This dissertation has been microfilmed exactly as received

OLSON, Karl William, 1936—
OPTIMUM QUANTIZER SYSTEM DESIGN,
The Ohio State University, Ph.D., 1965
Engineering, electrical

University Microfilms, Inc., Ann Arbor, Michigan
OPTIMUM QUANTIZER
SYSTEM DESIGN

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Karl William Olson, B.E.E., M.Sc.

The Ohio State University
1965

Approved by

[Signature]
Adviser
Department of Electrical Engineering
ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to Professor Robert L. Cosgriff for his guidance and encouragement throughout the duration of this study and also to Professors C. Earl Warren and William C. Davis for their criticisms and suggestions. The author also wishes to express his gratitude to Virginia L. Spragg for her patience in typing the manuscript and also to Evan L. Brill for his services as a digital computer programmer. Finally, the author is indebted to Kenneth E. Teeters and Thomas R. Mongold for their help in preparing the manuscript.

The work reported in this dissertation was supported in part by a contract between The Ohio State University Research Foundation and the Air Force Avionics Laboratory, Wright-Patterson Air Force Base, Ohio.
VITA

March 27, 1936  Born - Canton, Ohio

1959  B. E. E. and M. Sc., The Ohio State University, Columbus, Ohio

1960 - 1965  Research Associate, The Ohio State University, Department of Electrical Engineering, Columbus, Ohio

FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Mathematics. Professor Kurt Endl

Studies in Classical Physics. Professor Wave Shaffer

Studies in Nonlinear Control Theory. Professor Robert L. Cosgriff
CONTENTS

ACKNOWLEDGMENTS ii
VITA iii
CONTENTS iv
FIGURES vi
TABLES vii

Chapter

I. INTRODUCTION 1

II. RANDOM SIGNAL THEORY 5

Fourier Series Representation
Non-Periodic Series
Signal Moments
Random Number Set Theory

III. ORTHOGONAL FUNCTION DEVELOPMENT 18

Autocorrelation Function of \( f(x(t), q) \)
Derivation of Orthogonal Functions

IV. PROPERTIES OF HERMITE POLYNOMIALS 26

Convergence at a Point

V. OPTIMUM QUANTIZER LEVELS 34

Preliminary Discussion
Quantizing Error Minimization
The Direct Method
The Hermite Expansion Method
FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Two terminal network representation of signal quantizer.</td>
<td>34</td>
</tr>
<tr>
<td>2.</td>
<td>Typical quantizer input-output characteristic.</td>
<td>35</td>
</tr>
<tr>
<td>3.</td>
<td>Typical quantizing error function.</td>
<td>35</td>
</tr>
<tr>
<td>4.</td>
<td>Quantizing error with restriction imposed.</td>
<td>40</td>
</tr>
<tr>
<td>5.</td>
<td>Component function of quantizer output function.</td>
<td>43</td>
</tr>
<tr>
<td>6.</td>
<td>Discontinuity midpoints lie along 45° line.</td>
<td>48</td>
</tr>
<tr>
<td>7.</td>
<td>Filter optimizing configuration.</td>
<td>58</td>
</tr>
<tr>
<td>8.</td>
<td>Desired signal component of quantizer input and filter output signal.</td>
<td>85</td>
</tr>
<tr>
<td>9.</td>
<td>Noise component of quantizer input.</td>
<td>85</td>
</tr>
<tr>
<td>10.</td>
<td>Quantizer input signal.</td>
<td>85</td>
</tr>
<tr>
<td>11.</td>
<td>Quantizer output signal.</td>
<td>85</td>
</tr>
</tbody>
</table>
TABLES

