TRANSIENT CORRECTION BY MEANS OF ALL-PASS NETWORKS

JOHN C. PINSON

TECHNICAL REPORT 324

MAY 13, 1957

LOAN COPY

only

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

RESEARCH LABORATORY OF ELECTRONICS

Technical Report 324   May 13, 1957

TRANSIENT CORRECTION BY MEANS OF ALL-PASS NETWORKS

John C. Pinson

TRANSIENT CORRECTION BY MEANS OF ALL-PASS NETWORKS

by

JOHN CARVER PINSON

Submitted to the Department of Electrical Engineering on
May 13, 1957, in partial fulfillment of the requirements for
the degree of Doctor of Science.

ABSTRACT

Phase correction, realized in the form of an all-pass net-
work, is frequently used in order to improve the transient
response of a system. An investigation is made here to
determine the phase correction that should be used to achieve
the optimum corrected response for a given system. In general,
the ideally desired response cannot be obtained by means of
phase correction. Then an error criterion must be used to
define the corrected response that best approximates the desired
response. The phase correction which gives the corrected
response that approximates the desired response with minimum
integral square error is determined. For the particular class
of systems in which reproduction of the system input is desired,
it is found that the correction should linearize the phase of
the system in order to produce a corrected response with
minimum integral square error.

Phase correction is most commonly used to enable a system
to reproduce better a step function input. It is found that
the correction which yields the corrected response with
minimum integral square error does not provide the step response
with the shortest rise time. By consideration of a suitably
chosen weighted integral square error criterion, a phase
correction is derived which yields a corrected step response
with the shortest possible rise time. It is found, however,
that the realization of this correction requires an all-pass
network with an extremely large number of circuit elements.
When correction to reduce the overshoot of the step response of
a system is desired, it is found that correction for minimum
integral square error provides the optimum response.

Thesis Supervisor: Ernst A. Guillemin
Title: Professor of Electrical Communications

ACKNOWLEDGEMENT

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Brief Statement of the Problem

The problem to be considered here may be stated as follows: what can be done to change a given time domain transient to a different, given, transient using only an all-pass network? An all-pass network has a transfer characteristic whose zero locations on the complex s-plane mirror the pole locations about the imaginary axis. Thus the amplitude factor is a constant along the imaginary axis, while the phase is a function of frequency.

1.2 Origin and History of the Problem

It may happen that after design of a network having a desired amplitude characteristic, it is found that its response to some transient is undesirable. Then it would be advantageous to alter this transient response by cascading with the original network, a network with a constant amplitude transfer function. For example, government regulations place limits on the bandwidth allowed to television transmitters; full utilization of the allotted bandwidth would require a sharp cutoff of the amplitude spectrum near the band edge. But the output of such a sharp cutoff filter contains an undesirable ripple, or ringing component after a sharp change in input. Perhaps by cascading an all-pass network this ripple can be controlled.

In order to transform exactly one given transient to another, a network with both a specific amplitude characteristic

and a specific phase characteristic is required. The network may, or may not, be physically realizable. If the impulse response of the network is zero for negative times and bounded for positive times, then the network is realizable.[1,a]

The problem of correcting the distortions which occur in telephone lines has long been of interest. The effort of most of the early correction schemes was primarily directed toward flattening the amplitude of the transfer function. No attempt at phase correction was made because the effect of phase distortion on voice communication over relatively short lines was not important. As longer lines came into service, phase distortion became a serious problem and correction, in the form of an all-pass network, was used to linearize the phase. Recognizing the importance of the effect of phase distortion on transient response, researchers in the field endeavored to develop facile methods for evaluating this effect.[2,3] Working backwards from the results of analyses of particular cases, they were able to reach some general conclusions which served as guides in other problems of phase correction. But no reference in the literature has been found attempting to make a basic investigation of just how much can be done to change one transient to another by altering only the phase of its Fourier transform. It is not clear, for example, that correcting a telephone line or an amplifier to give linear phase leads to the best possible transient response.

---

[a] Superscripts refer to numbers in the Bibliography

CHAPTER II

PROPERTIES OF ALL-PASS NETWORKS

2.1 Phase of an All-Pass Network

An "all-pass" network is so named because the amplitude of the transfer function is a constant, unity, for all frequencies. These networks pass sinusoids of any frequency without attenuation but with a phase shift which is, in general, a function of frequency. Let the transfer function of an all-pass network be

$$\epsilon^{-j\theta(\omega)} \tag{1}$$

If an all-pass network is to be realized with lumped, linear circuit elements, its transfer function can have only poles as singularities in the complex frequency, or s-plane; if the network is to be stable, the poles must lie in the left half of the s-plane. Since the amplitude is a constant, it must be that the network has only right-half-plane zeros, and that these zeros are placed at locations which are mirror images about the imaginary axis of the pole locations. Figure 1 shows typical pole and zero locations of the transfer function of an all-pass network. The fact that the transfer function of an all-pass network must have a pattern of poles and zeros of this sort restricts the form of the phase which can be realized by these networks. It is evident, for example, that $\theta(\omega)$ must increase monotonically with frequency.

The restrictions on a phase function which can be realized
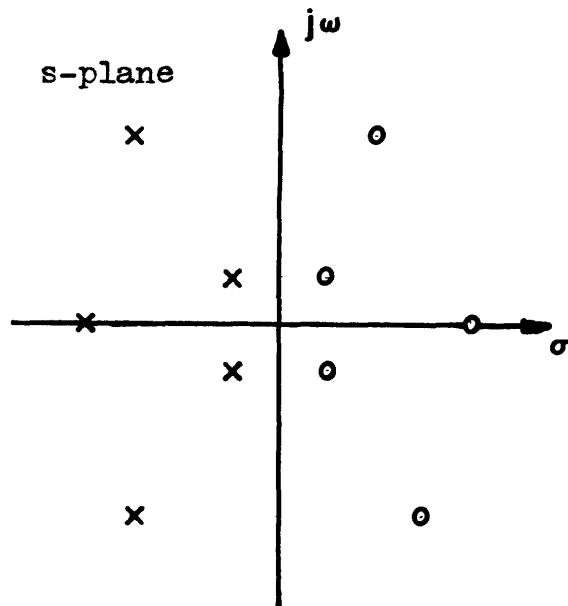
Fig. 1.
Transfer Function of
an All-Pass Network

as the phase of an all-pass network can also be made evident

by time domain considerations. For an arbitrarily specified

phase $\beta(\omega)$, the inverse Fourier transform of the function

$\epsilon^{-j\beta(\omega)}$, denoted $F^{-1}[\epsilon^{-j\beta(\omega)}]$, will not be zero for all nega-

tive time, in general. Obviously, only those phases for which

$F^{-1}[\epsilon^{-j\beta(\omega)}]$ is zero for all negative time can be physically

realized.

2.2 Approximation of Any Phase Function by All-Pass Network Phase

Although not all phase functions can be realized as the

phase of an all-pass network, it will now be shown that the

phase of an all-pass network, $\Theta(\omega)$, can approximate as closely

as desired a phase function $\beta(\omega) + \omega T$ over any finite frequency

range, however large, where $\beta(\omega)$ is any phase function, and ωT is a linear phase which must be added to $\beta(\omega)$ in order to allow realization in the form of an all-pass network. This means that we can approximate any desired phase plus the phase of a delay network by an all-pass network phase; we can use all-pass networks to realize a good approximation of _any_ phase if we are willing, in addition, to tolerate a delay, T.

In order to prove that the above statements are true, let us consider the phase of a single pole-zero pair. Figure 2(b) shows the phase of a pair whose zero lies at $s_1 = \sigma_1 + j\omega_1$. Note that the pair contributes a total phase shift of 2π over the range of frequencies $-\infty < \omega < \infty$.

(a)

(b)
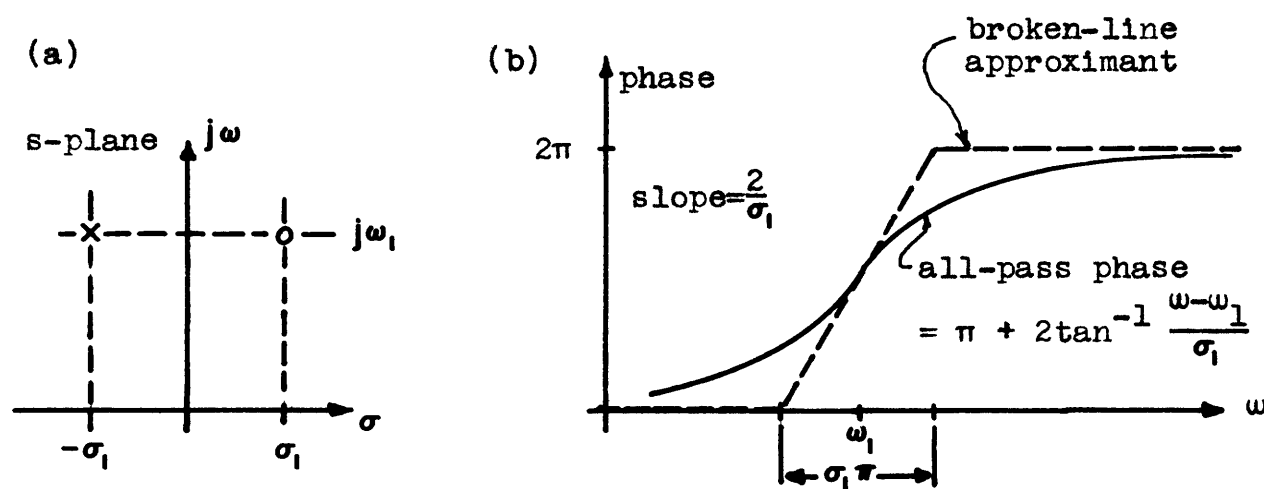
Fig. 2. Phase of $\dfrac{(s - \sigma_1) - j\omega_1}{(s + \sigma_1) - j\omega_1}$ and Broken-Line Approximant

6

The phase of this pair may be approximated by a broken-line
approximant of three segments, as shown in Fig. 2(b). The
broken-line approximant is tangent to the phase curve at
$\omega = \omega_1$ and at $\omega = \pm \infty$. The total phase of several pole-zero
pairs comprising a given all-pass characteristic can be
represented as a sum of these broken-line approximants.

For the purposes of this proof, let no pole-zero pair be
placed on the real axis. Further, if the zeros are located at
$s_1, s_2, \ldots, s_n, \ldots$, with $\omega_1 < \omega_2 < \ldots < \omega_n < \ldots$, let
these locations be chosen so that

$$\omega_{n+1} - \omega_n = \frac{\pi}{2} (\sigma_{n+1} + \sigma_n) \tag{2}$$

By choosing pair locations in this way, the second break point
of the phase approximant of the pair with the zero s_n, and the
first break point of the phase approximant of the pair with the
zero s_{n+1}, will occur at the same frequency. The approximant
to the phase of an all-pass network whose pair locations are so
chosen is shown in Fig. 3; the length of each line segment when
projected on the phase axis, or "phase length", is 2π. Now
it will be shown that by means of this broken-line phase
approximant of an all-pass network we can approximate $\beta(\omega) + \omega T$
for some choice of T.

Consider approximation of the curve $\beta(\omega)$ versus ω by a
curve of broken-line segments. Let us define a frequency
interval x which is small enough so that any broken-line
approximate makes a satisfactory approximation of $\beta(\omega)$ provided

Fig. 3.   The Broken-Line Approximant to a Possible Phase Shift
Characteristic of an All-Pass Network

that all of the line segments have a projected length on the
ω axis, or "ω length", that is no greater than x. If the least
slope of $\beta(\omega)$ is -m, let T be chosen large enough so that
xT - xm $\geq$ 2π. Now let the curve of $\beta(\omega)$ + ωT versus ω be
approximated by a broken-line with segments of phase length 2π;
this broken-line curve is also the approximant of some all-pass
network phase, $\Theta(\omega)$, as shown in Fig. 3. The line segment of
longest ω length will occur in the vicinity of the point of
minimum slope of $\beta(\omega)$ + ωT and, because of the choice of T
indicated above, will have an ω length equal to, or less than x.
By the original supposition of this paragraph, this is a good
approximation.

8

Because, in general, the function $\beta(\omega) + \omega T$ approaches infinity as ω approaches infinity, it would take an infinite number of line segments of phase length 2π to approximate the function over an infinite range of frequencies. But an all-pass network with a finite number of lumped circuit elements can have only a finite number of pole-zero pairs; thus it can have a phase characteristic which is approximated by only a finite number of line segments of phase length 2π. Then the phase of an all-pass network can be made to approximate any phase function, $\beta(\omega)$, plus some linear phase, ωT, over a range of frequencies $0 < \omega < \omega_o$, where ω_o is arbitrarily large but finite. If a given transfer function, $\epsilon^{-j\beta(\omega)}$, is needed to operate on a transient whose transform has no significant components above the frequency ω_o, an all-pass network can be realized whose transfer function closely approximates $\epsilon^{-j\beta(\omega)}\epsilon^{-j\omega T}$ over the important frequency range, zero to ω_o.

CHAPTER III

PHASE CORRECTION TO ACHIEVE TRANSIENT RESPONSE

WITH MINIMUM INTEGRAL SQUARE ERROR

3.1 Consequences of Allowing Unrestricted Choice of Phase Function

Chapter II was concerned with finding the class of phase characteristics which can be realized as the phase of the transfer function of an all-pass network. We found that any arbitrary phase characteristic can be approximated as closely as may be desired by the phase of an all-pass network, provided only that we are willing to accept, in addition, a time delay in the response of the all-pass network. In the transient correction problems being considered here, this delay is not objectionable; we conclude that we can realize any phase that we may need in order to accomplish a given transient correction.

The problem of this chapter, and of succeeding chapters, is to find the phase characteristic which does the best job of transient correction in a given problem. No physical realizability restrictions need to be imposed on the phase characteristic. We will simply look for the phase, $\Theta(\omega)$, which makes the best transient correction, where $\Theta(\omega)$ can be _any_ odd function of frequency. Now the transfer function $\epsilon^{-j\Theta(\omega)}$ can be considered as belonging to a fictitious, nonrealizable network but can no longer be associated with a realizable all-pass network. Remembering this, it should come as no surprise when, for example, the step response of some system which has been

cascaded with the phase correction network $\epsilon^{-j\theta(\omega)}$ has non-zero values before t = 0. Certainly this situation is not physically possible in the laboratory. However, we have shown that it is possible to realize an all-pass network for phase correction of the system which will give very nearly the same corrected system step response, except for a delay.

3.2 Choice of an Error Criterion

Having explored the consequences of allowing the phase correction, $\theta(\omega)$, to be any odd function, we proceed to the discussion of ways of choosing $\theta(\omega)$ in a given transient correction problem. Suppose that i(t) is the undesired transient that we propose to improve by means of phase correction. If possible, the phase correction should change i(t) into r(t), where r(t) is the desired transient. The network which changes i(t) to r(t) has the transfer function

$$G(\omega) = \frac{R(\omega)}{I(\omega)} = |G(\omega)| \; \epsilon^{-j\emptyset(\omega)} \tag{4}$$

where $R(\omega)$ and $I(\omega)$ are Fourier transforms of r(t) and i(t), and $|G(\omega)|$ and $\emptyset(\omega)$ are the magnitude and phase of $G(\omega)$. The convention implied here, that of using corresponding small and capital letters for time and frequency domain functions related by the Fourier transform, will be continued throughout and without further comment.

Clearly, unless $|G(\omega)| = 1$, a phase correction network, $\epsilon^{-j\theta(\omega)}$, cannot change i(t) to r(t). To allow more freedom in

the attempt to correct the undesired transient suppose that we also use an amplifier with a transfer characteristic K; a phase correction network with the transfer function $K\epsilon^{-j\Theta(\omega)}$ is now available for use in making the transient correction. Still, unless $|G(\omega)| = K$, the network $K\epsilon^{-j\Theta(\omega)}$ cannot accomplish the change to the desired transient. The best that can be done with phase correction is to change $i(t)$ to $r^*(t)$, where $r^*(t)$, the corrected transient, is an approximation of $r(t)$ and where

$$R^*(\omega) = KI(\omega)\epsilon^{-j\Theta(\omega)} \tag{5}$$

There is an error in the approximation

$$e(t) = r(t) - r^*(t) \tag{6}$$

Because the transient correction cannot be made without error, there is a question as to what constitutes the "best" correction. Some sort of error criterion must be imposed in order to define the $r^*(t)$ which best approximates $r(t)$. The use of the integral square error criterion furnishes an appropriate starting point in this investigation; the criterion makes good sense physically and, in addition, leads to a mathematical formulation which can be handled quite easily.

3.3 Minimization of Integral Square Error

The integral square error, $\overline{e^2}$, is given by

$$\overline{e^2} = \int_{-\infty}^{\infty} [r(t) - r^*(t)]^2 \, dt \tag{7}$$

By making use of Parseval's Theorem

$$\overline{e^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |R(\omega) - R^*(\omega)|^2 \, d\omega \qquad (8)$$

And since

$$R(\omega) - R^*(\omega) = I(\omega)[G(\omega) - K\epsilon^{-j\Theta(\omega)}] \qquad (9)$$

Equation 8 becomes

$$\overline{e^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |I(\omega)|^2 \, [G(\omega) - K\epsilon^{-j\Theta(\omega)}] \, [G(-\omega) - K\epsilon^{+j\Theta(\omega)}] \, d\omega \qquad (10)$$

We want to find the phase function $\Theta(\omega)$ and the gain factor K which minimize $\overline{e^2}$.

At this point the discussion will digress to introduce enough of the theory of calculus of variations to permit solution of this minimization problem. Knowledge of ordinary differential calculus allows us to find values of x for which $P(x)$ is a maximum or a minimum, or more directly, the values of x for which $P(x)$ has zero slope. But this knowledge does not tell us how to find the function $\Theta(\omega)$ which minimizes the integral

$$\overline{e^2} = \int_{-\infty}^{\infty} Q[\omega, \, \Theta(\omega)] \, d\omega \qquad (11)$$

However, a close analogy can be drawn between these two problems. We find values of x for which $P(x)$ has zero slope, and we find functions, $\Theta(\omega)$, called "extremals" of the integral for which the value of $\overline{e^2}$ is "stationary". At points of zero

slope of P(x), an infinitesimal change in x produces no change in the value of P(x). For extremals of $\overline{e^2}$, an infinitesimal change in the form of the function $\Theta(\omega)$ produces no change in the value of $\overline{e^2}$. Then extremals of $\overline{e^2}$ are the functions $\Theta(\omega)$ which satisfy the equation

$$0 = \frac{d}{d\alpha} \int_{-\infty}^{\infty} Q[\omega, \Theta(\omega) + \alpha \lambda (\omega)] \, d\omega \Bigg|_{\alpha = 0} \qquad (12)$$

where $\lambda(\omega)$ is any allowable variation of $\Theta(\omega)$. In the particular problem that is being considered in this section, $\lambda(\omega)$ can be any odd function of ω. Extremals of the integral of Eq. 11 give values of the integral which are stationary. Just as points of zero slope of P(x) occur where P(x) is a maximum, a minimum, or possibly where it is neither a maximum nor a minimum, stationary values of $\overline{e^2}$ may be values which are maxima, minima, or neither. Further testing of some sort is needed to determine which extremal gives the value of $\overline{e^2}$ which is <u>the</u> minimum that can be obtained.

Now we can proceed to find the phase function and the gain factor which minimize $\overline{e^2}$. The integral square error, as given by Eq. 10 must be minimized with respect to both $\Theta(\omega)$ and K. The problem will be attacked as follows. First the extremals, $\Theta(\omega)$, will be found with K as a parameter. Substitution of these extremals into Eq. 10 will leave an expression for $\overline{e^2}$ involving only K, which can then be minimized with respect to K. Proceeding according to this outline

14

$$0 = \frac{d}{d\alpha} \int_{-\infty}^{\infty} |I(\omega)|^2 (G(\omega) - K\epsilon^{-j[\Theta(\omega) + \alpha\lambda(\omega)]})$$

$$(G(-\omega) - K\epsilon^{j[\Theta(\omega) + \alpha\lambda(\omega)]}) \, d\omega \, \Big|_{\alpha = 0} \quad (13)$$

Performing the differentiation indicated in Eq. 13

$$0 = j \int_{-\infty}^{\infty} |I(\omega)|^2 \left\{ K\epsilon^{-j\Theta(\omega)}[G(-\omega) - K\epsilon^{j\Theta(\omega)}] \right.$$

$$\left. -K\epsilon^{j\Theta(\omega)}[G(\omega) - K\epsilon^{-j\Theta(\omega)}] \right\} \lambda(\omega) \, d\omega \quad (14)$$

With simplification of the curly bracketed term, Eq. 14 becomes

$$0 = \int_{-\infty}^{\infty} \lambda(\omega) \, |I(\omega)|^2 \left\{ Im[G(\omega)\epsilon^{+j\Theta(\omega)}] \right\} \, d\omega \quad (15)$$

Note that the integrand of Eq. 15 is an even function of ω. Because $\lambda(\omega)$ can be any, arbitrary odd function, if Eq. 15 is to be satisfied, it must be that

$$|I(\omega)|^2 \left\{ Im[G(\omega)\epsilon^{j\Theta(\omega)}] \right\} = 0 \quad (16)$$

Assume now that there are no frequency intervals over which either $|I(\omega)|$ or $|G(\omega)|$ is identically zero. This is certainly the usual case. Then using Eq. 4

$$\sin [\phi(\omega) - \Theta(\omega)] = 0 \quad (17)$$

The solution of Eq. 17 is not unique. The general solution is

$$\Theta(\omega) = \phi(\omega) + n(\omega)\pi \quad (18)$$

where n(ω) is any odd function of ω whose value is always an integer except at discontinuities.

Remember that the plan of attack on this minimization problem called first for finding $\Theta(\omega)$ with K as a parameter. However, according to Eq. 18, $\Theta(\omega)$ is independent of K. Then we can proceed immediately to minimize $\overline{e^2}$, as given in Eq. 10, with respect to K.

$$0 = \frac{d\,\overline{e^2}}{dK} \tag{19}$$

$$0 = \int_{-\infty}^{\infty} |I(\omega)|^2 \left\{ -\epsilon^{-j\Theta(\omega)}[H(-\omega) - K\epsilon^{j\Theta(\omega)}] \right.$$
$$\left. -\epsilon^{j\Theta(\omega)}[H(\omega) - K\epsilon^{-j\Theta(\omega)}] \right\}\,d\omega \tag{20}$$

$$0 = \int_{-\infty}^{\infty} |I(\omega)|^2 \left\{ K - \mathrm{Re}[\epsilon^{j\Theta(\omega)}H(\omega)] \right\}\,d\omega \tag{21}$$

Solving for K and making use of Eq. 18

$$K = \frac{\int_{-\infty}^{\infty} |I(\omega)|^2 |H(\omega)| (-1)^{n(\omega)}\,d\omega}{\int_{-\infty}^{\infty} |I(\omega)|^2\,d\omega} \tag{22}$$

Now there is the question of the function n(ω). For some one particular n(ω) function Eqs. 18 and 22 give the phase function and the gain factor which minimize $\overline{e^2}$. Substituting $\Theta(\omega)$ as in Eq. 18 into Eq. 10

$$\overline{e^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} |I(\omega)|^2 \left[|G(\omega)| - K(-1)^{n(\omega)} \right]^2\,d\omega \tag{23}$$

16

Evidently $\overline{e^2}$ is minimized when $n(\omega)$ identically equals zero (or some even integer since a change of 2π in a phase function is inconsequential). Now the solution of the minimization problem is complete.

$$\Theta(\omega) = \emptyset(\omega) \tag{24}$$

$$K = \frac{\int_{-\infty}^{\infty} |I(\omega)|^2 |H(\omega)| \, d\omega}{\int_{-\infty}^{\infty} |I(\omega)|^2 \, d\omega} \tag{25}$$

The minimum integral square error is

$$\overline{e^2} = \int_{-\infty}^{\infty} |I(\omega)|^2 \left[|G(\omega)| - K \right]^2 \, d\omega \tag{26}$$

Perhaps some discussion and interpretation of these results is in order. The specification of the phase correction needed to minimize integral square error is surprisingly simple. Put into words, Eq. 24 says that we should make the phase correction network the same as the phase of the network needed to make the transient correction without error. Further, this result is not dependent upon the transient to be corrected, $i(t)$, or upon the magnitude function of the network needed to make errorless correction, $|G(\omega)|$.

Application of the minimum integral square error criterion to the common practical problem of phase correcting an amplifier gives an interesting result. An amplified reproduction of the input is desired, thus $r(t) = Ci(t)$. Errorless correction

would be made by a network whose phase exactly cancels the phase shift of the amplifier; the phase correction network must also have this phase characteristic if its response is to have minimum $\overline{e^2}$. Thus, the over-all phase characteristic of the amplifier with phase correction must be made zero. But the network to make this phase correction is not physically realizable. Instead, we must realize a network which makes the phase of the corrected amplifier approximately linear. And this is just what is done in current engineering practice. It is not generally known, however, that this conclusion can be reached by requiring phase correction leading to minimum integral square error.

