Automatic Frequency-Domain Synthesis of Multiloop Control Systems

T. C. COFFEY*
The Aerospace Corporation, El Segundo, Calif.

A computerized algorithm to facilitate the automatic synthesis of time-invariant linear compensation for highly complex multiloop control systems is discussed. Performed in the frequency domain to attain desired open-loop frequency response characteristics, the algorithm is applicable equally to continuous-time systems in the S domain and sampled data systems in the W or Z domains. It is executed by a digital computer program, AUTO, which uses a gradient-search algorithm to select the coefficients of multiloop feedback compensation transfer functions. These coefficients provide an open-loop frequency response, optimum in the sense that its deviation from the one desired is minimal in the weighted least-squares sense. The selection of a method for incrementing compensation parameters that alleviates "ridge following" convergence problems is discussed, as are the geometric properties of the least-squares cost function itself. Examples of cost function geometries are presented to provide insight into their nature.

1. Introduction

THE means by which a digital computer program can be generated to implement conventional frequency domain, root locus, and time domain analyses of complex linear systems have been discussed in the literature.[1–3] However, even with such design aids, the engineer who wants his system to possess desirable frequency response characteristics must intervene and perform the task of selecting the compensation. This compensation must mold and blend the frequency responses from the various sensors attached to the plant into a total open-loop frequency response that agrees satisfactorily with the one desired.

A typically complex system, such as that of a ballistic missile, has a great many dynamic modes of oscillation distributed over a broad range of frequencies. Usually, these modes are such that the compensation that gives the best frequency response over one range of frequencies may be detrimental to the compensation of other modes. The problem is further complicated, for with multiloop feedback systems, the frequency responses of the various loops may have radically different phase relationships at different frequencies.

Described here is a means by which a digital computer can be programed in a straightforward manner to automate the synthesis of complex multiloop control systems. The particular program discussed, AUTO, applies a technique, known as gradient parameter optimization, to the problem of frequency-domain design.

AUTO is a digital program, which has as its inputs the frequency profiles of the various quantities that must be compensated, the desired open-loop frequency response, and the orders of the numerators and denominators of the transfer functions to be used as compensation for the feedback loops. AUTO is designed to synthesize compensation in the Laplace domain for continuous systems or in the w domain for sampled systems. AUTO has been used with great success on both types of systems.

Presented as Paper 69-941 at the AIAA Aerospace Computer Systems Conference, Los Angeles, Calif., September 8–10, 1969; submitted September 4, 1969; revision received March 4, 1970. The writer wishes to acknowledge the helpful suggestions in formulating and programing AUTO supplied by K. E. Hagen and R. J. Murabayashi of The Aerospace Corporation.

* Manager, Control Systems Analysis Section, Electronics Division.

The mathematical formulation of AUTO for continuous time systems will be considered here and simple modifications that extend its use to discrete time systems will be indicated. An illustration is given of the way in which AUTO is utilized to design w plane compensation for a ballistic missile having a digital autopilot.

2. Automatic Frequency-Domain Synthesis Problem

The program, AUTO, utilizes a gradient optimization technique[4,5] to implement the automatic frequency-domain synthesis of multiloop time-invariant (and constant-rate sampled) linear systems. The problem of automatic frequency-domain synthesis and its formulation to facilitate the use of the gradient optimization technique will be described.

Figure 1 is a block diagram of a typical unforced, continuous-time, multiloop control system. The plant $P(s)$ describes the dynamics of the portion of the system considered fixed in the sense that it is not subject to alteration by the controls engineer. It is assumed that the system equations have been written so that the plant has multiple outputs $x_j(s)$, but only one input $z(s)$; and also that the common loop is broken just prior to the plant, so that an open-loop transfer function

$$\hat{Y}(s) = \hat{y}(s)/z(s) \tag{1}$$

can be defined. If $z(s)$ is taken to be unity for all frequencies, $\hat{y}(s)$ represents the open-loop frequency response of the system.

The problem is to choose polynomial coefficients of the compensation elements $G_j(s)$ so that the open-loop frequency re-

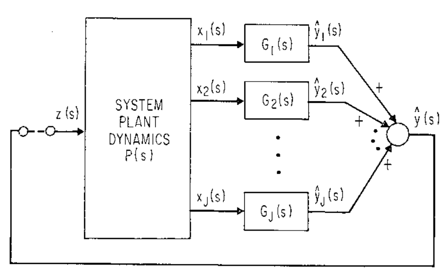

Fig. 1 Block diagram of a typical unforced, continuous-time, multiloop control system.

sponse of the total system $\hat{y}(s)$ corresponds as closely as possible to a desired open-loop frequency response $y(s)$, which is assumed to be known.

The frequency responses of the individual outputs of the plant $x_j(s)$ are also considered as known quantities. If a set of linear differential equations that describes $P(s)$ is known, other design tools can be used to find the $x_j(s)$. On the other hand, $P(s)$ may consist of existing hardware in which case frequency response tests may be performed to determine the $x_j(s)$. In addition, it is assumed that the orders of the numerator and denominator of the compensation transfer functions are known and fixed during the gradient search process.

2.1 Cost Function

As mentioned, the goal is to choose the compensation elements to make $\hat{y}(s)$ correspond as closely as possible to the desired frequency profile $y(s)$. The measure of closeness will be taken as the mean square difference of y and $\hat{y}$ over a set of frequency points ω_l, $l = 1, 2, \ldots, L$, for which a close fit is desired. Thus, the closeness of y to $\hat{y}$ can be described by a cost function in the form

$$J = \|(\mathbf{y}^* - \hat{\mathbf{y}}^*)^T \bar{W}^T \bar{W} (\mathbf{y} - \hat{\mathbf{y}})\| \tag{2}$$

where the asterisk (*) denotes the complex conjugate, and

$$\mathbf{y}^T = [y(s_1), y(s_2), \ldots, y(s_L)] \tag{3}$$

$$\hat{\mathbf{y}}^T = [\hat{y}(s_1), \hat{y}(s_2), \ldots, \hat{y}(s_L)] \tag{4}$$

$\bar{W}$ is a diagonal matrix of the form

$$\bar{W} = \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & & 0 \\ 0 & & 0 & w_L \end{bmatrix} \tag{5}$$

which can be used to assign different weights to the errors occurring at different frequency points. If all frequency points are to be weighted equally and $w_l = (1/L)^{1/2}$, the resultant J will represent a mean square error in the usual sense; thus, it is desired to minimize J as given by Eq. (2). Alternatively, the problem may be viewed as maximizing $-J$.

