Scientific Report No. 9
Control Theory Group

THE UNIVERSITY OF TENNESSEE

DEPARTMENT OF ELECTRICAL ENGINEERING

GEOMETRICAL INTERPRETATION AND GRAPHICAL SOLUTION TO MINIMUM ENERGY DISCRETE-DATA CONTROL

by

A. M. Revington and J. C. Hung

Supported by National Aeronautics and Space Administration
through Grant NsG-351

June 15, 1965

Scientific Report No. 9

Control Theory Group

GEOMETRICAL INTERPRETATION AND GRAPHICAL SOLUTION TO MINIMUM ENERGY DISCRETE-DATA CONTROL

by

A. M. Revington and J. C. Hung

Supported by National Aeronautics and Space Administration
through Grant NsG-351

June 15, 1965

Department of Electrical Engineering
The University of Tennessee

GEOMETRICAL INTERPRETATION AND GRAPHICAL SOLUTION TO MINIMUM ENERGY DISCRETE-DATA CONTROL

by

A. M. Revington and J. C. Hung

13558

SUMMARY

Compact and simple design equations have been previously obtained for the minimal energy design of a linear PAM regulator by working not in the state space, but in a new space, the "canonical space." The input sequence is split into two sections and the minimal energy condition is a simple relation between the two parts of the input. The relation is used here to give a direct method of finding what states can be taken to the origin, using the linearly designed input sequence, without violating a saturation constraint. For second order systems the technique is particularly useful in that a simple graphical technique is possible. An example demonstrates the method. author

INTRODUCTION

For a number of years the Z-transform was the only technique available to engineers to handle discrete regulators. This technique has many disadvantages when applied to the regulator problem, but it has the advantage that it is closely related to the familiar Laplace transform, and the root-locus and frequency domain methods can still be used. The introduction of state variable techniques has considerably clarified the analysis and design of both discrete and continuous systems, but unfortunately a knowledge of transform methods is not much help in understanding the state variable techniques[1]. However, using only elementary matrix algebra, remarkably compact results and a simple design procedure have been obtained for the pulse amplitude modulated (PAM) regulator[2,3].

It is well known[4,5] that for an n-th order linear PAM plant regulation can be achieved in, at most, n sampling periods. If the number of sampling periods used is increased beyond n then in general there are many possible input sequences that will regulate the plant. The problem of which sequence to choose and how to obtain it has received much attention; the usual approach is to make the system minimize some cost function while performing the primary task of regulation. We choose the cost function E given by

$$E = T \sum_{k=1}^{N} u^2(k). \qquad (1)$$

T is the sampling period, N the number of sampling periods used for regulation and u(k) is the input over the k-th sampling period. This cost function has several advantages. The system is intrinsically unlikely to saturate; it can be shown for stable systems that if N is large enough the linear design equations give an input sequence that enables the plant to operate in its linear region. The sequence is easy to calculate and is unique. Finally, the minimization of E gives a good approximation to the minimum fuel problem where the cost function is

$$T \sum_{k=1}^{N} |u(k)|$$

While the use of the linear design equations gives an input sequence which, because of the nature of the cost function E, is least likely to saturate the plant we would certainly like to know just when we can take advantage of the simplicity of the linear design equations. In other words for what set of initial states does a linear design give an input sequence satisfying a given saturation constraint M, that is $|u(k)| \leq M$, k=1, 2, ... N? Furthermore, suppose N sampling periods are not sufficient to allow a linear design then can we increase N and, if so by how much, so that the linearly designed sequence satisfies the saturation constraint?

In the next section the development of the linear design equations, using the "canonical vector space," is summarized. Then the set M_N is defined and generated via an intermediate set L_N. If the initial state is in M_N the linear design equations can be used.

MATHEMATICAL DEVELOPMENT

To keep the development as simple as possible we shall consider only a single input time-invariant plant. In discrete form the dynamics of such an n-th order system are described by the vector difference equation[6],

$$\underline{x}(k+1) = G(T)\underline{x}(k) + \underline{h}(T)u(k) \tag{2}$$

where $\underline{x}(k)$, an n-vector, is the state of the plant at the k-th sampling instant, G(T) is the n x n transition matrix and $\underline{h}(T)$ is the nx1 driving matrix.

Linear Design Equations

The disturbed state or initial condition x(0) can be represented[5] by

$$\underline{x}(0) = \sum_{k=1}^{N} \underline{r}_k \; u(k) \tag{3}$$

where $\underline{r}_k$ is the k-th "canonical vector" given by

$$\underline{r}_k = -G(-kT)\underline{h}(T) \tag{4}$$

The regulator problem is to take $\underline{x}(0)$ to the reference $\underline{x}(N) = 0$ in $N > n$ sampling periods while minimizing E. The solution to the problem is considerably simplified if, instead of working with the n-dimensional state space X, we work in an n-dimensional "canonical space" C. The basis vectors for points in C are the canonical vectors $\underline{r}_1, \underline{r}_2, \ldots \underline{r}_n$. Then corresponding to $\underline{x}$ in X we have $\underline{c}$ in C. The nonsingular linear transformation taking $\underline{c}$ in C to $\underline{x}$ in X is R, where

$$R = \left[\underline{r}_1, \underline{r}_2, \ldots \underline{r}_n\right]; \tag{5}$$