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>OPTIMUM PARAMETERS FOR A FOUR LEVEL SYMMETRICAL QUANTIZER</td>
<td>53</td>
</tr>
<tr>
<td>2</td>
<td>OPTIMUM PARAMETERS FOR A SEVEN LEVEL SYMMETRICAL QUANTIZER</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>HERMITE EXPANSION COEFFICIENTS FOR AN OPTIMIZED 7 LEVEL QUANTIZER FOR THE CASE WHERE $\sigma_x^2 = a + b = 1$</td>
<td>61</td>
</tr>
<tr>
<td>4</td>
<td>SIGNAL POWER COMPONENTS</td>
<td>74</td>
</tr>
<tr>
<td>5</td>
<td>MAGNITUDES OF COEFFICIENTS FOR RANDOM FUNCTION EXPRESSION</td>
<td>81</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

In general, all systems encountered in the real world are nonlinear. Even systems which are usually assumed to be linear are actually nonlinear to some degree. Consequently, the systems engineer is faced with the problem of considering nonlinearities in the analysis and synthesis of systems. The technique to be described permits the separation of the linear and nonlinear effects of systems where the system input signals are normally distributed.

It will be shown that the output of the nonlinear system in question is represented by a Hermite polynomial expansion in terms of its input signal, then the component of the output power spectrum due to the linear effect of the system can be separated from the component of the output power spectrum due to the nonlinear effect of the system. Furthermore, the portion of the system output signal which is attributed to the linear effect of the system is uncorrelated with the remaining portions of the output signal. Thus the Wiener optimum linear filter procedure can be conveniently applied to specify an optimum filter capable of reducing the components of the system output signal which are due to the nonlinear effects of the system. The nonlinear system treated in
this work is the signal quantizer; however the techniques presented can be applied to nonlinear devices in general provided that the output of the device is a single valued function of the input variable.

In many instances when an analog signal is to be stored or recorded and must be reproduced at a later time, or when the signal must be transmitted from one point to another, the analog signal is first quantized and then digitally coded. With the techniques of error detecting and correcting codes which are currently available, an accurate replica of the quantized signal can be recovered from the recording or at the receiving end of the transmission link.

To this point the error which has been introduced into the resulting signal, assuming errorless recovery of the quantized signal, is that due to the nonlinearity of the quantizer. If the statistical properties of the input signal are known, then the quantizer can be designed in such a way that the mean square quantizing error is minimized. This error can be further reduced by the use of a properly designed filter.

In order to design such a filter the power spectrum of the quantized waveform is required. This could be calculated by the method of taking the Fourier transform of the autocorrelation function of the quantized signal, however, by expressing the quantized signal in terms of a summation of uncorrelated components, the power spectrum can be more simply obtained. As an introduction to this technique, the representation of a random signal as a series of nonharmonically related
terms is discussed in Chapter II. In the following chapter the method
of construction of a set of orthogonal functions of the random signal of
Chapter II is given. These orthogonal functions are shown to be Hermite
polynomials. The power spectrum of the quantized signal is subsequently
obtained from the expansion of the quantizer output signal in terms of
Hermite polynomials. The material discussed in Chapters II and III
follows essentially unpublished class notes by Professor R. L. Cosgriff.

The power spectrum of the quantizer output as obtained by this
method is in the form of an infinite series of uncorrelated terms;
however, a method is given for the approximation of the series by the
use of only a finite number of terms plus a remainder.

A linear optimum filter is specified for the case where the
quantizer input consists of two independent, normally distributed func-
tions, one being considered as signal and the other as noise. The
criterion used for optimizing the filter is that the mean square of the
difference between the filtered output and the desired component of the
input signal to the quantizer input is a minimum.

