

A New Geometric Stability Margin for Feedback Systems

William H. Bennett

Systems Engineering, Inc.
Greenbelt, MD 20770

John S. Baras †

Systems Research Center
and
Department of Electrical Engineering
University of Maryland
College Park, MD 20742

ABSTRACT

This paper provides a new geometric concept for a stability margin for feedback control systems which is applicable for any number of inputs and outputs. The measure of stability margin employed has the advantage over standard measures of exposing certain additional internal stability properties of a feedback configuration employing dynamic compensation.

The stability margin is based on the essential topological features of an abstract Nyquist criterion proposed by Brockett and Byrnes. This abstract Nyquist criterion for multivariable systems involves the topology of intersections between certain pairs of subspaces of the direct sum of the system inputs and outputs. The measures employed here are derived from the idea of principal angles between a pair of subspaces.

Some examples are given which highlight the differences between these measures and other more standard measures.

contour suggested by Hermann and Martin as a "curve" on a Grassman manifold. The natural topology of the Grassman manifold involves angles between subspaces. We next provide a review of the gap-metric and the principal angles between a pair of subspaces. Based on this circle of ideas we define a geometric stability margin in terms of the gap-metric and discuss its properties for this application. Finally, a few simple examples serve to illustrate the properties of the geometric stability margin.

1.1. Background

Despite the popularity of the classical notions of gain and phase margin we will take the viewpoint (following [HE1]) that a slightly more conservative measure of stability margin is appropriate. Given the usual feedback equations

$$g(s)u(s) = y(s), \quad u(s) = -f y(s) \quad (1.1)$$

where $g(s)$ is a rational transfer function and f is a constant feedback gain.

Definition : The combined gain-phase margin, $g_{\theta m} \in \mathbb{R}_+$, is given by

$$g_{\theta m} \triangleq \inf_{\theta=j\omega} |1 + fg(s)| \quad (1.2)$$

Clearly $g_{\theta m}$ is just the euclidean distance between the Nyquist locus resulting from some loop breaking configuration and the "critical point" at -1. As a measure of stability margin $g_{\theta m}$ is more conservative than gain and phase margins since it accounts for possible simultaneous gain and phase variations. The practical extension of these ideas to MIMO systems is probably most meaningful by generalization of $g_{\theta m}$. This is at least in part due to lack of general significance of any other notion of multivariable phase which can take into account the individual phase of each scalar transfer function $g_{ij}(s)$ appearing in $G(s)$.

Some recent work on extending these notions of stability margins to MIMO feedback [DO2,LE1] has focused on the characterization of specific classes of allowable (non-destabilizing) perturbations in terms of a measure of stability employing the minimum singular value of a matrix return difference (say $I_m + FG(s)$) depending on loop breaking location. That singular value analysis is an appropriate tool in studying perturbations has been well known by numerical analysts (cf. [BJ1]). However for this analysis it leads to the possible definition of several different measures of stability margin: e.g.,

$$\begin{aligned} g_i &\triangleq \inf_{\theta=j\omega} \sigma_{\min}[I_m + FG(s)], \quad \dagger \\ g_o &\triangleq \inf_{\theta=j\omega} \sigma_{\min}[I_p + G(s)F], \\ \hat{g}_i &\triangleq \inf_{\theta=j\omega} \sigma_{\min}[I_m + \{FG(s)\}^{-1}], \\ \hat{g}_o &\triangleq \inf_{\theta=j\omega} \sigma_{\min}[I_p + \{G(s)F\}^{-1}]. \end{aligned} \quad (1.3)$$

† We use the notation $\sigma_{\min}(A)$ to denote the minimum singular value of the matrix A

1. Introduction

In classical design of feedback control the use of stability margins plays a role not only in determining system performance but also in providing guidelines for the choice of dynamic (lead/lag) compensation. In contrast existing stability margin concepts for MIMO systems focus on analysis of performance and robustness of control once the compensation has been chosen. This is accomplished by applying perturbation analysis for matrices (using for instance SVD analysis) to a matrix return difference.

In this paper we discuss an alternate construction of a stability margin for feedback control. This "geometric" stability margin is based on natural topological considerations for linear systems and can be readily computed using the notion of "principal angles" between a pair of subspaces. Moreover, unlike margins based on the return difference which reflect the loop transmission for some fixed loop breaking configuration, this geometric stability margin reflects a certain "distance" between two (possibly dynamic) transfer functions (one for the system or plant and one for compensation). Thus in certain cases where possible cancellations occur in forming the return difference this geometric stability margin can provide additional information about the potential for internal instabilities which would not be evident from the usual methods.

We begin by motivating the idea of a stability margin for feedback control in terms of the distance between a "critical point" and a Nyquist contour. We review the abstract Nyquist

† Research supported in part by NSF grant ECS-82-19123A01 at the University of Maryland, College Park.

each of which may lead to different characterization of allowable perturbations since in general these measures all achieve different values. These measures are however not completely independent as discussed by Safonov et al [SA1].

2. A Geometric View of MIMO Feedback

In this section we provide background on a particularly useful geometric construction of an abstract Nyquist contour for MIMO systems. We focus on some particularly salient properties of the abstract Nyquist contour for MIMO systems as discussed in [BR4]. We start with the general MIMO feedback equations with $u \in \mathbf{U} = \mathbf{C}^m$, $y \in \mathbf{Y} = \mathbf{C}^p$

$$G(s)u(s) = y(s), \quad u(s) = -Fy(s) \quad (2.1)$$

with $G(s) \in \mathbf{R}^{p \times m}(s)$, $F \in \mathbf{R}^{m \times p}$ which we write suggestively as

$$\begin{bmatrix} I_m & F \\ G(s) & -I_p \end{bmatrix} \begin{pmatrix} u(s) \\ y(s) \end{pmatrix} = 0. \quad (2.1')$$

The geometric theory of linear systems and feedback centers on the relative orientation of the two objects,

$$G(s) = \ker \begin{bmatrix} G(s) & -I_p \end{bmatrix} \quad (2.2)$$

and

$$F = \ker \begin{bmatrix} I_m & F \end{bmatrix} \quad (2.3)$$

which for any $s \in \mathbf{C}$ are a pair of subspaces in $\mathbf{U} \oplus \mathbf{Y} = \mathbf{C}^{p+m}$. Following Hermann and Martin [HE2] we can state the following.

Theorem 2.1: [HE2]. A complex number s_0 is a closed loop pole of the feedback equations (2.1) if and only if

$$\dim (G(s_0) \cap F) > 0. \quad (2.4)$$

Hermann and Martin [HE2] suggested the following definition of an abstract Nyquist locus.