CHAPTER IV

PHASE CORRECTION OF A SLUGGISH SYSTEM

4.1 Correction of a Sluggish System for Minimum Integral Square Error

A completely general specification of the phase correction needed to minimize $\overline{e^2}$ is given by Eqs. 24 and 25 of Chapter III. In this chapter we wish to use these results in a particular transient correction problem. This example will make clear some of the implications of using the integral square error criterion and will suggest a modification and extension in the mode of its application.

Phase correction finds its most common use with an amplifier, where it serves to improve the fidelity of the output of the system. The performance of such a system is usually judged by the accuracy with which it reproduces a unit step input. The definition of a unit step which will be used here is different from that used by most authors. A unit step will be defined as a function which is $-\frac{1}{2}$ for negative times and $+\frac{1}{2}$ for positive times with a discontinuity at $t = 0$. The Fourier transform of this unit step is $(j\omega)^{-1}$. The definition of the unit step is made in this way so that it will be a time function which is odd; its transform is purely imaginary and an odd function of ω.

Suppose that we want to correct a sluggish system having the transfer function

$$\frac{1}{(j\omega+1)^2} \tag{27}$$

so that its step response has minimum integral square error.
An ideal system would have a step response with the transform

$$R(\omega) = \frac{1}{j\omega} \tag{28}$$

The input to the phase correction network is

$$I(\omega) = \frac{1}{j\omega} \frac{1}{(j\omega+1)^2} \tag{29}$$

To correct for minimum $\overline{e^2}$, the phase correction must be such
that $R^*(\omega)$ has the same phase as $R(\omega)$. Thus

$$R^*(\omega) = \frac{1}{j\omega} \left| \frac{1}{j\omega+1} \right|^2 = \frac{1}{j\omega} \left(\frac{1}{\omega^2+1}\right) \tag{30}$$

The gain factor, K, needed to minimize $\overline{e^2}$ is obvious from
physical considerations; K = 1 is needed because this is the
gain for which $r^*(t)$ approaches the same final value as $r(t)$,
or $+\frac{1}{2}$.

Figure 4 shows $r^*(t)$, and $i(t)$ plotted with its time
origin moved back to t = -1.7 seconds. The rise time of the
system, defined as the time taken for the response to climb
from -0.4 to 0.4, is reduced somewhat by the phase correction;
from 3.25 seconds for $i(t)$ to 3.2 seconds for $r^*(t)$.

It is interesting to notice that $r^*(t)$ is an odd function
of time; $R^*(\omega)$ is purely imaginary. The fact that the correct-
ed response that yields the least integral square error turns
out to be an odd function seems very appropriate. Intuitively,
we might well have expected an odd function because in

$x^*(t)$ {zero-}{rise time 2.2 sec.}
{phase}{no overshoot}

$x(t)$ {minor-}{rise time 2.3 sec.}
{rected}{no overshoot}

Time in seconds
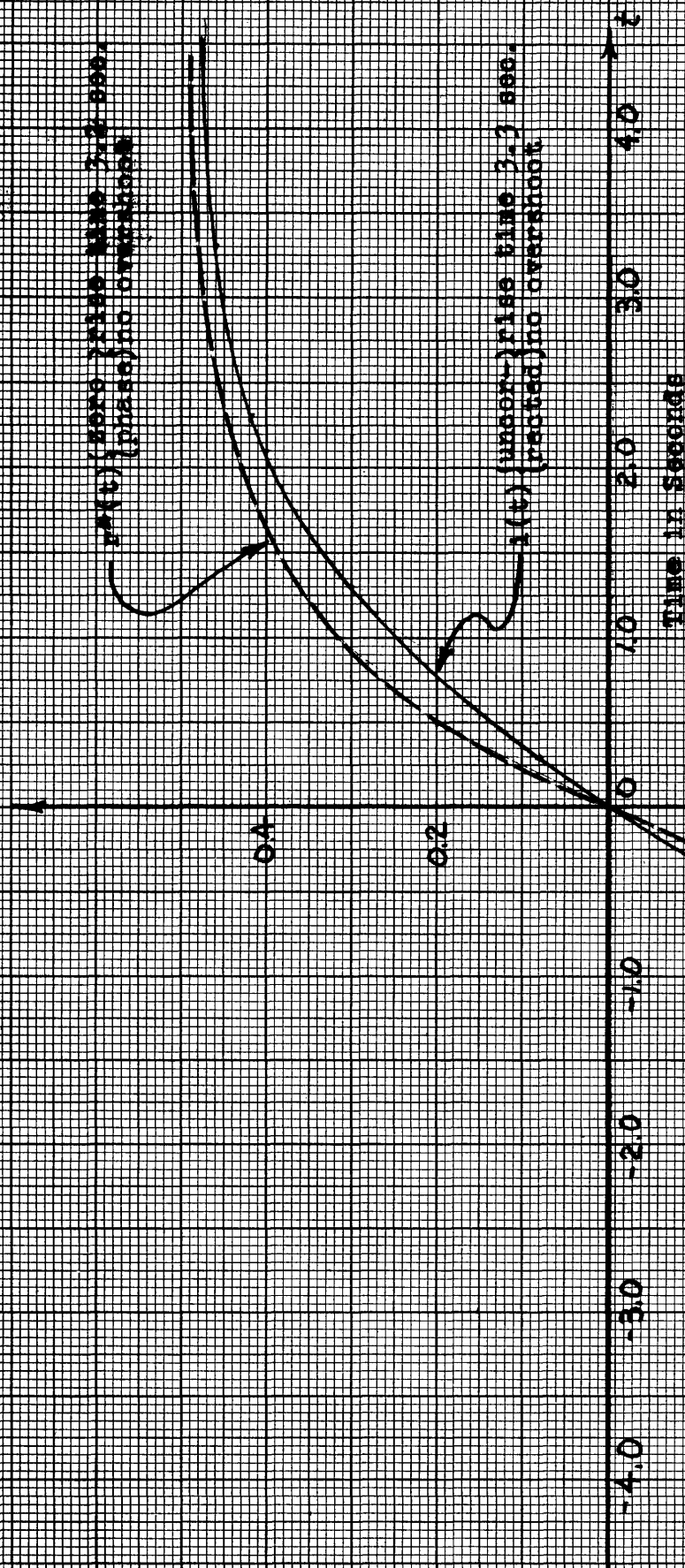
Fig. 4.
Step Response of System with
Transfer Characteristic

$$\frac{1}{(j\omega+1)^2}$$

0.4

0.2

0.2

0.4

1.0

2.0

3.0

4.0

-1.0

-2.0

-3.0

-4.0

1.0

2.0

3.0

4.0

evaluating $\overline{e^2}$, the error is considered as being equally unde-
sirable for both positive and negative times. That is, the
error is given equal weighting under the integral for all times.
Therefore it seems appropriate that the magnitude of the error
be the same for like values of positive and negative time.
Further, if r*(t) had contained both odd and even components,
its specification could not have been unique. Corrected
responses with the transforms $R_1^*(\omega) + jR_2^*(\omega)$ and $R_1^*(\omega) - jR_2^*(\omega)$
have the same integral square error. Neither of these argu-
ments is offered as anything conclusive, but they are intended
only to suggest that we might have guessed that an odd response
would minimize $\overline{e^2}$.

4.2 Extension of the Integral Square Minimization Method

Having solved in Chapter III the general problem of find-
ing the phase which is needed to make a minimum integral square
error correction, and having noted in the first section of
this chapter some of the implications of this solution, let us
now consider for a moment whether the use of the integral
square error criterion is really leading to the most desirable
corrections. Since we are unable to make an errorless trans-
formation of the undesired transient by means of phase correc-
tion, perhaps we should define a more realistic objective. For
the case of correcting the step response of the sluggish system
of section 4.1, we would like to have the shortest rise time
which can be obtained without excessive overshoot. In some
other problem where the step response of the system exhibits

overshoot, it may be that we want to reduce this overshoot by
the phase correction. Thus, while the minimum integral square
error criterion can be considered to give a broadly conceived
"best over-all transient correction", there is no reason to
believe that it leads to the best correction in terms of
decreasing rise time or overshoot. In fact, as we shall see,
it does not. But it is very difficult, if not impossible, to
formulate a mathematical error criterion which guarantees a
corrected step response with shortest rise time, or with
minimum overshoot. However, it is fruitful to look at the
results given by some other error criteria.

A simple extension of the integral square error criterion
proves useful. Although ideally we would like the corrected
step response of the sluggish system of section 4.1 to be a
step function, a better corrected step response can be obtained
by the somewhat devious procedure of designating a function
other than a step as the desired response, and correcting for
minimum integral square deviation from this function.

Suppose that we take as the desired response

$$r(t) = v(t) \tag{31}$$

where $v(t)$ is as shown in Fig. 5. The function $v(t)$ is an odd
function of time which is composed of three line segments, one
of which is tangent at $t = 0$ to the corrected response obtained
for the minimum $\overline{e^2}$ approximation of a step function, the curve
$r^*(t)$ of Fig. 4.

It is convenient to suppose that correction for the
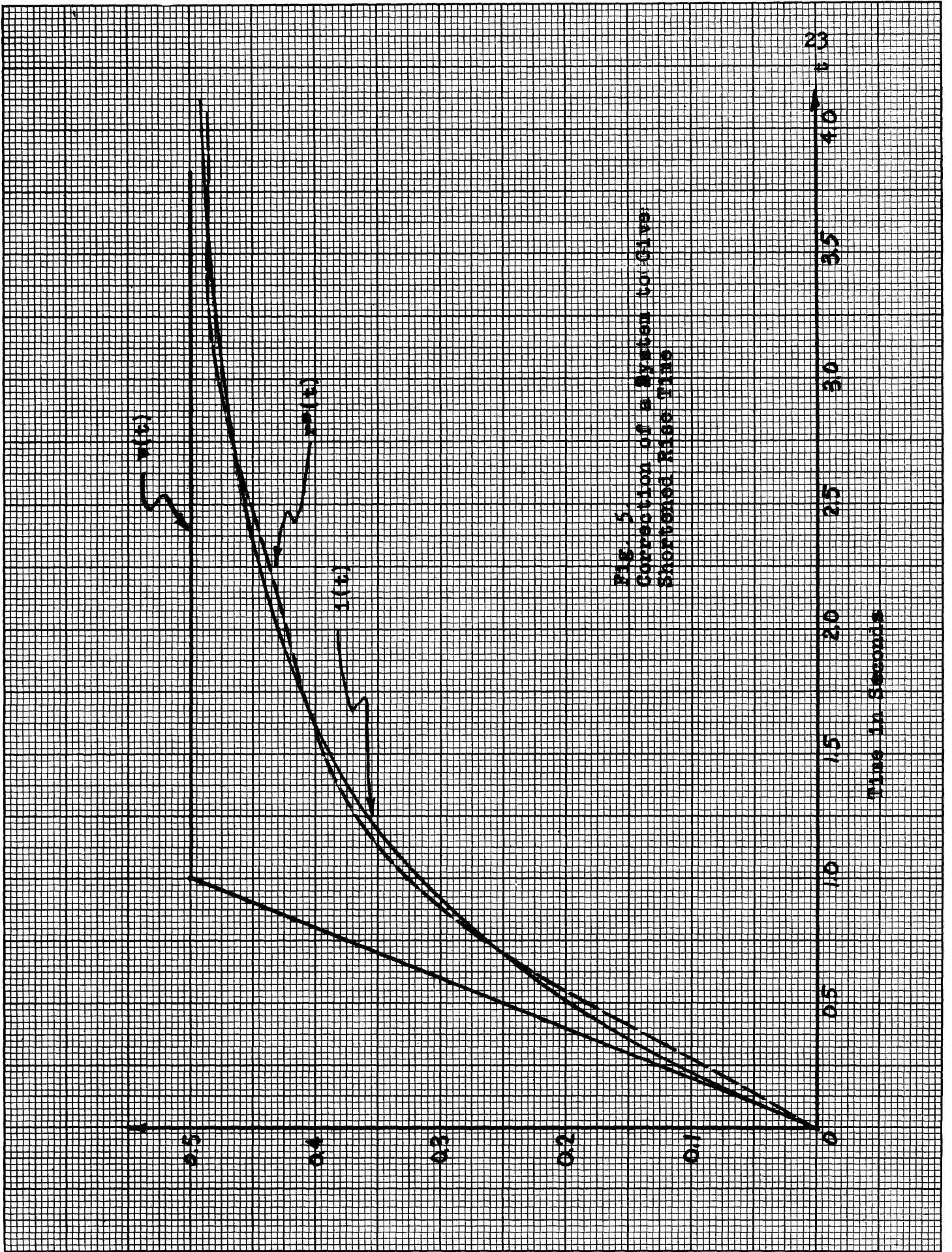
Fig. 5.
Correction of a System to Give
Shortened Rise Time

minimum $\overline{e^2}$ approximation of a step has already been made, and that we are adding additional correction in the form of a network with the transfer characteristic $\epsilon^{-j\Theta(\omega)}$ to further improve the transient response. Thus the input to the phase correction network is the curve $r^*(t)$ of Fig. 4, which has the transform

$$I(\omega) = \frac{1}{j\omega} \frac{1}{\omega^2+1} \tag{32}$$

We must now find the phase correction, $\Theta(\omega)$, which transforms $i(t)$ into the approximation of $v(t)$ having the least integral square error.

The transform of $r(t) = v(t)$ is

$$R(\omega) = \frac{1}{2(j\omega)^2} [\epsilon^{j\omega} - \epsilon^{-j\omega}] = \frac{1}{j\omega} \frac{\sin \omega}{\omega} \tag{33}$$

The phase correction must be chosen so that $R^*(\omega)$ has the same phase as $R(\omega)$. This phase is given by the curve of Fig. 6. The transfer function of the corrected system, $j\omega R^*(\omega)$, is a purely real function of ω and is as shown in Fig. 7.

At this point an argument that $r^*(t)$ must have a shorter rise time than $i(t)$ can be made. The corrected response, $r^*(t)$, is a better approximation, judged by $\overline{e^2}$, of $v(t)$ than is $i(t)$. In the region of time near $t = 0$, $r^*(t)$ has less slope than $i(t)$; the area under the curve of $j\omega R^*(\omega)$ in Fig. 6 is less than the area under a curve of $(\omega^2+1)^{-1}$. The function $v(t)$ is tangent to $i(t)$ at $t = 0$. Thus near $t = 0$, $i(t)$ is a better approximation of $v(t)$ than is $r^*(t)$. It must be that in some
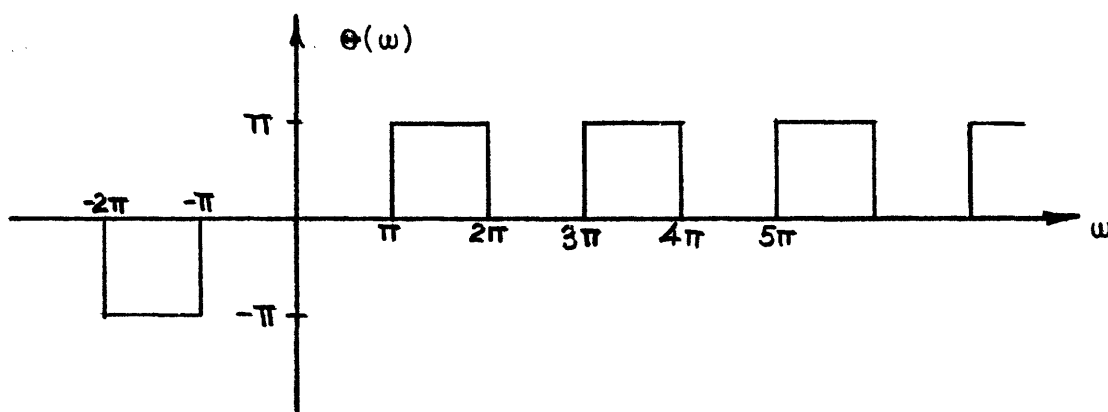
Fig. 6. Phase Correction Needed to Give Minimum $\overline{e^2}$ Approximation of v(t)
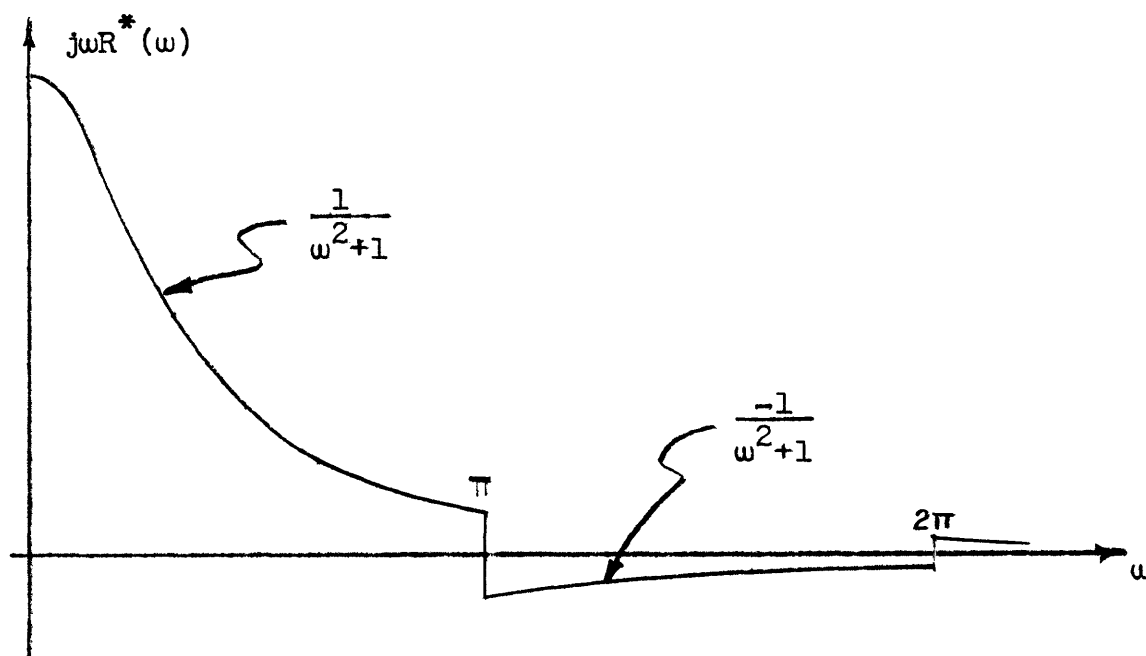
Fig. 7. Transfer Function of System Corrected to Give Approximation of v(t) with Mimimum $\overline{e^2}$

26

other region of time, r*(t) approximates v(t) more closely than does i(t). Looking at the curves of i(t) and v(t) in Fig. 5, it seems most likely that this occurs in the vicinity of t = 1. If this is so, it is evident that r*(t) should have a shorter rise time than i(t).

Having provided a justification of why phase correction designed to minimize integral square deviation from v(t) should give a corrected step response with a shorter rise time than that obtained for phase correction designed to minimize $\overline{e^2}$ deviation from the desired step function, we shall evaluate the response, r*(t), obtained in this manner. The function R*(ω) is shown in Fig. 8. Notice that beyond the first discontinuity in R*(ω) at ω = π, R*(ω) very nearly equals $\pm j\omega^{-3}$. The sum of the functions of Fig. 9(a), (b), (c), ... , closely approximates R*(ω). The inverse transform of the curve of Fig. 9(a) is the step response of the system corrected for minimum $\overline{e^2}$ deviation from a step function, which has already been calculated and plotted in Fig. 4. If the inverse transform of the functions $F_{\omega_0}^{(3)}(\omega)$, where

$$F_{\omega_0}^{(3)}(\omega) = \begin{cases} 0 \ , & |\omega| < \omega_0 \\ \dfrac{1}{(j\omega)^3} \ , & |\omega| > \omega_0 \end{cases} \tag{34}$$

were known for $\omega_0 = \pi$, 2π, ... , the job of calculating r*(t) could be completed rather simply. However, curves of this
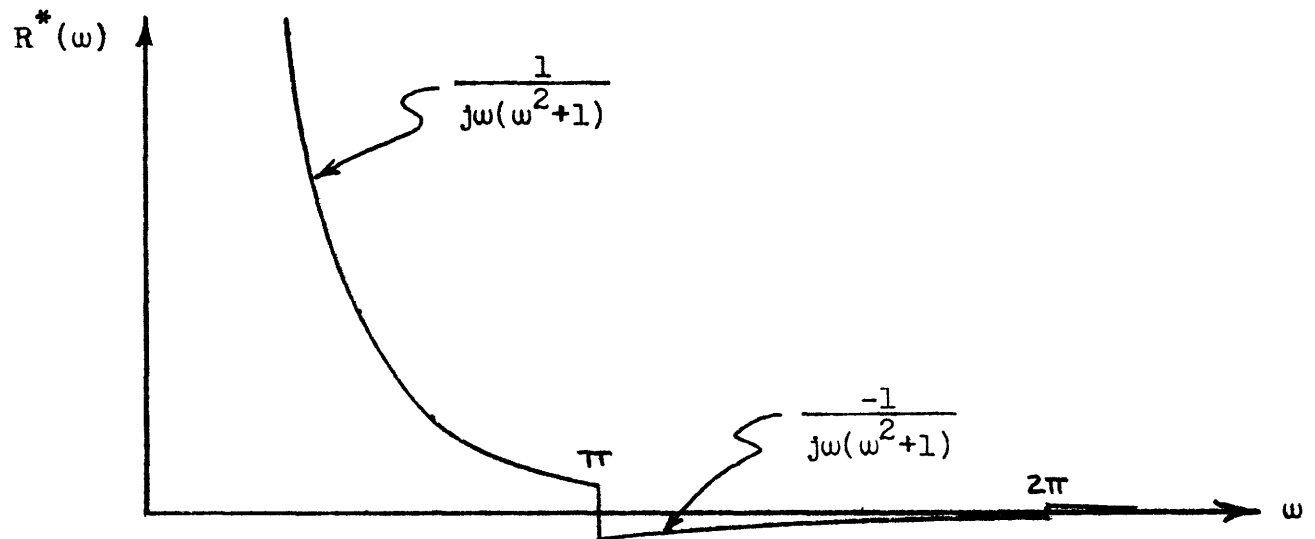
Fig. 8. Plot of $R^*(\omega)$ vs. ω
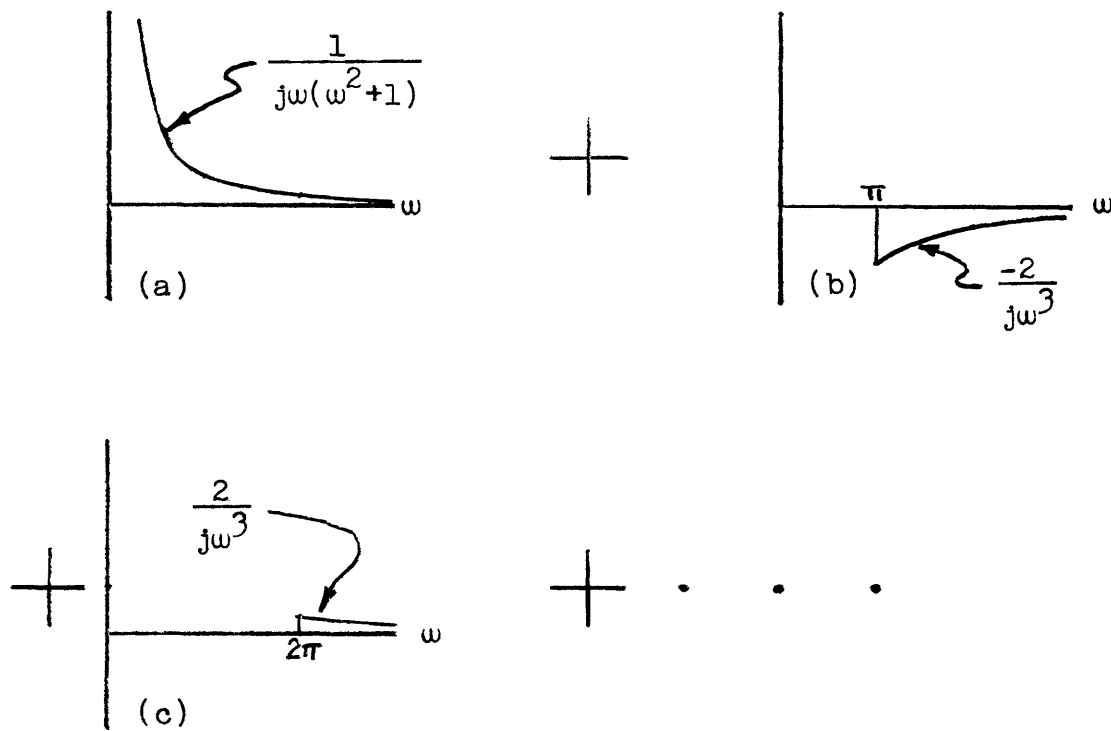
Fig. 9. The Series of Functions Whose Sum
Approximates $R^*(\omega)$

function could not be found in any of the mathematical publications and the task of calculating them had to be undertaken. Curves of the inverse transform of the function

$$F_1^{(n)}(\omega) = \begin{cases} 0 , & |\omega| < 1 \\ \dfrac{1}{(j\omega)^n} , & |\omega| > 1 \end{cases} \tag{35}$$

are presented in Figs. 10 through 16 for values of n from 2 to 8. The curves giving $f_1^{(n)}(t)$ can be extrapolated to larger values of time by noticing that for n even, $f_1^{(n)}(t)$ approaches $(-1)^{\frac{n}{2}} \dfrac{\sin t}{\pi\, t}$ for large t, and that for n odd, $f_1^{(n)}(t)$ approaches $(-1)^{\frac{n-1}{2}} \dfrac{\cos t}{\pi\, t}$ for large t. By a simple normalization of these curves, the inverse transform of $F_{\omega_0}^{(n)}(\omega)$ is found to be

$$f_{\omega_0}^{(n)}(t) = \frac{1}{(\omega_0)^{n-1}} \; f_1^{(n)}(\omega_0 t) \tag{36}$$

The curve of $f_1^{(3)}(t)$ of Fig. 11 is the one of immediate interest; the other curves are plotted for reference and will be used in a later chapter.