This configuration of input signals, quantizer, and filter has
been simulated on the digital computer, and strip charts of the results
are shown in Chapter VII.
Throughout this work the random signals involved have been assumed to be ergodic and therefore no distinction has been made between the symbols denoting time autocorrelation and statistical autocorrelation.
CHAPTER II
RANDOM SIGNAL THEORY

Fourier Series Representation

Rice,\(^1\) Williams,\(^2\) Ragazzini,\(^3\) and others have treated a random
time varying signal \(x\) by means of a Fourier series letting

\[
(2-1) \quad x = \sum_{m=-\infty}^{\infty} a_m \, e^{j m \Delta t}
\]

where the \(m\)'s are integers and \(T \Delta = 2\pi\) where \(T\) is the period of
the sample. Here

\[
(2-2) \quad a_m = \frac{1}{T} \int_{0}^{T} x(t) \, e^{-j m \Delta t} \, dt.
\]

If one has an ensemble of \(x\)'s which are stationary and have
a normal distribution, the ensemble average of any \(a_m\) is zero and
the mean square value of \(a_m\) will be equal to

\[
(2-3) \quad \overline{|a_m a_m^*|} = \frac{1}{2\pi} \int_{(m-1/2)\Delta}^{(m+1/2)\Delta} S_x(\omega) \, d\omega
\]

where \(S_x(\omega)\) is the power spectrum of \(x(t)\). In the limit as
\(T \to \infty\) and \(\Delta \to 0\), the ensemble average as specified by Eq. (2-3)
approaches the continuous spectrum

\[ \frac{1}{2\pi} S_x(m_\Delta) \]

If \( x \) has a normal distribution and if the \( a_m \)'s are represented by \( a_{mr} + j a_{mi} \), then \( a_{mr} \) and \( a_{mi} \) will both have zero mean and be normally distributed.

Non-periodic Series

The foregoing representation is very popular and well accepted because of the association with Fourier Series; however, a modification of this representation which is more convenient to use is given by

\[(2-4) \quad x = \lim_{\Delta \to 0} \sum_{n = -\infty}^{\infty} a_n e^{jn\Delta t}\]

where the \( n \)'s are not integers. It is desired that \( x \) be real; therefore

\[ a_n = a_n^* \]

The coefficients, \( a_n \), are related to the power spectrum of \( x \), which will be denoted by \( S_x(\omega) \), as follows:

\[(2-5) \quad |a_n|^2 = \lim_{\Delta \to 0} \frac{1}{2\pi} \int_{(k-1/2)\Delta}^{(k+1/2)\Delta} S_x(\omega) d\omega = \frac{1}{2\pi} \int_{(k-1/2)\Delta}^{(k+1/2)\Delta} S_x(\omega) d\omega \]
where \( k \) is the closest integer to the \( n \) in question.

The values of \( n \) for positive integer values of \( k \), zero inclusive, are chosen such that

\[(2-6) \quad k - 1/2 < n_k < k + 1/2 .\]

The values of \( n \) for negative integer values of the index \( k \) are chosen such that

\[(2-7) \quad n_k = - n_{-k} .\]

The set of numbers, \( n_k \), as defined for all integer values of \( k \) by Eq. (2-6) and (2-7) is said to have the property of cancellation, which means that

\[
\sum_{k=1}^{P} n_k = \sum_{k=1}^{P} n_k
\]

\( k \neq q \)

where

\( 1 \leq q \leq p \).

A restriction on the set of \( n_k \)'s will be given after two preliminary definitions. The sum \( T_1 \) is to consist of the sum of an arbitrary set of \( n_k \)'s after all cancelling numbers have been removed from the set. The sum \( T_2 \) consists of the sum of a second arbitrary set of non-cancelling numbers. The restriction, on the set of \( n_k \)'s defined by Eq. (2-6) and (2-7), is that the sum \( T_1 \) can be identical to the sum \( T_2 \) if and only if the set of numbers which is summed in sum \( T_1 \) is equal to the set of numbers summed in sum \( T_2 \). Thus if one positive \( n_k \)
is rational then all other positive \( n_k \)'s must be irrational. Throughout the remainder of this chapter and remaining chapters, references to the set of \( n \)'s will refer to that set defined by Eq. (2-6) and (2-7) and subject to the restriction given above.