Definition: The (abstract) Nyquist locus, Γ_G , of a $p \times m$ transfer function $G(s)$ is an algebraic "curve" given by the map

$$s \rightarrow \ker \begin{bmatrix} G(s) & -I_p \end{bmatrix}$$

as the image of the closed contour D . Γ_G is contained in the complex Grassman space consisting of all m -dimensional subspaces in \mathbf{C}^{p+m} . We consider a curve in a more general sense as an analytic map from a Riemann surface to a complex analytic manifold, viz., the complex Grassmanian. In this sense a curve has complex dimension one or real dimension two.

The complex Grassmanian is the set of all p -dimensional complex subspaces of \mathbf{C}^n , which we denote as $\text{Grass}(p, n)$. $\text{Grass}(p, n)$ admits the structure of an analytic manifold in this case of dimension $np - p^2$. A fundamental property of $\text{Grass}(p, n)$, which was successfully exploited in [BR4] toward the construction of a generalized Nyquist test, is the duality between $\text{Grass}(p, n)$ and $\text{Grass}(n-p, n)$. In particular, a canonical representation of a point $\mathbf{X} \in \text{Grass}(p, n)$ is as a hypersurface $\sigma(\mathbf{X}) \subseteq \text{Grass}(n-p, n)$. This so called Schubert hypersurface is given by

$$\sigma(\mathbf{X}) = \left\{ \mathbf{Y} \in \text{Grass}(n-p, n) : \dim(\mathbf{X} \cap \mathbf{Y}) \geq 1 \right\}; \quad (2.5)$$

i.e., all $\mathbf{Y} \in \text{Grass}(n-p, n)$ which intersect $\mathbf{X} \in \text{Grass}(p, n)$ non-trivially.

The most significant aspect of the abstract Nyquist locus, Γ_G , as constructed above is that it is fixed with respect to choice of feedback compensation, F , in contrast to methods which focus on eigenloci or determinants of a matrix return difference.

However, the construction is quite general and allows connections with more standard analyses. For instance, by change of basis in the space of inputs and outputs, $\mathbf{U} \oplus \mathbf{Y}$, one can generate a new "rotated" Nyquist contour, Γ_2 . In the particular case

$$\begin{bmatrix} G(s) & -I_p \end{bmatrix} \begin{bmatrix} I_m & F \\ G(s) & -I_p \end{bmatrix} = \begin{bmatrix} 0 & G(s)F + I_p \end{bmatrix} \quad (2.6)$$

reveals a "rotated" curve, Γ_2 as

$$\ker \begin{bmatrix} 0 & G(s)F + I_p \end{bmatrix} : D \rightarrow \Gamma_2.$$

However, the transformation

$$\begin{pmatrix} I_m & F \\ G(s) & -I_p \end{pmatrix}$$

represents a valid change of basis in \mathbf{C}^{p+m} only for s not a closed loop pole. To say this another way Γ_2 is a valid curve contained in $\text{Grass}(m, p+m)$ whenever s is not a closed loop pole for s on D . (Of course the standard construction of D does not guarantee this.) Thus, from this geometric point of view one can see the advantage of working in the "natural" basis (given by (2.1)) in defining a "legitimate" Nyquist contour.

A geometric viewpoint is taken in Brockett and Byrnes [BR4] in describing a generalized Nyquist criterion. Here for the first time the general case of $G(s) \in \mathbf{R}^{p \times m}(s)$ with $p \neq m$ is treated, although somewhat abstractly. Significantly, their approach avoids formulation of the return difference matrix and as a result allows the separate characterization of an abstract critical point (resulting from F) and a Nyquist locus (resulting from $G(j\omega)$). This formulation preserves most nearly the practical aspects of Nyquist criterion exploited in SISO system design [BR4].

Using this dual structure (cf. [BR4] and [BY1] for details) the following theorem is provided.

Theorem 2.2: (Generalized Nyquist Theorem)

Suppose $G(s)$ is a proper rational $p \times m$ transfer function matrix with no poles on $\text{Re } s = 0$. Suppose the abstract Nyquist locus Γ_G does not intersect the Schubert hypersurface $\sigma(FB)$ defined by the feedback matrix F . Let p_o be the number of open loop poles of $G(s)$ in the closed right half plane (CRHP) and p_c be the number of closed loop poles in CRHP. Then

$$N(\Gamma_G; \sigma(F)) = p_c - p_o \quad (2.7)$$

where $N(\bullet; \bullet)$ is the number of encirclements of the abstract Nyquist locus Γ_G about the Schubert hypersurface $\sigma(F)$ taken in a positive direction on the Grassman manifold.

Proof: (cf. [BR4]).

Clearly theorem 2.2 does not admit any readily obtainable graphical representation that would permit the determination of the winding number N (except in trivial cases). However the theorem does permit us to ascertain the stability of a MIMO feedback system involving a plant $G_1(s)$ with feedback F_1 by testing for homotopic equivalence with some other feedback system $G_2(s)$ with F_2 (of appropriate dimensions) which is known to be stable. To show such equivalence we will need a measure of how close a point $G(s) \in \text{Grass}(p, m+p)$ is to some Schubert hypersurface $\sigma(F) \in \text{Grass}(p, m+p)$.

We next consider a special class of transfer functions whose Nyquist loci have special properties.

Definition: A transfer function $G(s) \in \mathbf{R}^{p \times m}(s)$ is said to be degenerate if and only if there exists some Schubert hypersurface $\sigma(\mathbf{X}) \subseteq \text{Grass}(m, p+m)$ associated with some $\mathbf{X} \in \text{Grass}(p, m+p)$ which contains the abstract Nyquist locus, Γ_G .

According to [BR4] degenerate transfer functions are rather special.

Theorem 2.3: [BR4] Let $G(s)$ be strictly proper with McMillan degree n . If $mp \leq n$ then nondegeneracy is generic in the space of strictly proper transfer functions. If $mp > n$, then every $G(s)$ is degenerate \dagger .

Proof: (cf. [BR4], Thm. 4.2).

The notion of stability margins as discussed in [BR4] involves a measure of how close the Nyquist contour Γ_G is to the

\dagger In the context of this discussion, the statement that nondegeneracy is generic means that the set of all degenerate transfer functions can be defined by algebraic equations

"critical point", $\frac{-1}{f}$ in the complex plane. In this geometric setting we will be concerned with how nearly Γ_G intersects the appropriate fixed hypersurface $\sigma(F)$ for $F = \ker [I_m, F]$. In preparation for these results we discuss in the next section the topology of the Grassman manifold.