R^{-1} takes $\underline{x}$ in X to $\underline{c}$ in C. Furthermore, the input sequence is divided into two sections; thus

$$\underline{a}^t = \left[u(1), u(2), \ldots u(n)\right] \tag{6}$$

$$\underline{b}^t = \left[u(n+1), u(n+2), \ldots u(N)\right],$$

where t denotes the transpose of a matrix. Then Eq. (3) reduces to

$$\underline{c} = \underline{a} + H\underline{b} \tag{7}$$

where H, an nxN-n matrix, has as its columns the remaining N-n canonical vectors expressed in C: Eq. 1 is now

$$E = T\left[\underline{a}^t \underline{a} + \underline{b}^t \underline{b}\right]. \tag{8}$$

The condition that must be satisfied for regulation with minimum E is easily found to be[2]

$$\underline{b} = H^t \underline{a} \tag{9}$$

Eq. (7) and Eq. (9) give immediately the optimal input sequence $\underline{a}^o$ and $\underline{b}^o$ as

$$\underline{a}^o = \left[I + HH^t\right]^{-1} \underline{c} \tag{10}$$

$$\underline{b}^o = H^t \underline{a}^o \tag{11}$$

where I is the n x n unit matrix. Finally the least energy E^o is given by (letting T = 1)

$$E^o = \underline{c}^t \underline{a}^o \tag{12}$$

The Problem of Saturation

Normalizing the input sequence so that the saturation constraint takes the form

$$|u(k)| \leqslant 1\ , \ k=1,\ 2,\ \ldots N \tag{13}$$

We have

$$\begin{aligned} |a_i| &\leqslant 1 \quad i=1,\ 2,\ \ldots n, \\ |b_j| &\leqslant 1 \quad j=1,\ 2,\ \ldots N-n, \end{aligned} \tag{14}$$

where a_i and b_j are the components of $\underline{a}$ and $\underline{b}$.

Kalman[5] and others have considered the set Γ_N of all states $\underline{x}$ in X that can be taken to the origin in N sampling periods or less, subject to input saturation.

$$\Gamma_N = \left(\underline{x} \;\middle|\; \underline{x} = \sum_{i=1}^{N} \alpha_i \underline{r}_i ; \quad |\alpha_i| \leq 1,\ i=1,\ 2, \ldots N \right) \tag{15}$$

Read "Γ_N is the set of all states $\underline{x}$, where $\underline{x}$ is given by $\underline{x} = \sum_{i=1}^{N} \alpha_i r_i$ and $|\alpha_i| \leqslant 1$, $i=1,\ 2,\ \ldots N$." The α_i are scalars.

In the canonical space C, Γ_N becomes

$$\Gamma_N = \left(\underline{c} \;\middle|\; \underline{c} = \underline{a} + H\underline{b} ; \quad |a_i| \leqslant 1\ ,\ |b_j| \leqslant 1 \right) \tag{16}$$

Figures 1(a) and 1(b) below illustrate Γ_3 in X and C respectively for a second order system.

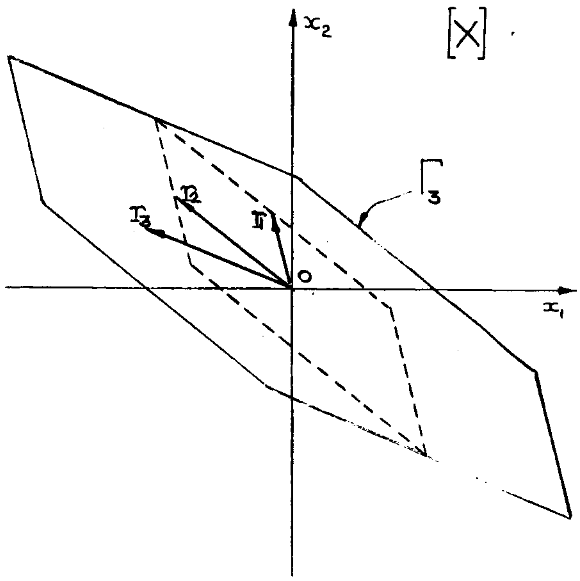

Fig. 1(a)

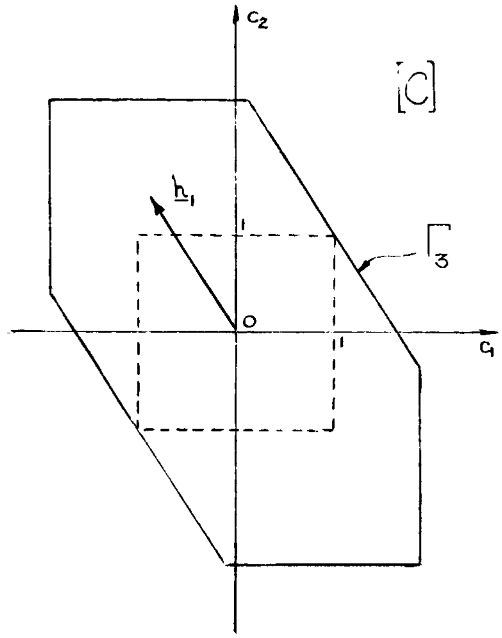

Fig. 1(b)

On the boundary of Γ_N there is a unique control sequence for regulation in N sampling periods[6]. In other words there is only one way to linearly combine the canonical vectors $\underline{r}_1, \underline{r}_2, \ldots \underline{r}_N$ to reach an initial state on the boundary of Γ_N. For an initial state inside Γ_N there are an infinite number of ways to represent the initial state. From these we have chosen that unique combination that minimizes the energy E. To find out just when the linear design can be applied (giving a non-saturating control sequence $|u(k)| \leqslant 1$, k=1, 2, ...N) we define the set M_N.