If an error vector

$$\boldsymbol{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} \tag{6}$$

is defined, Eq. (2) may be rewritten as

$$J = \|\boldsymbol{\varepsilon}^{*T} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}\| \tag{7}$$

The scalar cost function J is a function of the error vector $\boldsymbol{\varepsilon}$, which in turn is a function of the desired frequency response vector $\mathbf{y}$ and the approximating frequency response $\hat{\mathbf{y}}$.

2.2 Parameters to Be Varied

The approximating frequency response is given by

$$\hat{y}(s_l) = \sum_{j=1}^{J} G_j(s_l) x_j(s_l) \tag{8}$$

where J is the number of plant output variables to be compensated, and a typical transfer function is of the form

$$G_j(s_l) = \sum_{i=1}^{M_j} a_{ji}(s_l)^{i-1} \Big/ [1 + \sum_{i=2}^{N_j} b_{ji}(s_l)^{i-1}] \tag{9}$$

Notice that the coefficient of $(s_l)^0$ has been normalized to unity in the expression for the denominators of $G_j(s)$. This has been done to assure that a given $G_j(s)$ implies a unique set of coefficients a_{ji} and b_{ji}. If this or some similar normalization were not made, $k(a_{ji})$ and $k(b_{ji})$ would yield the same $G_j(s)$ as a_{ji} and b_{ji}.

2.3 An Expression for the Gradient Vector

To synthesize a set of compensating transfer functions $G_j(s)$ to minimize the cost function J, it is assumed that the orders of the numerator and denominator of each function are specified, and that the coefficients a_{ji} and b_{ji} are the only parameters subject to variation. Thus, J is ultimately a function of the compensation transfer function coefficients, and, with respect to these variable parameters, the gradient of J may be written as

$$(\nabla J)^T = \left(\frac{\partial J}{\partial \mathbf{c}}\right)^T = \left[\left(\frac{\partial J}{\partial \mathbf{c}_1}\right)^T, \left(\frac{\partial J}{\partial \mathbf{c}_2}\right)^T, \ldots, \left(\frac{\partial J}{\partial \mathbf{c}_{2j}}\right)^T\right] \tag{10}$$

where

$$(\partial J/\partial \mathbf{c}_{2j-1})^T = (\partial J/\partial a_{j1}, \partial J/\partial a_{j2}, \ldots \partial J/\partial a_{j,Mj}) \tag{11}$$

and

$$(\partial J/\partial \mathbf{c}_{2j})^T = (\partial J/\partial b_{j2}, \partial J/\partial b_{j3}, \ldots, \partial J/\partial b_{j,Nj}) \tag{12}$$

if the corresponding definitions

$$\mathbf{c}^T = (\mathbf{c}_1{}^T, \mathbf{c}_2{}^T, \ldots, \mathbf{c}_{2J}{}^T) \tag{13}$$

$$\mathbf{c}_{2j-1}{}^T = (a_{ji}, a_{j2}, \ldots, a_{j,M_j}) \tag{14}$$

and

$$\mathbf{c}_{2j}{}^T = (b_{j2}, b_{j3}, \ldots, b_{j,N_j}) \tag{15}$$

are made.

Equation (7) illustrates that J is a scalar function of the vector $\boldsymbol{\varepsilon}$. When the coefficients of the transfer functions are considered as a vector, as in Eqs. (13–15), $\hat{y}$ may be considered as a vector function of the vector $\mathbf{c}$. As a result of Eq. (6)

$$\partial \boldsymbol{\varepsilon}/\partial \mathbf{c} = -\partial \hat{\mathbf{y}}/\partial \mathbf{c} \tag{16}$$

where the quantities on both sides are matrices. By using the standard expression for the partial derivative of a vector with respect to a vector,

$$\frac{\partial \mathbf{y}}{\partial \mathbf{c}} = \begin{bmatrix} \mathbf{p}_{11} & \mathbf{p}_{12} & \ldots & \mathbf{p}_{1L} \\ \mathbf{p}_{21} & \mathbf{p}_{22} & \ldots & \vdots \\ \mathbf{p}_{2j,1} & \ldots & & \mathbf{p}_{2J,L} \end{bmatrix} \equiv \bar{P} \tag{17}$$

where

$$(\mathbf{p}_{2j-1,l})^T = (\partial \hat{y}_l/\partial a_{j1}, \partial \hat{y}_l/\partial a_{j2}, \ldots, \partial \hat{y}_l/\partial a_{j,M_j}) \tag{18}$$

and

$$(\mathbf{p}_{2j,l})^T = (\partial \hat{y}_l/\partial b_{j2}, \partial \hat{y}_l/\partial b_{j3}, \ldots, \partial \hat{y}_l/\partial b_{j,N_j}) \tag{19}$$

Applying the chain rule for derivatives to each component of the gradient vector and writing the result in vector matrix form give

$$\partial J/\partial \mathbf{c} = (\partial \boldsymbol{\varepsilon}^*/\partial \mathbf{c})\, \partial J/\partial \boldsymbol{\varepsilon}^* + (\partial \boldsymbol{\varepsilon}/\partial \mathbf{c})\, \partial J/\partial \boldsymbol{\varepsilon} \tag{20}$$

From Eq. (7)

$$\partial J/\partial \boldsymbol{\varepsilon}^* = \bar{W}^T \bar{W} \boldsymbol{\varepsilon} \text{ and } \partial J/\partial \boldsymbol{\varepsilon} = \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^* \tag{21}$$

Then combining Eqs. (7, 16, 20, and 21) produces the expression for the gradient vector as

$$\begin{aligned} \nabla J = \partial J/\partial \mathbf{c} &= -[\bar{P}^* \bar{W}^T \bar{W} \boldsymbol{\varepsilon} + \bar{P} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^*] \\ &= -[(\bar{P} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^*)^* + (\bar{P} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^*)] \\ &= -2Re\{\bar{P} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^*\} \end{aligned} \tag{22}$$

It will be recalled that the correction to the parameter vector $\mathbf{c}$ is proportional to the gradient vector and, since J is to be minimized, in the opposite sense. Thus

$$\Delta \mathbf{c} = KRe\{\bar{P} \bar{W}^T \bar{W} \boldsymbol{\varepsilon}^*\} \tag{23}$$

where the factor 2 has been absorbed in the constant of proportionality K. The value chosen for K determines the distance to proceed along the gradient vector. A standard doubling and halving algorithm was selected for incrementing K.