1. Angle Topology of the Grassman Space

1.1. Plücker Metric

As discussed above the Grassman space, $\text{Grass}(p, n)$, is the space of all p -dimensional subspaces of \mathbb{C}^n . Clearly for any $N \in \text{Grass}(p, n)$ we can express p basis vectors for N in terms of some coordinate system by writing an $n \times p$ matrix say B_1 . In terms of some other coordinate system we can write a new matrix B_2 . In this case there exists a nonsingular $p \times p$ matrix A such that $B_1 = B_2 A$ and conversely. Thus $N = \text{image}(B_1) = \text{image}(B_2)$.

Following Byrnes, et al [BY1] we state

Definition: The Plücker coordinates of a $p \times n$ matrix B is the $\binom{n}{p}$ dimensional vector of determinants of all $p \times p$ submatrices of B .

It is easy to show that for any B_1 and B_2 as above their respective Plücker coordinates will differ by a scalar multiple of each other; i.e., their Plücker coordinates are aligned. One way to provide a notion of distance on $\text{Grass}(p, n)$ is then to think of any two subspaces of dimension p say $N, M \in \text{Grass}(p, n)$ in terms of their Plücker coordinates (say $u_N, v_M \in \mathbb{C}^{\binom{n}{p}}$). Then the function

$$d(N, M) = \frac{(1 - u_N^* v_M)}{\|u_N\|_2 \|v_M\|_2} \quad (3.1)$$

is an "angle" metric and obeys the property $-1 < d(N, M) < 1$.

3.2. Gap Metric

Another sort of angle metric can be described by thinking of points in $\text{Grass}(p, n)$ in terms of the p -dimensional subspaces of \mathbb{C}^n that they represent. Here we take an abstract "basis-free" viewpoint in describing the subspaces.

Let $M, N \subseteq \mathbb{C}^n$ be subspaces. Let $\dim M = p$ and $\dim N = m$. The following function called the gap (or aperture) between M and N is defined in Kato [KA1, pp.197].

Definition: The gap between M and N is

$$\delta(M, N) = \max \left\{ \sup_{\|x\|=1} \inf_{y \in N} \|x - y\|, \sup_{\|y\|=1} \inf_{x \in M} \|y - x\| \right\} \quad (3.2)$$

The gap function obeys the following properties which follow immediately from the definition.

- (P1) $\delta(M, N) = 0$ if and only if $M = N$.
- (P2) $\delta(M, N) = \delta(N, M)$.
- (P3) $0 \leq \delta(M, N) \leq 1$.

Furthermore, the gap function obeys the following property which will be significant for us:

- (P4) $\delta(M, N) < 1$ if and only if $\dim(M) = \dim(N)$ (cf. [KA1, pg. 200]).

In general, the gap, $\delta(M, N)$, is not a metric.† However, after modifying the definition slightly by taking infimums over the appropriate unit ball in each subspace, the resulting modified gap, $\gamma(M, N)$, obeys the triangle inequality and thus becomes a metric.

Property (P4) clearly indicates that using this modified "gap-metric" one can actually form a basis for a topology of the Grassman space via neighborhoods of the form

† In the sequel we will consider the gap based on the euclidean vector norm. In this case the gap is a metric.

$$B_\epsilon(M) = \{ N \in \text{Grass}(p, n) : \delta(M, N) \leq \epsilon < 1 \}$$

3.3. Orthogonal Projections in Unitary Space and the Gap Metric

In a unitary space, E^n , we can employ the notion of an orthogonal projector to represent a subspace $M \subseteq E^n$; e.g., take $E^n = \mathbb{C}^n$ and the natural inner product $\langle x, y \rangle = \bar{x}^* y$. If P_M (resp. P_N) is an orthogonal projector whose range is the subspace $M \subseteq E^n$ (resp. $N \subseteq E^n$) then using the natural Euclidean norm we can state the following:

Theorem 3.1:

$$\delta(M, N) = \|P_M - P_N\|_2$$

Proof: (cf. Kato [KA1]).

Property (P4) of the gap is then related to the following fact.

Theorem 3.2: Any two orthogonal projectors P_M, P_N which satisfy $\|P_M - P_N\| < 1$ are unitarily equivalent. That is, there exists a unitary transformation U such that $U P_M U^* = P_N$ (U is unitary if $U^* U = I$).

Proof: (cf. Kato [KA1]).

Unitary transformations have an intuitive geometric appeal because they represent orthogonal rotations of the given vector space coordinate system. Thus we see that with the structure of a unitary space the gap-metric (here the gap function $\delta(\bullet, \bullet)$ becomes naturally a metric) takes on a particularly natural geometric appeal. Indeed, in finite dimensional unitary spaces, for which we have interest, the transformation U of theorem 3.2 can be represented by an easily computable matrix. In Kato [KA1] these results (and others) are used to study perturbations of linear operations on infinite dimensional spaces. In Stewart [ST1] similar ideas are applied to certain numerical problems in the computation of invariant subspaces for matrix (finite dimensional) operators. As we discuss in the subsequent sections our concern is slightly different but will follow along the same line of reasoning.

3.4. Near Intersection Between Subspaces and the Minimum Gap Function

From the statement of the generalized Nyquist criterion above it is clear that we will be interested in characterizing the "near" intersection between certain pairs of subspaces. On the Grassmanian manifold this is characterized by near intersection between a point $M \in \text{Grass}(p, n)$ and a Schubert hypersurface $\sigma(N) \subseteq \text{Grass}(p, n)$ associated with the subspace $N \in \text{Grass}(n-p, n)$.

Towards this end we provide the following:

Definition: Let $M \subseteq \mathbb{C}^n$ be a p -dimensional subspace and $N \subseteq \mathbb{C}^n$ an m -dimensional subspace. Then the *minimum gap* (or *min-gap*) between M and N in \mathbb{C}^n is given by.

$$\gamma(M, N) = \min \left\{ \inf_{\|x\|=1} \inf_{y \in N} \|x - y\|, \inf_{\|y\|=1} \inf_{x \in M} \|y - x\| \right\} \quad (3.3)$$

Obviously, the minimum gap satisfies the properties:

- (P1) $0 \leq \gamma(M, N) \leq \delta(M, N)$
- (P2) $\gamma(M, N) = 0$ if and only if $\dim(M \cap N) > 0$.

Based on (P2) it is clear that, for the abstract Nyquist criterion described in section 2, the min-gap can provide a measure of distance between the abstract Nyquist contour Γ_G and the abstract critical point $\sigma(F)$ as

$$\min_{\text{Re } s = 0} \gamma(G(s), F) \quad (3.4)$$

where $G(s) \in \text{Grass}(m, p+m)$ and $F \in \text{Grass}(p, n)$. In section 4.0 we consider this further.

Following the line of reasoning of section 3.3 we make the following claim.