<u>Definition</u>. The set of initial states that can be taken to the origin in exactly N sampling periods with minimal energy $E^o = \underline{c}^t \underline{a}^o$ and satisfying the saturation constraint is called M_N.

Thus in the canonical space,

$$M_N = \left(\underline{c} \;\middle|\; \underline{c} = \underline{a} + H\underline{b},\ \underline{b} = H^t \underline{a};\ |a_i| \leqslant 1,\ |b_j| \leqslant 1 \right) \quad (16)$$

In general M_N is a proper subset of Γ_N. This is because of the extra condition $\underline{b} = H^t \underline{a}$ that is added to Γ_N to give M_N. To be able to discuss M_N we need the term "free minimal energy input sequence."

Definition. The vectors $\underline{a}^o$ and $\underline{b}^o$, calculated from Eq. 10 and Eq. 11 giving the minimal energy $E^o = \underline{c}^t \underline{a}^o$, constitute the "free minimal energy input sequence," $u^o(k)$, k=1, 2, ...N, i.e., the amplitude is not constrained.

In this sequence we may well have $|u^o(k)| > 1$ for some value of k. By definition therefore the set M_N contains only those initial states whose free optimum input sequences satisfy the saturation constraint. States in Γ_N but not in M_N have free minimal energy input sequences that violate the saturation constraint, so that if the input sequence is constrained by saturation, the energy needed for regulation is greater than E^o. The problem of finding this constrained input sequence, one or more members of the input sequence need to be forced back to the saturation limit, is not directly considered here.

THE GENERATION OF M_N

From Eq. 10 and Eq. 11 there are N equations of the form

$$\sum_{i=1}^{n} \alpha_{ji} c_i = u(j) \ , \quad j=1,\ 2,\ \ldots N \tag{17}$$

where the c_i are the n components of the initial state $\underline{c}$ and the α_{ji} are scalars. We could then set $u(j) \geqslant +1$ and $u(j) \leqslant -1$ to obtain 2N n-dimensional half spaces in C-space. The convex set formed by the intersection of these half spaces would then be M_N. This method is quite correct but fails to give a clear picture of what is happening. The set M_N can be found quite simply via consideration of the intermediate set L_N which will be developed in the following paragraphs.

Consider Eq. 9, which can be rewritten as

$$b_j^o = h_j^t a^o,\ j = 1,\ 2,\ \ldots\ N\text{-}n, \tag{18}$$

where $\underline{h}_j$ is the j-th column vector of H and b_j^o is the j^{th} component of $\underline{b}^o$, j = 1, 2, ...N-n. The vectors $\underline{h}_j$ are the canonical vectors $\underline{r}_{n+j}$ expressed in C-space ($\underline{h}_1$ is shown in Fig. 1b.). So that

$$\underline{h}_j = R^{-1}\underline{r}_{n+j} \ . \tag{19}$$

Eq. 18 represents N-n n-dimensional hyperplanes, normal to $\underline{h}_j$ in C-space. It should be noted that the vector $\underline{a}^o$ does not correspond to the initial state $\underline{c} = \underline{a}^o$ (unless N=n) even though the coordinates of a point $\underline{c}$ in C are usually considered to be the components of some initial state $\underline{c}$. The coordinates of C-space are considered to have a dual meaning for the generation of M_N; they can represent the components of an initial state $\underline{c}$ or the components of the input vector $\underline{a}^o$. To demonstrate this consider a second order system (n=2) with $\underline{h}_1$ as shown in Fig. 2. Consider the straight line $\underline{h}_1^t\ \underline{a}^o = b_1$. For $b_1 = 0$ the line passes through the origin, normal to $\underline{h}_1$. For $b_1 > 0$ it moves parallel to itself in the direction of + $\underline{h}_1$ and for $b_1 < 0$ in the direction $-\underline{h}_1$. When $|b_1| = 1$ the perpendicular distance from the line to the origin is $1/\|h_1\|$. $\|h_1\|$ is the Euclidean norm or the length of $\underline{h}_1$. These properties hold for $\underline{h}_j$ in general, but are most useful when n = 2 so that hyperplanes are straight lines.