This restriction upon the \( n \)'s is such that it allows a simplified representation for the moments of the signal of Eq. (2-4) and most important it allows functions of the signals given by Eq. (2-4) to be constructed which are orthogonal and which have convenient characteristics for the representation of a nonlinear process.

One scheme which could be used for selecting the members of set \( n_k \) is

\[
n_k = \sqrt{k^2 + \frac{1}{P_k}}
\]

where the \( P_k \)'s are prime numbers, that is,

\[
1, 2, 3, 5, 7 \cdots .
\]

Throughout the present chapter only time averages will be considered. The time average of \( x \) will be denoted by \( <x> \).

Signal Moments

It has been shown\(^5\) that if a random variable, \( x(t) \), has moments of the form

\[
(2-8) \quad <x^p> = 0
\]

for \( p \) equal to odd integers and
for \( p \) equal to even integers, then \( x(t) \) is normally distributed.

The moments of the signal specified by

\[
x(t) = \lim_{\Delta \to 0} \sum_{n=-\infty}^{\infty} a_n e^{j n \Delta t}
\]

will be calculated and will be shown to be of the same form as Eq. (2-8) and (2-9).

It will be assumed that \( a_0 \to 0 \) as \( \Delta \to 0 \); therefore, the average value of \( x \) is zero. The time average of \( x^2 \) is obtained from

\[
x^2 = \sum_{n_1=-\infty}^{\infty} a_{n_1} a_{n_2} e^{j(n_1+n_2) \Delta t}
\]

and is given by the sum of all zero frequency terms of \( x^2 \). These can occur only when \( n_1 = -n_2 \); therefore,

\[
\langle x^2 \rangle = \sum_{-\infty}^{\infty} a_{n_1} a_{-n_1} \lim_{\Delta \to 0} \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) \, d\omega = \sigma_x^2.
\]
The average value of $x^4$, where

$$x^4 = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \sum_{n_3=-\infty}^{\infty} \sum_{n_4=-\infty}^{\infty} a_{n_1} a_{n_2} a_{n_3} a_{n_4} e^{j(n_1+n_2+n_3+n_4)\Delta t}$$

is determined by finding all of the possible ways in which the indices can be paired such that

$$n_1 + n_2 + n_3 + n_4 = 0.$$ 

This complete cancellation can be achieved in the following three ways:

(2-13) $n_1 = -n_2$ and $n_3 = -n_4$, or

(2-14) $n_1 = -n_3$ and $n_2 = -n_4$, or

(2-15) $n_1 = -n_4$ and $n_2 = -n_3$.

The selection rules indicated above were first proposed by Levadi.6

The average value of $x^4$ is, therefore, equal to the sum of three identical double summations, or simply

$$\langle x^4 \rangle = 3 \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} |a_{n_1}|^2 |a_{n_2}|^2 = 3 \sigma_x^4$$

since

$$\sum_{n_i=-\infty}^{\infty} |a_{n_i}|^2 = \sigma_x^2.$$ 

Some terms of the quadruple summation of Eq. (2-12) have been included twice in the expression given by Eq. (2-16). This can be seen by rewriting Eq. (2-16) as three separate terms resulting from Eq. (2-13),
(2-14) and (2-15) respectively. Thus

\[
\langle x^4 \rangle = \sum_{n_1=-n_2}^{n_1=n_2} \sum_{n_3=-n_4}^{n_3=n_4} |a_{n_1}|^2 |a_{n_3}|^2 + \sum_{n_1=-n_3}^{n_1=n_3} \sum_{n_2=-n_4}^{n_2=n_4} |a_{n_1}|^2 |a_{n_2}|^2
\]

+ \sum_{n_1=-n_4}^{n_1=n_4} \sum_{n_2=-n_3}^{n_2=n_3} |a_{n_1}|^2 |a_{n_2}|^2.