Corollary 9.3: If P_M and P_N are both orthogonal projectors in C^n with $\text{image}(P_M) = M$, $\text{image}(P_N) = N$ then

$$\gamma(M, N) = \|P_M - P_N\|_2 \quad (3.5)$$

Remark: Here we introduce the notation $\|A\|$. Let $|\bullet|$ denote a vector norm on the subspace X of C^n . Then denote the matrix infimum or reciprocal norm as:

$$\|A\| = \inf_{\substack{x \in X \\ x \neq 0}} \frac{|Ax|}{|x|}$$

Clearly if A is nonsingular then

$$\|A\| = \|A^{-1}\|^{-1}$$

3.5. Canonical Angles Between Subspaces

There is a natural notion of angles between pairs of subspaces in a unitary space. In finite dimensional spaces these angles can be computed from singular values of a particular matrix. If we let M, N be a pair of subspaces of C^n with $\dim M = p$, $\dim N = m$. Assume $m > p$. Then we say the smallest angle between M and N (cf. [BJ1]), $\theta_1(M, N) = \theta_1 \in [0, \frac{\pi}{2}]$, is given by

$$\cos \theta_1 = \max_{\substack{u \in M \\ \|u\|_2=1}} \max_{\substack{v \in N \\ \|v\|_2=1}} u^* v \quad (3.6)$$

Following Björck and Golub [BJ1] we define recursively the *principal angles*, θ_k , $k=1, \dots, p$ as follows.

Definition: The principal angles $\theta_k \in [0, \frac{\pi}{2}]$ between M and N are given recursively for $k=1, 2, \dots, p$ by

$$\cos \theta_k = \max_{\substack{u \in M \\ \|u\|_2=1}} \max_{\substack{v \in N \\ \|v\|_2=1}} u^* v = u_k^* v_k \quad (3.7)$$

subject to the constraints

$$u_j^* u = 0, \quad v_j^* v = 0 \quad (3.8)$$

for $j=1, \dots, k-1$. We call the set of vectors $\{u_1, \dots, u_p, v_1, \dots, v_p\}$ the *principal vectors* for the pair of subspaces.

In this section we review how the principal angles can be computed for a pair of subspaces. The relation between certain principal angles and the gap will be clarified using orthogonal projectors. The result will be a computational procedure for determining the gap, $\delta(\bullet, \bullet)$, and the min-gap, $\gamma(\bullet, \bullet)$, between a pair of subspaces M, N . Moreover using the principal vectors we can compute a basis for the intersection, $M \cap N$. For the problem of multivariable feedback such a basis can be used to describe how certain modal behavior of the system is reflected from an input-output view point.

The main computational result which we exploit requires that we have a unitary basis for each of the subspaces M and N . Since this can be obtained conceptually using a Gram-Schmidt procedure (and in practice using Householder reflections) we assume that we have a pair of matrices $Q_M \in C^{n \times m}$, $Q_N \in C^{n \times p}$ with $Q_N^* Q_N = I_m$ and $Q_M^* Q_M = I_p$ such that $\text{image}(Q_M) = M$ and $\text{image}(Q_N) = N$.

Theorem 9.4: Given Q_M and Q_N such that $\text{image}(Q_M) = M$ and $\text{image}(Q_N) = N$ each a subspace of C^n . Compute the singular value decomposition (SVD) of

$$Q_M^* Q_N = Y_M C Y_N^* \quad (3.9)$$

where

$$Y_M^* Y_M = Y_M Y_M^* = Y_N^* Y_N = I_p \quad (3.10)$$

$$C = \cos \Theta = \text{diag} \{ \sigma_1, \dots, \sigma_p \}$$

with singular values $\sigma_1 \geq \dots \geq \sigma_p$ and

$$\Theta = \text{diag} \{ \theta_1, \dots, \theta_p \}$$

† Note that in finite dimensional unitary spaces that the right hand side of (3.5) is just the minimum singular value of the matrix difference

Then $\theta_1 \leq \dots \leq \theta_p$ are the principal angles between M and N . The columns of $U = Q_M Y_M$, $V = Q_N Y_N$ are the principal vectors.

Proof: (cf. [BJ1, thm.1]).

Corollary 9.5: Let $P_M = Q_M Q_M^*$ be an orthogonal projector of M . Then compute the SVD of

$$(I_n - P_M) Q_N = W_M S Y_N^* \quad (3.11)$$

where $S = \sin \Theta$. Here W_M gives the principal vectors in the orthogonal complement, M^{per} , associated with the pair of subspaces M, N .

Theorem 9.6: As above, let P_M and P_N be orthogonal projectors on M and N respectively. Then the nonzero eigenvalues of $P_M - P_N$ are $\pm \sin \theta_i$, for $i=1, \dots, p$.

Proof: (cf. [ST1, thm.2.5]).

Finally we can state as a corollary to theorem 3.6:

Corollary 9.7: With the above notation

$$\delta(M, N) = \|P_M - P_N\|_2 = \sin \theta_p \quad (3.12)$$

$$\gamma(M, N) = \|P_M - P_N\|_2 = \sin \theta_1 \quad (3.13)$$

Proof: See Theorem 3.1 and Corollary 3.3.

3.6. Computational Procedures for Obtaining the Gap and Min-Gap

Following the procedure suggested by corollaries 3.7 and 3.5 we can provide a straightforward, numerically stable, procedure for computing the gap or min-gap between a pair of finite dimensional subspaces $M, N \in C^n$ in terms of some matrix representations. In this most general form, the procedure is computationally intensive.

Let M, N be represented as $M = \text{image}(M)$, $N = \text{image}(N)$ where $M \in C^{n \times p}$, $N \in C^{n \times m}$. The following procedure can be coded directly for computer solution using, for instance, LINPACK routines [DO1].

Procedure for Computing the Gap or Min-Gap

Given: M, N , a pair of $n \times p$ (resp. $n \times m$) matrices

Step₁: Obtain a unitary basis for M . Conceptually, this is done by obtaining a QR factorization of M

$$M = [Q_M, Z_M] \begin{bmatrix} R_M \\ 0 \end{bmatrix}$$

where R_M is right triangular matrix and $[Q_M, Z_M]$ is unitary. Then Q_M is the required $n \times p$ matrix of unitary bases for M .

Step₂: Obtain a unitary basis for N . Again, employ a QR factorization as

$$N = [Q_N, Z_N] \begin{bmatrix} R_N \\ 0 \end{bmatrix}$$

with R_N right triangular. Now, Z_N is the $n \times (n-p)$ matrix of bases for M^{per} .

Step₃: To compute the gap, $\delta(M, N)$, (resp. min-gap, $\gamma(M, N)$) obtain the maximum (resp. minimum) singular value of the $(n-p) \times m$ matrix

$$Z_N^* Q_M$$

Remark: The QR factorization outlined here can be performed using a numerically stable algorithm involving the use of Householder reflections to compute the transformations. This has been implemented efficiently in LINPACK routine CQRDC [DO1].