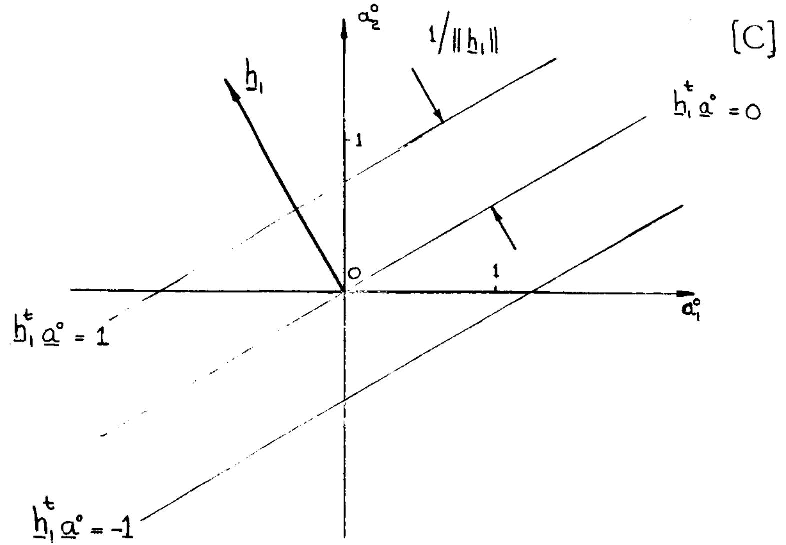

Fig. 2

The admissible (non-saturating) set of a_1^o and a_2^o lie in or on the square shown in Fig. 3. For $|b_1^o| \leqslant 1$ the $\underline{a}^o$ must also lie in the space between or on the lines,

$$\pm 1 = b_1^o = \underline{h}_1^t \, \underline{a}^o ,$$

so that any $\underline{a}^o$ that is to satisfy $|a_i^o| \leqslant 1$ and $|b_1^o| \leqslant 1$ must lie in the cross-hatched region, which we call L_3.

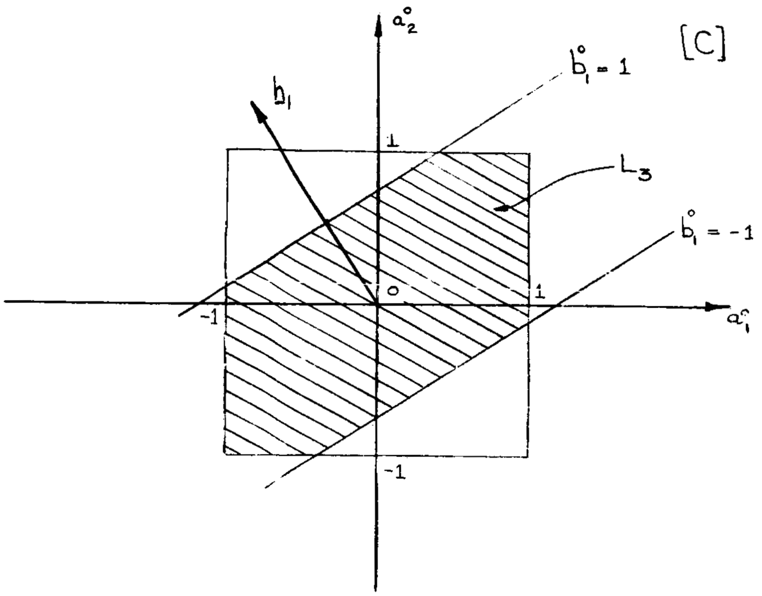

Fig. 3

When N = 4 we consider also b_2^o . The set of admissible $\underline{a}^o$ now must be the intersection of L_3 and the space between the lines

$$\pm 1 = b_2^o = \underline{h}_2^t \, \underline{a}^o .$$

This is shown in Fig. 4 and is called L_4.