The step response of the corrected system, $r^*(t)$, is shown in Fig. 5. Only the first two terms of the infinite series indicated in Fig. 9 were taken. The inverse transform of the function of Fig. 9(b) is the difference between the curves $r^*(t)$ and $1(t)$ of Fig. 5. This difference is small and according to Eq. 36, the amount which would be contributed by

the remaining terms of Fig. 9 is even smaller. Again, as we might have guessed, the correct response is odd. By the additional phase correction $\Theta(\omega)$ of Fig. 6, we have managed to decrease the rise time of the step response from 3.2 seconds to 3.15 seconds with negligible overshoot. Although this improvement in rise time is small, it is now quite clear that correction for minimum $\overline{e^2}$ deviation from a step function does not lead to the shortest possible rise time. Other phase corrections, such as the one presented in this section, should be investigated.
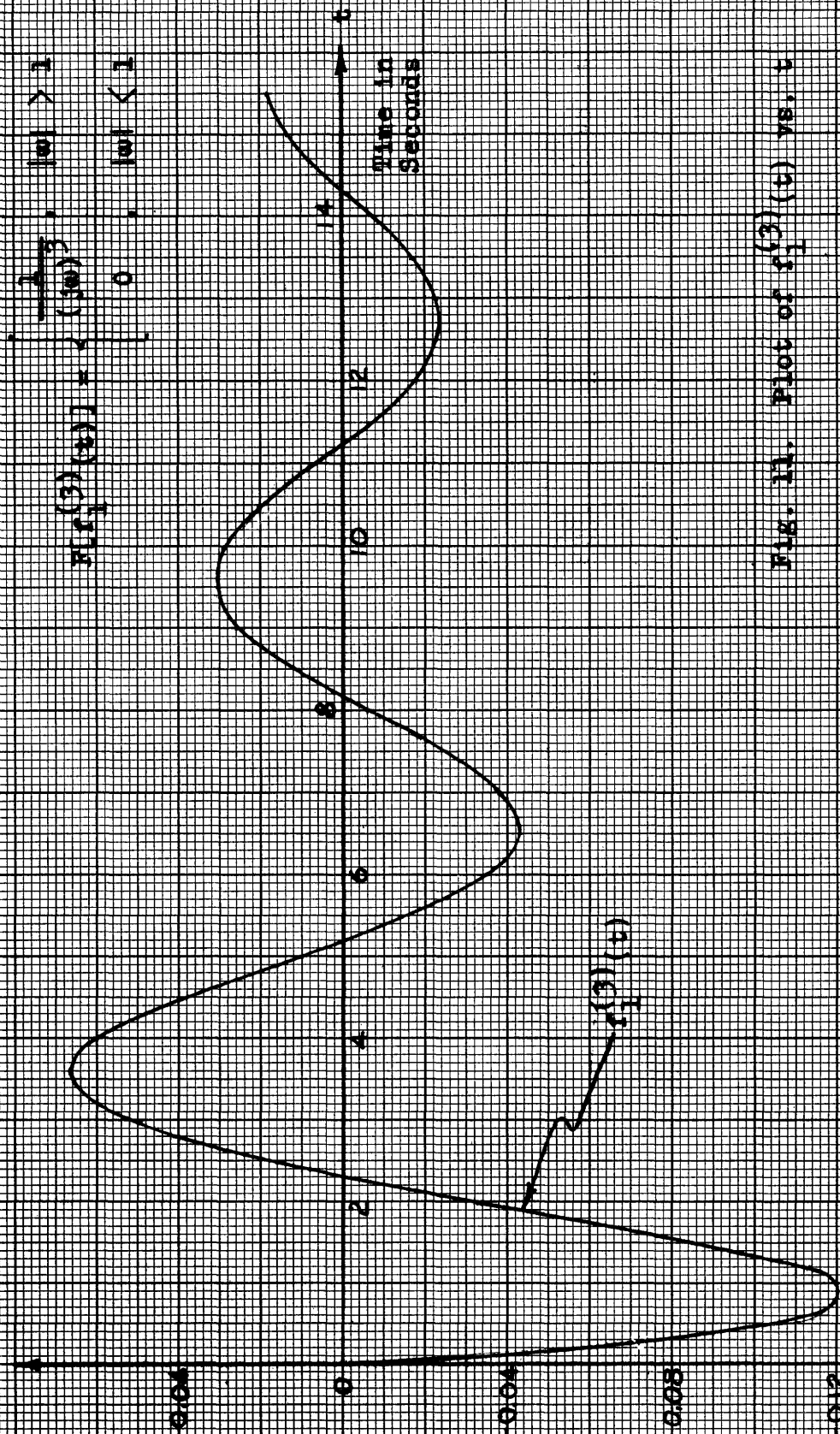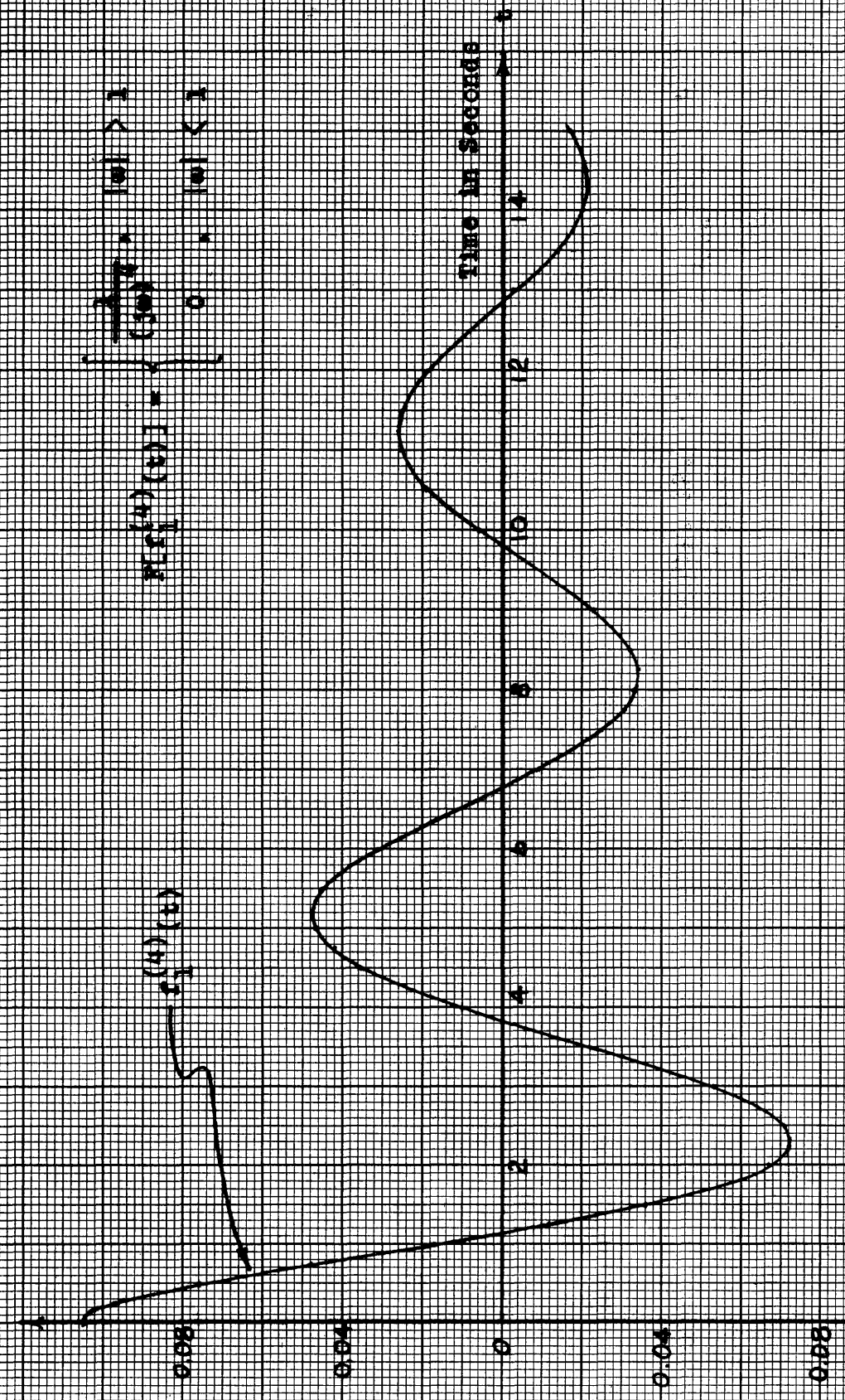
Time in Seconds

$$F[f_{(2)}(t)] = |f_1|$$

$$F[f_{(2)}(t)] = |f_1| = \sqrt{-\frac{1}{2(j\omega)^2}}, \quad |\omega| < 2$$

$$0, \quad |\omega| > 2$$

$f_{(2)}(t)$

Fig. 10. Plot of $f_{(2)}(t)$ of Eq. (4.9a).

Fig. 11. Plot of $f_1^{(3)}(t)$ vs. t

$F[f_1^{(3)}(t)] = \begin{cases} 1, & |\omega| > 1 \\ 0, & |\omega| < 1 \end{cases}$

Time in Seconds

$f_1^{(3)}(t)$

31

$$f_1^{(4)}(t) = \begin{cases} \dfrac{1}{\pi} \sqrt{\dfrac{1+\alpha}{1-\alpha}} & |\alpha| < 1 \\[2mm] 0 & |\alpha| > 1 \end{cases}$$

$f_1^{(4)}(t)$

Time in Seconds

Fig. 12. Plot of $f_1^{(4)}(t)$.

33

Fig. 13. Plot of $f_1^{(5)}(t)$ vs. t

$$F[f_1^{(5)}(t)] = \begin{cases} \dfrac{1}{5}, & |\omega| < 1 \\ 0, & |\omega| > 1 \end{cases}$$

Time in Seconds

$$F_1 f_1^{(6)}(t) = F[f_1^{(6)}(t)] = \begin{cases} \frac{1}{2}, & |m| > 6 \\ 0, & |m| = 6 \\ -\frac{1}{2}, & |m| < 6 \end{cases}$$

$-f_1^{(6)}(t)$

Time in Seconds

0.08   0.04   0   -0.04   -0.08

2   4   6   8   10   12   14
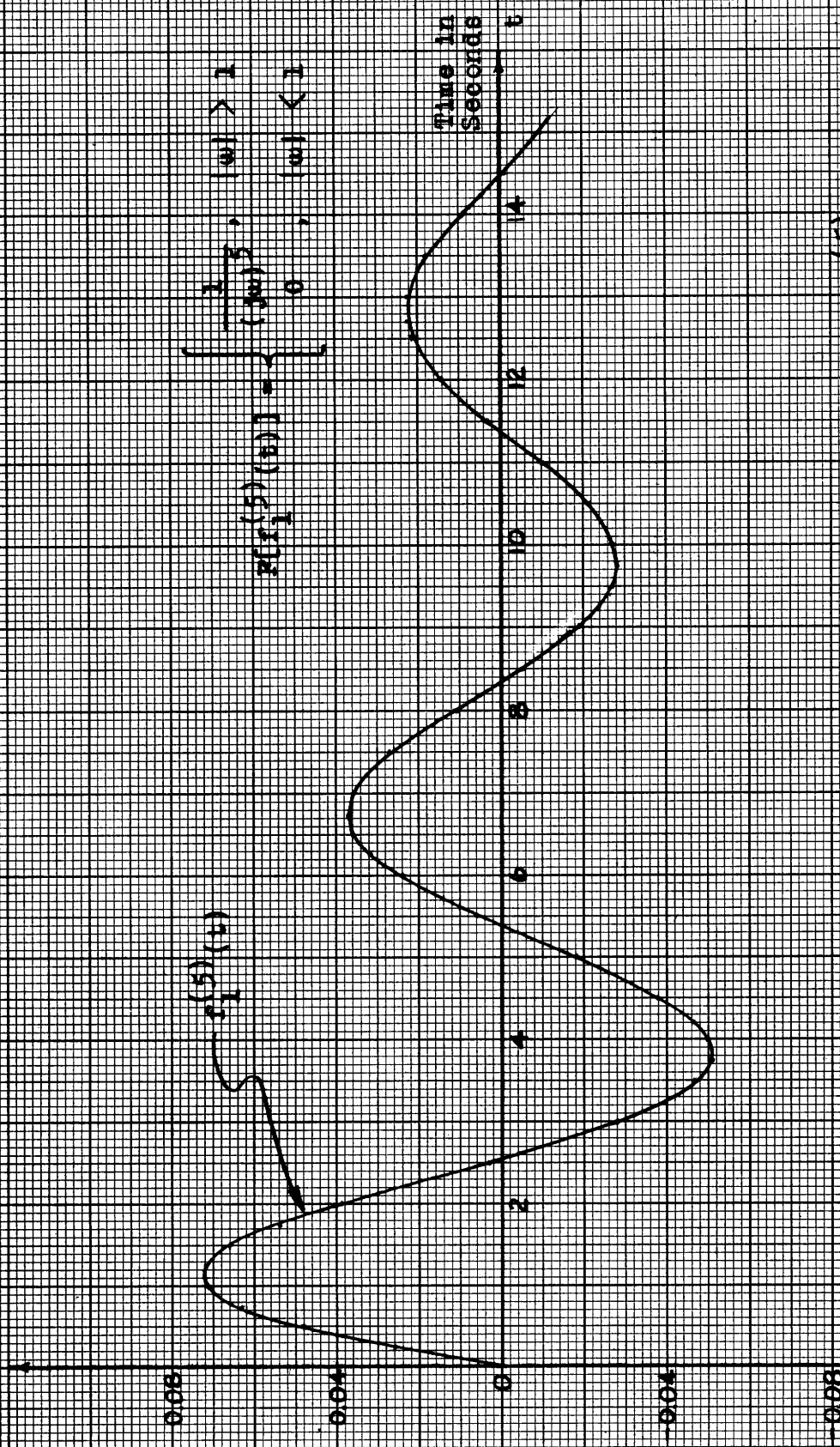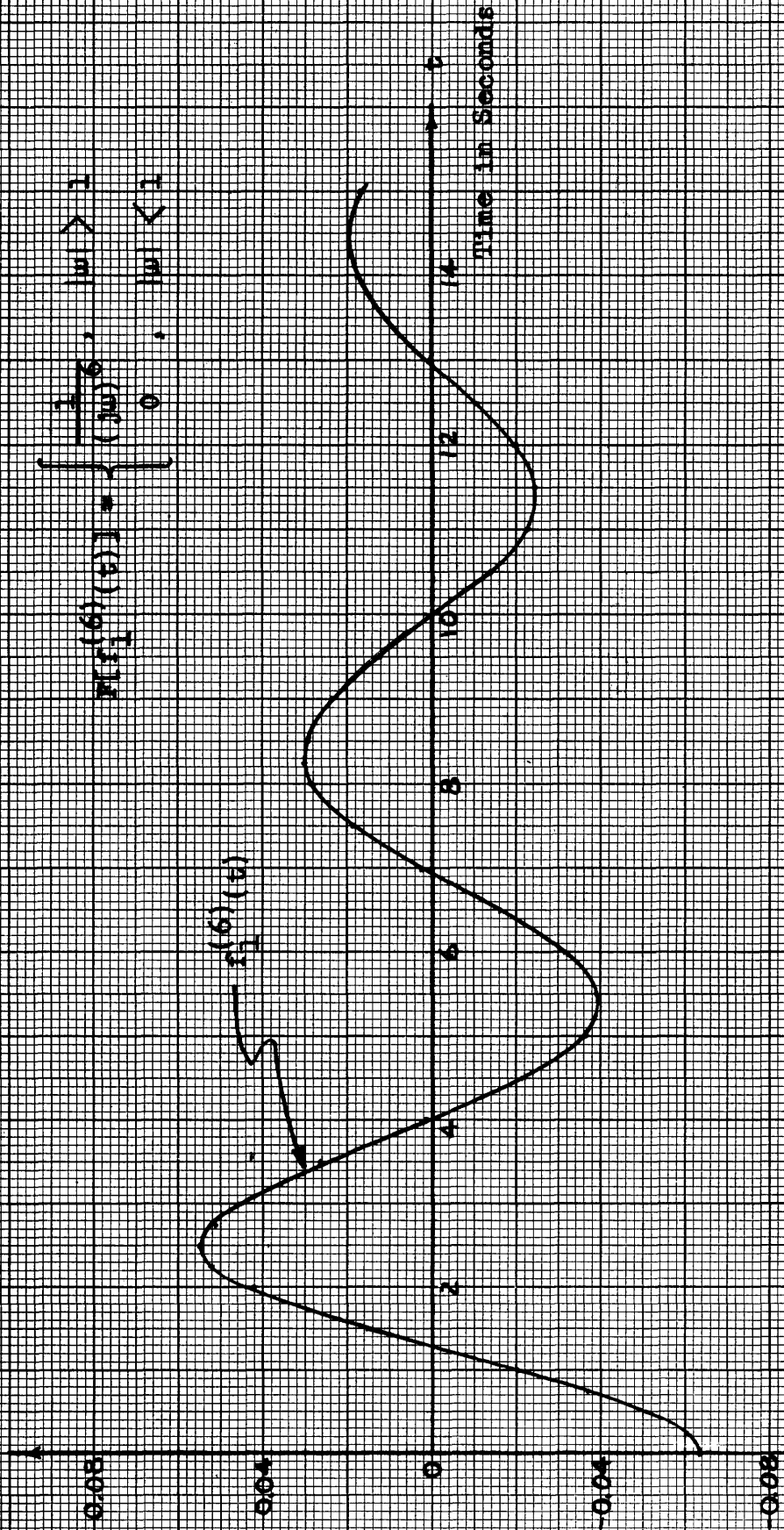
Fig. 14. Plot of $f_1^{(6)}(t)$ vs. t.

$$H\{f_1^{(7)}(t)\} = F_1^{(7)}(\omega) = \begin{cases} \dfrac{1}{(3\omega)^7}, & |\omega| > 7 \\ 0, & |\omega| < 7 \end{cases}$$

$f_1^{(7)}(t)$

Time in Seconds
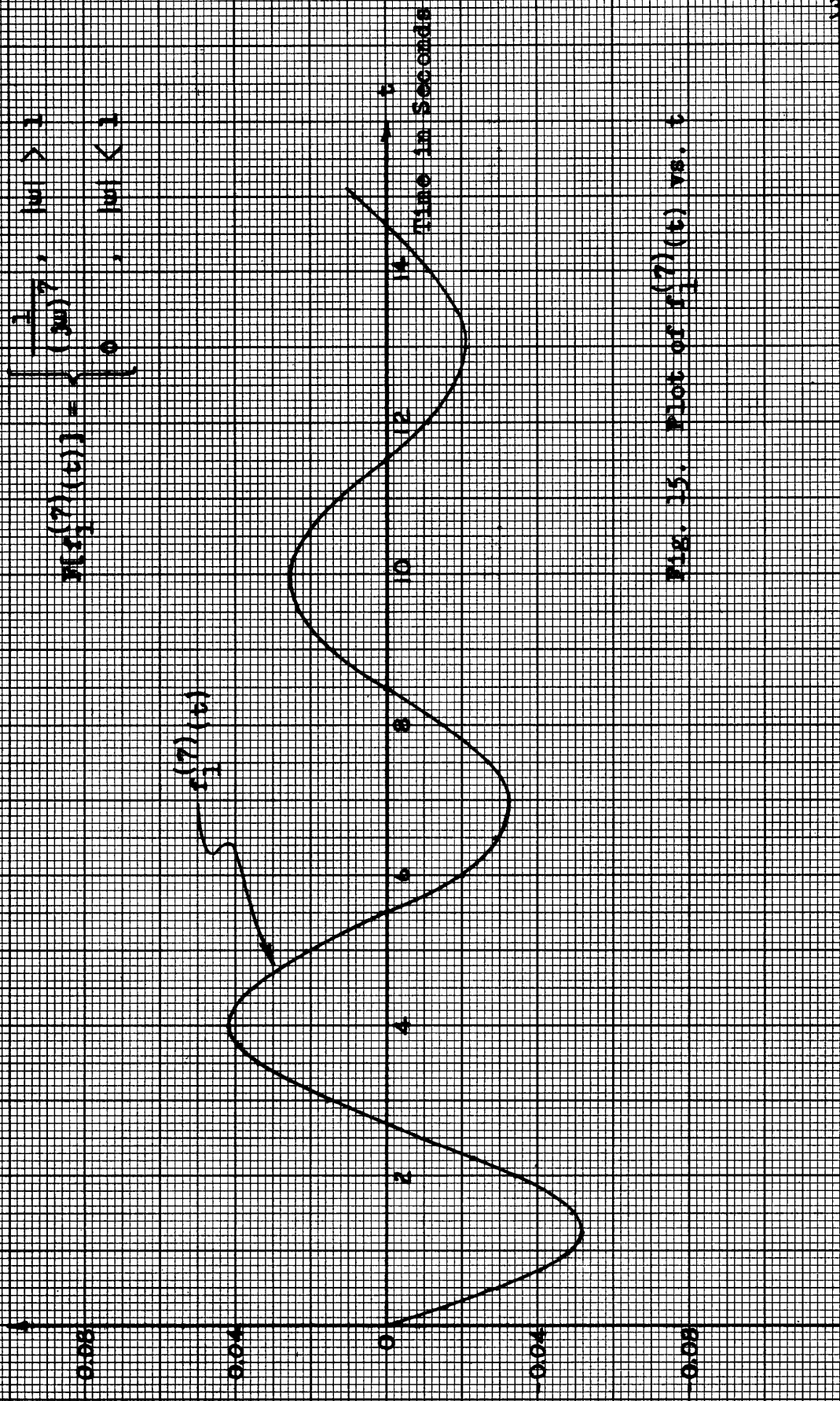
Fig. 15. Plot of $f_1^{(7)}(t)$ vs. t

$$F[f_1^{(8)}(t)] = \begin{cases} \dfrac{1}{(j\omega)} \dfrac{1}{8} \cdot \dfrac{1}{\omega} > 1 \\ 0 \quad , \quad |\omega| < 1 \end{cases}$$

Fig. 16. Plot of $f_1^{(8)}(t)$ vs. t.

36

CHAPTER V

PHASE CORRECTION TO ACHIEVE

TRANSIENT RESPONSE WITH MINIMUM

WEIGHTED INTEGRAL SQUARE ERROR

5.1 Derivation of the Equation Satisfied by the Phase

Correction which Minimizes Weighted Integral Square Error

In section 4.2 it was established that correction for

minimum integral square deviation of the corrected response

from the desired response may not lead to the best step

response in terms of shortened rise time or suppressed over-

shoot. Thus we are led to consider other error criteria. A

simple extension of the integral square error criterion has

been discussed in which the subterfuge was used of considering

the integral square deviation from some function other than

the desired response. This criterion gave a slightly improved

corrected step response for the system of section 4.1. The

principal fault of the extended $\overline{e^2}$ criterion is that it is

difficult to use with much real perception; that is, it is

hard to know how to choose $r(t)$ so as to improve $r^*(t)$. For

example, it was not obvious that by choosing $r(t) = v(t)$, a

better step response would be obtained. The choice of $r(t)$

which is needed to give the best possible corrected step

response is not at all clear.