Remark: The product of unitary matrices can be obtained in a numerically stable way. Then application of a standard algorithm can provide the required singular value. The routine CSVDC is available in LINPACK for computing these quantities [DO1].

4. Generalized Stability Margins from a Geometric Viewpoint

For the purposes of designing feedback compensation, it is usually more convenient to consider a slightly different stability margin; viz.,

$$g'_{\theta m} = \inf_{s \in D} |f^{-1} + g(s)|,$$

which is the euclidean distance between the Nyquist contour, Γ_g (fixed with respect to $g(s)$) and a critical point, $s = -\frac{1}{f}$, depending on the choice of feedback. Furthermore, insight can be gained for the design of dynamic (lead/lag) feedback compensation in this setup by considering a locus of critical points, Γ_f where $-f^{-1}(s): D \rightarrow \Gamma_f$ [RO1, pp. 56-59].

Clearly any generalization of $g'_{\theta m}$ to the case of MIMO feedback based on regularity of the matrix function $F^{-1} + G(s)$ on D can be meaningful only in special cases (e.g. $p=m$ and F nonsingular). However, the geometric Nyquist criterion discussed in section 2.3.2 suggests applying the topology of the Grassman space to construct a measure of how nearly the subspaces $G(s) = \ker [G(s), -I_p]$ and $F = \ker [I_m, F(s)]$ intersect in $U\theta Y$.

4.1. Definition and Properties of a Geometric Stability Margin

We employ the minimum gap function, $\gamma(N, M)$, between a pair of subspaces N and M of a unitary space to measure the distance between the abstract Nyquist locus, $\Gamma_G \subseteq \text{Grass}(m, p+m)$ and the Schubert hypersurface, $\sigma(F)$, representing the fixed critical point $F = \ker [I_m, F] \in \text{Grass}(p, m+p)$.

Definition (Geometric Stability Margin): Given $G(s) \in \mathbb{R}^{p \times m}(s)$ and $F \in \mathbb{R}^{m \times p}$ the *geometric stability margin*, $g_{\theta m}$, is a real number $0 \leq g_{\theta m} \leq 1$ given by,

$$g_{\theta m} \triangleq \inf_{s \in D} \gamma(G(s), F). \quad (4.1)$$

In this section we focus on some properties of the geometric stability margin which clarify its relation to more traditional stability margins.

Theorem 4.1: A point $s_0 \in \mathbb{C}$ is a closed loop pole of the feedback equations (2.1) if and only if either of the following holds

$$(i) \quad \det [I_p + G(s_0)F] = \det [I_m + FG(s_0)] = 0 \quad (4.2)$$

$$(ii) \quad \gamma(\ker [G(s_0), -I_p], \ker [I_m, F]) = 0 \quad (4.3)$$

Next we would like to show that for,

$$\begin{aligned} \phi_1(s) &= \sigma_{\min} [I_m + FG(s)], \\ \phi_2(s) &= \sigma_{\min} [I_p + G(s)F], \\ \phi_3(s) &= \gamma(G(s), F), \end{aligned}$$

which map the closed contour $D \subseteq \mathbb{C}$ into \mathbb{R}_+ (resp. $[0,1] \in \mathbb{R}$ for ϕ_3), if for some s_1 , $\phi_1(s_1)$ achieves its minimum on D then $\phi_2(s_1)$ and $\phi_3(s_1)$ also achieve their respective minimums at s_1 . To do this we consider some further aspects of the topology of Grassman manifolds.

Lemma 4.2: With $Y \in \text{Grass}(p, m+p)$ the set

$$B_\epsilon(Y) = \left\{ X \in \text{Grass}(p, m+p) : \delta(X, Y) < \epsilon \right\}. \quad (4.4)$$

is a convex neighborhood of $\text{Grass}(p, m+p)$ for $\epsilon < 1$.

Lemma 4.3: Let $\Gamma_G \subseteq \text{Grass}(m, p+m)$ be the abstract Nyquist contour associated with the $G(s) \in \mathbb{R}^{p \times m}(s)$ and $F = \ker [I_m, F]$. For any $X \in B_{g_{\theta m}}(F) \subseteq \text{Grass}(p, m+p)$

$$N(\Gamma_G; \sigma(F)) = N(\Gamma_G; \sigma(X)). \quad (4.5)$$

Proof: (cf. [BE2])

Finally, we clarify the extent to which $g_{\theta m}$ provides similar information with respect to gain variations in F . With respect to an appropriately constructed contour D (which avoids poles of $G(s)$ on $j\omega$ axis) the values $\gamma(G(s), F)$ on D form a proper subset of $[0,1] \subseteq \mathbb{R}$ which is both closed and bounded. Therefore, we can replace the definition (4.1) with

$$g_{\theta m} = \min_{s \in D} \gamma(G(s), F). \quad (4.1')$$

In the following theorem we will consider the case (which is most typical in practice) when the set

$$\arg \min_{s \in D} \gamma(G(s), F)$$

consists of a single point $s^* \in D$. More generally, this set will consist of a countable number of points on D .

Theorem 4.4: If

$$g_{\theta m} = \min_{s \in D} \gamma(G(s), F) > 0, \quad (4.7)$$

and

$$s^* = \arg \min_{s \in D} \gamma(G(s), F), \quad (4.8)$$

then there exists a "gain" $K_1 \in \mathbb{C}^{p \times p}$ such that

$$\min_{s \in D} \gamma(G(s), X) = \gamma(G(s^*), X) = 0, \quad (4.9)$$

for some $X = \ker [I_m, FK_1]$ if and only if there exists a $K_2 \in \mathbb{C}^{p \times p}$ such that

$$\inf_{s \in D} \det [I_p + G(s)FK_2] = \det [I_p + G(s^*)FK_2] = 0. \quad (4.10)$$

Moreover, K_1 and K_2 both satisfy

$$\|K_i\| \geq \frac{1}{\|G(s^*)F\|}. \quad (4.11)$$

Proof: (cf. [BE2])

5. Application of the Geometric Stability Margin and Some Examples

In this section we seek to demonstrate some salient features of the geometric analysis of stability margins for feedback systems. We indicate, by way of illustration, that the geometric stability margin proposed in the previous section has some peculiar properties which can extend its application to more general settings than the classical case. Indeed, even for SISO analysis, the geometric analysis can provide additional useful information which can be lost using the classical approach.

