest subspace containing $\mathcal{N}_H$ for all H making $D\Gamma(V^*)[H] \geq 0$. We remark that replacing $\mathcal{N}_H$ by $\mathcal{N}_*$ in (POS$_1$) appears to make it an overly lenient test for optimality. However, the next lemma concerning $\mathcal{N}_*$ makes the test sharp. Set $\widetilde{\mathcal{N}}_* := \mathcal{N} \ominus \mathcal{N}_* = \mathcal{N} \cap (\mathcal{N}_*)^\perp$

Clearly $\mathcal{N}_H \subset \mathcal{N}_*$, so $\mathcal{H}\Gamma(V^*)[H]$ compressed to $\mathcal{N}_*$ must have a negative eigenvalue if its compression to $\mathcal{N}_H$ does. Thus the condition (POS$_1$) is equivalent to (POS) although we have not yet introduced ψ, which merely is used to rephrase (POS) in dual terms, thereby obtaining (7). •

Lemma 4 *If there is an H_0 making $D\Gamma(V^*)[H_0] \leq 0$ and $\mathcal{H}\Gamma(V^*)[H_0]$ negative on $\mathcal{N}_*$, then there is an H_1 for which $D\Gamma(V^*)[H_1] < 0$ on $\widetilde{\mathcal{N}}_*$ and $\mathcal{H}\Gamma(V^*)[H_1] \ < 0$ on $\mathcal{N}_*$.*

This lemma combined with (NEG$_1$) yields the condition (NONNEG) of Theorem 4, where we take ρ to be the projection of $\mathbf{C}^n$ onto $\mathcal{N}_*$. Later we shall relate this to maximal rank ψ.

Proof of Lemma 4 By definition of $\mathcal{N}_*$, the matrix $\rho D\Gamma(V^*)[H_0]\rho \leq 0$ has range contained in $\widetilde{\mathcal{N}}_*$ and we now suppose its range is not onto $\widetilde{\mathcal{N}}_*$. By maximality of $\mathcal{N}_*$ there is an H, such that $\rho D\Gamma(V^*)[H]\rho$ has range in $\widetilde{\mathcal{N}}_*$, but not contained in Range $\rho D\Gamma(V^*)[H_0]\rho \leq 0$. Consider $H_1 = H_0 + sH$ with $s > 0$. Linearity of $D\Gamma(V^*)$ implies that $\rho D\Gamma(V^*)[H_1]\rho$ is non-positive and that its range is strictly bigger than Range $\rho D\Gamma(V^*)[H_0]\rho$. Moreover, small enough s insures that $\mathcal{H}\Gamma(V^*)[H_0]$ is so close to $\mathcal{H}\Gamma(V^*)[H_1]$ that it too is a negative matrix when compressed to $\mathcal{N}_*$. The lemma follows immediately since we have shown that there is no H_0 producing $\rho D\Gamma(V^*)[H_0]\rho$ with maximal range. •

6.3 Proofs with Duality

Negativity and positivity conditions on $L[H] := D\Gamma(V^*)[H]$ are elegantly dualized using the Hahn-Banach Theorem. It implies the space Range L and the cone $\rho P_{n\times n}\rho$ are separated by a strictly positive linear functional, $tr\ \psi$, if and only if they do not intersect[2] at more than $\{0\}$. That is, $\psi > 0$ on $\mathcal{N}$ and $tr\ \psi L[H] = 0$ all H, is equivalent to $\rho L(H)\rho \leq 0$ having only $L(H) = 0$ as a solution.

Likewise $\psi \geq 0$ prevents strict inequality in (15) from being possible. Logical semantics imply, if the range of a maximal rank ψ is properly contained in $\mathcal{N}$, then there is a non-zero $L(H) \leq 0$. Also

$$\textit{Range } L(H) \perp \textit{Range } \psi$$

follows from $tr\ \psi L(H) = 0$. Also any maximal rank ψ has range equal to $\mathcal{N}_*$. This latter fact is immediate, because $tr\ \psi_\ell L(H) = 0$ all H implies $tr(\psi_1 + \psi_2)L(H) = 0$. Thus the sum of ψ_ℓ's is also a ψ.

In the previous subsection we proved a version of Theorem 4 which did not involve duality. Combine that with the properties of ψ and maximal ψ in this section to prove Theorem 4 as it is stated.

[2]Even though the convex sets being separated are not compact their intersections with the closed unit ball are compact and so these can be separated. Since both of the original sets are cones, this proves that they can be separated.

7 Matrix Convex Performance Functions

The selfadjoint performance function Γ will be called *matrix convex* provided for all scalars $0 \leq \alpha \leq 1$ we have that

$$\Gamma(\alpha V_1 + (1-\alpha)V_2) \leq \alpha\Gamma(V_1) + (1-\alpha)\Gamma(V_2).$$

The function Γ is *strictly matrix convex* if the inequality is strict for $0 < t < 1$. The reverse inequality characterizes *matrix concave.*

Lemma 5 *If Γ is smooth, then it is matrix convex (respectively strictly concave) if and only if*

$$\mathcal{H}\Gamma(V)[H] \geq 0 \tag{16}$$

(respectively ≤ 0) for all H and V.