Considering the above situation, it is evident that the

need is for an error criterion which is more flexible than the

simple integral square error criterion, but one which gives a

38

greater prior indication of the resulting response than does the extended integral square error criterion. A useful criterion is weighted integral square error, $\overline{ae^2}$, where

$$\overline{ae^2} = \int_{-\infty}^{\infty} a(t)[r(t) - r^*(t)]^2 \, dt \tag{37}$$

and where $a(t)$, the weighting function, can be any desired function of time. The weighted integral square error criterion meets the requirements of flexibility and perceptiveness mentioned above. In correcting for minimum $\overline{ae^2}$, the weighting function can be used to gain greatest fidelity in important regions of time, while de-emphasizing errors in less important regions.

If we are to use weighted integral square error as the basis for correction, we must solve the problem of finding the phase correction, $\Theta(\omega)$, and the gain constant, K, which minimize $\overline{ae^2}$ in a given problem. The mathematics of solving this minimization problem are more involved than they were in solving the $\overline{e^2}$ minimization problem of section 3.3, but the ideas are exactly the same. Unfortunately, the resulting two equations which must be satisfied by $\Theta(\omega)$ and K cannot be solved, in general. However, with suitable specialization, the equations yield information which is very useful in making phase corrections.

Having made these preliminary remarks, let us proceed with the solution of the minimization problem. The plan of attack will be first to find the values of K for which the

rate of change of $\overline{ae^2}$ with respect to K is zero, then to eliminate K and find, by the calculus of variations technique, the equation satisfied by the extremals, $\Theta(\omega)$, of $\overline{ae^2}$. Equation 37 may be written

$$\overline{ae^2} = \int_{-\infty}^{\infty} a(t)[e(t)]^2 \, dt \qquad (38)$$

If Eq. 38 is written in terms of $E(\omega)$, the transform of $e(t)$, the result is

$$\overline{ae^2} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} a(t) \left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(x)E(y)\epsilon^{j(x+y)t} \, dx \, dy \right] dt \qquad (39)$$

Interchanging orders of integration

$$\overline{ae^2} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(x)E(y) \int_{-\infty}^{\infty} a(t)\epsilon^{j(x+y)t} \, dt \, dx \, dy \qquad (40)$$

$$= \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E(x)E(y) \, A(-x \, -y) \, dx \, dy \qquad (41)$$

Written in terms of $\Theta(\omega)$ and K, Eq. 41 becomes

$$\overline{ae^2} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y) \, [K\epsilon^{-j\Theta(x)} - G(x)]$$

$$[K\epsilon^{-j\Theta(y)} - G(y)] \, A(-x \, -y) \, dx \, dy \qquad (42)$$

In order to minimize $\overline{ae^2}$ with respect to K, we set

$$O = \frac{d \; \overline{ae^2}}{dK} \tag{43}$$

$$O = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y) \; A(-x \; -y)\epsilon^{-j\theta(x)} \; [K\epsilon^{-j\theta(y)} - G(y)]dx \; dy \tag{44}$$

Solving (44) for K

$$K = \frac{N}{D} \tag{45}$$

where

$$N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y) \; A(-x \; -y)\epsilon^{-j\theta(x)} \; G(y) \; dx \; dy \tag{46}$$

$$D = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y) \; A(-x \; -y)\epsilon^{-j\theta(x)} \; \epsilon^{-j\theta(y)} \; dx \; dy \tag{47}$$

By means of Eqs. 45, 46 and 47, Eq. 42 may be written

$$\overline{ae^2} = \frac{1}{4\pi^2} [K^2 D - 2KN + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y) \; A(-x \; -y)G(x)G(y) \; dx \; dy] \tag{48}$$

$$\overline{ae^2} = \frac{1}{4\pi^2} [\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x)I(y)A(-x-y)G(x)G(y) \; dx \; dy - \frac{N^2}{D}] \tag{49}$$

Now, having an expression for $\overline{ae^2}$ from which K has been elim-
inated, we can find the extremals of this integral. In order
to simplify the writing of the equations which are to follow,
the definitions will be adopted

$$\delta \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[x,y,\Theta(x),\Theta(y)] \, dx \, dy$$

$$= \frac{d}{d\alpha} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[x,y,\Theta(x) + \alpha\lambda(x), \Theta(y) + \alpha\lambda(y)] \, dx \, dy \bigg|_{\alpha = 0}$$

$$(50)$$

$$\delta_x \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[x,y,\Theta(x), \Theta(y)] \, dx \, dy$$

$$= \frac{d}{d\alpha} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P[x,y,\Theta(x) + \alpha\lambda(x), \Theta(y)] \, dx \, dy \bigg|_{\alpha = 0} \qquad (51)$$

Extremals of $\overline{ae^2}$ must satisfy the equation

$$0 = \delta\,(\overline{ae^2}) \qquad (52)$$

From Eq. 49

$$0 = \delta\,(\frac{N^2}{D}) = 2\,\frac{N}{D}\delta N - \frac{N^2}{D^2}\delta D \qquad (53)$$

Obviously

$$0 = \frac{N}{D} = K \qquad (54)$$

is not the solution of Eq. 53 which is of interest. Thus

$$0 = 2\,\delta N - K\,\delta D \qquad (55)$$

Noticing that

$$\delta N = \delta_x\,N \qquad (56)$$

$$\delta D = \delta_x\,D + \delta_y D = 2\,\delta_x\,D \qquad (57)$$

Eq. 55 becomes

$$0 = \delta_x N - K \delta_x D \tag{58}$$

Now the quantities $\delta_x N$ and $\delta_x D$ must be evaluated

$$\delta_x N = -j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x)I(y)A(-x-y)\lambda(x)\epsilon^{-j\Theta(x)}G(y) \, dx \, dy \tag{59}$$

$$\delta_x D = -j \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x)I(y)A(-x-y)\lambda(x)\epsilon^{-j\Theta(x)}\epsilon^{-j\Theta(y)} \, dx \, dy \tag{60}$$

Substituting Eqs. 59 and 60 into Eq. 58

$$0 = \int_{-\infty}^{\infty} I(x)\lambda(x)\epsilon^{-j\Theta(x)} \left\{ \int_{-\infty}^{\infty} I(y)A(-x-y)[G(y) - K\epsilon^{-j\Theta(y)}] \, dy \right\} \, dx \tag{61}$$

Grouping some of the terms and substituting ω for the dummy variable x

$$0 = \int_{-\infty}^{\infty} R^*(\omega)\lambda(\omega) \left\{ \int_{-\infty}^{\infty} A(-\omega-y)[R(y) - R^*(y)] \, dy \right\} \, d\omega \tag{62}$$

Since $\lambda(\omega)$ can be chosen as any odd function of ω, if Eq. 62 is to be satisfied for all possible choices of $\lambda(\omega)$ it must be that

$$0 = Im[R^*(\omega) \int_{-\infty}^{\infty} A(-\omega-y)(R(y)-R^*(y)) \, dy] \tag{63}$$

Then

$$\arg R^*(\omega) + \arg \int_{-\infty}^{\infty} A(-\omega-y)[R(y) - R^*(y)] \, dy = n(\omega)\pi \qquad (64)$$

where $n(\omega)$ is any odd function of ω whose value is always an integer except at discontinuities. The usual abbreviation for argument, arg, is used in Eq. 64; thus $\arg [c\epsilon^{jd}]$ is d. It follows from Eq. 64 that

$$0 = \arg [(-1)^{n(\omega)}R^*(\omega)] + \arg \int_{-\infty}^{\infty} A(-\omega-y)[R(y)-R^*(y)] \, dy \qquad (65)$$

Since the second term of Eq. 65 could represent the phase of a real time function, this term must be an odd function of ω. Making use of this fact

$$\arg [(-1)^{n(\omega)}R^*(\omega)] = \arg \int_{-\infty}^{\infty} A(\omega-y)[R(y)-R^*(y)] \, dy \qquad (66)$$

There is a question as to the choice of the function $n(\omega)$, for it is unspecified as yet. Remember that an integral may have a number of extremals, some of which yield maxima, some minima, and some neither. The choice of $n(\omega)$ determines which of these extremals satisfies Eq. 66. But we do not know which choice of $n(\omega)$ allows the extremal that minimizes $\overline{ae^2}$ to satisfy the equation.

There is something else which must be noted about Eq. 66. Extremals of $\overline{ae^2}$ must satisfy Eq. 66, but the equation, being a non-linear integral equation, cannot be solved, in general, to find these extremals. The analysis to find the phase correction which minimizes weighted integral square error has

been carried out, but the equation which results cannot be
solved when it is formulated in completely general terms. How-
ever, for the particular cases where phase correction is
usually most desired, specialization of Eq. 66 simplifies it
sufficiently so that a solution can be obtained.

To preserve complete generality in this section, the gain
factor K was allowed to vary in order to find the value which
minimizes $\overline{ae^2}$. Equation 45 gives this value. However, in a
large number of the practical problems of phase correction,
the necessary gain factor is obvious from physical considera-
tions. In order to simplify the notation in the succeeding
sections of this chapter, it will be assumed that the desired
response has been normalized so that K = 1.

5.2 Minimizing $\overline{ae^2}$ When the Desired Response Is Either Odd or
   Even

Since Eq. 66, which must be satisfied by the phase that
minimizes $\overline{ae^2}$, cannot be solved in its most general form, let
us make some appropriate specializations which make solution
possible. In most phase correction problems, the desired
response, r(t), is a step function, which is a purely odd
function of time. Occasionally, the desired response may be
an even function of time, such as a pulse. Thus, desired
responses are mostly either odd or even functions of time. If
the desired response is an odd or an even time function, the
weighting function, a(t), should be chosen as an even time
function, since errors at like values of time before and

after time t = 0 are equally undesirable. If we adopt the convention that the subscripts 1 and 2 are associated with the even and odd parts of a time function, or the real and imaginary parts of a frequency function, the special cases of r(t) odd or even may be written

r(t) odd:
$$r(t) = r_2(t) \tag{67}$$

$$R(\omega) = jR_2(\omega) \tag{68}$$

$$a(t) = a_1(t) \tag{69}$$

$$A(\omega) = A_1(\omega) \tag{70}$$

$$\overline{a_1 e^2} = \int_{-\infty}^{\infty} a_1(t) \, [r_2(t) - r_2^*(t)]^2 \, dt + \int_{-\infty}^{\infty} a_1(t)[r_1^*(t)]^2 \, dt \tag{71}$$

or

r(t) even:
$$r(t) = r_1(t) \tag{72}$$

$$R(\omega) = R_1(\omega) \tag{73}$$

$$a(t) = a_1(t) \tag{74}$$

$$A(\omega) = A_1(\omega) \tag{75}$$

$$\overline{a_1 e^2} = \int_{-\infty}^{\infty} a_1(t)[r_1(t) - r_1^*(t)]^2 \, dt + \int_{-\infty}^{\infty} a_1(t)[r_2^*(t)]^2 \, dt \tag{76}$$

Let us specialize Eq. 66 for the case where $r(t)$ is an odd time function, such as a step function; notice that a parallel development for the case of $r(t)$ even could be made throughout. If $r(t)$ is odd, Eq. 66 becomes

$$\arg\left\{(-1)^{n(\omega)}[R_1^*(\omega) + jR_2^*(\omega)]\right\}$$

$$= \arg\int_{-\infty}^{\infty} A_1(\omega-y)[jR_2(y)-R_1^*(y)-jR_2^*(y)]\,dy \qquad (77)$$

The requirement of Eq. 77 may be stated as two requirements: (1) the imaginary part of $(-1)^{n(\omega)}R^*(\omega)$ must have the same algebraic sign as that of the integral expression of Eq. 77 and, (2) the ratio of the real and imaginary parts of $(-1)^{n(\omega)}R^*(\omega)$ must be the same as the ratio of the real and imaginary parts of the integral expression. These requirements may be written

$$(-1)^{n(\omega)}R_2^*(\omega) \quad \text{with same sign as}$$

$$\int_{-\infty}^{\infty} A_1(\omega-y)[R_2(y)-R_2^*(y)]\,dy \qquad (78)$$

and

$$\frac{R_1^*(\omega)}{R_2^*(\omega)} = \frac{-\int_{-\infty}^{\infty} A_1(\omega-y)R_1^*(y)\,dy}{\int_{-\infty}^{\infty} A_1(\omega-y)[R_2(y)-R_2^*(y)]\,dy} \qquad (79)$$

The extremals of $\overline{ae^2}$ are those functions, $\Theta(\omega)$, that

satisfy Eqs. 78 and 79, where the relationships

$$R_1^*(\omega) = \text{Re} \left[I(\omega) \epsilon^{-j\Theta(\omega)} \right] \tag{80}$$

and

$$R_2^*(\omega) = \text{Im} \left[I(\omega) \epsilon^{-j\Theta(\omega)} \right] \tag{81}$$

are implied, of course. At the moment, however, the corrected responses which correspond to these extremals, rather than the extremals themselves, are of primary concern. Therefore, let us consider solving Eqs. 78 and 79 for $R_1^*(\omega)$ and $R_2^*(\omega)$ directly.

Since we have complete freedom in choosing the function $n(\omega)$, Eq. 78 can always be satisfied for any $R^*(\omega)$ and the equation really has no meaning. Thus, Eq. 79 is equivalent to Eq. 77, but with the troublesome function $n(\omega)$ eliminated.

At this point, it may appear doubtful that we have made any real progress toward solving the minimization problem by specialization of Eq. 66 for we are still faced with a non-linear integral equation, Eq. 79. Nevertheless, Eq. 79 has an obvious solution

$$R_1^*(\omega) = 0 \tag{82}$$

From Eq. 82 we see that any odd frequency function with the magnitude, $|I(\omega)|$, satisfies Eq. 79. Thus

$$R^*(\omega) = jR_2^*(\omega) = j(-1)^{m(\omega)} |I(\omega)| \tag{83}$$

where $m(\omega)$ may be chosen as any odd frequency function whose
value is always an integer, except at discontinuities. It
should be noted that there may very well be other solutions of
Eq. 79 which are not given by Eq. 83; however, we are unable
to solve Eq. 79 to find them. One of the solutions of Eq. 79,
either a solution given by Eq. 83, or possibly some other
solution, is the transform of the corrected response that
minimizes $\overline{a_1 e^2}$.

On the basis of our past experience with the integral
square error criterion, we might guess that one of the
solutions given by Eq. 83 minimizes $\overline{a_1 e^2}$; this would mean that
the corrected response that best approximates an odd desired
response, judged by the weighted integral square error
criterion, is an odd function. However, this conclusion is
very difficult to support with proof. Remember that extremals
of an integral give stationary values of that integral. These
stationary values may be maxima, minima, or neither. When
given a complete set of extremals of an integral, in order to
find which of the extremals produces the minimum value of the
integral, we simply substitute each extremal into the integral
expression and note which one yields the smallest value. The
difficulty in trying to prove that one of the functions given
by Eq. 83 minimizes $\overline{a_1 e^2}$ lies in the fact that we are not sure
that we know a complete set of solutions of Eq. 79. It is
conceivable that there might exist a solution of Eq. 79, not
given by Eq. 83, which minimizes $\overline{a_1 e^2}$.

We can show, however, that in all cases of any practical

interest a purely odd corrected response does minimize $\overline{a_1 e^2}$.
Because of the difficulties explained above, the proof of this
fact is rather involved and has been relegated to the Appendix.

5.3 Further Consideration of the $\overline{ae^2}$ Minimization Problem

Let us turn again to consideration of the general problem
of minimizing $\overline{ae^2}$ that is discussed in section 5.1. In that
section we attempted to find those phase functions, $\Theta(\omega)$, that
are extremals of the integral expression in Eq. 42, or $\overline{ae^2}$.
In this section we wish to find the corrected response time
functions, $r^*(t)$, that are the extremals of the integral
expression in Eq. 37, or $\overline{ae^2}$, subject to the constraint that

$$\left| \int_{-\infty}^{\infty} r^*(t)\, \epsilon^{-j\omega t}\, dt \right| = |I(\omega)| \tag{84}$$

To make the distinction clear, let us call the former, phase
extremals of $\overline{ae^2}$, and the latter, response extremals of $\overline{ae^2}$.
We expect to find a close relationship between the response
extremals and the corrected responses related to the phase
extremals by

$$R^*(\omega) = I(\omega)\, \epsilon^{-j\Theta(\omega)} \tag{85}$$

Certainly we know that the corrected response related to the
phase extremal that minimizes $\overline{ae^2}$ must be the same as the
response extremal which minimizes $\overline{ae^2}$. The discussion in this
section will show that every response extremal is identical to
a corrected response related to some phase extremal by Eq. 84,

but that the converse of this statement is not true. In other words, we will show that the number of phase extremals exceeds the number of response extremals. The response extremals must satisfy Eq. 66, but must satisfy some other restriction as well. It is worthwhile to investigate the response extremals simply because they are fewer in number than the phase extremals. Thus the task of recognizing the one function that minimizes $\overline{ae^2}$ from among all of the extremals is made easier.

Let us recall the definition of an extremal of an integral. An extremal is a function which has the property that an infinitesimal change in the form of the function makes no change in the value of the integral. Let us examine the procedure used in finding the extremals of an integral by reviewing the derivation of section 5.1 where we attempted to find the phase extremals of $\overline{ae^2}$. We allowed an infinitesimal change, or a variation, $\alpha \lambda(\omega)$, in the form of $\Theta(\omega)$. The function $\lambda(\omega)$ can be any function that is an allowable variation of $\Theta(\omega)$; thus $\lambda(\omega)$ can be any odd function of frequency. We made the substitution

$$\Theta(\omega) \longrightarrow \Theta(\omega) + \alpha \lambda(\omega) \tag{86}$$

in an integral expression giving $\overline{ae^2}$ in terms of $\Theta(\omega)$. Then, by considering the change in the value of $\overline{ae^2}$ caused by an infinitesimal change in α from $\alpha = 0$ for all of the possible $\lambda(\omega)$ functions, we derived Eq. 66 which must be satisfied by the phase extremals.

Let us now consider the problem of finding the response

extremals of $\overline{ae^2}$. Suppose that we allow a variation in $r^*(t)$ of the form $pq(t)$. The function $q(t)$ can be any function that satisfies the relation

$$\left| \int_{-\infty}^{\infty} [r^*(t) + pq(t)]\epsilon^{-j\omega t} \, dt \right| = |I(\omega)| \tag{87}$$

for infinitesimal values of p. Then we make the substitution

$$r^*(t) \longrightarrow r^*(t) + pq(t) \tag{88}$$

in an integral expressing $\overline{ae^2}$ in terms of $r^*(t)$. By considering the changes in the value of $\overline{ae^2}$ caused by an infinitesimal change in p from $p = 0$ for all of the allowable $q(t)$ functions, we derive the requirements to be satisfied by the response extremals of $\overline{ae^2}$. However, it is very difficult to make this derivation due to the difficulty in specifying the allowable $q(t)$ functions. Rather than attempting this derivation, let us seek another approach which avoids this difficulty.

Notice what a variation in $\Theta(\omega)$, $\alpha \lambda(\omega)$, means in terms of a variation of $r^*(t)$, $pq(t)$.

$$r^*(t) = F^{-1}\left\{ I(\omega)\epsilon^{-j\Theta(\omega)} \right\} \tag{89}$$

Making the substitution indicated by Eq. 86, where α is infinitesimal

$$r^*(t) \longrightarrow F^{-1}\left\{ I(\omega)\epsilon^{-j\Theta(\omega)}\epsilon^{-j\alpha \lambda(\omega)} \right\} \tag{90}$$

$$r^*(t) \longrightarrow F^{-1} \left\{ I(\omega) \epsilon^{-j\Theta(\omega)} [1 - j\alpha \lambda(\omega)] \right\} \tag{91}$$

$$r^*(t) \longrightarrow r^*(t) + \alpha F^{-1}[-j\lambda(\omega)R^*(\omega)] \tag{92}$$

Thus we can make a variation of $r^*(t)$ by making a variation of $\Theta(\omega)$. Every variation of $\Theta(\omega)$, $\alpha \lambda(\omega)$, exactly corresponds to a variation of $r^*(t)$

$$pq(t) = \alpha F^{-1}[-j\lambda(\omega)R^*(\omega)] \tag{93}$$

for the infinitesimal values of α and p that are of interest. We deduce that every response extremal is identical to a corrected response related to some phase extremal by Eq. 85. Thus the response extremals must satisfy Eq. 66 just as the phase extremals must satisfy Eq. 66.

We should now investigate the generality of the variation of $r^*(t)$ that can be accomplished by means of a phase variation. Can every possible variation, $pq(t)$, subject to the constraint of Eq. 87, be accomplished by introduction of a variation of phase? As we shall see, the answer is no. We can demonstrate this by consideration of an example. Consider two purely odd response functions $r_2^*(t)$ and $r_2^{*'}(t)$. The transforms of the two responses, $jR_2^*(\omega)$ and $jR_2^{*'}(\omega)$, have the same magnitude. Both $R_2^*(\omega)$ and $R_2^{*'}(\omega)$ are discontinuous functions, and $R_2^{*'}(\omega)$ has points of discontinuity that are

located at frequencies which differ from the $R_2^*(\omega)$ discontinuity location frequencies by an infinitesimal amount. We can take as an allowable variation of $r_2^*(t)$ the purely odd, infinitesimal time function

$$pq(t) = r_2^{*'}(t) - r_2^*(t) \tag{94}$$

However, we notice that, according to Eq. 93, for a function $r^*(t)$ which is purely odd, we can create only even variations through a variation of phase, $\alpha\,\lambda(\omega)$. Thus we are unable to create all allowable response variations, $pq(t)$, by means of the phase variation, $\alpha\,\lambda(\omega)$.

Although the phase variation, $\alpha\,\lambda(\omega)$, fails to provide a general response variation, the concept of using phase manipulations to create response variations is very attractive. The fulfillment of the magnitude constraint in Eq. 87 is guaranteed. The phase variation, $\alpha\,\lambda(\omega)$, fails to provide a general response variation because it cannot shift the discontinuity locations of $\Theta(\omega)$. Thus we are led to consider the substitution

$$\Theta(\omega) \longrightarrow \Theta[\omega + \beta\mu(\omega)] \tag{95}$$

where $\Theta(\omega)$ may have discontinuities and $\mu(\omega)$ is any continuous odd function. Making the substitution for $\Theta(\omega)$ indicated in Eq. 95, the response extremal becomes

$$r^*(t) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{\infty} I(\omega)\epsilon^{-j\Theta[\omega+\beta\mu(\omega)]}\epsilon^{j\omega t}\,d\omega \tag{96}$$

Since β is an infinitesimal quantity, we can make the change of variable

$$x = \omega + \beta\mu(\omega) \tag{97}$$

$$\omega = x - \beta\mu(x) \tag{98}$$

Equation 96 becomes

$$r^*(t) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{\infty} I[x-\beta\mu(x)]\epsilon^{-j\theta(x)}\epsilon^{+j[x-\beta\mu(x)]t}\,dx \tag{99}$$

$$r^*(t) \longrightarrow \frac{1}{2\pi}\int_{-\infty}^{\infty} [I(x) - \beta\mu(x)\frac{d}{dx} I(x)]\epsilon^{-j\theta(x)}\epsilon^{jxt}[1-j\beta\mu(x)t]\,dx \tag{100}$$

$$r^*(t) \longrightarrow r^*(t) + \beta\left\{ -F^{-1}[\mu(\omega)\epsilon^{-j\theta(\omega)}\frac{d}{d\omega} I(\omega)] \right.$$

$$\left. - \frac{d}{dt} F^{-1}[\mu(\omega)\epsilon^{-j\theta(\omega)}I(\omega)]\right\} \tag{101}$$

Thus the substitution for $\theta(\omega)$ that is indicated in Eq. 95 creates the response variation

$$pq(t) = \beta\left\{ F^{-1}[\mu(\omega)\epsilon^{-j\theta(\omega)}\frac{d}{d\omega} I(\omega)] - \frac{d}{dt} F^{-1}[\mu(\omega)\epsilon^{-j\theta(\omega)}I(\omega)]\right\} \tag{102}$$

It will be noticed, however, that this response variation is not completely general, just as the variation in Eq. 93 is not.

Consider the case where both i(t) and r*(t) are odd time functions; the variation given by Eq. 102 can only be purely odd. But according to Eq. 93, even variations are also allowable in this case.

A completely general response variation can be created by means of phase manipulation by the substitution for $\Theta(\omega)$

$$\Theta(\omega) \longrightarrow \Theta[\omega + \beta\mu(\omega)] + \alpha \lambda(\omega) \qquad (103)$$

in Eq. 42, giving $\overline{ae^2}$. The response extremals of $\overline{ae^2}$ must satisfy both of the equations

$$\left. \frac{d}{d\alpha} \overline{ae^2} \right|_{\substack{\alpha = 0 \\ \beta = 0}} = 0 \qquad (104)$$

$$\left. \frac{d}{d\beta} \overline{ae^2} \right|_{\substack{\alpha = 0 \\ \beta = 0}} = 0 \qquad (105)$$

We notice, however, that Eq. 104 has already been treated in section 5.1; Eq. 66 is the result. We have now established that the response extremals must satisfy not only Eq. 66, but also Eq. 105. In solving Eq. 105, we need to make only the substitution for $\Theta(\omega)$ given in Eq. 95 since Eq. 103 reduces to Eq. 95 for $\alpha = 0$.