2.4 Matrix of Partial Derivatives

With a cost function as simple as the one at hand, it is easiest to obtain the gradient analytically (as opposed to employing numerical differencing techniques). This is done by computing the expressions for the needed partial derivatives contained in the matrix P, as given by Eqs. (17–19), and storing them in the computer. The elements of P are of the types $\partial\hat{y}(s_l)/\partial a_{ji}$ and $\partial\hat{y}(s_l)/\partial b_{ji}$.

From Eqs. (8) and (9), the expressions for these partial derivatives may be shown to be

$$\partial\hat{y}(s_l)/\partial a_{ji} = (s_l)^{i-1}x_j \Big/ \left[1 + \sum_{k=2}^{Nj} b_{jk}(s_l)^{k-1}\right] \quad (24)$$

and

$$\partial y(s_l)/\partial b_{ji} = -(s_l)^{i-1}x_j \sum_{k=1}^{Mj} a_{jk}(s_l)^{i-1} \Big/ \left[1 + \sum_{k=2}^{Nj} b_{jk}(s_l)^{k-1}\right]^2 \quad (25)$$

Equations (24) and (25) can be readily evaluated with a digital computer.

2.5 Improving Convergence of the Gradient Search

It is well known[4] that sharp ridges (valleys) in the cost function J cause extremely slow convergence with a standard gradient search technique. Ridge-following problems can be detected by examining the components of $\mathbf{\Delta c}$ along the gradient path. If convergence is slow and the magnitude of at least one component of $\mathbf{\Delta c}$ is relatively very large and of alternating sign, a ridge is being transversed.

The change made in each component of the parameter vector $\mathbf{c}$ depends very strongly upon the values of the partial derivatives of $\hat{y}$ with respect to each of the components of $\mathbf{c}$. From Eqs. (24) and (25), it can be seen that these partial derivatives are proportional to $(s_l)^{i-1}x_j$. The values of the x_j may vary greatly in magnitude for each of the loops to be compensated. This is particularly true if the x_j do not represent quantities that are physically similar, as would be the case if x_j were in radians and x_2 were in in./sec². In this case, no meaningful measure of the distance between points in the parameter space is defined, and the space is said to be non-Euclidean. In addition, for the frequency range of concern, the values of $s_l = j\omega_l$ may produce values of $(s_l)^{i-1}$ that are very much smaller or larger in magnitude than unity. Consequently, convergence problems are anticipated.

One would expect that for a suitable fit of $\hat{y}$ to y, all coefficients would be equally important, because every term of each transfer function can significantly affect the value of y at some frequency so as to reduce $\epsilon = y - \hat{y}$ to a small value. In accordance with this admittedly heuristic argument, the partial derivatives should have maximum values of approximately the same magnitudes, though not necessarily at the same frequencies.

Towards this end, the average values of x_j over the frequency range of interest can be made unity by the selection of appropriate scale factors. The resulting values of a_{ji} can automatically be unscaled by multiplying them by the same factors. That is,

$$\alpha_j \triangleq 1 \Big/ \sum_{l=1}^{L} x_j(l) \quad (26)$$

and

$$X_j(l) \triangleq \alpha_j x_j(l) \quad (27)$$

then

$$a_{ji} = \alpha_j A_{ji}; \quad j = 1, 2, \ldots, M_j \quad (28)$$

where A_{ji} are the unscaled coefficients returned by AUTO.

In addition, the partial derivatives can be made to contain a scaling factor proportional to the contribution of a typical term, say $a_{ji}(s_l)^{i-1}$ rather than $(s_l)^{i-1}$ alone, by making percentage rather than absolute changes in the coefficients of the compensation transfer functions.

In order to capitalize on this observation, a new parameter vector $\boldsymbol{\delta}$ is chosen so that each of its components represents a percentage change in one of the transfer function coefficients. The vector $\boldsymbol{\delta}$ is represented as

$$\boldsymbol{\delta}^T = (\boldsymbol{\delta}_1{}^T, \boldsymbol{\delta}_2{}^T, \ldots, \boldsymbol{\delta}_{2J}{}^T) \quad (29)$$

where

$$\begin{aligned} \boldsymbol{\delta}_{2j-1}{}^T &= (\delta_{j1}{}^a, \delta_{j2}{}^a, \ldots, \delta_{j,M_j}{}^a) \\ \boldsymbol{\delta}_{2j}{}^T &= (\delta_{j2}{}^b, \delta_{j3}{}^b, \ldots, \delta_{j,N_j}{}^b) \end{aligned} \quad (30)$$

and

$$\begin{aligned} a_{ji} &= a_{ji}(1 + \delta_{ji}{}^a) \text{ for } \delta_{ji}{}^a = 0 \\ b_{ji} &= b_{ji}(1 + \delta_{ji}{}^b) \text{ for } \delta_{ji}{}^b = 0 \end{aligned} \quad (31)$$

The gradient vector to be used in the search is taken with respect to $\boldsymbol{\delta}$. It will be seen that if the a_{ji} and b_{ji} in Eq. (9) are replaced with $a_{ji}(1 + \delta_{ji}{}^a)$ and $b_{ji}(1 + \delta_{ji}{}^b)$, respectively, the partial matrix $\bar{P}$ will be composed of partial derivatives of the form

$$\partial\hat{y}_l/\partial\delta_{ji}{}^a = a_{ji}(s_l)^{i-1}x_j \Big/ \left[1 + \sum_{k=2}^{Nj} b_{jk}(s_l)^{k-1}\right] \quad (32)$$

and

$$\partial\hat{y}_l/\partial\delta_{ji}{}^b = -b_{ji}(s_l)^{i-1}x_j \sum_{k=1}^{Mj} a_{jk}(s_l)^{k-1} \Big/ \left[1 + \sum_{k=2}^{Nj} b_{jk}(s_l)^{k-1}\right]^2 \quad (33)$$

The partial derivatives of Eqs. (32) and (33) will tend to be more similar in magnitude than those of Eqs. (24) and (25) and thus the ridge problem will be greatly alleviated. This is borne out by experience. The scheme of changing the parameter vector $\mathbf{c}$ by absolute increments $\Delta\mathbf{c}$ was found to be completely unworkable in AUTO due to ridge problems. By contrast, percentage corrections give very satisfactory results. They are implemented by finding the percentage correction vector $\boldsymbol{\delta}$ through the gradient approach and adjusting each transfer function coefficient c_k according to the relationship

$$c_k = c_{k-1}(1 + \delta_k) \quad (34)$$

This essentially eliminates ridge problems.