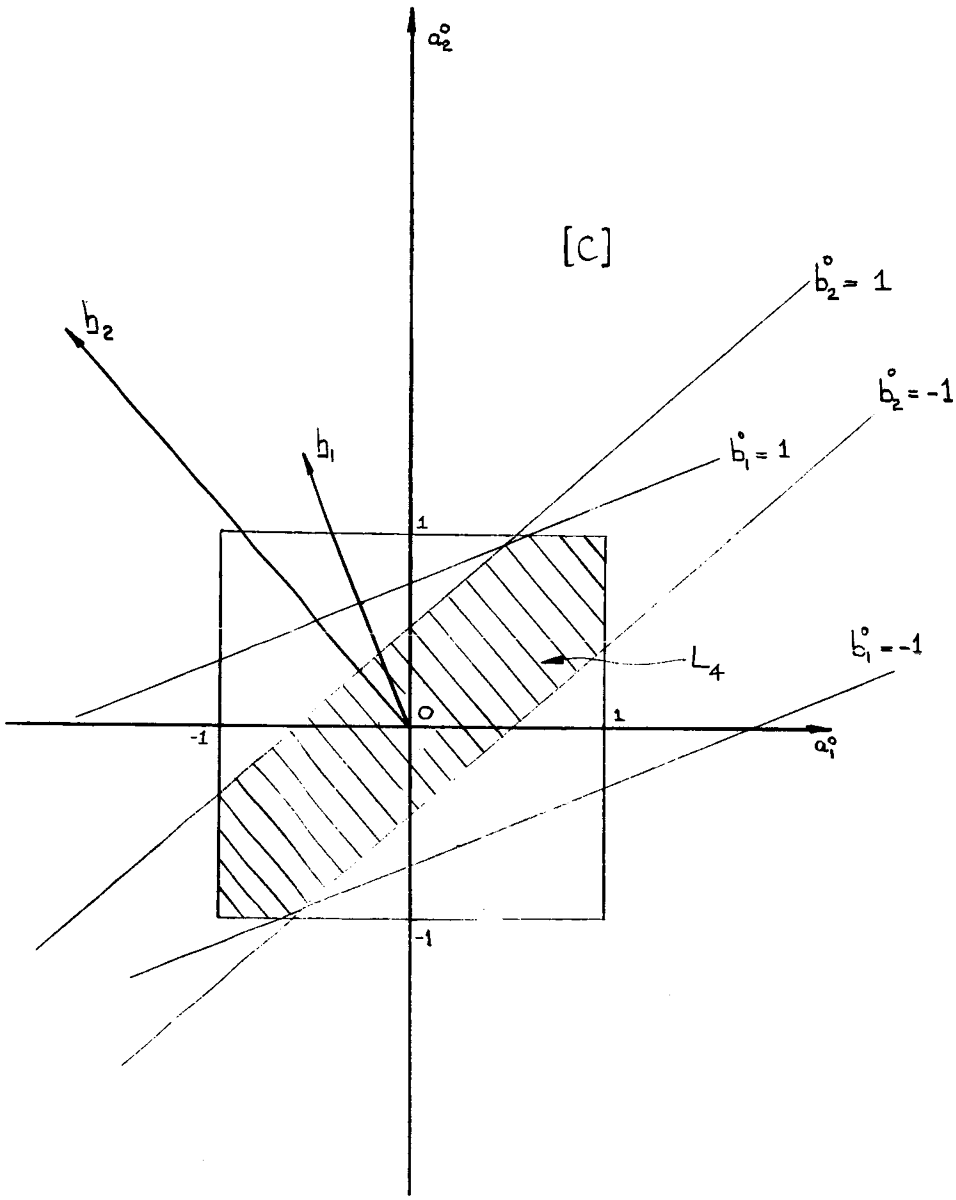

Fig. 4

For the general case the set L_N is defined.

Definition. The set L_N is defined as the intersection of the 2n half-spaces $a_i^o \leqslant 1$, $a_i^o \geqslant -1$ and the 2(N-n) half-spaces $\underline{h}_j^t \underline{a}^o \leqslant 1$, $\underline{h}_j^t \underline{a}^o \geqslant -1$, i = 1, 2, ...n, j = 1, 2, ..,N-n in the n-space with Cartesian coordinates $a_1^o, a_2^o, .., a_n^o$.

Clearly L_N is convex and L_{N+1} is a subset of L_N. The reason for defining L_N is twofold.

(a) We only need to calculate $\underline{a}^o$, if we have L_N, to know if the free minimal energy input sequence exceeds the saturation constraint. In other words if $\underline{a}^o$ is in L_N then $\underline{c}$ is in M_N.

(b) We can generate M_N directly from L_N without inverting any matrices.

Generation of M_N from L_N

There is clearly a one to one correspondence between points $\underline{a}^o$ in L_N and initial states $\underline{c}$ in M_N. In particular, if $\underline{a}^o$ is on the boundary of L_N then $\underline{c}$ is on the boundary of M_N. Thus by moving along the boundary of L_N we trace out the boundary of M_N. We have

$$\underline{c} = \underline{a}^o + Hb^o$$

$$= \underline{a}^o + b_1^o h_1 + b_2^o \underline{h}_2 + \ldots\ldots + b_{N-n}^o \underline{h}_{N-n} \qquad (20)$$

Consider the point A at a vertex of L_3 in Fig. 5. Adding $\underline{h}_1$ alone to this $\underline{a}^o$ gives a vertex of M_3, A', defined by the intersection of the lines $b_1^o = 1$, $a_2^o = 1$. Similarly B in L_3 gives B', an initial state, on the boundary of M_3 with $b_1^o = 0$ and $a_2^o = -1$. (We added nothing to $\underline{a}^o$ at B.) Other points on M_3 can be obtained just as easily.

We note that in Fig. 5 the axes are now labelled c_1 and c_2, the coordinates of the initial state. Thus for example the lines $b_1^o = \pm 1$ are now the bounding lines of M_3 (see Fig. 3). Strictly speaking, while generating L_3 we should label the axes a_1^o and a_2^o; but while generating and using M_3 we should use c_1 and c_2.