We will now solve Eq. 105, not for the most general case,

but with appropriate specializing assumptions. Let us assume
that neither $I(\omega)$ nor $\int_{-\infty}^{\infty} A(\omega-x)[R(x) - R^*(x)]\, dx$ has
discontinuities. These assumptions will not always be valid.
As we shall see, however, there is only one problem in which
a solution of Eq. 105 is useful; this is the problem of
correcting a sluggish system to obtain the step response with
the shortest possible rise time. These assumptions will prove
justified here.

Using the substitution given by Eq. 95 in Eq. 42

$$\overline{ae^2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(-x-y)[R(x)-I(x)\epsilon^{-j\theta(x+\beta\mu(x))}]$$
$$[R(y)-I(y)\epsilon^{-j\theta(y+\beta\mu(y))}]\, dx\, dy \qquad (106)$$

Differentiating $\overline{ae^2}$ with respect to β

$$0 = \frac{d}{d\beta}\overline{ae^2}\Bigg|_{\beta=0}$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(-x-y)[R(x)-R^*(x)][-I(y)\mu(y)\frac{d}{dy}\epsilon^{-j\theta(y)}]\, dx\, dy \qquad (107)$$

Let $\theta(y)$ be divided into a sum of continuous and discontinuous
parts

$$\theta(y) = \theta_c(y) + \theta_d(y) \qquad (108)$$

The function $\theta_d(y)$ contains all of the discontinuities of
$\theta(y)$ and has zero slope between discontinuities.

Then

$$\frac{d}{dy} \epsilon^{-j\Theta(y)} dy = -j\epsilon^{-j\Theta(y)} \frac{d}{dy} \Theta_c(y) dy + \epsilon^{-j\Theta_c(y)} d[\epsilon^{-j\Theta_d(y)}] \quad (109)$$

Equation 109 substituted into Eq. 107 gives

$$0 = j\int_{-\infty}^{\infty} dy \ \mu(y)[\frac{d}{dy} \Theta_c(y)]R*(y)\int_{-\infty}^{\infty} dx \ A(-x-y)[R(x)-R*(x)]$$

$$-\int_{-\infty}^{\infty} d[\epsilon^{-j\Theta_d(y)}] \ \mu(y)I(y)\epsilon^{-j\Theta_c(y)}\int_{-\infty}^{\infty} dx \ A(-x-y)[R(x)-R*(x)]$$

$$(110)$$

Let us recognize that the first term of Eq. 110 has exactly the same form as the right-hand member of Eq. 61 and, therefore, must equal zero. An interpretation of Eq. 110 is possible. The first term accounts for the change in $\overline{ae}^2$ caused by the change in $\Theta(\omega)$ between points of discontinuity. The second term gives the effect on $\overline{ae}^2$ caused by the discontinuity shifting alone. Now we have

$$0 = \int_{-\infty}^{\infty} d[\epsilon^{-j\Theta_d(y)}] \ \mu(y)I(y)\epsilon^{-j\Theta_c(y)}\int_{-\infty}^{\infty} dx \ A(-x-y)[R(x)-R*(x)]$$

$$(111)$$

Suppose that the points of discontinuity of $\Theta(y)$ occur at $y = y_k$, for $k = -N, \ldots, -1, +1, \ldots, N$, and that in going from $y_k -$ to $y_k +$, $\epsilon^{-j\Theta_d(y)}$ changes by $\Delta[\epsilon^{-j\Theta_d(y_k)}]$. From Eq. 111

58

$$0 = \sum_{k=-N}^{N} I(y_k)\mu(y_k)\; \epsilon^{-j\Theta_c(y_k)}\; \Delta[\epsilon^{-j\Theta_d(y_k)}]$$

$$\times \int_{-\infty}^{\infty} A(-x-y_k)[R(x) - R^*(x)]\,dx \quad (112)$$

Noticing that

$$I(y_k)\; \epsilon^{-j\Theta_c(y_k)}\; \Delta[\epsilon^{-j\Theta_d(y_k)}] = \Delta R^*\,(y_k) \quad (113)$$

where $\Delta R^*(y_k)$ is the change in $R^*(y)$ that occurs in going from y_k- to y_k+, Eq. 112 becomes

$$0 = \sum_{-N}^{N} \mu(y_k)\; \Delta R^*(y_k) \int_{-\infty}^{\infty} A(-x-y_k)[R(x)-R^*(x)]\,dx \quad (114)$$

Because $\mu(y)$ can be any odd function, if Eq. 114 is to be satisfied for every choice of $\mu(y)$ it must be that

$$\Delta R^*(y_k) \int_{-\infty}^{\infty} A(-x-y_k)[R(x)-R^*(x)]\,dx$$

$$= \Delta R^*(y_{-k}) \int_{-\infty}^{\infty} A(-x-y_{-k})[R(x)-R^*(x)]\,dx \quad (115)$$

But the right-hand member of Eq. 115 is the negative conjugate of the left-hand member. Then finally

$$0 = \mathrm{Re}\left\{ \Delta R^*(\omega_k) \int_{-\infty}^{\infty} A(\omega_k-x)[R(x)-R^*(x)]\,dx \right\} \quad (116)$$

Equation 116 is the result of the derivation that started with

Eq. 105. We now know that response extremals of $\overline{ae^2}$ must

satisfy not only Eq. 66 but also Eq. 116. But since Eq. 116

is a non-linear integral equation, we are unable to solve it

for the extremals. However, for the special class of correction

problems considered in section 5.2, a simplification is possible.

It is given that $A(\omega) = A_1(\omega)$ and that $R(\omega) = jR_2(\omega)$, and it

is shown in the Appendix that $R^*(\omega) = jR_2^*(\omega)$. Therefore

Eq. 116 can be written

$$\int_{-\infty}^{\infty} A_1(\omega_k-x)[R_2(x) - R_2^*(x)]\, dx = 0 \tag{117}$$

where the discontinuities of $\Theta(\omega)$ occur at the frequencies ω_k.

The response extremal that minimizes $\overline{a_1 e^2}$ must satisfy not

only Eq. 79, but also Eq. 117.

5.4 Correction of a System to Give a Step Response with the Shortest Possible Rise Time.

The previous sections of Chapter V have been devoted to

deriving the equations that must be satisfied by a corrected

response that minimizes weighted integral square error. Section

5.2 was concerned with the class of problems in which the

desired response is a purely odd time function. In this section

we will investigate a specific problem within this class. By

minimizing $\overline{a_1 e^2}$ for an appropriately chosen weighting function,

we want to shorten the rise time of the step response of a

sluggish system. The problem of correcting the step response

of a sluggish system for minimum (unweighted) integral square error has been considered previously in Chapter IV. It is convenient here to suppose that the system has already been corrected for minimum $\overline{e^2}$, and that we are making an additional correction to minimize $\overline{a_1 e^2}$. Thus both $r(t)$ and $i(t)$ are odd time functions.

$$r(t) = u_{-1}(t) \tag{118}$$

$$R(\omega) = \frac{1}{j\omega} \tag{119}$$

$$i(t) = i_2(t) \tag{120}$$

$$I(\omega) = jI_2(\omega) \tag{121}$$

Furthermore, we know from the work of Chapter IV that $I_2(\omega)$ equals $-|I(\omega)|$ for $\omega > 0$ and $+|I(\omega)|$ for $\omega < 0$; let us assume that $|I(\omega)|$ has no discontinuities.

Since the desired response is an odd function, the weighting function should be chosen as an even function. Then the corrected response that minimizes $\overline{a_1 e^2}$ will be some odd time function.

$$r^*(t) = r_2^*(t) \tag{122}$$

$$R^*(\omega) = jR_2^*(\omega) \tag{123}$$

We shall see that with a suitable choice of the weighting function, the corrected response which minimizes $\overline{a_1 e^2}$ is
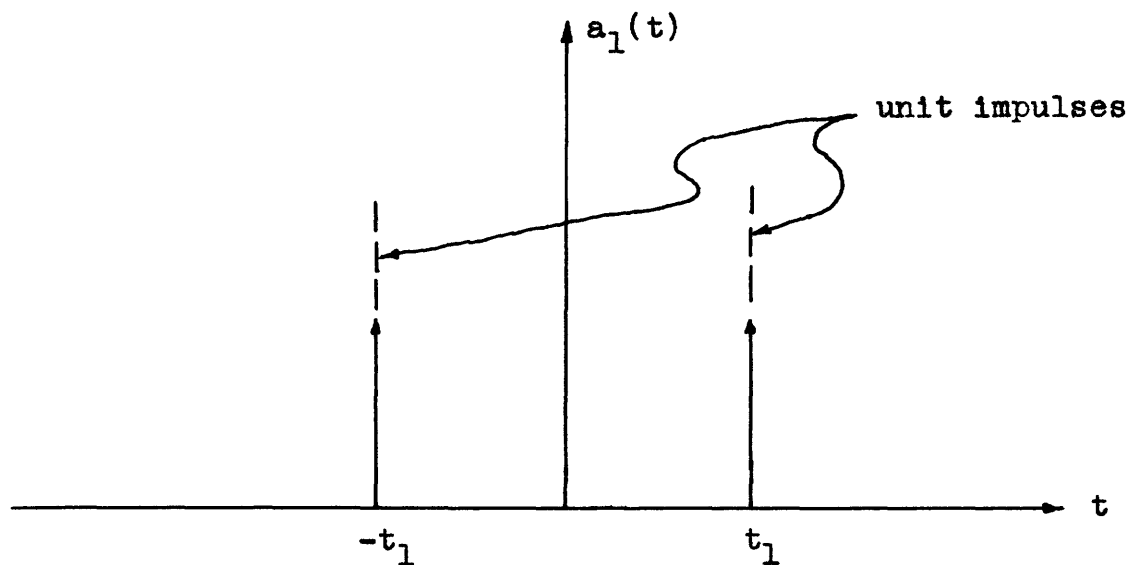
Fig. 17. An Appropriate Weighting Function

guaranteed to have the shortest rise time that can possibly
be obtained. Rise time will be defined as the time required
for the response to rise from the value -c to the value c.
In practice, the value of c is usually taken as 0.4 or 0.45 .

Let us now consider the choice of an appropriate weight-
ing function, $a_1(t)$. Let us choose

$$a_1(t) = u_o(t-t_1) + u_o(t+t_1) \tag{124}$$

where $u_o(t)$ is a unit impulse at $t = 0$. This weighting
function is shown in Fig. 17. Thus the weighted integral
square error is simply the sum of the squares of the errors

at the times t_1 and $-t_1$.

$$\overline{a_1 e^2} = e^2(-t_1) + e^2(t_1) \tag{125}$$

Since we know that the corrected response that minimizes $\overline{a_1 e^2}$ is an odd function

$$\overline{a_1 e^2} = 2e^2(t_1) = 2[\tfrac{1}{2} - r_2^*(t_1)]^2 \tag{126}$$

for this response.

Now let us imagine that we have determined the corrected responses, $r_{2m}^*(t)$, which minimize $\overline{a_1 e^2}$ for a number of different choices of t_1. Suppose that we have found the small-est value of t_1 for which

$$r_{2m}^*(t_1) = c \tag{127}$$

and that we denote this value of t_1 as $\dfrac{t_r}{2}$. Clearly, the corrected response that minimizes $\overline{a_1 e^2}$ when $t_1 = \dfrac{t_r}{2}$ is the corrected response with the shortest possible rise time; if there exists a response with a shorter rise time, then t_1, does not have the smallest value for which Eq. 127 is satisfied, and we have stated that t_1 is chosen to have this smallest value. Then the shortest step response rise time that can be achieved by means of phase correction is t_r .

Having shown the usefulness of the weighting function

given by Eq. 124, we must determine the corrected response, $r_{2m}^*(t)$, that minimizes $\overline{a_1 e^2}$ for a given value of t_1, and we must develop an easy method of finding the smallest value of t_1 for which Eq. 127 is satisfied. First, suppose that t_1 is given; let us find the corrected response that minimizes $\overline{a_1 e^2}$. From the work of the previous sections of this chapter, we know that this response is purely odd. Furthermore, we know that the transform of this response must satisfy Eq. 117. Transforming both sides of Eq. 124, we find that

$$A_1(\omega) = 2 \cos \omega t_1 \tag{128}$$

For $A_1(\omega)$ as in Eq. 128, we may write Eq. 117 as

$$0 = \int_{-\infty}^{\infty} \cos (\omega_k - y) t_1 [R_2(y) - R_2^*(y)] \, dy \tag{129}$$

$$0 = \sin \omega_k t_1 \int_{-\infty}^{\infty} \sin y t_1 [R_2(y) - R_2^*(y)] \, dy \tag{130}$$

$$0 = [r_2(t_1) - r_2^*(t_1)] \sin \omega_k t_1 \tag{131}$$

where ω_k is a frequency at which $\Theta(\omega)$, and thus $R_2^*(\omega)$, has a discontinuity. According to Eq. 131, the response extremals of $\overline{a_1 e^2}$ are all odd functions whose transforms, $jR_2^*(\omega)$, have the magnitude $|I(\omega)|$ and have discontinuities only at frequencies that are integral multiples of $\frac{\pi}{t_1}$. Let us now select the one of these response extremals that minimizes $\overline{a_1 e^2}$. The transform

of this minimizing extremal may have discontinuities at any of the frequencies that are integral multiples of $\frac{\pi}{t_1}$; we must determine the frequencies at which discontinuities actually occur.

Remember that $\overline{a_1 e^2}$ is minimized when $e(t_1)$ is minimized. For the desired choice of t_1 where $t_1 = \frac{t_r}{2}$, $e(t_1)$ is minimized when $r_2^*(t)$ is maximized. The value of $r_2^*(t_1)$ is given by

$$r_2^*(t_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R_2^*(\omega) \sin \omega t_1 \, d\omega \qquad (132)$$

It is obvious that $r_2^*(t_1)$ has the maximum possible value when $R_2^*(\omega)$ has a change of algebraic sign at each frequency for which $\sin \omega t_1$ changes sign. Then the transform of the extremal that minimizes $\overline{a_1 e^2}$ has a discontinuity at each of the frequencies

$$\omega_k = \frac{k\pi}{t_1} \qquad (133)$$

where k assumes every integer value from $-\infty$ to ∞. The only remaining question is how to choose $t_1 = \frac{t_r}{2}$ without actually calculating a series of trial solutions.

The proper choice of t_1 can be made quite simply. This fact is best illustrated by taking an example. Let us consider again the sluggish system that was examined in Chapter IV. The system, when it has been corrected to give the step

response having minimum $\overline{e^2}$, has the transfer function $(\omega^2+1)^{-1}$. The step response of this system, $i(t)$, is shown in Fig. 18. Let us define rise time as the time required for the corrected response to rise from -0.4 to +0.4. Looking at the curve of $i(t)$ versus t in Fig. 18, we guess that t_1 should have a value of about 1.25 seconds. This would place the first discontinuity of $R_2^*(\omega)$ at about $\omega_1 = \frac{\pi}{t_1} = 2.5$. This frequency is large enough so that we can utilize the same procedure for evaluating the corrected response that was used in Chapter IV. Thus we note that

$$\frac{1}{j\omega} \frac{1}{\omega^2+1} \cong \frac{1}{j\omega_3} \, , \quad |\omega| > \omega_1 \qquad (134)$$

Making use of the series expansion of $[j\omega(\omega^2+1)]^{-1}$ analogous to that of Fig. 9, we find that

$$r_{2m}^*(t) = i(t) + 2 \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{(\frac{k\pi}{t_1})^2} f_1^{(3)} (\frac{k\pi}{t_1} t) \qquad (135)$$

where the transform of $f_1^{(3)}(t)$ is given by Eq. 35, and $f_1^{(3)}(t)$ is plotted in Fig. 11. Since the series in Eq. 135 converges rapidly, we will take only the first two terms

$$r_{2m}^*(t) = i(t) + 2 \frac{t_1^2}{\pi^2} f_1^{(3)} (\frac{\pi}{t_1} t) \qquad (136)$$

Fig. 18.
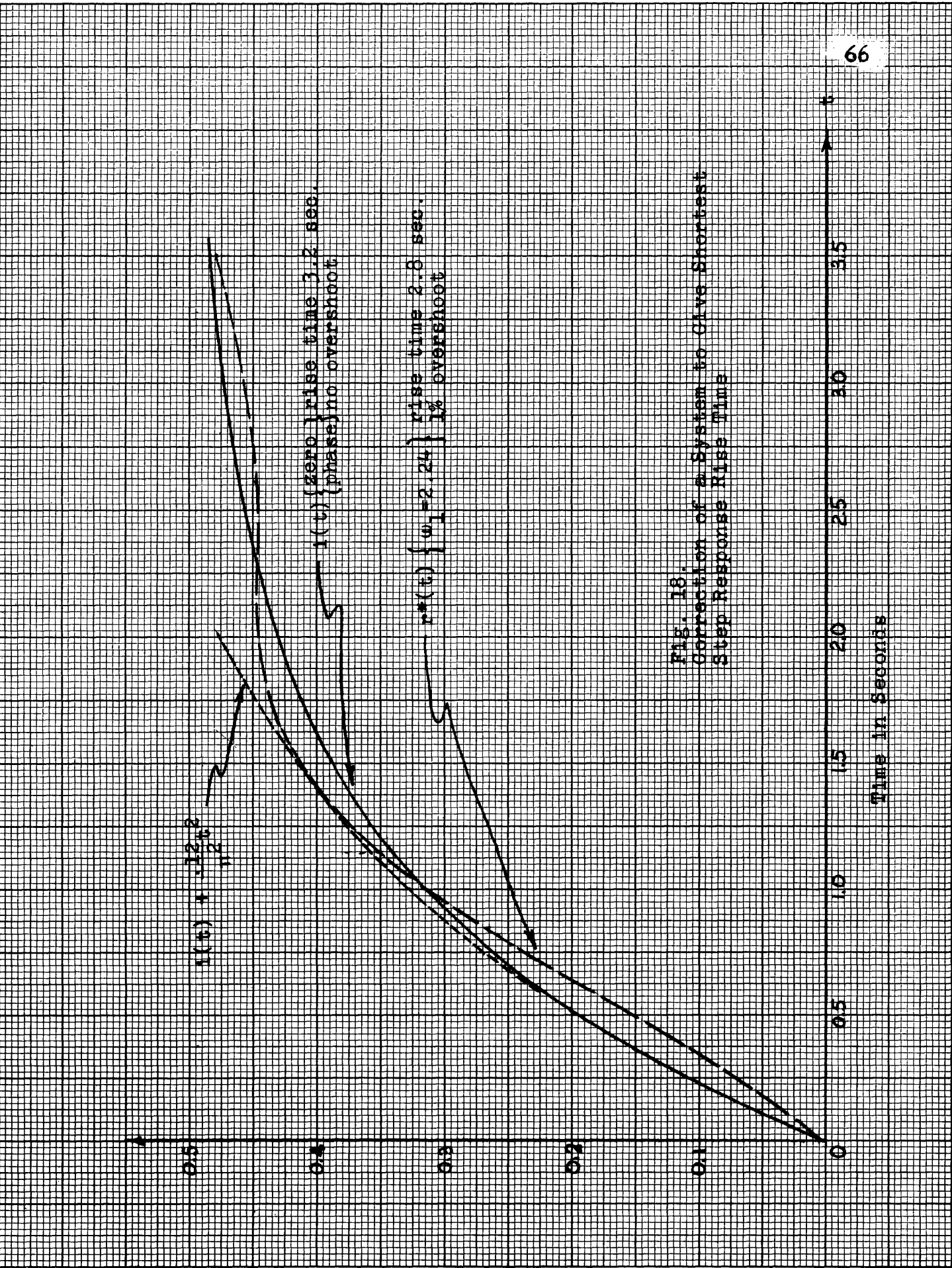Correction of a System to Give Shortest
Step Response Rise Time

The termination of the series in this way corresponds to allowing $R_{2m}^{*}(\omega)$ to have a discontinuity only at $\omega = \frac{\pi}{t_1}$ for $\omega > 0$, rather than having an infinite number of discontinuities at $\omega = \frac{k\pi}{t_1}$. However, the termination makes very little difference in the time response, and the functions given by Eqs. 135 and 136 are nearly the same.

Now we are able to find the value of t_1 that makes the corrected response that minimizes $\overline{a_1 e^2}$ have the shortest possible rise time. This is the value for which $r_{2m}^{*}(t_1) = 0.4$. From Fig. 11 we note that

$$f_1^{(3)}(\pi) = 0.06 \tag{137}$$

According to Eq. 136, t_1 should have the value for which

$$0.4 = i(t_1) + 2\frac{t_1^2}{\pi^2} \, .06 \tag{138}$$

Suppose that we plot a curve of $i(t) + \frac{.12}{\pi^2} t^2$ versus t, as shown in Fig. 18. We note that the curve has the value 0.4 for

$$t = 1.4 = t_1 = \frac{t_r}{2} \tag{139}$$

Thus $R_{2m}^{*}(\omega)$ has a discontinuity at $\omega = \frac{\pi}{t_1} = 2.24$. Using the value of t_1 given by Eq. 139, we can calculate the corrected response $r_{2m}^{*}(t)$, as given by Eq. 136, that has the shortest

possible rise time. This function is shown in Fig. 18.

Let us review that which we have been able to accomplish by means of phase correction.

Response of system with	Rise time	Rise time shortened by
(1) no correction	3.3 sec.	—
(2) correction for minimum $\overline{e^2}$, or linear phase	3.2 sec.	3 %
(3) correction for shortest possible rise time	2.8 sec.	15 %

By means of phase correction, the rise time of the step response of this sluggish system can be reduced by as much as (but by no more than) 15 %. It should be noticed however that the optimum transient correction requires a discontinuous phase correction characteristic. We can approximate such a characteristic by the phase of an all-pass network, but the cost, in terms of complexity of the network, is very high indeed. This matter will be discussed in Chapter VI.

In the derivation of the correction giving the shortest possible step response rise time, we have given no consideration to the overshoot that this correction may create. In the example discussed here, this correction causes an overshoot
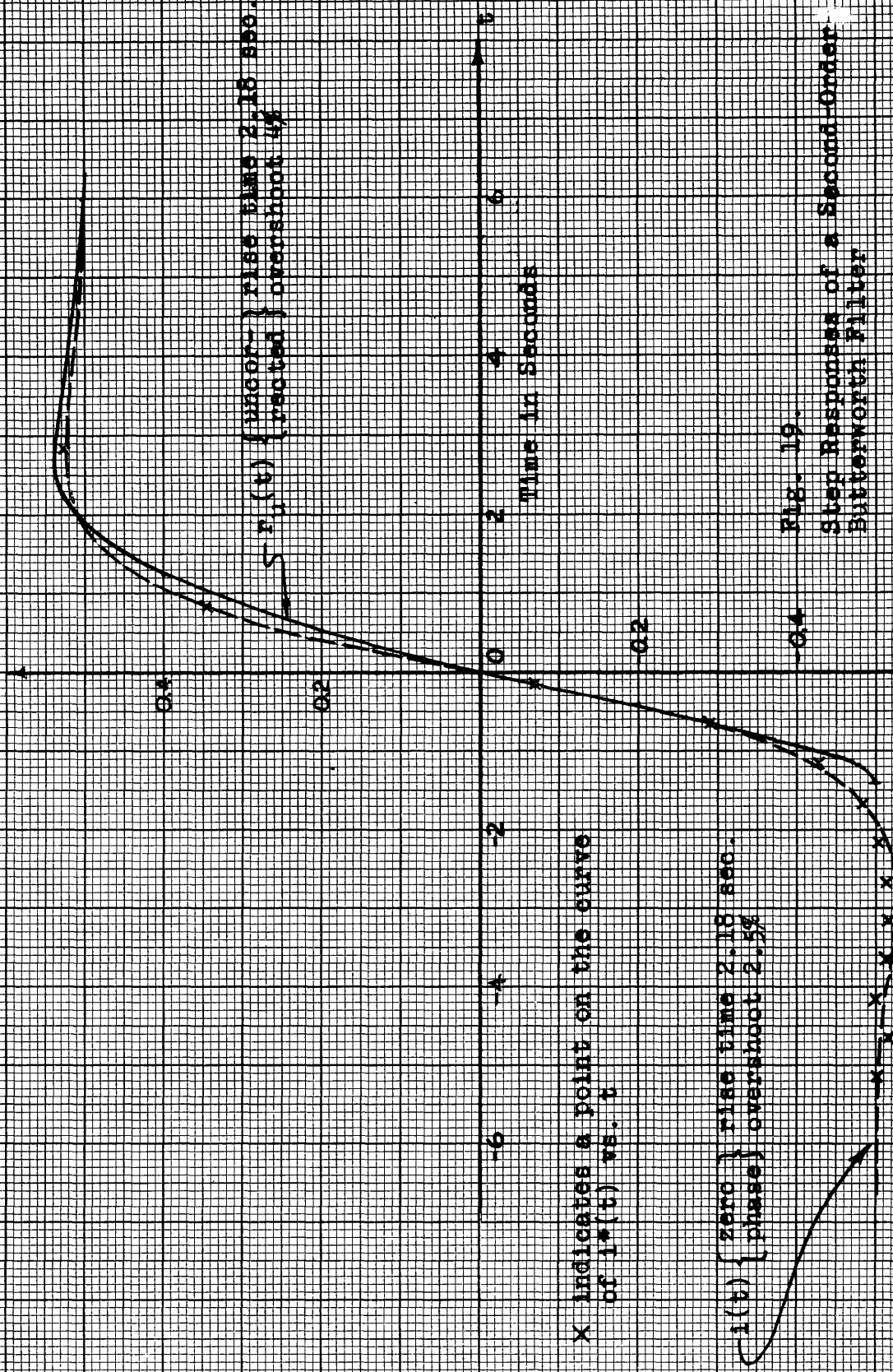
Time in Seconds

$r_0(t)$ $\left\{\begin{array}{l}\text{uncor-}\\\text{rected}\end{array}\right\}$ rise time 2.15 sec. overshoot 4%

$r_1(t)$ $\left\{\begin{array}{l}\text{zero}\\\text{phase}\end{array}\right\}$ rise time 2.18 sec. overshoot 2.5%

× indicates a point or record of $r_1''(t)$ vs. t

Fig. 19.