2.6 Geometrical Properties of the Cost Function Used in AUTO

Not all cost functions have geometric properties that insure convergence to a unique minimum or maximum. In such cases, where the minimum or maximum cannot be reached by a gradient search initiated from any point in the parameter space, it is necessary to have some knowledge of the cost function's geometric properties. With this knowledge, the search can be started at a point that will insure convergence to the desired minimum or maximum.

The cost function used in AUTO is geometrically very complicated. It will be demonstrated that the cost function J of Eq. (2) does not always have the properties necessary to insure convergence to its minimum via a gradient search from every region of the parameter space. This is best demonstrated by the following simple example.

Consider the case where $\hat{y}$ is given by

$$\hat{y}_l = [a/(bs_l + 1)]x_l \quad (35)$$

and the desired value of y is unity for all frequencies. Further assume that the values of x_l correspond to the frequency re-

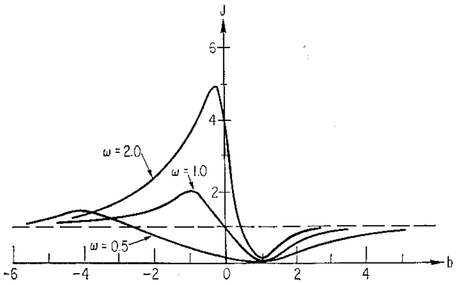

Fig. 2 Intersection of the plane $a = 1.0$ with the cost function of Eq. (38).

sponse of the simple transfer functions $(s + 1)$; that is, that

$$x_l = s_l + 1 \tag{36}$$

In this situation, if all of the weighting factors w_l are unity, the cost function is given by

$$J = \|1 - [a/(bs_l + 1)][s_l + 1]\| \tag{37}$$

or, equivalently, by

$$J = \sum_{l=1}^{L} \left\{ \left[1 - \frac{a(1 + b\omega_l^2)}{1 + (b\omega_l)^2} \right]^2 + \left[\frac{a\omega_l(1 - b)}{1 + (b\omega_l)^2} \right]^2 \right\} \tag{38}$$

The cost function J in Eqs. (37) and (38) is a function of the two parameters, a and b. Once the values of ω_l are specified, it is easy to see that if b is fixed as a constant, the cost function is an upturned parabola in the parameter a. Thus, the cost function cross sections, generated by the intersections of the plane b equals a constant with J, are all upturned parabolas.

It is not as easy to visualize the intersections of the plane a equals a constant with J. Assume that it is desired to obtain a best fit of $\hat{y}$ to y for only one frequency point. If a is fixed at 1.0, it can be seen that J is described by the curves of the form shown in Fig. 2, where each curve corresponds to a different choice of the value of ω_l.

Notice in the figure that in all cases an exact fit is attained (i.e., $J = 0$) for $b = 1.0$, as is also obvious from Eq. (37). If the initial value of b is chosen to be greater than unity, the gradient search will always converge to the minimum at $a = 1$ and $b = 1$, regardless of the value of ω_l. On the other hand, if the initial value of b is more negative than -4.0 the gradient search will diverge towards a value of $J = 1.0$, corresponding to a value of $b = -\infty$, regardless of the value of ω_l for all three cases shown.

If a fit of $\hat{y}$ to y is specified for all three frequencies simultaneously, it can be seen from Eq. (38) that the cost function

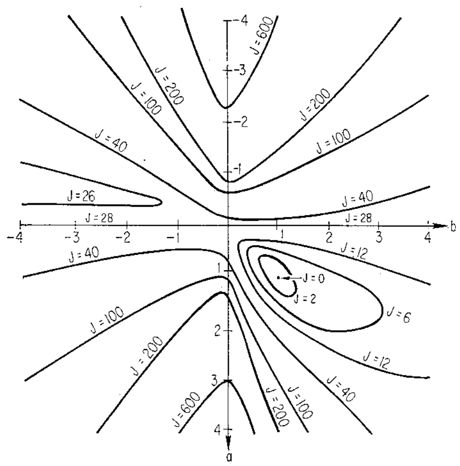

Fig. 3 Constant cost contours of cost function for Eq. (38).

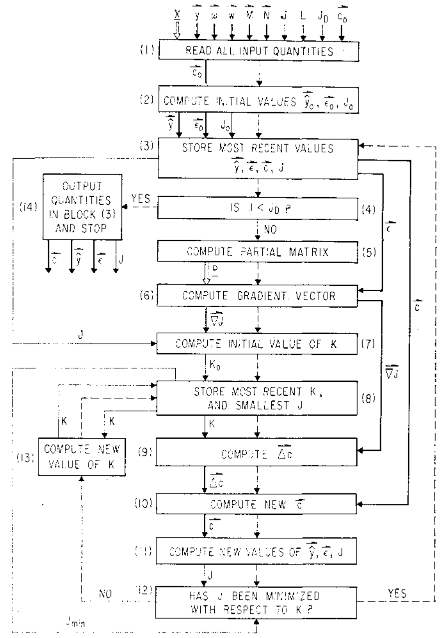

Fig. 4 Block diagram of automatic frequency-domain synthesis program.

is the sum of the three curves of Fig. 2 and thus is qualitatively the same in its geometric properties as are the individual curves.