The terms of Eq. (2-12) for which

\[
(2-18) \quad n_1 = -n_2 = -n_3 = n_4
\]

are accounted for once by the first term of Eq. (2-17) in the cases where \( n_1 = -n_3 \), thus \( n_1 = -n_2 = -n_3 = n_4 \), and a second time by the second term of Eq. (2-17) in the cases where \( n_1 = -n_2 \), thus \( n_1 = -n_3 = -n_2 = n_4 \). There are two other combinations of the four indices for which terms of Eq. (2-12) are accounted for twice in Eq. (2-16). These combinations are

\[
(2-19) \quad n_1 = -n_2 = n_3 = -n_4
\]

and

\[
(2-20) \quad n_1 = n_2 = -n_3 = -n_4.
\]

In spite of the fact that technically Eq. (2-16) is not an exact representation of the average value of Eq. (2-12), it does yield the exact value for \( \langle x^4 \rangle \) as \( \Delta \) approaches zero. This can be proven by showing that the average value of the terms which are considered twice is zero in the limit as \( \Delta \) approaches zero. From Eq. (2-12) and (2-5) the average value of the terms specified by each one of Eq.
(2-18), (2-19) and (2-20) can be seen to be

\[
(2-21) \quad \sum_{n_1=-\infty}^{\infty} |a_{n_1}|^2 |a_{-n_1}|^2 = \frac{\Delta}{2\pi} \sum_{n_1=-\infty}^{\infty} \frac{\Delta}{2\pi} [S_x(n_1, \Delta)]^2
\]

where \( \bar{n}_1 \) is the closest integer to \( n_1 \). As is indicated by Eq. (2-5), in the limit as \( \Delta \to 0 \), the sum of Eq. (2-21) can be written as an integral and therefore

\[
(2-22) \quad \sum_{n_1} |a_{n_1}|^2 |a_{-n_1}|^2 = \lim_{\Delta \to 0} \left( \frac{\Delta}{2\pi} \right) \frac{1}{2\pi} \int_{-\infty}^{\infty} [S_x(\omega)]^2 d\omega
\]

\[
\leq \lim_{\Delta \to 0} \left( \frac{\Delta}{2\pi} \right) S_{x_{\text{max}}} \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(\omega) d\omega
\]

where \( S_{x_{\text{max}}} \) is the maximum value of \( S_x(\omega) \). Finally, from Eq. (2-11), the above expression can be seen to be

\[
(2-23) \quad \sum_{n_1=-\infty}^{\infty} |a_{n_1}|^2 |a_{n_1}|^2 \leq \lim_{\Delta \to 0} \frac{\Delta}{2\pi} S_{x_{\text{max}}} \Delta = 0
\]

Thus the proof is complete for the case of \( <x^4> \). A similar situation exists for each higher order even moment and the same type of argument can be applied to these cases.

In general, the average value of \( x^p \) where \( p \) is an even integer and where
is determined by finding the number of possible ways of matching the indices \( n_1 \) through \( n_p \) in such a way that their sum is zero. This can be accomplished in the following manner. Call one of the \( p \) indices \( n_a \) and remove it from the available list which leaves \( p-1 \) indices. Thus there are \( p-1 \) ways of choosing an index to be set equal to \(-n_a\). Call one of the remaining \( p-2 \) indices \( n_b \) and remove it from the available list of indices which leaves \( p-3 \). There are \( p-3 \) ways of choosing an index to be set equal to \(-n_b\). This process is continued until all \( p \) indices have been paired. The number of ways of making the indicated choices is

\[
(p-1) (p-3) (p-5) \cdots 3 \cdot 1
\]

Therefore, the average value of \( x^p \) can be expressed simply as

\[
\langle x^p \rangle = \left[ (p-1)(p-3) \cdots 3 \cdot 1 \right] \sum_{n_a=-\infty}^{\infty} \sum_{n_b=-\infty}^{\infty} \cdots \sum_{n_{p/2}=-\infty}^{\infty} |a_{n_a}|^2 |a_{n_b}|^2 \cdots |a_{n_{p/2}}|^2
\]