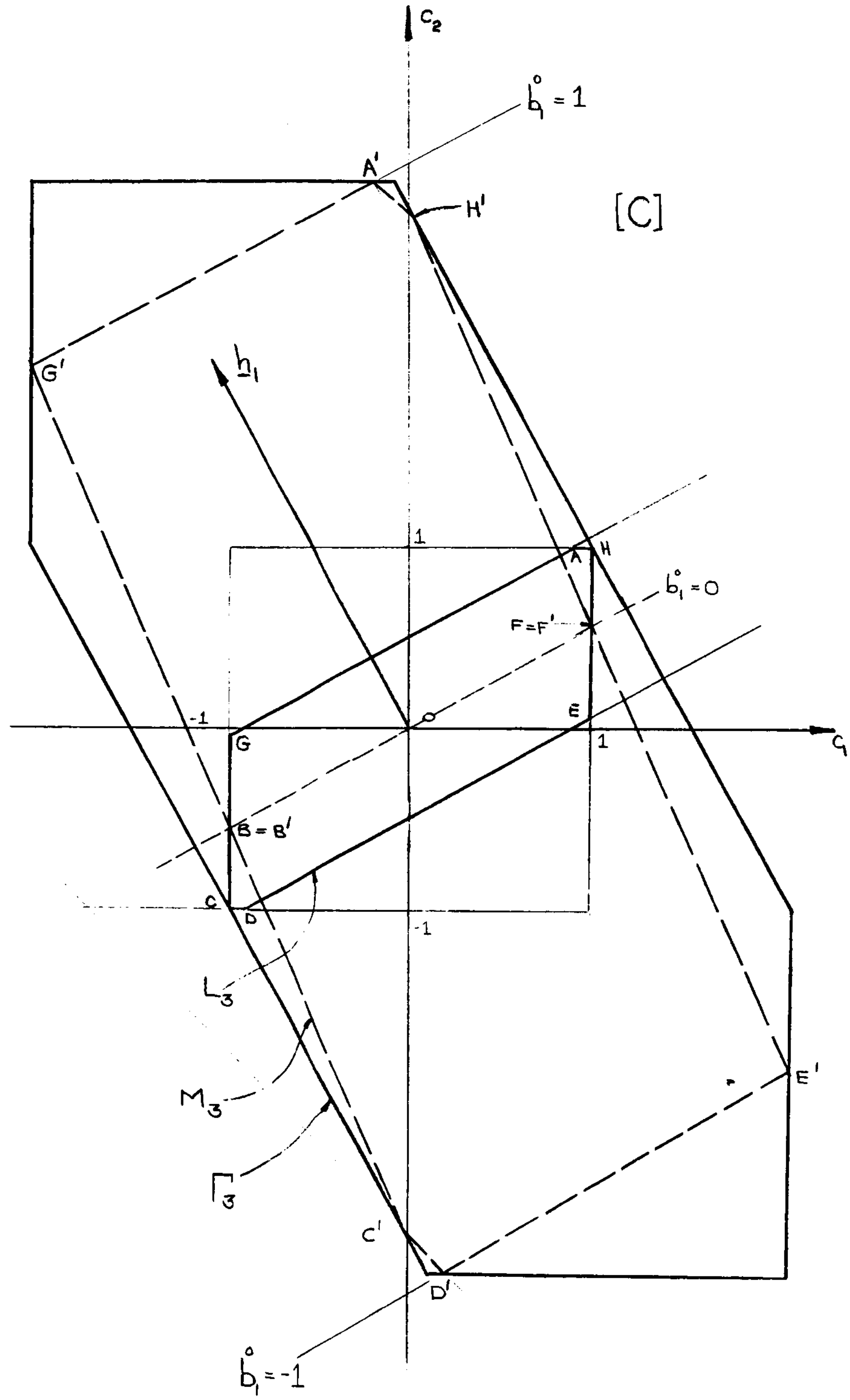

Fig. 5

The set M_3 has been obtained by using only a set square and dividers. Points A through H in L_3 correspond to points A' through H' in M_3.

In general M_N can be found in exactly the same way from the set L_N. For $n > 2$ simple graphical techniques are no longer possible. For stable systems $\|h_i\| \geqslant \|h_{i-1}\|$, i = 1, 2, ... N, so that, for N sufficiently large, any initial state can be taken to the origin with a non-saturating free optimum input sequence.

We have shown how to generate M_N from L_N. Geometrical properties of L_N give the geometrical properties of M_N. The vertices of L_N correspond to the vertices of M_N and the bounding faces of L_N correspond to the bounding faces of M_N.

An example is now given for a second order system, which will demonstrate several particular uses of L_N.

EXAMPLE

Consider the pure inertial plant $1/s^2$. With a sampling period of T secs the vector difference for this system is

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} \frac{T^2}{2} \\ T \end{bmatrix} u(k) \qquad (21)$$

where $x_1(k)$ is the output (position) at the k-th sampling instant and $x_2(k)$ is the time derivative of the output (velocity) at the k-th sampling instant. We soon find

$$\underline{r}_k = \begin{bmatrix} \frac{T^2}{2}(1-2k) \\ T \end{bmatrix} \quad , k = 1, 2, \ldots . \qquad (22)$$

and then

$$R = \begin{bmatrix} -\frac{T^2}{2} & -\frac{3}{2}T^2 \\ T & T \end{bmatrix} \tag{23}$$

giving

$$\underline{h}_p = \begin{bmatrix} -p \\ p+1 \end{bmatrix} \qquad p = 1, 2, \ldots N-n. \tag{24}$$

For this plant, the canonical vectors in C are independent of the sampling period.

The lines $\underline{h}_p^t \, \underline{a}^o = b_p^o$ are shown for $b_p^o = 0, +1, -1$, p = 1, 2, 3 in Fig. 6. For $b_p^o = \pm 1$ these lines intersect $a_1^o = \pm 1$ at $a_2^o = \pm 1$; they intersect $a_1^o = \pm 1$ at $a_2^o = \pm \frac{p-1}{p+1}$.