The constant cost contours of Fig. 3 indicate the nature of J for all values of the parameters, a and b. The figure represents the cost function of Eq. (38) corresponding to a fit specified for 29 values of ω, from $\omega = 0.1$ to $\omega = 5.0$. It can be seen that regardless of the value of a, if the initial value of b is more than slightly negative, the gradient search will never reach the minimum of $J = 0.0$ at $a = 1.0$ and $b = 1.0$, but will proceed instead to the region denoted by $J = 26$. If b is positive, the gradient search will always converge to the minimum value of $J = 0.0$.

It is usually desirable to have the compensation poles in the left-half plane. In this case, it has been observed that if the initial transfer function denominator polynomials also have roots in the left-half plane, convergence to a suitable minimum value of J is achieved. This is independent of the initial choice of the numerator polynomial coefficients.

3. Automatic Frequency-Domain Synthesis Program

The relationships described in the preceding sections can be combined into an iterative scheme for automatic frequency-domain synthesis, as indicated in block diagram form in Fig. 4. Table 1 summarizes the input and output quantities of AUTO. The operations shown can be programed on a general-purpose digital computer with a scientific language such as FORTRAN. The specifications of the computer program will depend upon the particular computer and language available.

Figure 4 illustrates the order of the operations performed in AUTO; the flow from one operation to another is indicated by broken arrows. The various types of solid arrows indicate the transfer of quantities computed in one block to other blocks where they are required for further computations. For the sake of clarity, every input quantity required for a given operation is not shown entering the corresponding block. The quantities required for a given operation appear in the associated equations previously developed. An effort was made to concentrate on the flow of the additional important quantities generated during the course of a computer

Table 1 Summary of AUTO input and output quantities

Inputs	
J	= a real scalar whose value denotes the number of compensation transfer functions being considered
L	= a real scalar whose value denotes the number of frequency values for which the desired open-loop frequency response is specified
$\bar{X}(s)$	= a complex $J \times L$ matrix whose elements $x_{jl}(s)$ correspond to the value of the frequency response input to the jth compensation transfer function for the lth value of frequency
$\mathbf{y}(s)$	= a complex L vector whose elements $y_l(s)$ correspond to the desired open-loop frequency response of the total system for the lth value of frequency
$\mathbf{w}(s)$	= a real L vector whose elements w_l are the weighting factors for the error $\boldsymbol{\varepsilon} = (\mathbf{y} - \hat{\mathbf{y}})$ at the lth value of frequency
$\boldsymbol{\omega}$	= a real L vector whose elements ω_l correspond to the lth value of frequency associated with $s_l = j\omega_l$
$\mathbf{M}$	= a real J vector whose elements M_j represent the number of terms in the numerator of the jth compensation transfer function
$\mathbf{N}$	= a real J vector whose elements N_j represent the number of terms in the denominator of the jth compensation transfer function
$\mathbf{c}_0$	= a real $\sum_{j=1}^{J} (M_j + N_j - 1)$ vector whose elements c_k are the initial estimates of the coefficients of the compensation transfer function. [The arrangements of the elements of this vector are indicated in Eqs. (13–15).]
J_D	= a real scalar whose value is the average acceptable error $(\mathbf{y} - \hat{\mathbf{y}})^2$ at each value of frequency ω_l
Outputs	
$\mathbf{c}$	= a vector similar to $\mathbf{c}_0$ but containing the final design values of the transfer function coefficients chosen by AUTO
$\hat{\mathbf{y}}$	= a vector similar to $\mathbf{y}$ but containing the total open-loop frequency response value corresponding to the coefficient vector $\mathbf{c}$
$\boldsymbol{\varepsilon}$	= a real L vector whose elements ϵ_l are the magnitudes of $\boldsymbol{\varepsilon} = (\mathbf{y} - \hat{\mathbf{y}})$ corresponding to the coefficient vector $\mathbf{c}$
J	= a real scalar whose value is that of the cost function J corresponding to the coefficient vector $\mathbf{c}$

run. The operations blocks are numbered in the order of execution in a typical run. It should be particularly noted that there are two operational loops in AUTO.

The outer, or major, loop begins at Block (3), continues to Block (12), and then goes back to Block (3). This loop contains all the operations necessary to compute the direction of the gradient vector of the cost function, and the distance to proceed along the gradient vector to obtain the largest possible reduction in J afforded by the computational algorithm used.

The inner, or minor loop, contains all the operations necessary to compute the value of K corresponding to the largest reduction in J that can be obtained along a particular gradient direction, consistent with the algorithm used. In a typical run, the major loop may be executed 50 or more times, and the minor loop an average of 5 to 10 times during one major loop.

Once the cost function J is less than the desired value J_D, the program outputs the final design parameters, the resulting open-loop frequency response, and its deviation from the desired response. If it is discovered that AUTO encountered difficulty in providing the desired frequency response in some particular frequency regime, the weighting factors on the errors corresponding to the frequency points within this regime are increased, and another computer run is made. If an increase of the weighting factors for one frequency regime results in an undesirable response at other frequency points, it may be necessary to increase the order of the transfer function numerators and denominators. This would provide the

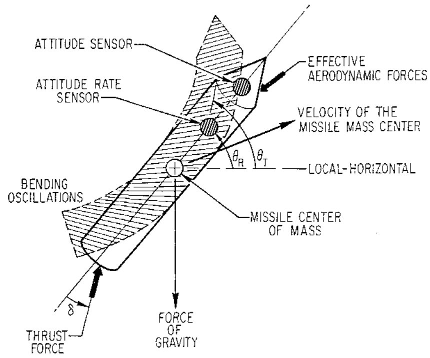

Fig. 5 Typical ballistic missile in flight.

compensation with enough flexibility to allow a frequency response satisfactory over the entire frequency range of interest.

It should be mentioned that AUTO has been able to design compensation superior to that generated with conventional methods by engineers with years of experience. In fact, it was found that AUTO can provide frequency responses as good as, or better than, those previously obtainable, while restricted to using lower-order compensation.

4. An Example of Automatic Frequency-Domain Design

4.1 System

As an illustration of the automatic frequency-domain design using AUTO, consider the ballistic missile system in Fig. 5. Figure 6 is a simplified block diagram of the missile and its digital attitude control system.

The controlled variable is the nominal missile angular inclination θ_R to the horizon in the plane of its flight path. The attitude θ_T is measured by the angular sensor attached to it. Since it is not perfectly rigid, θ_T is composed of two contributions: θ_R, the attitude that would be sensed if the missile were rigid, and θ_B, the attitude sensed because it is flexible and oscillates (or bends) about its centerline, similar to a plucked violin string.