Since

\[
(p-1)(p-3) \cdots 3 \cdot 1 = \frac{p!}{(p/2)^{p/2}}
\]

then finally

\[
(2-26) \quad \frac{p!}{(p/2)^{p/2}}
\]
for \( p \) equal to the even integers. The average value of \( x^m \), where \( m \) is an odd integer, is zero since the sum of an odd number of elements of the set \( n \) cannot be made to cancel. That is, after all possible pairs of \( n \)'s cancel, at least one \( n \) always remains.

Thus the signal \( x(t) \) as defined by Eq. (2-4) has been shown to have moments which are equal to those given by Eqs. (2-8) and (2-9). The signal \( x(t) \) has, therefore, a normal distribution.

Random Number Set Theory

In the discussion to follow a method of constructing mutually exclusive sets of numbers from the elements of the set of \( n \)'s, which was defined in the section on Nonperiodic Series, will be given. These mutually exclusive sets will be defined as \( S_k \) where \( 0 \leq k < \infty \). The frequencies obtained by multiplying members of the set \( S_k \) by \( \Delta \) will be described as belonging to the set \( S_k \). Subsequently an ensemble of functions of \( x(t) \), which will be denoted as \( f(x(t), k) \), will be constructed in such a way that the frequency of each term of \( f(x(t), k) \) is a member of only set \( S_k \). (The \( x(t) \) referred to in the previous statement is defined by Eq. (2-4).)
The set \( S_0 \) has only the zero element corresponding to zero frequency. The set \( S_1 \) is equal to the set of \( n \)'s as defined in the section on Nonperiodic Series. The elements of the set \( S_2 \) are equal to \( n_1 + n_2 \) where \( n_1 \) and \( n_2 \) each independently assume all possible values of \( n \) with the exception that

\[
n_1 \neq -n_2
\]

thus zero is not a member of set \( S_2 \). Due to the restriction on the values of \( n \), given in the section on Nonperiodic Series, no element of \( S_2 \) is identical to any element of \( S_1 \) or \( S_0 \). The elements of the set \( S_q \) are equal to

\[
\sum_{m=1}^{q} n_m
\]

where each \( n_m \) assumes all possible values of \( n \) but restricted in such a way that no two \( n_m \)'s cancel or add to zero since a cancellation of two of the \( n_m \)'s would cause the sum of the remaining number to fall into set \( S_{q-2} \). Therefore the signal

\[
X(t) = \lim_{\Delta \to 0} \sum_{n=-\infty}^{\infty} a_n e^{jn \Delta t}
\]

contains terms at frequencies which are members of set \( S_1 \) only.

Whereas the signal
contains terms at frequencies which are members of set $S_2$, covering the cases for which $n_1 \neq -n_2$. The signal $x^2(t)$ also contains terms at zero frequency which is the sole member of set $S_0$, covering the remaining cases for which $n_1 = -n_2$.

Since the signals $x(t)$ and $x^2(t)$ contain no common frequencies they can be shown to be uncorrelated. To illustrate this point, consider the two functions

$$b_1 e^{jn_1 \Delta t} \text{ and } b_2 e^{jn_2 \Delta t}$$

where $n_1$ and $n_2$ are single elements of sets $S_1$ and $S_2$ respectively. The time cross correlation of these two functions is

$$(2-30) \quad \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} (b_1 e^{jn_1 \Delta t})(b_2 e^{jn_2 \Delta (t+\tau)}) \, dt.$$ 