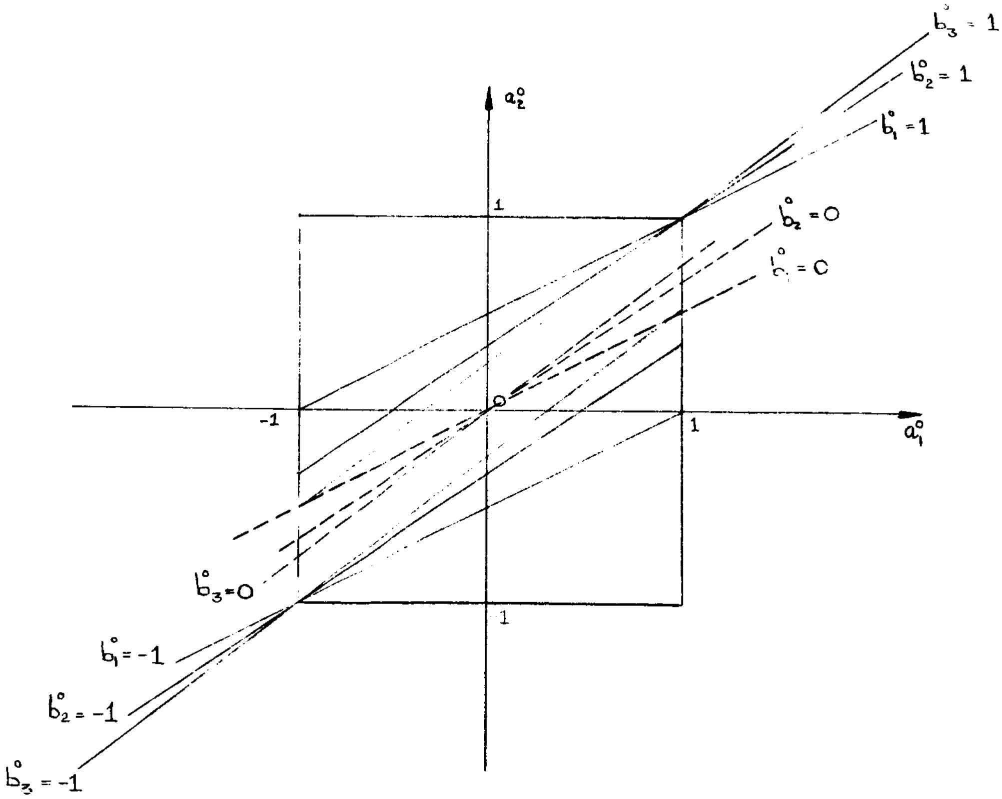

Fig. 6

L_3 is bound by $a_1^o = \pm 1$ and $b_1^o = \pm 1$. In general L_{p+2} is bounded by $a_1^o = \pm 1$ and $b_p^o = \pm 1$. Then we know that M_{p+2} has only 4 vertices and is bounded by the lines $a_1^o = \pm 1$ and $b_p^o = \pm 1$. Thus if $N = p+2$, and if the first and last inputs are within the saturation limit then all the other inputs are also with the saturation limit.

M_3 and M_4 are shown in Fig. 7. Clearly M_3 is not a subset of M_4, even though M_4 covers a large portion of Γ_3 that M_3 did not. Comparing the areas of M_3 and Γ_3, M_4 and Γ_4 we see that a large proportion of initial states in Γ_3 and Γ_4 are included in M_3 and M_4. This will remain true as p is increased.

The lines of Eq. 17 can be found from L_{p+2} by finding two points on the lines $a_i^o = \pm 1$, i=1, 2, ...n, or $b_j^o = \pm 1$, j = 1, 2, ... N-2. These lines all intersect at the vertex of Γ_N given by

$\underline{x} = \sum_{i=1}^{N} \underline{r}_i$, but it is evident that only the first and last, a_1^o and b_{N-2}^o pass into Γ_N. Then we have the following results for this plant.

1. A sufficient condition for $\underline{x}$ to be in Γ_N is that a_1^o and b_{N-2}^o be non-saturating (since $\underline{x}$ is in M_N).

2. A necessary condition for $\underline{x}$ to be in Γ_N is that none of the inputs a_2^o, b_j^o, j = 1, 2, ... N-3 saturate.

N.B. If a_1^o and b_{N-2}^o saturate we may still have $\underline{x}$ in Γ_N, see for example Fig. 7.

These properties of M_N, even though for a particularly simple plant, do illustrate how the set $L_N(M_N)$ can be useful in attempting to solve the saturation problem in this indirect way.