Usually, a ballistic missile is unstable without feedback; the aerodynamic forces on it exert a net force normally in front of the missile's center of mass, which produces a destabilizing moment that tends to turn the missile end-over-end. This same effect can be observed by shooting an arrow into the air feathers-first. The arrow will flip over due to the aerodynamic forces on its feathers, which are located in front of the center of mass near the relatively heavy arrowhead. The variable δ represents the deflection of the thrust force exerted at the rear of the missile, in accordance with a feedback law, to maintain the desired attitude while propelling the missile forward.

The bending dynamics tend to destabilize the missile when the feedback loop is closed. Forces exerted on the missile create bending vibrations that are superimposed on the quantities detected by the feedback sensors. If not properly filtered, these vibrations will cause variations of the same frequency in the thrust forces. This can result in further excitation of the bending oscillations thus giving rise to a possible unstable oscillation.

Two feedback quantities are used to stabilize the missile. The total sensed attitude θ_T is subtracted from the commanded attitude θ_c to form the attitude error $\theta\boldsymbol{\varepsilon}$. The attitude error is compensated and summed with a compensated attitude rate signal to form the variable δ_c, which is the commanded thrust vector deflection. The mechanism that controls the positioning of the thrust vector has its own dynamics, which are included in the model. The attitude rate sensor, which is mounted on the missile structure, also senses

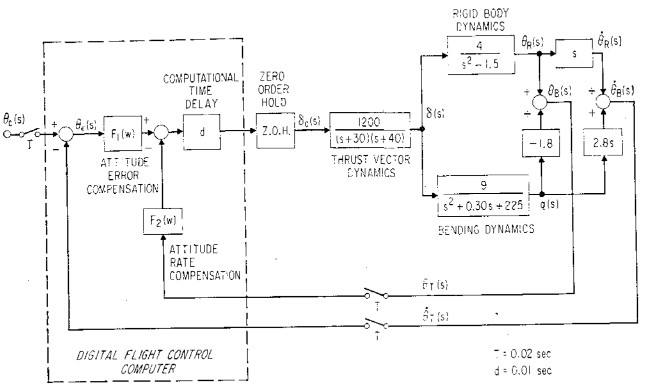

Fig. 6 Block diagram of a ballistic missile with a digital attitude control system.

bending vibrations. Thus, the total attitude rate $\dot{\theta}_T$ is affected by the bending dynamics as shown in Fig. 6.

Without dwelling further on the system model, one can summarize as follows. The plant of the missile is composed of the rigid body dynamics, bending dynamics, and thrust vector deflection dynamics, the models of which are assumed to be given. The two feedback quantities, θ_T and $\dot{\theta}_T$, are to be compensated in order to stabilize the missile.

4.2 Use of AUTO to Design Feedback Compensation

The design of suitable digital compensation consists of the following steps: 1) Write the system plant equations, indicated in the block diagram of Fig. 6, in vector-matrix form. 2) Consider the control loop broken in the common path at δ_c and use an analysis program to generate open-loop frequency responses from δ_c to θ_T, and from δ_c to $\dot{\theta}_T$. 3) Input the uncompensated frequency responses to AUTO, along with the desired compensated frequency response and form of w plane compensation, and allow AUTO to design the compensation. The results of these steps will now be discussed.

4.3 Vector Matrix Form of the Missile Dynamics

The relations indicated in Fig. 6 are written in vector-matrix form as Eq. (39):

$$\begin{bmatrix} s^2 + 70s + 1200 & 0 & 0 & 0 & 0 \\ -4 & s^2 - 1.5 & 0 & 0 & 0 \\ -9 & 0 & s^2 + 0.30s + 225 & 0 & 0 \\ 0 & -1 & 1.8 & 1 & 0 \\ 0 & -s & -2.8 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \delta \\ \theta_R \\ q \\ \theta_T \\ \dot{\theta}_T \end{bmatrix} = \begin{bmatrix} 1200 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \delta_c \quad (39)$$

4.4 Uncompensated Frequency Response of the Feedback Paths

The system matrix of Eq. (39) is input to a computerized algorithm to obtain the frequency responses from δ_c to θ_T and from δ_c to $\dot{\theta}_T$. The frequency responses of θ_T and $\dot{\theta}_T$ will be denoted by $x_1(s)$ and $x_2(s)$, which is consistent with the notation of Fig. 1.

AUTO will accept values for $x_1(s)$, $x_2(s)$, and the desired open-loop frequency response $y(s)$ in either rectangular or polar form. In this example, $x_1(s)$ and $x_2(s)$ were input in rectangular form and $y(s)$ was input in polar form.

4.5 Desired Open-Loop Frequency Response

The desired compensated open-loop frequency response from δ_c to δ_c (the loop is broken at δ_c) will be denoted by $y(s)$ according to convention. From experience, it has been determined that a desirable Nichols plot for $y(s)$ is of the form shown in Fig. 7.

It is desired that the digital compensation have no poles in the right-half w plane. The open-loop transfer function will have only one pole in that half-plane. This pole is contained in the rigid body dynamics, as can be seen in Fig. 6. Consequently, from the Nyquist theory, only one encirclement of the critical point is desired for stability. Thus, it is desirable that the frequency response remain outside the shaded region of Fig. 7.

To insure a stable corridor in the presence of uncertainties n the values of the system parameters, it is desirable to have high- and low-frequency gain margins of about 6 db, a low-frequency phase margin of at least 30°, and a high-frequency phase margin of at least 60°. The larger high-frequency phase margin is necessitated by a relatively large uncertainty in the model for the bending dynamics, which produces a resonance at 15 rad/sec. Uncertainty in the parameters of the bending model also makes it desirable to keep the resonance peak reasonably low. In addition, it is desirable to have a gain margin of 10 db or more for frequencies greater than the first bending mode resonance frequency. This is to ensure that effects from any high-frequency dynamics that have been excluded from the missile model are minimized. The values of y input to AUTO correspond to gain and phase margins somewhat in excess of those just discussed.