Step Response of a Second-Order Butterworth Filter

Since $I(\omega)$ has branch points in the complex s-plane, an approximation method, or rather a combination of several approximation methods, will be used to calculate $i(t)$. For $0 < \omega < 0.8$, we represent $I(\omega)$ by

$$I(\omega) = \frac{1}{j\omega} (1-0.4\omega^4) \ , \ 0 < \omega < 0.8 \tag{142}$$

To this portion of the frequency function there corresponds a time response

$$\left[\frac{1}{\pi} \int_0^{0.8t} \frac{\sin x}{x} \ dx - (0.4)(0.8)^4 \ \frac{d^3}{dx^3} \ \frac{\sin x}{x} \bigg|_{x=0.8t} \right] \tag{143}$$

Both of the functions needed in calculating this portion of the response are tabulated.[4,5] In the frequency range $0.8 < \omega < 2$, $I(\omega)$ is approximated by three straight line segments. The contribution to $i(t)$ made by these line segments may be readily calculated.[6] For $\omega > 2$, $I(\omega)$ is approximated by $-j\omega^{-3}$. The part of the time response that corresponds to this portion of the frequency function may be found using the curve of $f_1^{(3)}(t)$ given in Fig. 11. The result of these computations is the curve of $i(t)$ versus t shown in Fig. 19 and also in Fig. 20. This function is the corrected response that is the minimum $\overline{e^2}$ approximation of a step function. We notice that this corrected step response has an overshoot of 2.5 %, as compared to an overshoot of 4 % for the step response of the uncorrected system, $r_u(t)$.

$r_{2BI}^*(t)$ $\left\{\begin{array}{l}w_1 = 2.36\end{array}\right\}$ rise time 2.12 sec., overshoot 2.5 %

$r_{2B}^*(t)$ $\left\{w_1 = 3.14\right\}$ rise time 2.03 sec., overshoot 3.9 %

$i(t)$ $\left\{\begin{array}{l}\text{zero} \\ \text{phase}\end{array}\right\}$ rise time 2.18 sec., overshoot 2.5 %
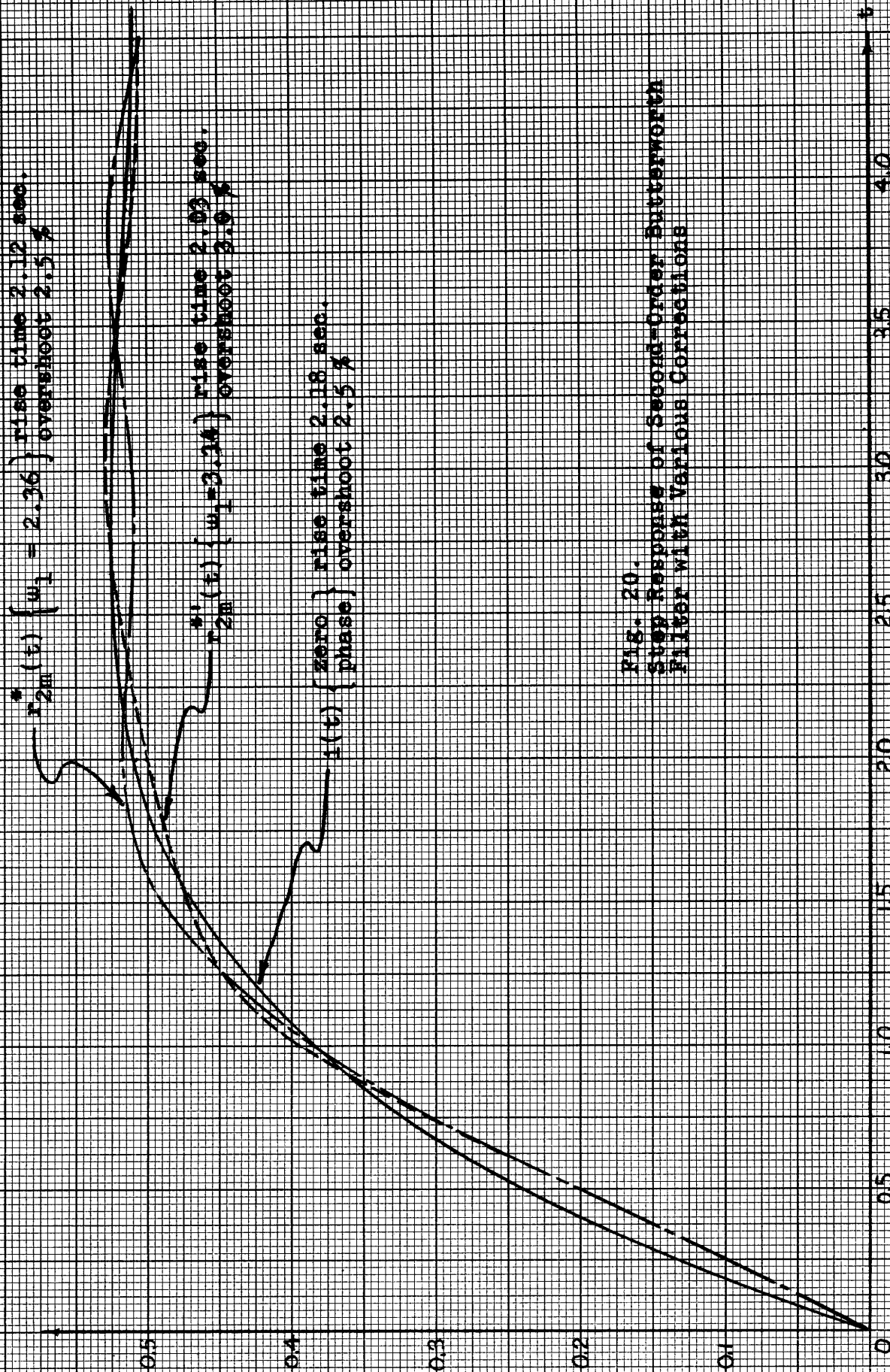
Fig. 20.
Step Response of Second-Order Butterworth
Filter with Various Corrections

Time in Seconds

Now let us investigate the possibility of making an additional correction to obtain a corrected step response that has less overshoot than does $i(t)$. It has already been mentioned that the straightforward approach, wherein we solve Eq. 117 to find the corrected response that minimizes $\overline{a_1 e^2}$, is impossible in this problem. But we know that the straightforward approach would lead to a corrected response that is some odd function of time; this fact is verified in the Appendix. Suppose that this odd function is denoted as $r_{2m}^*(t)$. The transform of the corrected response, $jR_{2m}^*(\omega)$, has the magnitude function $|I(\omega)|$. Only the locations of the discontinuities of $R_{2m}^*(\omega)$ are unspecified, and these can be determined quite easily.

Note that the discontinuities of $R_{2m}^*(\omega)$ cannot be located at low frequencies; the first discontinuity must be located at a frequency that is well above the cutoff frequency of the filter, $\omega_c = 1$. A discontinuity located at low frequencies would create a very large, low frequency ripple component in $r_{2m}^*(t)$ that could not possibly cause a decrease in the overshoot. The same argument can be stated in a slightly different form. It is known that the step response, $i(t)$, has overshoot because $I(\omega)$ exhibits a sharp decrease in the region near the cutoff frequency. The addition of a discontinuity, a sign reversal of $I(\omega)$, in this region could only increase the overshoot.

The first discontinuity of $R_{2m}^*(\omega)$ is located at a

frequency, ω_1, that is well above $\omega_c = 1$. Let us suppose that $R_{2m}^*(\omega)$ has only one discontinuity for $\omega > 0$. By making the approximation

$$I(\omega) = \frac{1}{j\omega^3} \quad \omega > \omega_1 \qquad (144)$$

we can write

$$r_{2m}^*(t) = i(t) + \frac{2}{(\omega_1)^2} f_1^{(3)}(\omega_1 t) \qquad (145)$$

where $f_1^{(3)}(t)$ is plotted in Fig. 11. Notice that the second term of Eq. 145 is simply the function $f_1^{(3)}(t)$ with an appropriate normalization of the time scale and of the constant multiplier to account for the choice of ω_1.

In our attempt to suppress the overshoot of $i(t)$, we have developed Eq. 145 which gives the corrected response, $r_{2m}^*(t)$ as the sum of $i(t)$ and a normalized form of the function $f_1^{(3)}(t)$. Now we simply look at Fig. 11 and determine by inspection the most desirable normalization, thus determining the value of ω_1. The peak of the overshoot of $i(t)$ occurs for $t = 2.75$. The normalization of $f_1^{(3)}(t)$ should be chosen so that one of the negative peaks of the normalized function falls at $t = 2.75$. The first negative peak of $f_1^{(3)}(t)$ occurs at $t = 1$. To move this peak to $t = 2.75$ by normalization requires that $\omega_1 = \frac{1}{2.75} = 0.364$. But we have previously noted that in order

to make the desired correction, ω_1 must be chosen greater than
1. Then let us consider the normalization that causes the
second negative peak of $f_1^{(3)}(t)$ to occur at $t = 2.75$. This
requires that

$$\omega_1 = \frac{6.5}{2.75} = 2.36 \tag{146}$$

Using this value of ω_1, Eq. 145 becomes

$$r_{2m}^*(t) = i(t) + 0.36 \, f_1^{(3)}(2.36t) \tag{147}$$

This function is plotted in Fig. 20. The overshoot at $t = 2.75$
has been suppressed by this correction, just as desired. How-
ever, note that $r_{2m}^*(t)$ has the same overshoot as $i(t)$. The
fact is that the curve of $i(t)$ versus t is so nearly flat in
the region near its peak overshoot that the positive peak of
$f_1^{(3)}(2.36t)$ at $t = 3.4$ causes $r_{2m}^*(3.4)$ to be as great as $i(2.75)$.
However, $r_{2m}^*(t)$ does have a slightly smaller rise time than
does $i(t)$, and with no increase in overshoot.

Perhaps by taking additional discontinuities in the
function $R_{2m}^*(\omega)$, a small decrease in the overshoot of $i(t)$ can
be obtained, but it is clear that very little can be done to
decrease the overshoot of the step response of a system that
has already been corrected for minimum $\overline{e^2}$. The correction for
minimum $\overline{e^2}$ effected a sizable decrease in overshoot (from
4 % to 2.5 %) but little more can be accomplished by further
correction.

of perhaps 1 %. The problem of making a phase correction in order to decrease overshoot is considered in the next section.

5.5 Phase Correction of a System to Reduce Step Response Overshoot.

In the previous sections we determined the phase correction that leads to the step response with the shortest possible rise time for a given system. The solving of this problem was straightforward and the solution was unique. In this section we will consider the possibility of correcting a system step response so as to decrease the amount of overshoot, or so as to decrease both rise time and overshoot. The solving of these problems is more difficult and the solutions are less clear-cut than for the problem of correction to reduce rise time.

First let us consider the choice of the weighting function that is needed in order to make the corrected response that minimizes $\overline{a_1 e}^2$ to have the least possible overshoot. As usual it is convenient to assume that the system has already been corrected to minimize $\overline{e}^2$. We might think that the weighting function given by Eq. 124 would lead to the response with the least overshoot if t_1 were chosen equal to the time at which the largest peak of $i(t)$ occurs. Correction to minimize $\overline{a_1 e}^2$ would certainly suppress the peak at $t = t_1$; unfortunately, this correction also causes peaks to crop up elsewhere. We could add other pairs of impulses to the $a_1(t)$ function in an attempt to suppress all of the peaks of the corrected response simultaneously, but the result of such an endeavor is

uncertain. We must concede that it is difficult to know how to choose the weighting function that leads to the corrected response with the least possible overshoot.

Even if the proper weighting function were known, the solution of the $\overline{a_1e}^2$ minimization problem is not easy. The transform of the corrected response that minimizes $\overline{a_1e}^2$ must be an odd function that satisfies Eq. 117. But it is very difficult to solve Eq. 117 for any weighting function except the simple one of Eq. 124. However, we shall develop a method for the direct synthesis of corrected responses that avoids both the question of choosing $a_1(t)$ and the difficulty of solving Eq. 117. This method can best be demonstrated in an example.

Let us consider the step response, $r_u(t)$, of a second-order Butterworth filter. The transform of this response is

$$R_u(\omega) = \frac{1}{j\omega} \frac{1}{-\omega^2+j\sqrt{2}\ \omega+1} \tag{140}$$

The uncorrected step response, $r_u(t)$, is plotted in Fig. 19 with its time origin shifted back to $t = -1.4$ seconds.

We must now make the correction that minimizes $\overline{e}^2$ for this filter. The step response of the filter with correction for minimum $\overline{e}^2$ is denoted $i(t)$, and

$$I(\omega) = \frac{1}{j\omega}\left| \frac{1}{-\omega^2+j\sqrt{2}\ \omega+1} \right| = \frac{1}{j\omega}\frac{1}{\sqrt{1+\omega^4}} \tag{141}$$

Also shown in Fig. 20 is the step response of the system when corrected for shortest possible rise time. This function is denoted as $r_{2m}^{*\prime}(t)$.

5.6 Phase Correction of a Fourth-Order Butterworth Filter

Let us consider the step response, $r_u(t)$, of a fourth-order Butterworth filter. The transform of this response is

$$R_u(\omega) = \frac{1}{j\omega\,(j\omega-e^{-j\frac{5}{8}\pi})(j\omega-e^{-j\frac{7}{8}\pi})(j\omega-e^{-j\frac{9}{8}\pi})(j\omega-e^{-j\frac{11}{8}\pi})} \qquad (148)$$

The uncorrected step response, $r_u(t)$, is plotted in Fig. 21 with its time origin shifted back to $t = -2.85$ seconds.

The transform of the system step response, when corrected for minimum $\overline{e^2}$ is

$$I(\omega) = \frac{1}{j\omega}\,\frac{1}{\sqrt{1+\omega^8}} \qquad (149)$$

The function $i(t)$ was calculated by approximation methods like those described in the previous section; the curve of $i(t)$ versus t is shown in Fig. 21. Notice that correction for minimum $\overline{e^2}$ has reduced the overshoot from 11 % to 6 %, with no change in rise time.

No significant improvement in step response can be obtained by additional phase correction. For example, shown in Fig. 21 are points on the curve of the system step response
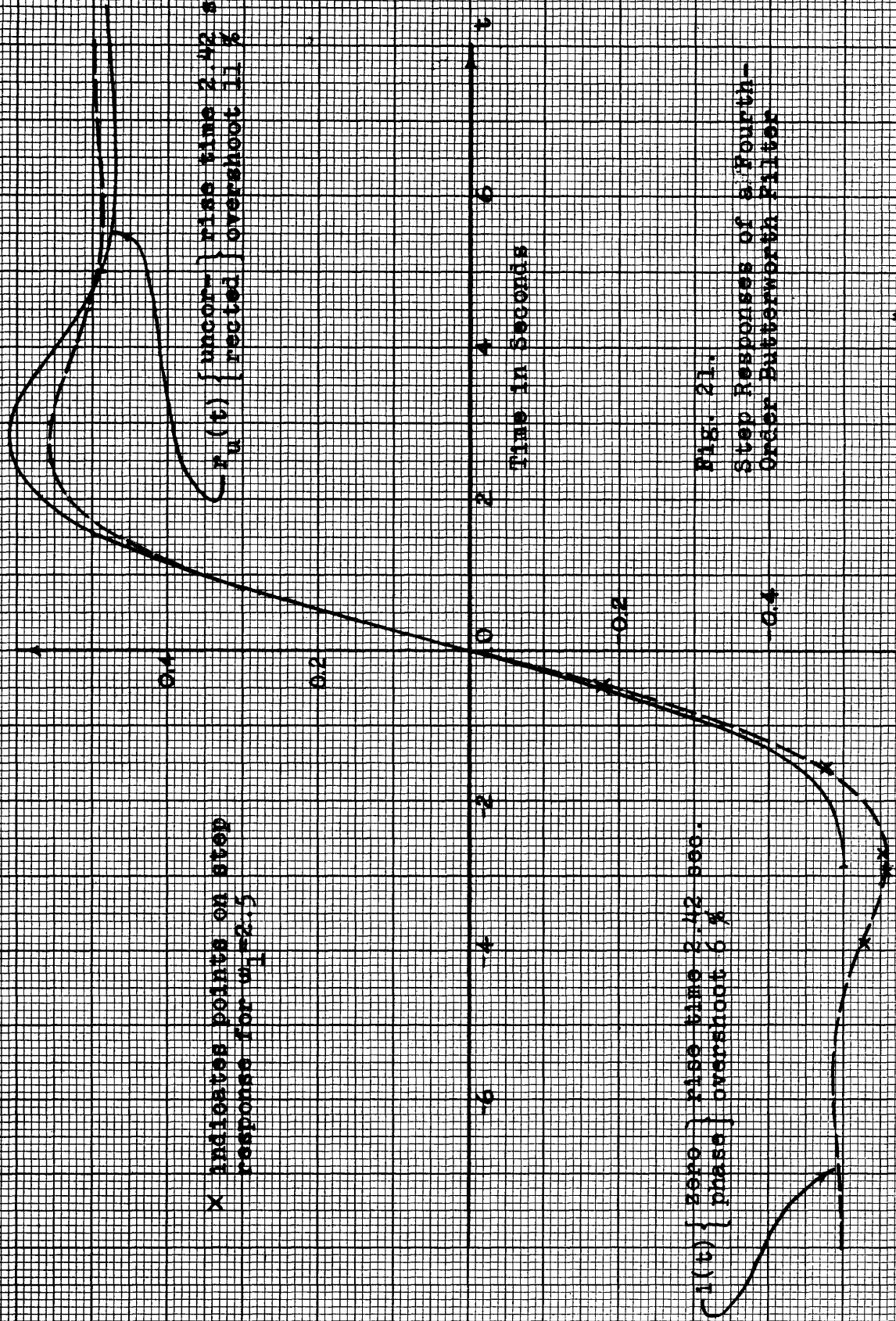
Fig. 21.
Step Responses of a Fourth-Order Butterworth Filter

when corrected for shortest rise time. The amount of deviation from the i(t) curve is barely perceptible. The reason for the failure of additional correction attempts is clear. These corrections always involve the placement of discontinuities at frequencies above the cutoff frequency of the filter. Since the magnitude of $I(\omega)$ behaves like ω^{-5} above the cutoff frequency, $|I(\omega)|$ is so small that the addition of discontinuities in this region can cause little change in the response.

In sections 5.4 and 5.5, and in the present section, we have investigated the effect of phase correction on the step responses of various systems. Let us summarize the findings. Correction to obtain the step response with minimum integral square error produces a decrease in overshoot for those systems whose uncorrected step response exhibit this characteristic. However, this correction makes very little change in the rise time of the step response of any system. Additional phase correction involving phase discontinuities can be used to reduce rise time, especially for a sluggish system, but cannot be used to decrease overshoot.

CHAPTER VI

REALIZATION OF DESIRED PHASE

CORRECTIONS BY ALL-PASS NETWORKS

6.1. Possible Circuits for Realizing an All-Pass Network.

Figure 1 shows the pole-zero pattern in the s-plane of a typical all-pass network transfer function. It consists of a pole-zero pair on the real axis and two pole-zero quadruplets. In the realization of an all-pass network, it is convenient to divide the network into all-pass "sections", where each section realizes a pole-zero pair or quadruplet. The all-pass sections are cascaded to obtain the desired all-pass network transfer characteristic. It is expedient to realize these all-pass sections in the form of constant resistance networks. A constant resistance network has the property that when it is terminated in the characteristic resistance R_1, the input impedance of the network is R_1 for all frequencies. Thus, all-pass sections realized in this form can be cascaded directly.

The realization of the all-pass network with the transfer function

$$z_{12}(\omega) = -\tfrac{1}{2} \frac{j\omega - \sigma_1}{j\omega + \sigma_1}$$

(150)

can be made as shown in Fig. 22.

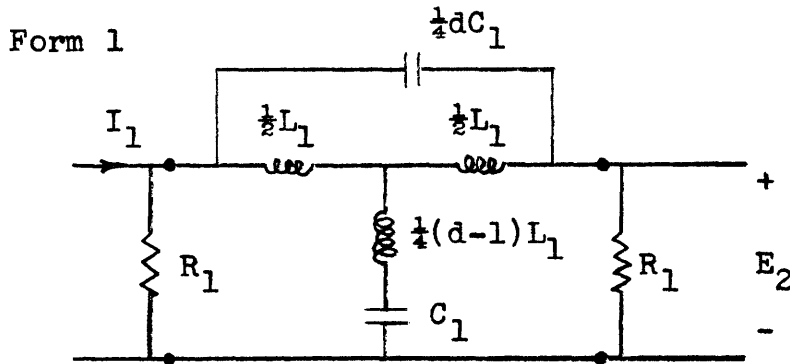Two possible forms of realization of the all-pass network with the transfer function

$$z_{12}(\omega) = \tfrac{1}{2} \frac{(j\omega - \sigma_1 - j\omega_1)(j\omega - \sigma_1 + j\omega_1)}{(j\omega + \sigma_1 - j\omega_1)(j\omega + \sigma_1 + j\omega_1)}$$

(151)
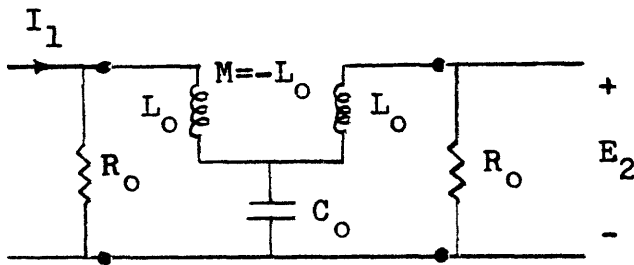
Fig. 22. A Constant Resistance Network with $Z_{12}(\omega) = -\frac{1}{2} \dfrac{j\omega - \sigma_1}{j\omega + \sigma_1}$

$$L_o = \frac{R_o}{2\sigma_1}$$

$$C_o = \frac{2}{\sigma_1 R_o}$$

Form 1

$$d = \frac{1}{4}\left[1 + \left(\frac{\omega_1}{\sigma_1}\right)^2\right]$$

$$L_1 = \frac{R_1}{d\,\sigma_1}$$

$$C_1 = \frac{L_1}{R_1^{\,2}}$$
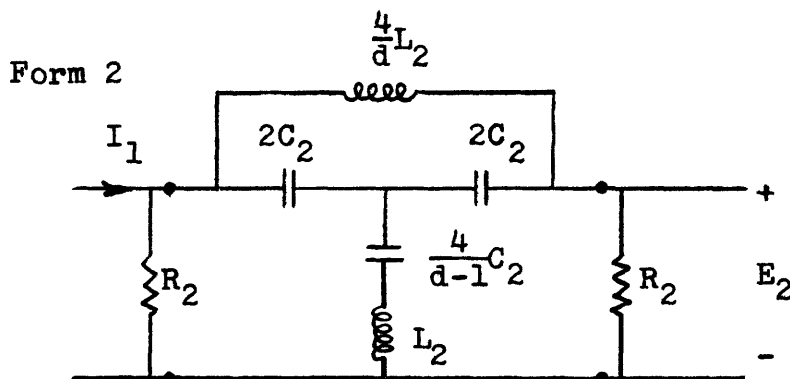
Form 2

$$d = \frac{1}{4}\left[1 + \left(\frac{\omega_1}{\sigma_1}\right)^2\right]$$

$$L_2 = \frac{R_2}{4\sigma_1}$$

$$C_2 = \frac{L_2}{R_2^{\,2}}$$

Fig. 23. Two Constant Resistance Networks with

$$Z_{12}(\omega) = \frac{1}{2}\,\frac{(j\omega - \sigma_1 - j\omega_1)(j\omega - \sigma_1 + j\omega_1)}{(j\omega + \sigma_1 - j\omega_1)(j\omega + \sigma_1 + j\omega_1)}$$

are shown in Fig. 23. The network of Form 1 can always be realized, and when d > 1, the network of Form 2 can be realized.