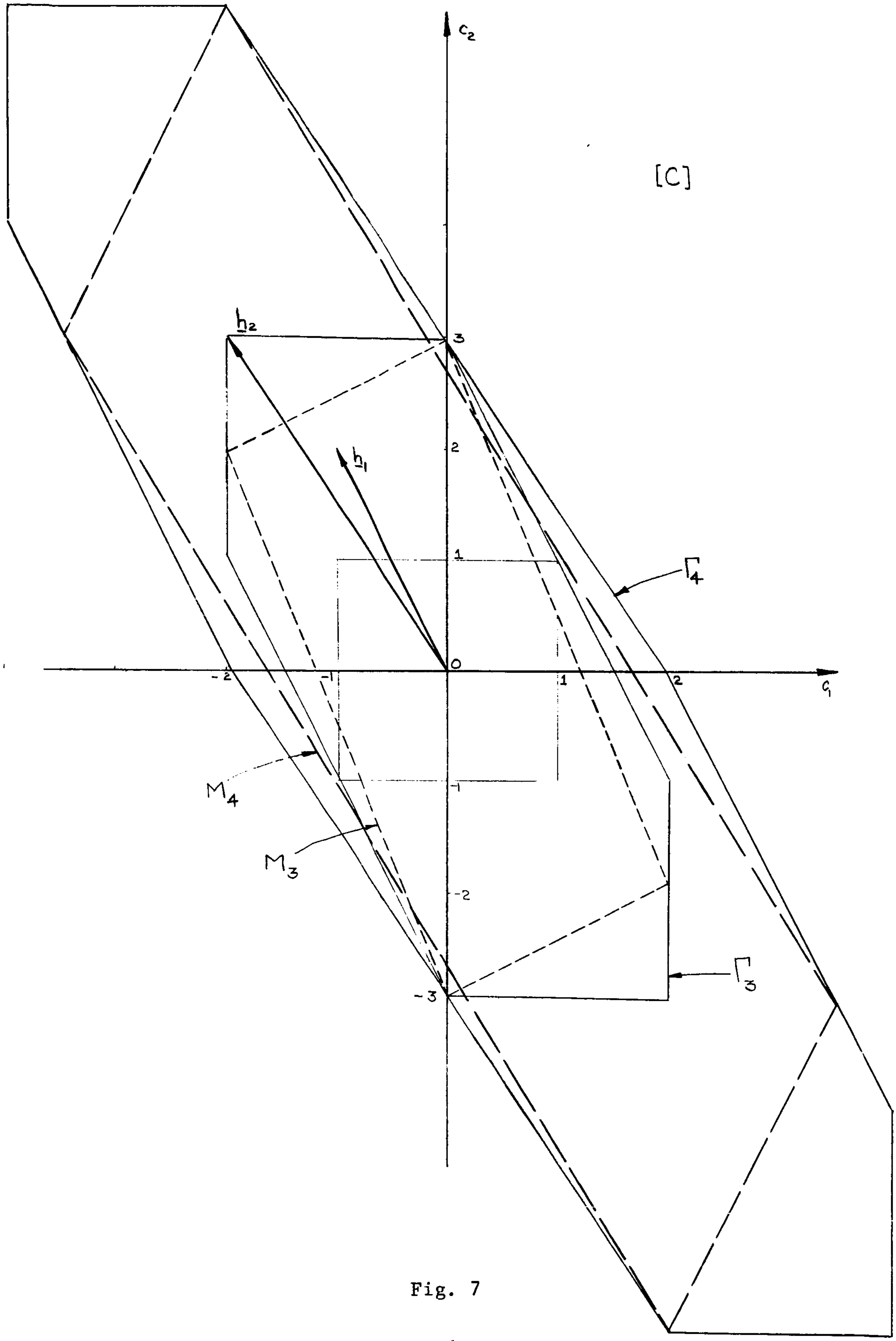

Fig. 7

CONCLUSIONS

The simple and compact design equations previously obtained for the minimal energy linear PAM regulator have been reviewed. It was shown that if the input sequence is split into two components these two components are related by a linear transformation. Using this property, a method has been presented to find those states, M_N, in Γ_N that can be taken to the origin using only the linear design equations. While the geometrical concepts and properties of this method are true in general, for second order systems the technique can be used to find these states graphically.

ACKNOWLEDGMENT

The work in this report was supported by the National Aeronautics and Space Administration through Grant NsG-351.

REFERENCES

1. Zadeh, L. A., and C. A. Desoer, "Linear System Theory," McGraw-Hill Book Co., Inc., New York, 1963.

2. Revington, A. M., and J. C. Hung, "Design of Minimum Energy Discrete-Data Control Systems," Scientific Report No. 7, Control Theory Group, Dept. of Electrical Engineering, University of Tennessee, January 1965.

3. Revington, A. M., and J. C. Hung, "Design of Time Weighted Minimum Energy Discrete-Data Control Systems," Scientific Report No. 8, Control Theory Group, Dept. of Electrical Engineering, University of Tennessee, February 1965.

4. Kalman, R. E., and J. E. Bertram, "General Synthesis Procedure for Computer Control of Single and Multiloop Linear Systems," Trans. AIEE, Vol. 77, Part II, 1958, pp. 602-609.

5. Kalman, R. E., "Optimal Nonlinear Control of Saturating Systems by Intermittent Action," IRE WESCON Convention Record, Vol. 1, pg. 4, 1957, pp. 130-135.

6. Desoer, C. A., and J. Wing, "An Optimal Strategy for a Saturating Sampled-Data System," IRE Trans. on Automatic Control, Vol. AC-6, February, 1961, pp. 5-15.