The system being studied is fairly difficult to compensate due to the conflicting requirements of maintaining rather large phase margins at both the rigid-body ($\simeq$2 rad/sec) and bending-resonance ($\simeq$15 rad/sec) frequencies. Simple compensation to add phase lag increases the high-frequency phase margin. The opposite is true of phase lead compensation. In addition, phase lead, which is needed to obtain rigid-body phase margin, increases the value of the bending-mode resonance peak. It is difficult to compensate the rigid body and bending dynamics independently because they have conflicting requirements. A successful compensation design depends on a skillful shaping and blending of the frequency response characteristics of the attitude and attitude-rate loop responses. Such a design is achieved by AUTO.

4.6 Results Produced by AUTO

AUTO is programed to design s domain compensation for continuous systems. For sampled systems, AUTO will design compensation in either the w, z, or inverse z domain. The only difference in these three options is the manner in which the frequency-dependent variable is used in evaluating the compensation transfer functions. Once the pertinent values of ω are input, AUTO evaluates the corresponding values of $s = j\omega$, $w = \tan\omega T/2$, $z = e^{j\omega T}$, or $z^{-1} = e^{-j\omega T}$, depending on the mode of operation specified. In the example used here the design is made in the w plane. Once the desired orders of the w plane compensation transfer function numerator and denominator polynomials have been selected, initial values of the transfer function coefficients must be entered into AUTO.

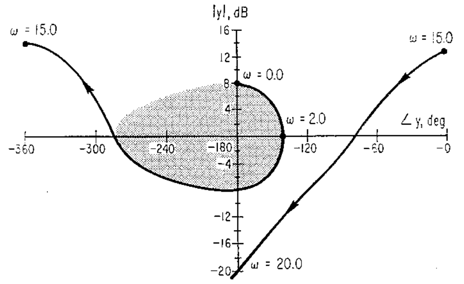

Fig. 7 Nichols plot of the desired frequency response.

To ascertain the difficulty of compensating the system in Fig. 6, pure gain compensation is tried first. To initialize the gradient search, gains of unity were entered for both compensation transfer functions, resulting in an initial cost function value of $J = 2904.3$. AUTO was allowed to run until no further decrease in the cost function J was discernible. The value obtained after approximately 50 gradient iterations was $J = 19.4$. The corresponding gains were 0.878 for the attitude feedback compensation (transfer function 1) and 0.0315 for the attitude-rate feedback compensation (transfer function 2). The values of the desired frequency response y and the actual response $\hat{y}$ obtained for pure gain compensation are shown in Fig. 8. It may be seen that in combining the feedback variables so as to keep the error between y and $\hat{y}$ at the bending-mode frequency ($\omega = 15.0$ rad/sec) small, it was impossible to match the desired rigid-body ($\omega \simeq 2.0$ rad/sec) frequency response. In fact, in this instance, the rigid-body response represents an unstable system.

It would be possible to reweight the various frequency points in an attempt to obtain a stable rigid-body frequency response at the expense of a larger bending-mode resonance peak. However, uncertainties in the bending-mode dynamics data make large peaks undesirable in practice.

It should be noted that if equal weighting factors are applied to all frequency points, the $\hat{y}$ generated by AUTO will match y most closely at values of y having large positive decibel values. This is due to the fact that errors between y and $\hat{y}$ at these values represent large errors in rectangular coordinates. The cost function used by AUTO is formulated in rectangular coordinates. It has since been reformulated in polar coordinates (with magnitude in decibels and phase in degrees) so that equal weighting factors would imply an equally close fit of $\hat{y}$ and y for all regions of a Nichols plot. However, since decibels and degrees are physically different, a numerical equivalence must be established between magnitude and phase errors. For instance, the penalty for four degrees of phase error would be equivalent to that for one decibel of magnitude error.

Since pure gain compensation was not found to be satisfactory, another computer run was made with AUTO. For the second run, the orders of both compensation transfer functions were increased to allow for two zeros and three poles in each. To initialize the gradient search, poles were desired in the left-half w plane so that exactly one encirclement of the critical point would be required in the Nyquist plot. As there was no better information on which to base the selection of the initial function coefficients, the four zeros and six poles were chosen to have values of minus one, giving the initial w plane

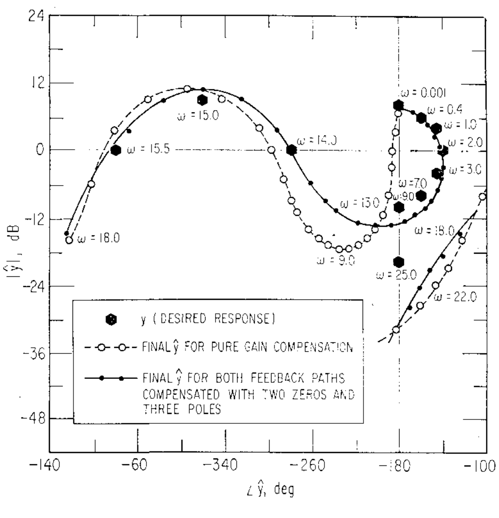

Fig. 8 Final open-loop frequency response generated by AUTO.

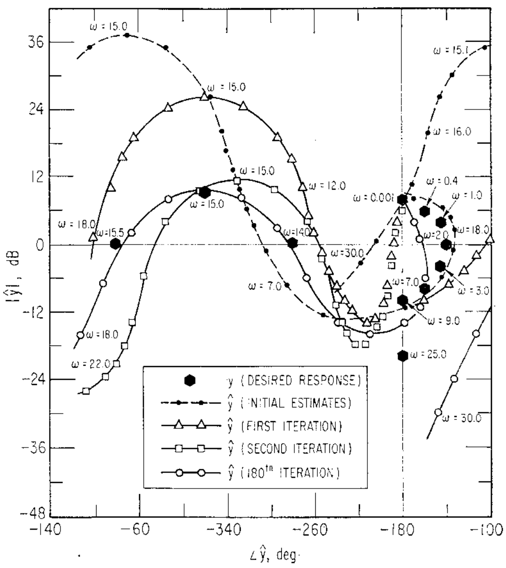

Fig. 9 Intermediate open-loop frequency responses generated by AUTO.

transfer functions

$$0_{F_1(w)} = 0_{F_2(w)} = (w^2 + 2w + 1)/(w^3 + 3w^2 + 3w + 1) \quad (40)$$

The corresponding initial value of the cost function was $J = 2873.2$. The frequency response $\hat{y}$ corresponding to the initial compensation is shown in Fig. 9. It can be seen that the rigid-body frequency response is quite suitable, but the bending-mode peaks are at the extremely large value of 36 db. The figure also shows $\hat{y}$ after 1, 2, and 180 gradient search iterations. Note that the first two iterations succeed in reducing the large error at the bending-mode resonance to a very small value at the expense of destabilizing the rigid-body frequency response. After the second iteration, the value of the cost function was $J = 25.8$.