Proof of Lemma 5

Positive linear functionals on the space of $n \times n$ selfadjoint matrices are those for which $\ell(X) \geq 0$ if $X \geq 0$. It can be shown that positive linear functionals have the representation $\ell(X) = tr(X\psi)$ for a $\psi \geq 0$. Also, it is easy to check that $\ell(X) \geq 0$ for all positive linear functionals ℓ if and only if $\psi \geq 0$. From this it follows immediately that Γ is matrix convex if and only if the scalar valued function $\ell \circ \Gamma$ is convex in the traditional sense for each positive functional ℓ. It is standard that the lemma is true for scalar valued functions, cf. [K]. Thus it is true for $\ell \circ \Gamma$ where ℓ is any positive linear functional. From this and the fact

$$\ell \circ \mathcal{H}\Gamma(V)[H] = \mathcal{H}\ell \circ \Gamma(V)[H] \tag{17}$$

proved in the next paragraph one easily obtains that matrix convexity of Γ is equivalent to (16).

Set

$$G_{j,k} := \frac{\partial^2\Gamma(V)}{\partial v_j \partial v_k} h_j \overline{h_k} + 2Re\frac{\partial^2\Gamma(V)}{\partial v_j \partial v_k} h_j h_k. \tag{18}$$

Write $\mathcal{H}\Gamma(V)[H] = \sum_{j,k}^{L} G_{jk}$ and so

$$\ell \circ \mathcal{H}\Gamma(V)[H] = tr \sum_{j,k}^{L} G_{jk} X$$

for some X which in turn equals each of the following:

$$tr \sum_{j,k}^{L} \left\{ \frac{\partial^2\Gamma(V)}{\partial v_j \partial v_k} X h_j \overline{h_k} + 2Re\frac{\partial^2\Gamma(V)}{\partial v_j \partial^2 v_k} X h_j h_k \right\}$$

$$\sum_{j,k}^{L} \left\{ \frac{\partial^2 \ell \circ \Gamma(V)}{\partial v_j \partial v_k} h_j \overline{h_k} + 2Re\frac{\partial^2 \ell \circ \Gamma(V)}{\partial v_j \partial v_k} h_j h_k \right\} \tag{19}$$

Indeed we have proved equation 17 as was required in the previous paragraph to finish the proof. $\bullet$

We call a matrix function Γ *partly (respectively strictly) convex* if for all V and all H

$$\lambda(\mathcal{H}\Gamma(V)[H]) \geq 0.$$

(respectively > 0). Likewise *partly concave* refers to the bottom eigenvalue being negative. Corollary 2 with this language implies the following result.

Corollary 3 *Suppose condtions (7) (8) (??) on $\gamma^*, V^*\psi \geq 0$ hold. If Γ is partly strictly convex, near V^*, then V^* is a local minimizer for Γ.*

8 Coordinate Optimization in H^∞ Optimization

In this section we discuss coordinate optimization in the context of H^∞ Optimization. For simplicity of presentation we discuss the special case of optimization over two (rather than N) analytic functions. For more general results see [HMer], [HMW1], [HMW2].

Denote with $\mathcal{A}$ the vector space of all complex valued functions f of a complex variable z which are continuous on the closed unit disk, and such that the restriction of f to the open disk is an analytic function. We denote with $\mathcal{A}_2$ the vector space of all pairs $f = (f_1, f_2)$ where f_ℓ is in $\mathcal{A}$, $\ell = 1, 2$. The space $\mathcal{A}_2$ is a complete normed space (or Banach space), equipped with the norm

$$||f||_\infty = \max_\theta ||f(e^{i\theta})||_2,$$

where $||z||_2$ represents the euclidean norm on C^2, that is, for $z = (z_1, z_2)$, we have

$$||z||_2 = \left\{|z_1|^2 + |z_2|^2\right\}^{\frac{1}{2}}$$

A *performance function* is a nonnegative valued function Γ on $\partial D \times C^2$. We will use the notations $\Gamma(e^{i\theta}, z)$ and $\Gamma(e^{i\theta}, z_1, z_2)$ to denote the same object. The optimization problem that we consider is

OPT

Given a performance function $\Gamma(e^{i\theta}, z)$,

find $\gamma^* \geq 0$ and f^* in $\mathcal{A}_2$ (if it exists) such that

$$\gamma^* = \inf_{f \in \mathcal{A}_2} \sup_\theta \Gamma(e^{i\theta}, f(e^{i\theta})) = \sup_\theta \Gamma(e^{i\theta}, f^*(e^{i\theta}))$$

The function f^* in the statement of OPT is called a *solution* or *minimizer*. We call $(f_1^*, f_2^*) = f^* \in \mathcal{A}_2$ a *coordinate descent solution to* OPT provided

(CDS)

(1) $\quad \inf_{f_1} \sup_\theta \Gamma(e^{j\theta}, f_1, f_2^*) = ||\Gamma(\,\cdot\,, f^*)||_\infty$

(2) $\quad \inf_{f_2} \sup_\theta \Gamma(e^{j\theta}, f_1^*, f_2) = ||\Gamma(\,\cdot\,, f^*)||_\infty\,.$

hold locally. Clearly any local solution to OPT is a coordinate descent solution to OPT.