As in the previous section, our discussion focuses on the properties of the maps

$$\phi_1(s) = \sigma_{\min} [I_m + FG(s)],$$

$$\phi_2(s) = \sigma_{\min} [I_p + G(s)F],$$

$$\phi_3(s) = \gamma(G(s), F);$$

viz., their respective local minima on D ,

$$g_{\theta m 1} = \inf_{s \in D} \phi_1(s),$$

$$g_{\theta m 2} = \inf_{s \in D} \phi_2(s),$$

$$g_{\theta m 3} = \inf_{s \in D} \phi_3(s).$$

In the SISO (classical) case where $p=m=1$ we get

$$\phi_1(s) = 1 + fg(s) = \phi_2(s)$$

regardless of where the single loop is broken. However, $\phi_3(s)$ is fundamentally different from $\phi_1(s) = \phi_2(s)$ even in this case.

Corollary 5.1: In the SISO case where $p=m=1$ the geometric stability margin becomes,

$$g_{\theta m} = \inf_{s \in D} \phi_3(s) \quad (5.1)$$

for each $s \in \mathbb{C} \cup \{\infty\}$ in terms of the set of all possible ordered pairs in $\mathbf{U} \times \mathbf{Y}$; i.e.,

$$\text{graph}\{G(s)\} = \left\{ \left(u(s), G(s)u(s) \right) \in \mathbf{U} \times \mathbf{Y} \right\}.$$

So the geometric viewpoint focuses on identifying the equivalence between $\text{graph}\{G(s)\}$ and $\ker [G(s), -I_p]$ for $s \in \mathbb{C} \cup \{\infty\}$.

Next we remark that given a (left) coprime factorization

$$G(s) = D^{-1}(s)N(s)$$

then clearly

$$\mathbf{G}(s) = \ker [G(s), -I_p] = \ker [N(s), -D(s)]$$

has rank m over the field of rational functions. This means that $\text{rank } \mathbf{G}(s) = m$ for $s \in \mathbb{C} \cup \{\infty\} - \{p_{oi}\}$. Thus $\mathbf{G}(s)$ can be thought of as an element of $\text{Grass}(m, p+m)$ for $s \in \mathbb{C} \cup \{\infty\} - \{p_{oi}\}$; but for some sequence $s_n \rightarrow p_{oi}$ the sequence $\mathbf{G}(s_n)$ does not have a limit on $\text{Grass}(m, p+m)$. Nevertheless, if we think in terms of $\text{graph}(G(s))$ it is clear that

$$\mathbf{X}_{oi} = \lim_{n \rightarrow \infty} \text{graph}\{G(s_n)\} \subseteq \mathbf{U} \oplus \mathbf{Y}$$

is a subspace of $\mathbf{U} \oplus \mathbf{Y}$ such that

$$\dim(\mathbf{X}_{oi} \cap \mathbf{Y}) \geq 1.$$

Thus we make the following observation based on the angle topology of the gap.

Theorem 5.2: Let $p_{oi} \in \{p_{oi}\}$ be an internal closed loop pole of (2.1). Consider any sequence $s_n \in \mathbb{C}$ which approaches p_{oi} . Then

$$\lim_{n \rightarrow \infty} \gamma(\mathbf{G}(s_n), \mathbf{F}(s_n)) = 0.$$

Proof: (cf. [BE2]).

Example 5.2: Consider again example 5.1 where we reveal another possible internal loop breaking configuration as illustrated in Figure 5.4. Here $H(s)$ is the 2×1 transfer function

$$H(s) = \begin{bmatrix} \frac{2s+1}{s[(s+.025)^2+.01]} \\ \frac{1}{2s+1} \end{bmatrix}$$

and

$$F(s) = \left[\frac{[(s+.025)^2+.01]}{s+1}, -4 \right].$$

A hidden mode is revealed. We plot in figure 5.5 the curve $\gamma(H(s), F(s))$ for $s = j\omega$ with $\omega \in [0.1, 8]$.

Finally, we state a caveat. In the case that a hidden mode exists for some dynamic feedback configuration; e.g., $f(s) = f'(s)/a(s)$ and $g(s) = a(s)g'(s)$, then the function,

$$\left| \frac{1}{f(s)} + g(s) \right| = |a(s)| \left| \frac{1}{f'(s)} + g'(s) \right|,$$

amounts to a scaling which can degrade the numerical conditioning of the computational problem. Use of the min-gap function does not mitigate this problem. Indeed,

$$\begin{bmatrix} 1/a(s) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & f'(s) \\ g'(s) & -1 \end{bmatrix} \begin{bmatrix} a(s) & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & f(s) \\ g(s) & -1 \end{bmatrix}.$$

is again a scaling of the feedback equations which can degrade the numerical conditioning of the problem of computing the principle angles via singular value analysis in the neighborhood of a root of $a(s)$.

6. Conclusions

We have presented a new concept for a stability margin for feedback structures which is broadly applicable without regard for the number of inputs or outputs. The geometric stability margin is fundamentally different from the standard measures used even in the SISO case. It appears that this stability margin may be useful in many cases where $p \neq m$ and where guidance is sought for the choice of dynamic compensation. The metric employed can be readily computed using numerically stable algorithms.

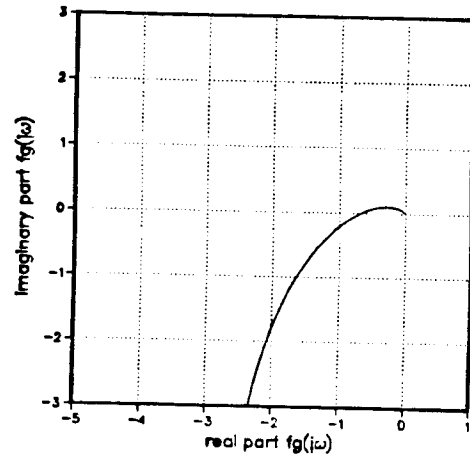


Figure 5.1: The Nyquist plot for Example 5.1.

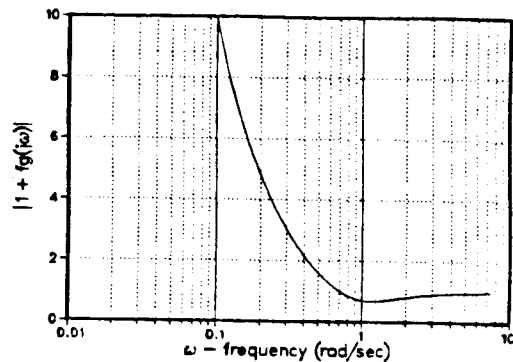


Figure 5.2: $\phi_2(j\omega)$ vs. ω for example 5.1.