6.2 Number of All-Pass Sections Required to Linearize Phase

In Chapter II we showed that any phase function plus the phase of a delay network, ωT, can be closely approximated by the phase of a realizable all-pass network. But we gave no attention to the practical problem of realizing the all-pass network. Let us consider now the number of all-pass sections that are needed to realize the various phase corrections that have been derived in Chapters IV and V.

First, let us determine the number of all-pass sections needed to approximate the phase characteristic that corrects a system for minimum $\overline{e^2}$ step response. This requires correction to linearize the phase of the corrected system. As a specific example let us attempt to linearize the phase of the second-order Butterworth filter of section 5.5 . In practice, correction is usually made to flatten the corrected group delay characteristic rather than to linearize the phase, where group delay is defined as the derivative of phase with respect to ω. Then we want to design a correction that makes the group delay of the Butterworth filter, T_{gB}, plus the group delay of the all-pass network, T_{gAP}, nearly equal to a constant. Shown in Fig. 24 is a curve of $5.3 - T_{gB}$. Also shown is the curve of T_{gAP} that is obtained for a three section all-pass network where the transfer functions of the sections have the zero locations

section 1: $\sigma_1 = 0.96$, $\omega_1 = 0$

section 2: $\sigma_2 = 0.85$, $\omega_2 = \pm 1.5$

section 3: $\sigma_3 = 0.72$, $\omega_3 = \pm 2.315$

We note that $T_{gB} + T_{gAP}$ very nearly equals 5.3 seconds for frequencies less than $\omega = 2$. Let us denote the step response of the Butterworth filter with this correction as $1^*(t)$. The function $1^*(t)$ should be a good approximation to the corrected response with minimum $\overline{e^2}$, or the function $1(t)$ shown in Fig. 19, except for a delay of about 5.3 seconds. Shown in Fig. 19 are points on the curve of $1^*(t)$ versus t when plotted with its time origin shifted back to $t = -5.15$ seconds. The approximation of $1(t)$ by $1^*(t)$ is very good indeed.

Our experience with this example leads us to believe that a satisfactory correction to linearize the phase of a system can be accomplished with a reasonable number of all-pass sections, no more than four or five at the most. Then if substantial improvement in the step response of a system can be obtained by correction to minimize $\overline{e^2}$, it may be worthwhile to construct an all-pass network to achieve phase linearization of the system.
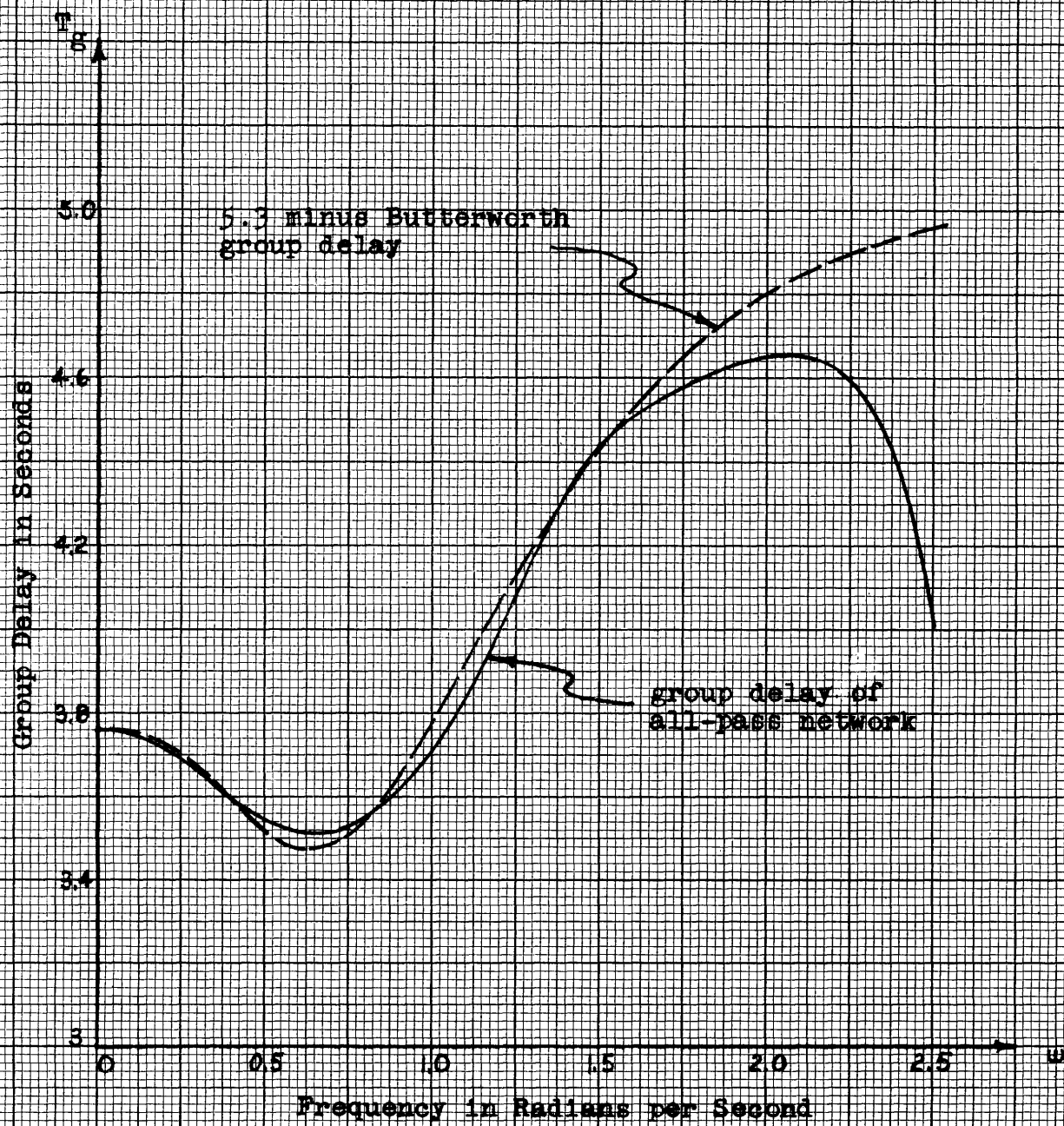
Fig. 24. Group Delay of an All-Pass Network

6.3 Number of All-Pass Sections Needed to Permit Approximation of a Discontinuous Phase Function by the Phase of an All-Pass Network.

Correction to linearize the phase of a system can be accomplished rather easily, as was indicated in the previous section. But some desirable phase corrections require a discontinuous phase function. Correction of a system to give the step response with the shortest possible rise time requires a phase correction with discontinuities; the value of the phase correction function must change by an amount π at the discontinuities. It is shown in Chapter II that the phase of an all-pass network, $\Theta(\omega)$, can closely approximate the function $\beta(\omega) + \omega T$ over any finite frequency range, where $\beta(\omega)$ can be any odd frequency function. As we shall see, however, if the function $\beta(\omega)$ is discontinuous, the number of all-pass sections needed to make this approximation is very large.

Let us make an estimation of the number of all-pass sections needed to correct the sluggish system of Chapter IV so as to obtain the corrected step response with the shortest possible rise time. We have shown in Chapter V that the phase correction needed to accomplish this must linearize the phase of the sluggish system and have a discontinuous change in value of amount π at $\omega = 2.24$. However, the phase linearization can be made quite easily with the number of all-pass sections that is required to approximate the discontinuity. Therefore, let us assume that the desired correction is simply

a linear frequency function except at ω = 2.24, where there is a discontinuity. This function is shown in Fig. 25. Let us represent the phase of an all-pass network by a series of line segments of phase length 2π, just as was done in Fig. 3. The approximation of the desired phase correction that is made by the all-pass network phase is shown in Fig. 25. By imagining a shift of origin to t = T in the time domain, we eliminate the linear component, ωT, of the phase functions shown in Fig. 25. Now, corresponding to the difference between the desired phase and the all-pass network phase, there exists an error in the real and imaginary parts of the transform of the corrected response, as is shown in Fig. 26. The error in the corrected time response due to the approximation of the discontinuous desired phase by the all-pass network phase must be less than the total area enclosed by the two curves plotted in Fig. 26. The total of this enclosed area amounts to

$$\frac{2}{T}(1 + 0.571) \, |I(2.24)| = \frac{0.44}{T} \tag{152}$$

Suppose that we want the time domain error in the corrected response to always be less than 0.01. Then we must choose

$$T = 44 \text{ seconds} \tag{153}$$

Remember that we can approximate the desired phase correction by the phase of an all-pass network for only a finite frequency range. Suppose that for frequencies greater
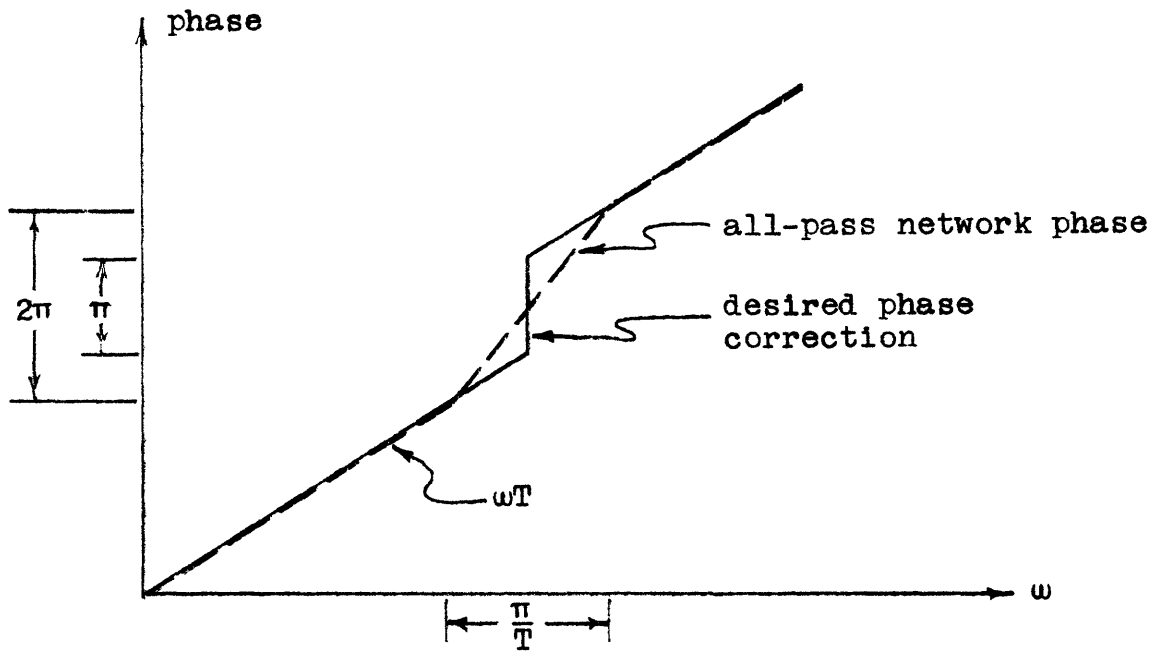
Fig. 25. Approximation of a Discontinuous Phase Function by the Phase of an All-Pass Network
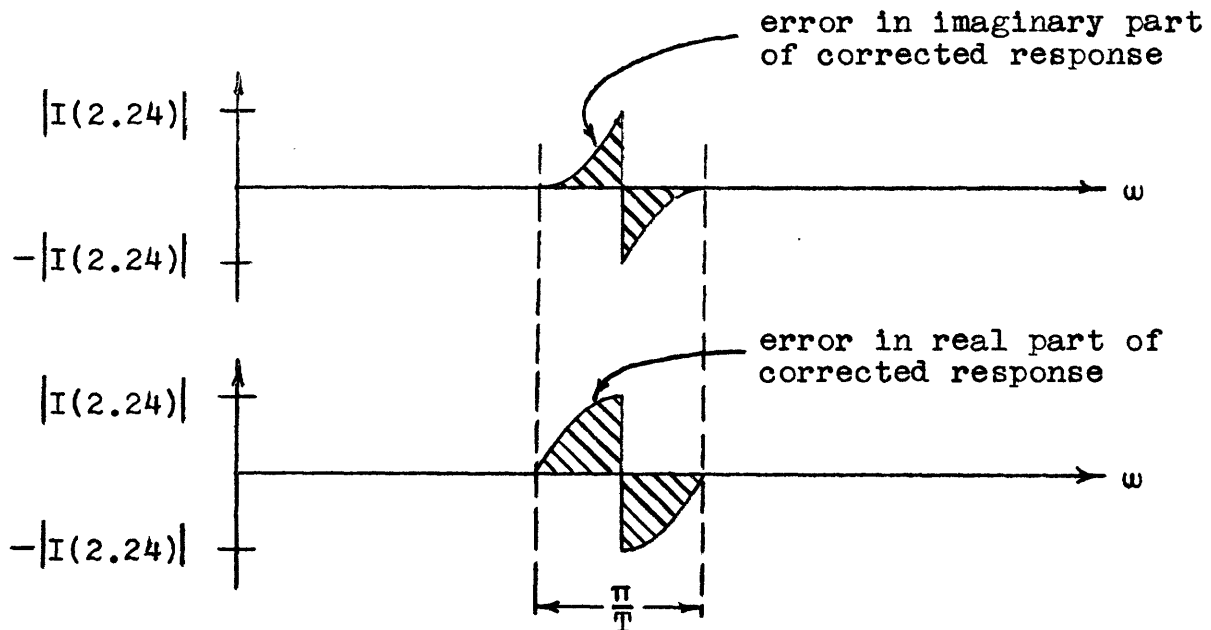

Fig. 26. Errors Occurring in Real and Imaginary Parts of the Corrected Response

than $\omega = 5$, we decide that we will no longer attempt to make the approximation. Then we can see from Fig. 25 that the total phase shift contributed by the all-pass network as ω varies from zero to infinity is about

$$5T = 220 \text{ radians} \tag{154}$$

Each all-pass section contributes a phase shift of 2π over this frequency range; then the number of all-pass sections must be

$$\frac{220}{2\pi} \cong 35 \text{ all-pass sections} \tag{155}$$

The above considerations give us at least a rough estimate of the number of all-pass sections that would be needed in order to correct this sluggish system to give the step response with the shortest possible rise time. The use of 35 all-pass sections to accomplish a 15 % decrease in rise time hardly seems justified.

APPENDIX

PROOF THAT AN ODD CORRECTED RESPONSE

APPROXIMATES AN ODD DESIRED RESPONSE WITH MINIMUM $\overline{a_1 e^2}$

A.1 Outline of the Proof and Assumptions to be Made

Since we will be making frequent references to the solutions of Eq. 79 in this appendix, when we mention a "solution", Eq. 79 is referred to unless otherwise stated.

In section 5.2 we have shown that any odd frequency function that has the proper magnitude is a solution of Eq. 79. We noted, however, that there may be other solutions for which $R_1^*(\omega) \neq 0$; such functions, when they exist, will be called "degenerate solutions". The task of this appendix is to show that a purely odd solution, rather than a degenerate solution, minimizes $\overline{a_1 e^2}$. Because Eq. 79 cannot be solved for the degenerate solutions, if there are any, this must be shown in a way that does not require knowledge of the complete set of extremals of $\overline{a_1 e^2}$.

The method of proof will be as follows. We will postulate that Eq. 79 does have a degenerate solution. We will then construct a non-degenerate, or purely odd, solution in such a way that it can be shown that this solution yields at least as small a value of $\overline{a_1 e^2}$ as does the degenerate solution. The non-degenerate solution which will be constructed in this way will not, in general, be the solution which minimizes $\overline{a_1 e^2}$. But we will have proved that some non-degenerate solution minimizes $\overline{a_1 e^2}$, since we will have shown that we can always

construct a non-degenerate solution which gives at least as small a value of $\overline{a_1 e^2}$ as is given by a specified degenerate solution.

In order to fix ideas, let $r_2(t)$ be taken as a unit step function in this proof. The same proof, with only minor changes, could be made using any other odd function as the desired response.

It is convenient to suppose that phase correction has already been made to minimize $\overline{e^2}$, and that the additional phase correction, $\Theta(\omega)$, will be used to minimize $\overline{a_1 e^2}$. Thus $i(t)$ is odd and $I(\omega) = jI_2(\omega)$.

We must assume that it is possible to make a phase correction for which $\overline{a_1 e^2}$ is finite; otherwise, any consideration of minimizing $\overline{a_1 e^2}$ would be meaningless. Thus, for example, if $a_1(t)$ approaches a constant at large times, $i(t)$ must approach the same final value as the desired step function, $r_2(\infty) = \frac{1}{2}$.

Knowledge of the sort of errors to be corrected in this problem restricts the class of weighting functions, $a_1(t)$, which need to be considered. Because the system has already been corrected for minimum $\overline{e^2}$, the additional correction cannot reduce the error over one time interval without increasing the error elsewhere. The additional correction, which is designed to minimize $\overline{a_1 e^2}$, may be considered to shift the error from one time interval to another. It is known that in all physically motivated problems, the principal part of the error, be it due to sluggishness or overshoot, occurs near the time origin. In
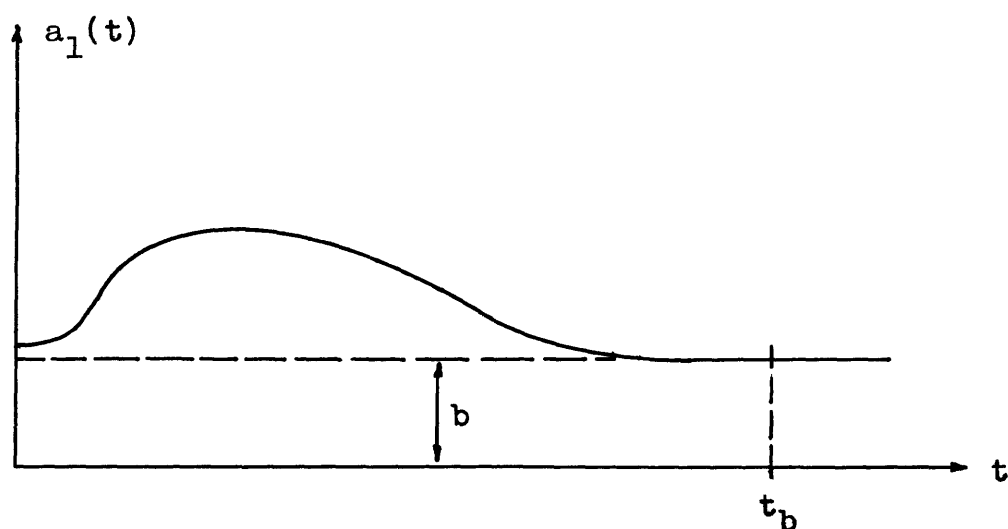
Fig. Al.   Example of a Suitable Weighting Function.

general, it is desired to shift some of the error from the vicinity of the time origin to later times where the error is not so large. In order to accomplish this, the weighting function should be like the one shown in Fig. Al. It should be a function which is always greater than its asymptotic value b. A weighting function of this form accomplishes the desired shift of error away from the time origin by weighting early errors more heavily than later errors. The exact form of the weighting function that is needed depends on the particular problem, but it should always have this characteristic. It will be supposed that by the time t_b, $a_1(t)$ has reached the level b.

$$a_1(t) = b \quad , \quad |t| > t_b \qquad (A1)$$

A.2 The Value of $\overline{a_1 e^2}$ Given by a Degenerate Solution of Eq. 79.

The equation giving $\overline{a_1 e^2}$ may be written as in Eq. 71. The equation is repeated here for convenience.

$$\overline{a_1 e^2} = \int_{-\infty}^{\infty} a_1(t)[r_2(t) - r_2^*(t)]^2 \, dt + \int_{-\infty}^{\infty} a_1(t)[r_1^*(t)]^2 \, dt \tag{A2}$$

The relation

$$|R_1^*(\omega)|^2 + |R_2^*(\omega)|^2 = |I(\omega)|^2 \tag{A3}$$

must exist between the transforms of $r_1^*(t)$ and $r_2^*(t)$.

Equation A2 indicates that $\overline{a_1 e^2}$ equals to sum of two terms, the weighted integral square of the difference between the desired response and the odd part of the corrected response plus the weighted integral square of the even part of the corrected response. For a corrected response that is the inverse transform of a degenerate solution, the second term of Eq. A2 is non-zero and positive; for a non-degenerate solution the second term is zero.

Now suppose that there does exist a degenerate solution of Eq. 79

$$R_d^*(\omega) = R_{1d}^*(\omega) + jR_{2d}^*(\omega) \tag{A4}$$

and thus

$$r_d^*(t) = r_{1d}^*(t) + r_{2d}^*(t) \tag{A5}$$

The even part of this response, $r_{1d}^*(t)$, always makes a positive contribution to $\overline{a_1 e^2}$. However, since the magnitudes of the transforms of $r_{1d}^*(t)$ and $r_{2d}^*(t)$ are related as in Eq. A3, it may be that the first term of Eq. A2 is smaller than the minimum value of $\overline{a_1 e^2}$ that can be obtained for a non-degenerate solution. It is not possible to tell by this discussion whether or not a degenerate solution can minimize $\overline{a_1 e^2}$. But it is obvious that if a given degenerate solution can possibly minimize $\overline{a_1 e^2}$, $R_{1d}^*(\omega)$ must be the one function among all of the real functions with the magnitude $|R_{1d}^*(\omega)|$ that minimizes the second term of Eq. A2. Although $R_{1d}^*(\omega)$ will not satisfy this requirement, in general, let us assume for the sake of argument that it does. It is now possible to determine some of the characteristics of $R_{1d}^*(\omega)$ and its inverse transform $r_{1d}^*(t)$.