A word of caution with respect to notation is necessary. The symbol Γ was used in previous sections to denote a matrix valued function of vectors or matrices. In this section Γ denotes a real valued function of $e^{i\theta}$, z_1 and z_2, and the optimization problem requires that Γ be evaluated at $z_1 = f_1(e^{i\theta})$ and $z_2 = f_2(e^{i\theta})$ where f_1 and f_2 are analytic functions. Hence the setups of this section and of previous sections are not the same. However, there is a strong analogy that becomes apparent when the problem OPT is stated as a semidefinite programming problem (see [HMW2].)

The theoretical study of OPT with most practical implications is the characterization of local solutions, as conditions for optimality lead quickly to practical tests for optimality, and to algorithms [HMer], [HMW1], [HMW2].

We now present a simple (practical) test, Theorem 5 to determine when the natural coordinate descent algorithm for "solving" OPT stops. As we see the conditions for coordinate descent to stop appear so much weaker than for OPT to have a true solution that we conjecture

Conjecture 1 *For generic* Γ *coordinate descent does not obtain a true local optimum. In other words, with probability one coordinate descent gives the wrong answer.*

Theorem 5 *Let Γ be of class C^3 and $f^* = (f_1, f_2)$ in $\mathcal{A}_2$ be such that neither entry of $\left(\frac{\partial \Gamma}{\partial z_1}(\cdot, f^*), \frac{\partial \Gamma}{\partial z_2}(\cdot, f^*)\right)$ is ever equal to 0. If the function f^* is a coordinate descent solution to OPT then conditions (I) and (II) hold:*

I *The function*

$$\Gamma(e^{i\theta}, f^*(e^{i\theta})) \text{ is constant in } e^{i\theta}. \qquad \textbf{(Flatness)}$$

II *There exist $F = (F_1, F_2)$ in the Hardy space H^1_2 and two nonegative measurable functions ψ_1, ψ_2 of $e^{i\theta} \in \partial D$ such that $\int \psi_\ell \frac{d\theta}{2\pi} = 1$ and*

$$\begin{aligned} \frac{\partial \Gamma}{\partial z_1}(e^{i\theta}, f^*(e^{i\theta}))\psi_1(e^{i\theta}) &= e^{i\theta} F_1(e^{i\theta}) \\ \frac{\partial \Gamma}{\partial z_2}(e^{i\theta}, f^*(e^{i\theta}))\psi_2(e^{i\theta}) &= e^{i\theta} F_2(e^{i\theta}) \end{aligned} \qquad \begin{pmatrix} \textbf{Gradient} \\ \textbf{Alignment} \end{pmatrix}$$

If f^* is a local solution to OPT, then statements I and II with $\psi_1 = \psi_2$ hold.

The key issue is: Does CD give local solutions to OPT?

To answer this we compare II of Theorem 5 above which asserts two seemingly unrelated functions ψ_1 and ψ_2 exist when f^* is a coordinate descent solution to OPT and the condition $\psi_1 = \psi_2$ required by true local solutions to OPT. Now given two functions ψ_1 and ψ_2 the probability that they are equal is 0. Thus if indeed there is no hidden relationship between the functions ψ_1 and ψ_2, the conjecture is true. This agrees with experiments (see [HMer]).

References

[DGZ] J. Doyle, K. Glover and K. Zhou. Robust and Optimal Control. Prentice Hall, NJ, 1996

[GSL] K. C. Goh, M. G. Safonov, and J. H. Ly, Robust system design via bilinear matrix inequalities. To appear in: Int. J. Control.

[HMer] J. W. Helton and O. Merino. Conditions for Optimality over $\mathcal{H}^\infty$. SIAM J. Cont. Opt., Vol. 31, No. 6, November 1993.

[HMW1] J. W. Helton, O. Merino and T. Walker, Algorithms for Optimizing over Analytic Functions, Indiana U. Math. J., Vol. 42, No. 3 (1993).

[HMW2] J. W. Helton, O. Merino and T. Walker, H^∞ Optimization with Plant Uncertainty and Semidefinite Programming. To appear in J. Nonlinear and Robust Control.

[HO] J.-P.A. Haeverly and M. L. Overton, Optimizing Eigenvalues of Symmetric Definite Pencils. Proceedings of the American Control Conference, Baltimore, July 1994.

[K] Krantz, S., Function Theory of Several Complex Variables. Wiley. 1982

[SI] Robert E. Skelton and T. Iwasaki. "Increased Roles of Linear Algebra in Control Education." IEEE Control Systems, 8/95, pp. 76–90.

[VB] L. Vandeberghe and S. Boyd, Semidefinite Programming. SIAM Review Mar 1996