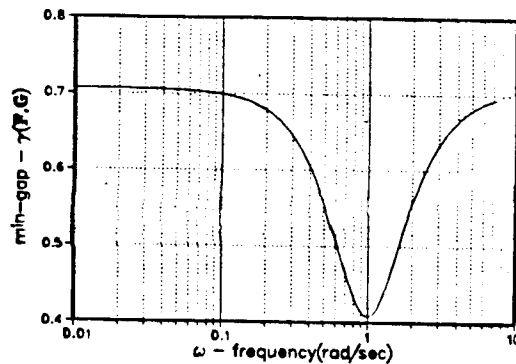


Figure 5.3: $\phi_3(j\omega)$ vs. ω for example 5.1.

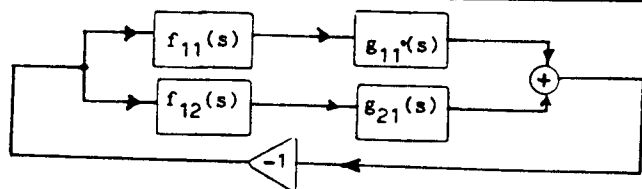


Figure 5.4: The Loop Breaking Configuration for Example 5.2.

where

$$\phi_3(s) = \frac{|g(s)f + 1|}{\sqrt{(1 + |g(s)|^2)(1 + |f|^2)}} \quad (5.2)$$

Proof: We consider a constructive approach based on the computational procedure given in section 3.6. We can represent the two subspaces alternately as,

$$\mathbf{G}(s) = \ker |g(s), -1| = \text{image} \begin{bmatrix} 1 \\ g(s) \end{bmatrix},$$

and

$$\mathbf{F} = \ker |1, f| = \text{image} \begin{bmatrix} -f \\ 1 \end{bmatrix}.$$

Thus we obtain normalized basis vectors for the pair of 1-dimensional subspaces as.

$$\mathbf{F} = \text{image} \frac{\begin{bmatrix} -f \\ 1 \end{bmatrix}}{\sqrt{1 + |f|^2}}$$

and for the orthogonal complement,

$$\mathbf{G}^{per} = \text{image} \frac{\begin{bmatrix} -\bar{g}(s) \\ 1 \end{bmatrix}}{\sqrt{1 + |g(s)|^2}}$$

Then (cf. (3.13)) from step 3 of the computational procedure we obtain

$$\gamma(\mathbf{G}(s), \mathbf{F}) = \sin \theta = \frac{\left| \begin{bmatrix} -\bar{g}(s) \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -f \\ 1 \end{bmatrix} \right|}{\sqrt{(1 + |g(s)|^2)(1 + |f|^2)}}$$

and the result follows.

Thus we see that $\phi_3(s)$ in this case involves a "normalized" return difference in order to make $0 \leq \phi_3(s) \leq 1$. In the general MIMO case ($p \neq m$) the procedure of orthonormalization is more complex. Computation of the matrix product reveals,

$$\begin{bmatrix} -G^*(s) \\ I_m \end{bmatrix}^* \begin{bmatrix} -F \\ I_p \end{bmatrix} = \begin{bmatrix} -G(s) & I_p \end{bmatrix} \begin{bmatrix} -F \\ I_p \end{bmatrix} = G(s)F + I_p.$$

Then as the computational procedure suggests, obtain QR factorizations,

$$\begin{bmatrix} -G^*(s) \\ I_m \end{bmatrix} = \begin{bmatrix} Q_G(s) & Z_G(s) \end{bmatrix} \begin{bmatrix} R_G(s) \\ 0 \end{bmatrix} = Q_G(s)R_G(s),$$

where for each s on D , $Q_G(s)$ is $(p+m) \times p$ and $R_G(s)$ a right triangular $p \times p$ matrix.

$$\begin{bmatrix} -F \\ I_p \end{bmatrix} = \begin{bmatrix} Q_F & Z_F \end{bmatrix} \begin{bmatrix} R_F \\ 0 \end{bmatrix} = Q_F R_F,$$

with Q_F is $(p+m) \times m$ and R_F a right triangular $m \times m$ matrix. Thus we can compare $\phi_2(s)$ and $\phi_3(s)$ by comparing the singular values of $Q_G^* Q_F$ with those of

$$G(s)F + I_p = R_G^*(s)Q_G^*(s)Q_F R_F. \quad (5.4)$$

(The relation (5.4) amounts to nothing more than an algebraic statement of the computational procedure used.)

Example 5.1: To illustrate these ideas we consider a simple SISO example. For some loop breaking let the loop transmission be

$$fg(s) = \frac{1}{s(s+1)(2s+1)}.$$

The resulting Nyquist contour for $s = j\omega$ with $\omega \in [0.25, 8.0]$ is displayed in Figure 5.1. The relevant euclidean distance between this curve and the critical point at $s = -1$ is given by $\phi_2(j\omega)$ which is displayed in Figure 5.2 giving $g_{sm} = 0.7$ occurring at $\omega^* = 1$ rad/sec. The curve $\phi_3(j\omega)$ is displayed in figure 5.3 giving $g_{sm} = 0.41$ at $\omega^* = 1$ rad/sec. The value $g_{sm} = 0.41$ suggests a minimum principle angle between

$$\ker \begin{bmatrix} 1 & 1 \end{bmatrix}$$

and

$$\ker \left[\frac{1}{s(s+1)(2s+1)}, -1 \right]$$

for $s = j\omega$ of $\theta = 24.2$ degrees. The significance of the asymptotic value $\phi_3(0) = 0.707$ (or $\theta = 45$ degrees) comes from the following observation. With reference to (2.1') for $g(s) = 1/s(s+1)(2s+1)$ we see that $g(s)$ has a pole at the origin. (Of course D would be appropriately indented to avoid this pole.) Thus $g(j\omega) \rightarrow \infty$ as $\omega \rightarrow 0$ which in the geometric picture of (2.1') implies that

$$\ker |g(j\omega), -1| \rightarrow \mathbf{Y}$$

in terms of the angle (gap) metric. Thus 45 degrees is just the angle between $\mathbf{U} = \ker |1, 0|$ and $\mathbf{F} = \ker |1, 1|$.

It is important to recognize that the geometric stability margin analysis we are discussing appropriately generalizes, from classical frequency domain analysis, the notion of "distance" between a Nyquist contour and a fixed point *without* employing the return difference. Horowitz has observed [HO1, HO2] that any physical feedback control system will involve *dynamic* compensation with a *dynamic* plant (i.e., $g(s), f(s) \in \mathbf{R}_{sp}(s)$). The complication for stability analysis involves the possible existence of *internal poles* of the resulting closed loop transfer function. Such internal poles exist due to possible cancellations in forming the loop transmission $F(s)G(s)$ when the McMillan degree of $F(s)G(s)$ is strictly less than the sum of the respective McMillan degrees of $F(s)$ and $G(s)$.