Let us recall that the integral square of any time function is dependent only upon the magnitude of the transform of the function. Thus all time functions with transforms having the magnitude $|R_{1d}^*(\omega)|$ have the same integral square value. When we note that

$$\left.\begin{array}{l} a_1(t) \geq b \ , \quad |t| < t_b \\[2em] a_1(t) = b \ , \quad |t| > t_b \end{array}\right\} \qquad \text{(A6)}$$

it is evident that the second term of Eq. A2, $\overline{a_1(t)[r_{1d}^*(t)]^2}$, has its smallest value when $r_{1d}^*(t)$ is a function that is zero,

or nearly zero for $|t| < t_b$, if there can be such a function. And such a function always does exist. Suppose, for example, that $|R^*_{1d}(\omega)|$ is as shown in Fig. A2(a). $R^*_{1d}(\omega)$ can then have either the value $|R^*_{1d}(\omega)|$ or minus $|R^*_{1d}(\omega)|$ at a given frequency. Suppose that the curve of $R^*_{1d}(\omega)$ has very closely spaced, regular, discontinuities, as shown in Fig. A2(b), with

$$\Delta\omega << \frac{\pi}{t_b} \tag{A7}$$

where $\Delta\omega$ is the frequency interval between discontinuities. Thus, the inverse transform of $R^*_{1d}(\omega)$ is a function with

$$r^*_{1d}(t) \cong 0 \quad , \quad |t| < t_b \tag{A8}$$

as desired. This is the one function among all of the functions with the transform magnitude $|R^*_{1d}(\omega)|$ that has the smallest weighted integral square. The weighted integral square of this function is

$$\overline{a_1(t)[r^*_{1d}(t)]^2} \cong \overline{b[r^*_{1d}(t)]^2} \tag{A9}$$

If a degenerate solution of Eq. 79 is to yield the minimum $\overline{a_1 e^2}$, then it must be that $R^*_{1d}(\omega)$ has very closely spaced discontinuities, as shown in Fig. A2(b). The inverse transform of this degenerate solution approximates the desired response with a weighted integral square error of
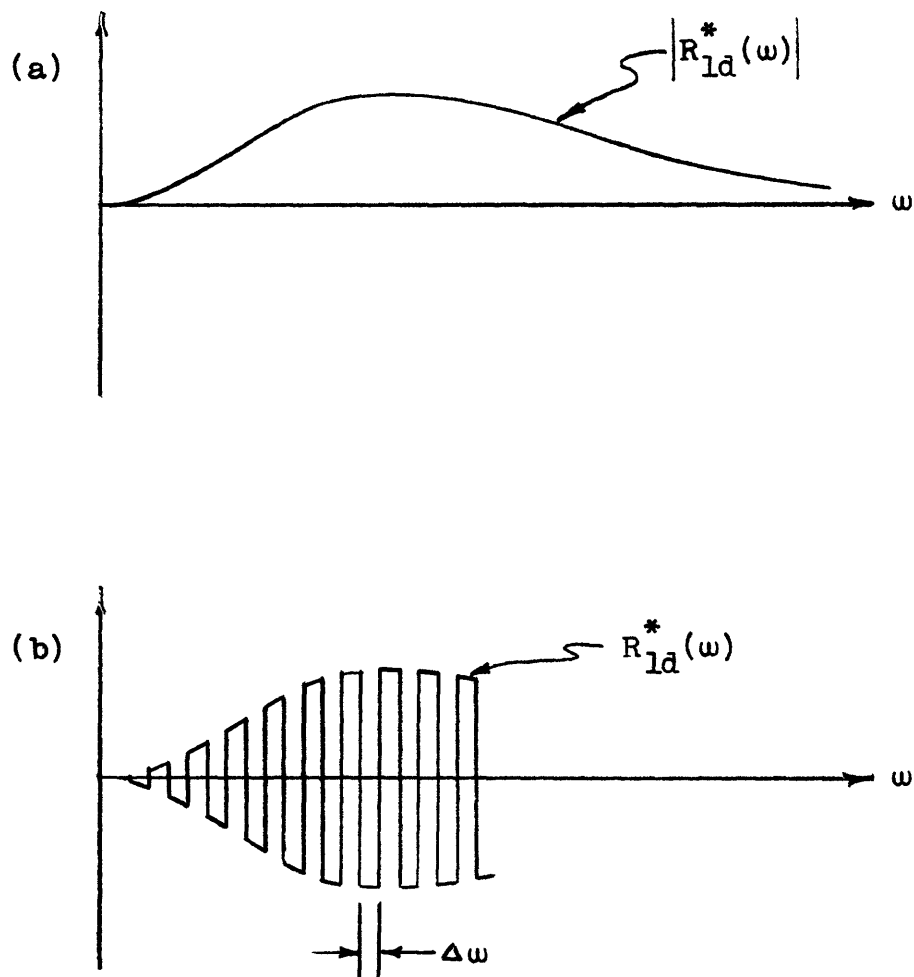
Fig. A2. Characteristics of the Function $R^*_{1d}(\omega)$

$$\overline{a_1 e^2} \cong \overline{a_1(t)[r_2(t) - r_{2d}^*(t)]^2} + \overline{b[r_{1d}^*(t)]^2} \tag{A10}$$

The "approximately equals" sign appears in Eqs. 8-10 and will appear in the equations of the next section. It should be noted that in each case the two quantities related by the sign can be made as nearly equal as desired by carrying toward the limit some process involved in the derivation of the equation. For example, the members of Eqs. 8-10 can be made as nearly equal as desired by making $\Delta\omega$ sufficiently small. Then we can interpret the approximately equals sign as meaning equals since we know that it can mean "as nearly equal as demanded".

A.3 Generation of a Non-Degenerate Solution of Eq. 79 with Smaller $\overline{a_1 e^2}$ Than That of any Degenerate Solution

In this section we shall show that it is possible to construct a non-degenerate solution, $R_n^*(\omega) = jR_{2n}^*(\omega)$, which has the required magnitude, $|I(\omega)|$, and whose inverse transform approximates the desired response with a weighted integral square error which is at least as small as that of any degenerate solution. Let us express $R_{2n}^*(\omega)$ as the sum of two functions: the imaginary part of the degenerate solution of section A.2, $R_{2d}^*(\omega)$, plus a function $R_{2c}^*(\omega)$ which will be defined presently.

$$R_{2n}^*(\omega) = R_{2d}^*(\omega) + R_{2c}^*(\omega) \tag{A11}$$

The function $R_{2n}^*(\omega)$ can have either the value $|I(\omega)|$ or minus $|I(\omega)|$ at a given frequency.

$$R_{2n}^*(\omega) = (-1)^{n(\omega)} |I(\omega)| \qquad (A12)$$

where $n(\omega)$ is any odd function of frequency whose value is always an integer, except at discontinuities. Through our choice of $n(\omega)$ we are able to choose the value of $R_{2c}^*(\omega)$ at any frequency as either of the values allowed by

$$R_{2c}^*(\omega) = (-1)^{n(\omega)} |I(\omega)| - R_{2d}^*(\omega) \qquad (A13)$$

And because

$$|R_{2d}^*(\omega)| \leq |I(\omega)| \qquad (A14)$$

for all ω, the two possible values of $R_{2c}^*(\omega)$ are never of the same algebraic sign.

Suppose, for example, that the two possible values of $R_{2c}^*(\omega)$ are as shown in Fig. A3. Now we propose to choose $R_{2c}^*(\omega)$ in such a way that $r_{2n}^*(t)$ is a corrected response whose weighted integral square error is at least as small as that of the degenerate solution, $r_d^*(t)$. Let $R_{2c}^*(\omega)$ be chosen as shown in Fig. A4. Choose $R_{2c}^*(\omega)$ positive for $0 < \omega < \omega_1$, negative for $\omega_1 < \omega < \omega_2$, negative for $\omega_2 < \omega < \omega_3$, positive for $\omega_3 < \omega < \omega_4$, etc., as shown in Fig. A4. The width of the frequency intervals in which $R_{2c}^*(\omega)$ is either wholly positive
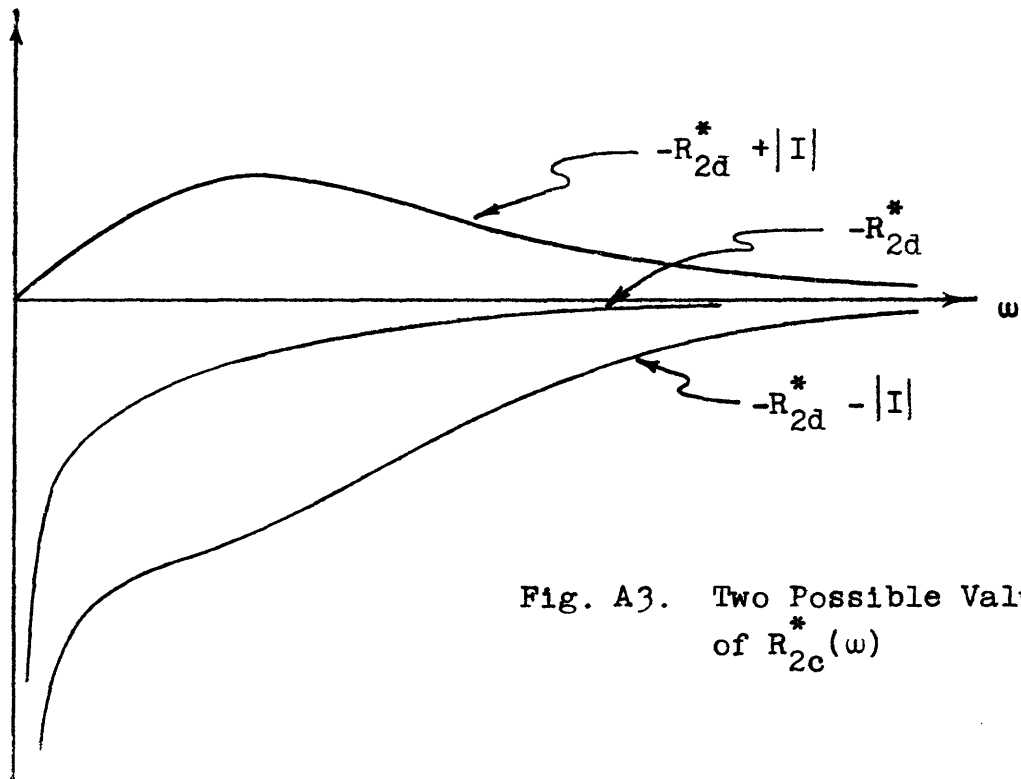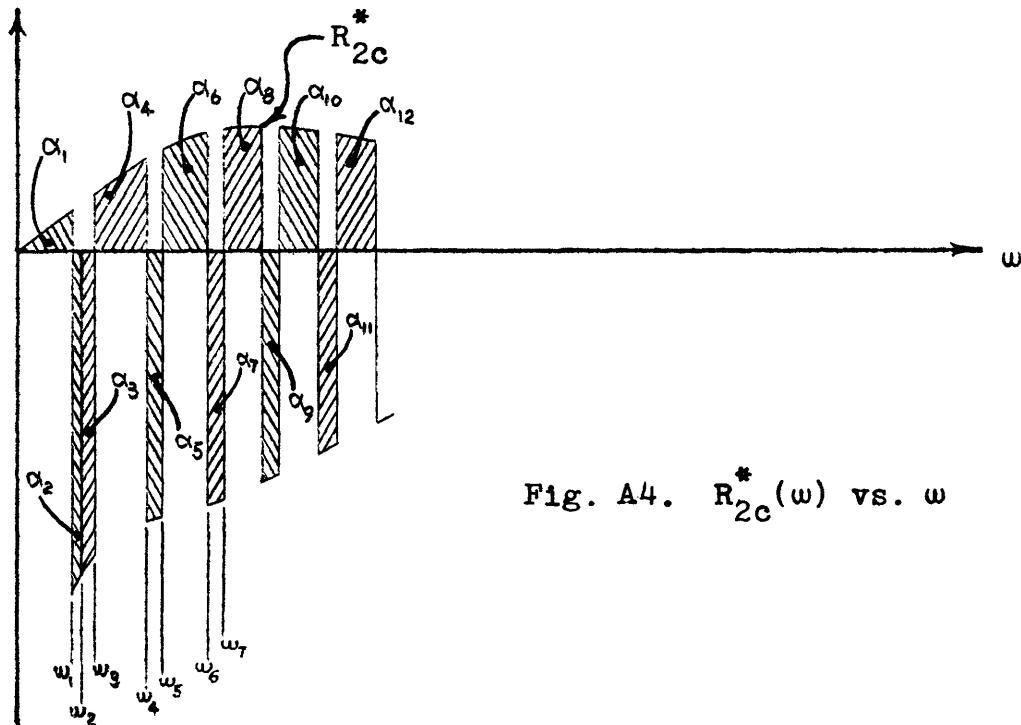
Fig. A3. Two Possible Values of $R_{2c}^*(\omega)$

Fig. A4. $R_{2c}^*(\omega)$ vs. ω

or negative is always much, much smaller than $\frac{\pi}{t_b}$. Thus the curve of $R_{2c}^*(\omega)$ versus ω is composed of many very short segments of the curves of $-R_{2d}^*(\omega) + |I(\omega)|$ and $-R_{2d}^*(\omega) - \lceil I(\omega)|$. Further, the frequencies ω_1, ω_2, ... , are chosen in such a way that the areas under pairs of adjacent segments are equal. In the notation of Fig. A4, $\alpha_1 = \alpha_2$, $\alpha_3 = \alpha_4$, etc. However, it is not necessary that $\alpha_1 = \alpha_3 = \alpha_5 = \ldots$. Having so chosen $R_{2c}^*(\omega)$, we note that

$$ r_{2c}^*(t) \cong 0 , \quad |t| < t_b \tag{A15} $$

and

$$ \overline{a_1(t)[r_{2c}^*(t)]^2} \cong \overline{b[r_{2c}^*(t)]^2} \tag{A16} $$

The inverse transform of the non-degenerate solution, $R_n^*(\omega)$, approximates the desired response with

$$ \overline{a_1 e^2} = \int_{-\infty}^{\infty} a_1(t) \left\{ r_2(t) - [r_{2d}^*(t) + r_{2c}^*(t)] \right\}^2 dt \tag{A17} $$

$$ = \overline{a_1(t)[r_2(t) - r_{2d}^*(t)]^2} - \overline{2a_1(t)r_{2c}^*(t)[r_2(t) - r_{2d}^*(t)]} $$

$$ + \overline{a_1(t)[r_{2c}^*(t)]^2} \tag{A18} $$

We must now treat two separate cases, depending on whether or not $b = 0$. First, suppose that $b = 0$. The second term of Eq. A18 is almost zero because $r_{2c}^*(t)$ is almost zero for $|t| < t_b$ and $a_1(t) = b = 0$ for $|t| > t_b$. Further, using Eq. A16, the third term of Eq. 18 is almost zero. Thus

$$\overline{a_1 e^2} \cong \overline{a_1(t)[r_2(t) - r_{2d}^*(t)]^2} \qquad (A19)$$

But for $b = 0$, Eq. A10 is the same as Eq. A19. We have been able to show that the weighted integral square error given by the non-degenerate solution generated in this section is the same as that of the arbitrarily specified degenerate solution in section A.2 .

Treatment of the case for which $b \neq 0$ is more difficult. As it was remarked in section A.1, if $b \neq 0$, $|I(\omega)|$ must be of a form which makes it possible for the corrected transient to approach the same final value as the desired step function, $r_2(\infty) = \frac{1}{2}$; otherwise the consideration of minimizing $\overline{a_1 e^2}$ has no meaning. Thus $I(\omega)$ must behave like $(j\omega)^{-1}$ near $\omega = 0$, and must have no other poles on the $j\omega$ axis of the s-plane.

An additional assumption about the choice of the time t_b in Fig. A1 is needed. It will be assumed that the time t_b is chosen large enough so that for $t > t_b$, $r_{2d}^*(t)$ has practically reached its final value of $r_2(\infty)$. This choice can always be made, and the forms of $R_{1d}^*(\omega)$ and $R_{2c}^*(\omega)$ readjusted to satisfy

Eq. A8 and Eq. A15, if necessary. But now the second term of Eq. A19 is almost zero because $r_{2c}^*(t)$ is almost zero for $|t| < t_b$ and $r_2(t) - r_{2d}^*(t)$ is almost zero for $|t| > t_b$. Thus

$$\overline{a_1 e^2} \cong \overline{a_1(t)[r_2(t)-r_{2d}^*(t)]^2} + \overline{b[r_{2c}^*(t)]^2} \tag{A20}$$

Comparing Eq. A20 with Eq. A10, we can see that $r_{2n}^*(t)$ yields less weighted integral square error than does $r_d^*(t)$ if

$$\overline{[r_{2c}^*(t)]^2} < \overline{[r_{1d}^*(t)]^2} \, .$$

Remembering that the value of the integral square of any time function is dependent only upon the magnitude of its transform, we can write

$$\overline{[r_{2n}^*(t)]^2} = \overline{[r_d^*(t)]^2} \tag{A21}$$

or

$$\overline{[r_{2d}^*(t)]^2} + 2\,\overline{r_{2d}^*(t)r_{2c}^*(t)} + \overline{[r_{2c}^*(t)]^2} = \overline{[r_{2d}^*(t)]^2} + \overline{[r_{1d}^*(t)]^2} \tag{A22}$$

Some care should be used in the interpretation of the above equations, for both sides of Eq. A21 are infinite and the meaning of the equation is somewhat in doubt. For the equations to have a clear meaning, it must be imagined for the moment that the functions $r_2(t)$, $r_{2n}^*(t)$ and $r_d^*(t)$ all approach zero rather

than a final value at large times, but that they converge
toward zero so slowly that all phenomena of interest have
taken place long before $r_2(t)$ shows an appreciable decrease.
After all, this is what must be assumed if these functions are
to be Fourier transformable, in the strict sense. Then, if
we consider the members of Eq. 21 as being extremely large,
but finite, the Eqs. A21 and A22 are unambiguous. This
distinction is mentioned only to avoid possible confusion and
is really of no concern.

From Eq. A22, we notice that we can show that
$\overline{[r_{2c}^*(t)]^2} < \overline{[r_{1d}^*(t)]^2}$ by showing that $\overline{r_{2d}^*(t)r_{2c}^*(t)}$ is
positive. First we must establish the fact that the function
$R_{2c}^*(\omega)$ does indeed approach zero as ω approaches zero just as
shown in Fig. A4. This can be proved most easily by assuming
that $R_{2c}^*(\omega)$ does not approach zero and deducing a contradiction.
Assume, for example, that $R_{2c}^*(\omega)$ behaves like the step func-
tion, $2ku_{-1}(\omega)$, in the vicinity of $\omega = 0$. Noticing that

$$|I(\omega)|^2 = [R_{2d}^*(\omega)]^2 + [R_{1d}^*(\omega)]^2 = [R_{2d}^*(\omega) + R_{2c}^*(\omega)]^2 \qquad (A23)$$

and making use of the assumed form of $R_{2c}^*(\omega)$, we find that for
small ω

$$[R_{2d}^*(\omega)]^2 + [R_{1d}^*(\omega)]^2 = [\,|R_{2d}^*(\omega)| + k]^2 \qquad (A24)$$
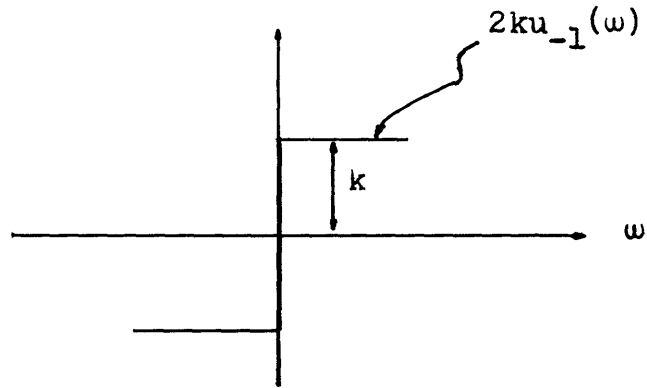
Fig. A5. Assumed Behavior of $R_{2c}^{*}(\omega)$ Near $\omega = 0$

or

$$[R_{1d}^{*}(\omega)]^2 = 2k \, |R_{2d}^{*}(\omega)| \; + k^2 \tag{A25}$$

Since we know that $r_{d}^{*}(t)$ approaches $r_2(\infty)$ as its final value, it must be that in the vicinity of $\omega = 0$

$$jR_{2d}^{*}(\omega) = (j\omega)^{-1} + \left\{ \begin{array}{l} \text{other terms involving} \\ \omega \text{ raised to powers} \\ \text{greater than -1} \end{array} \right\} \tag{A26}$$

Considering now the integral square of $r_{1d}^{*}(t)$

$$\overline{[r_{1d}^{*}(t)]^2} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [R_{1d}^{*}(\omega)]^2 \, d\omega \tag{A27}$$

it follows from Eqs. A25 and A26 that $\overline{[r_{1d}^{*}(t)]^2}$ is infinite. Since $a_1(t)$ has the form shown in Fig. A1 with $b \neq 0$, it must

104

be that $\overline{a_1(t)[r_{1d}^*(t)]^2}$ is infinite and thus from Eq. A10

that $\overline{a_1 e^2}$ for the degenerate solution, $R_d^*(\omega)$, is infinite. A

degenerate solution which yields infinite weighted integral

square error certainly cannot minimize $\overline{a_1 e^2}$. It is

necessary to conclude that $R_{2c}^*(\omega)$ does not have the behavior

that we assumed above, and that it does approach zero as ω

approaches zero, just as in Fig. A4.

Having disposed of this matter, we can now evaluate the

quantity $\overline{r_{2d}^*(t)r_{2c}^*(t)}$ to see whether or not it is positive.

Because $r_{2c}^*(t)$ is almost zero for $|t| < t_b$ and $r_{2d}^*(t) \cong r_2(\infty)$

for $|t| > t_b$

$$\overline{r_{2d}^*(t)r_{2c}^*(t)} \cong 2 r_2(\infty) \int_0^\infty r_{2c}^*(t) \, dt \qquad (A28)$$

$$\cong 2 r_2(\infty) \overline{[u_{-1}(t) r_{2c}^*(t)]} \qquad (A29)$$

where $u_{-1}(t)$ is a unit step function. Equation A29 can be

written in terms of the transforms of $u_{-1}(t)$ and $r_{2c}^*(t)$ as

$$\overline{r_{2d}^*(t)r_{2c}^*(t)} = \frac{r_2(\infty)}{\pi} \int_{-\infty}^\infty (-\frac{1}{j\omega}) j R_{2c}^*(\omega) \, d\omega \qquad (A30)$$

$$= -\frac{2r_2(\infty)}{\pi} \int_0^\infty \frac{R_{2c}^*(\omega)}{\omega} \, d\omega \qquad (A31)$$

Let us determine the algebraic sign of the integral expression

$$\int_0^\infty \frac{R_{2c}^*(\omega)}{\omega} \, d\omega \qquad\qquad (A32)$$

in Eq. A31. Remember that the curve of $R_{2c}^*(\omega)$ versus ω is composed of many short segments of the curves $-R_{2d}^*(\omega) \pm |I(\omega)|$, and that adjacent pairs of segments of $R_{2c}^*(\omega)$ enclose equal areas. But considering now the integrand of Eq.A32, the areas enclosed by adjacent pairs of segments of $\omega^{-1}R_{2c}^*(\omega)$ are not equal; the net contribution of each pair of segments of $\omega^{-1}R_{2c}^*(\omega)$ to the integral is of the same sign as the first segment of the pair. We see in Fig. A4 that the first pair of segments come in the order, positive then negative. All succeeding pairs of segments come in the order, negative then positive. The first pair of segments of $\omega^{-1}R_{2c}^*(\omega)$ makes a positive contribution to the integral in Eq. A32. Since $R_{2c}^*(\omega)$ approaches zero as ω approaches zero, this contribution is finite and can be made arbitrarily small by choosing ω_1 and ω_2 sufficiently small. All succeeding pairs of segments of $\omega^{-1}R_{2c}^*(\omega)$ make negative contributions to the integral. Let us suppose that the frequencies ω_1 and ω_2 have been chosen small enough so that the integral expression of Eq. A32 is negative; according to Eq. A31 $\overline{r_{2d}^*(t)r_{2c}^*(t)}$ is positive. From Eq. A22,

$$\overline{[r_{1d}^*(t)]^2} > \overline{[r_{2c}^*(t)]^2} \qquad\qquad (A33)$$

Comparing Eqs. A10 and A20, we see that the weighted integral square error given by the inverse transform of the non-degenerate solution $R_{2n}^*(t)$ is less than that given by the inverse transform of the degenerate solution $R_d^*(t)$. The proof for the case where $b \neq 0$ is complete.

It should not be inferred that the particular non-degenerate solution derived in this section, $R_{2n}^*(\omega)$, is the non-degenerate solution which minimizes $\overline{a_1 e^2}$. The function $R_{2n}^*(\omega)$ is simply a non-degenerate solution which was generated in such a way as to permit easy proof that its weighted integral square error is at least as small as that of the degenerate solution, $R_d^*(\omega)$. Since we can always obtain a non-degenerate solution which yields a value of $\overline{a_1 e^2}$ at least as small as the $\overline{a_1 e^2}$ of any given degenerate solution, it must be concluded that the function which minimizes $\overline{a_1 e^2}$ is a non-degenerate solution. But there is no reason to believe that the minimizing solution is one which can be generated as $R_{2n}^*(\omega)$ was. In fact, $r_{2n}^*(t)$ exhibits a very undesirable behavior. It approaches $r_2(t)$ nicely, just as $r_{2d}^*(t)$ does, until after a large time t_b when appreciable error again occurs due to $r_{2c}^*(t)$. If it should happen that a corrected response with this behavior does minimize $\overline{a_1 e^2}$, then $a_1(t)$ has been poorly chosen. In this case, the level b in Fig. A1 should be increased in order to weight more heavily the later errors. There are certainly appropriate weighting functions that give rise to a minimizing response which does not exhibit

this undesirable behavior. In the limiting case of raising the
level b to higher and higher values, the weighting function
becomes a constant, and then the minimizing response certainly
does not have the undesirable characteristic of $r_{2n}^*(t)$.

We have assumed a particular desired response here, a
unit step function, in order to facilitate the discussion.
However, when we review the proof we realize that it could as
well have been made for any other odd function that might be
desired as the response in a physically motivated problem.

The conclusion that has been reached is this: in any
practical phase correction problem, the corrected response
which approximates any odd desired response with the least
weighted integral square error is an odd time function.

BIBLIOGRAPHY

1. Cerrillo, M.V., Basis Existance Theorems, Quarterly Progress Report, Research Laboratory of Electronics, M.I.T., (Jan. 15, 1952)

2. Wheeler, H.A., The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, Proc. I.R.E., 27, (June, 1939)

3. Di Toro, M.J., Phase and Amplitude Distortion in Linear Networks, Proc. I.R.E., 36, (Jan., 1948)

4. Jahnke and Emde, Tables of Functions, Dover Publications, Inc., New York, (1945)

5. Tables of the Function $\frac{\sin \phi}{\phi}$ and of Its First Eleven Derivatives, The Annals of the Computation Laboratory of Harvard University, Harvard University Press, Cambridge, Mass., (1949), vol. XXII

6. Guillemin, E.A., Computational Techniques Which Simplify the Correlation Between Steady State and Transient Response of Filters and Other Networks, Proceedings of the National Electronic Conference, vol. 9, (Feb., 1954)