By iteration 180, the values of the transfer functions have changed significantly and rate loop dc gain has increased to 0.1235730, which provides some rigid-body stability margin. The corresponding value of the cost function was $J = 7.3$.

At the end of 200 gradient iterations, the cost function is reduced to a value of 1.9, and any further decrease in J was imperceptible. The corresponding frequency response $\hat{y}$ is shown in Fig. 8 to be stable, with very large gain and phase margins. The compensation transfer functions that AUTO generated to yield this response are

$$200_{F_1(w)} = \frac{0.9878175w^2 + 31.47857w + 0.8710548}{0.9782988w^3 + 2.973800w^2 + 1.758666w + 1.0}$$

$$= 1.009730 \times \frac{(w + 0.02770)(w + 31.83908)}{(w + 2.48131)(w^2 + 0.55845w + 0.41197)} \quad (41)$$

and

$$200_{F_2(w)} = \frac{0.3432731w^2 + 0.2968132w + 0.2287250}{1.000785w^3 + 2.904175w^2 + 2.813753w + 1.0}$$

$$= 0.3430035 \times \frac{(w^2 + 0.86464w + 0.66629)}{(w + 1.41166)(w^2 + 1.49023w + 0.70783)} \quad (42)$$

It should be mentioned that if the closed-loop time response is required to have a certain steady-state error performance, the appropriate coefficients of the compensation transfer functions may be fixed at the correct values and excluded from adjustment by the gradient algorithm.

5. Summary and Conclusions

Equations have been presented for a gradient search algorithm to design feedback compensation transfer functions for general linear time-invariant systems. The organization of these equations to produce a computer program for automatic frequency-domain synthesis was discussed.

It was shown that a suitable method for avoiding convergence difficulties in minimizing the particular cost function treated involves the use of percentage changes in the transfer function coefficients.

A design problem involving the stabilization of a complex ballistic missile control system was chosen to demonstrate the performance of the design algorithm implemented in the computer program AUTO.

Practical experience indicates that for the ratio of polynomials in a complex variable, a cost function expressed in polar coordinates is preferable to its equivalent in rectangular coordinates. The polar formulation yields faster convergence and, in some cases, can be shown mathematically to eliminate local minima, which might prevent convergence to the true minimum where a rectangular cost function is utilized. These observations should also stimulate further research into the nature of such cost functions.

References

[1] Coffey, T. C. and Williams, I. J., "Stability Analysis of Multiloop, Multirate Sampled Systems," *AIAA Journal*, Vol. 4, No. 12, Dec. 1966, pp. 2178–2190.

[2] Kuo, F. F. and Kaiser, J. F., *System Analysis by Digital Computer*, Wiley, New York, 1966.

[3] Saucedo, R. and Schiring, E. E., *Introduction to Continuous and Digital Control Systems*, Chap. 14, "The Application of Modern Computing Technology to Control System Analysis and Design," by T. C. Coffey, Macmillan, New York, 1968.

[4] Wilde, D. J., *Foundations of Optimization Methods*, Prentice-Hall, Englewood Cliffs, N. J., 1967.

[5] Sage, A. P., *Optimum Systems Control*, Prentice-Hall, Englewood Cliffs, N. J., 1968.

OCTOBER 1970 AIAA JOURNAL VOL. 8, NO. 10

Structural Optimization in the Dynamics Response Regime: A Computational Approach

R. L. Fox*
Case Western Reserve University, Cleveland, Ohio

AND

M. P. Kapoor†
Indian Institute of Technology, Kanpur, India

A structural optimization problem is considered in which the design requirements include restrictions on the dynamic response and frequency characteristics of the structure. One of the central concerns of this phase of the work has been to overcome the problems inherent in treating systems with many degrees of freedom. The simple planar truss-frame with both distributed and concentrated mass is the model upon which this exploratory study is based. Limitations have been imposed upon maximum dynamic stresses and displacements (handled by the shock spectral approach) as well as on the natural frequencies of the structure. A "direct" optimization method (the method of feasible directions) which consists of a design-analysis cycle was used. Computationally efficient schemes are given for the necessary derivatives of maximum response and natural frequency. Numerical examples are given and computational effectiveness is indicated.

1. Introduction

THE importance of dynamic response in current structural problems has prompted an increased attention to the synthesis (automated optimum design) of structures for which the dynamic response will be a controlling criterion. This paper reports upon work that was undertaken to demonstrate the feasibility of using a dynamics technology within the structural synthesis framework. The design restrictions imposed include limitations on the dynamic displacements, stresses, and ranges in which the natural frequencies of the structure are allowed to fall.

As a first phase, the automated minimum weight design of the general planar truss-frame system‡ has been undertaken. The dynamic response is assumed to be linear, undamped, and the result of known externally applied exciting forces at the joints or of foundation displacements.

The structural design problems that arise within several of the more common design philosophies have been put in the form of mathematical programing problems. In the present work the minimum weight design of general planar truss-frame systems, subject to dynamic response limitations, has been cast as a mathematical programing problem. The resulting mathematical programing problem has been solved using the method of feasible directions.[1]

Presented at the AIAA Structural Dynamics and Aeroelasticity Specialist Conference, New Orleans, La., April 16–17, 1969 (no paper number; published in bound volume of conference papers); submitted April 29, 1969; revision received April 3, 1970. This work was supported in part by the National Aeronautics and Space Administration under research grant NsG 110-61.

* Associate Professor of Engineering.

† Assistant Professor of Civil Engineering.

‡ The nomenclature "truss-frame" as used in this paper denotes a structure consisting of elements capable of axial and lateral deformations. Such elements may be rigidly connected or pinned together.