In terms of the geometric picture of feedback,

$$\begin{bmatrix} I_m & F(s) \\ G(s) & -I_p \end{bmatrix} \begin{bmatrix} u(s) \\ y(s) \end{bmatrix} = 0, \quad (5.5)$$

we see that we may have difficulty in estimating the regularity, $|\bullet|$, for all s on D of

$$\Sigma = \begin{bmatrix} I_m & F(s) \\ G(s) & -I_p \end{bmatrix}$$

as a map on $\mathbf{U} \oplus \mathbf{Y}$ in terms of a transformed basis,

$$\Sigma^2 = \begin{bmatrix} I_m + F(s)G(s) & 0 \\ 0 & I_p + G(s)F(s) \end{bmatrix}. \quad (5.6)$$

The basis given by the right hand side of (5.6) suggests we can estimate $|\Sigma|$ in terms of the regularity of an operator on \mathbf{U} or \mathbf{Y} . However, for s_1 in the neighborhood of an internal pole the transformation

$$\begin{bmatrix} I_m & F(s_1) \\ G(s_1) & -I_p \end{bmatrix}$$

will be poorly conditioned despite the fact that $I_m + F(s_1)G(s_1)$ and $I_p + G(s_1)F(s_1)$ may be relatively well conditioned. Moreover, this observation holds despite the fact that,

$$-\det \begin{bmatrix} I_m & F(s) \\ G(s) & -I_p \end{bmatrix} = \det [I_m + F(s)G(s)] = \det [I_p + G(s)F(s)].$$

We remark that $|\det(\bullet)|$ (in contrast to the matrix infimum, $|\bullet|$) is *not* a measure of regularity which admits any useful perturbation analysis [DA1, DO2-4].

Before we consider the question of dynamic compensation further we discuss some aspects of the geometric theory of linear systems with respect to a characterization of open loop poles. First, we recognize that the map

$$s \rightarrow \ker [G(s), -I_p]$$

is well defined on the domain $\mathbf{C} \cup \{\infty\} - \{p_{oi}\}$, where $\{p_{oi}\}$ is the set of n open loop poles of $G(s) \in \mathbf{R}^p \times \mathbf{R}^m(s)$. More generally, we can consider the equation

$$y(s) = G(s)u(s)$$

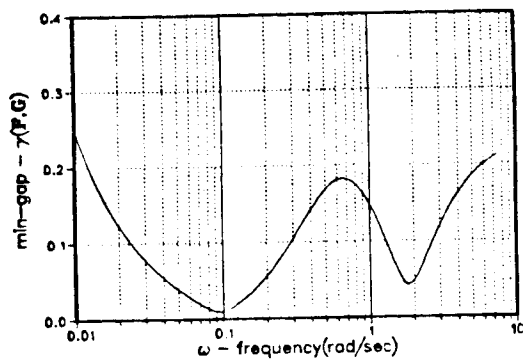


Figure 5.5: $\phi_3(j\omega)$ vs. ω for example 5.2.

- [RO1] H. H. Rosenbrock, *Computer-Aided Control System Design*, Academic Press, Inc., London, 1974.
- [SA1] M. G. Safonov, A. J. Laub, and G. L. Hartman, "Feedback Properties of Multivariable Systems: The Role and Use of the Return Difference Matrix", *IEEE Trans. on Automatic Control*: Special Issue on Linear Multivariable Control Systems, Vol. AC-26, No.1, 1980.
- [ST1] G.W. Stewart, "Error and Perturbation Bounds for Subspaces Associated with Certain Eigenvalue Problems," *SIAM Review*, Vol. 15, No. 4, Oct. 1973.

7. References

- [BE1] W.H. Bennett, "Decentralized Control System Design: Dynamic Weak Coupling from a Geometric Viewpoint", *Ph.D. Dissertation Univ. of Md.*, 1984.
- [BE2] W.H. Bennett and J.S. Baras, "A New Geometric Stability Margin for Multivariable Feedback Systems", submitted for publication, *Systems Research Center Tech Rep. No. TR 85-15*, University of Maryland, August, 1985.
- [BJ1] A. Björck and G. H. Golub, "Numerical Methods for Computing Angles Between Linear Subspaces," *Math of Comp.*, Vol. 27, No. 123, July 1973, pp. 579-594.
- [BR4] R.W. Brockett and C.I. Byrnes, "Multivariable Nyquist Criteria, Root Loci, and Pole Placement: A Geometric Viewpoint", *IEEE Trans. on Auto. Control*, Vol. AC-26, No.1, Feb.1981, pp. 271-283.
- [BY1] C.Byrnes, M.Hazewinkel, C.Martin, & Y.Rouchaleau, "Introduction to Geometrical Methods for the Theory of Linear Systems", in *Geometric Methods for the Theory of Linear Systems*, D.Reichel, Dordrecht, 1980.
- [DA1] G. Dahlquist and A. Björck, *Numerical Methods*, Prentice Hall, Inc., 1974.
- [DO1] J.J.Dongarra, C.B.Moler, J.R.Bunch, & G.W.Stewart, *LINPACK Users Guide*, SIAM Press, 1979.
- [DO2] J.C. Doyle, "Robustness of Multiloop Linear Feedback Systems", *IEEE Conf. on Dec. & Control*, Jan. 1979, San Diego, CA., pp. 12-18.
- [DO3] J.C. Doyle and G. Stein, "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis", *IEEE Trans. on Auto. Control*, Vol. AC-26, No. 1, pp. 4-16.
- [HE1] J. W. Helton, "An H^∞ Approach to Control", *Proc. IEEE Control Decision Conf.*, Dec. 1983, San Antonio, TX., pp. 607-611.
- [HE2] R. Hermann and C.F. Martin, "Application of Algebraic Geometry to Linear Systems Theory", *IEEE Trans. on Auto. Control*, Vol. AC-22, 1977, pp.19-25.
- [HO1] I. M. Horowitz, *Synthesis of Feedback Systems*, Academic Press, 1963.
- [HO2] I. M. Horowitz and U. Shaked, "Superiority of Transfer Function Over State-Variable Methods in Linear Time-Invariant Feedback System Design", *IEEE Trans. on Auto. Control*, Vol. AC-20, No. 1, Feb. 1975, pp. 84-97.
- [KA1] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, 1976.
- [LE1] N.A.Lehtomaki, N.R.Sandell, & M.Athans, "Robustness Results in Linear-Quadratic Gaussian Based Multivariable Control Designs", *IEEE Trans. on Auto. Control*, Vol. AC-26, No.1, Feb.1981, pp.75-92. Vol. AC-20, No. 1, pp. 32-46.