It will be shown that since $n_1$ and $n_2$ are elements from different sets, which implies that $n_1 \neq \pm n_2$, this cross correlation function approaches zero as $T$ approaches infinity for all values of $\tau$. That is

$$(2-31) \quad \lim_{T \to \infty} \frac{b_1 b_2}{2T} e^{jn_2 \Delta \tau} \int_{-T}^{T} e^{j(n_1+n_2) \Delta t} \, dt$$

$$= \lim_{T \to \infty} \frac{b_1 b_2}{2T} e^{jn_2 \Delta \tau} \frac{2 \sin(n_1+n_2) \Delta T}{(n_1+n_2) \Delta} = 0 \quad n_1 \neq -n_2.$$
The preceding argument is of course equally applicable in the general case where \( n_1 \) and \( n_2 \) are single elements of sets \( S_k \) and \( S_m \) respectively, where \( k \) is not equal to \( m \). In general the signal \( x^k(t) \) has terms at frequencies which are members of sets \( S_k \), \( S_{k-2} \), \( S_{k-4} \), \( \cdots \), \( S_1 \) if \( k \) is an odd integer. If \( k \) is an even integer the frequencies are members of sets \( S_k \), \( S_{k-2} \), \( S_{k-4} \), \( \cdots \), \( S_2 \), \( S_0 \).
CHAPTER III

ORTHOGONAL FUNCTION DEVELOPMENT

Now that the properties of the mutually exclusive sets of numbers, $S_k$, have been defined, it is possible to construct a function, $f(x(t), q)$, which is composed of a linear combination of $x^q(t)$, $x^{q-2}(t)$, $x^{q-4}(t)$, etc., in such a way that the frequency of each term of $f(x(t), q)$ is a member of set $S_q$ only. (All references to $x(t)$ in this chapter refer to the function defined by Eq. (2-4).) The function which is constructed such that the frequency of each of its terms is a member of set $S_k$ where $k \neq q$ is denoted by $f(x(t), k)$. Since the elements of set $S_q$ and $S_k$ are mutually exclusive then $f(x(t), k)$ is uncorrelated with $f(x(t), q)$. That is

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x(t), k) f(x(t+\tau), q) \, dt = 0$$

for $k \neq q$. Thus the functions, $f(x(t), q)$ are orthogonal. It will be shown that $f(x(t), q)$ is equivalent to the $q$th order Hermite polynomial.
Autocorrelation Function of $f(x(t), q)$

As a preliminary step to constructing the functions $f(x(t), q)$, the autocorrelation function of $f(x(t), q)$ will be obtained by considering the equivalent definition of $f(x(t), q)$, that is, by considering only those terms of $x^q(t)$ at frequencies which are members of set $S_q$.

Thus

$$f(x(t), q) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \cdots \sum_{n_q=-\infty}^{\infty} a_{n_1} a_{n_2} \cdots a_{n_q} e^{j(n_1+n_2+\cdots+n_q) \Delta t}.$$  \hspace{1cm} (3-2)

Since $(n_1 + n_2 + \cdots + n_q)$ is by definition an element of set $S_q$, then these indices are restricted by $(q-1)!$ equations of the form

$$n_1 \neq n_2$$
$$n_1 \neq n_3 \quad n_2 \neq n_3$$
$$\cdot \quad \cdot \quad \quad \text{etc.},$$
$$\cdot \quad \cdot \quad \cdot$$
$$n_1 \neq n_q \quad n_2 \neq n_q \quad \cdot \cdot \cdot$$

The autocorrelation function of $f(x(t), q)$ is

$$R_{f(x,q)}(\tau) = \left\langle f(x(t), q) \ f(x(t+\tau), q) \right\rangle$$

$$= \left\langle \sum_{n_{11}} \sum_{n_{21}} \cdots \sum_{n_{q1}} a_{n_{11}} a_{n_{21}} \cdots a_{n_{q1}} e^{j(n_{11}+n_{21}+\cdots+n_{q1}) \Delta t} \right\rangle$$

$$\sum_{n_{12}} \sum_{n_{22}} \cdots \sum_{n_{q2}} a_{n_{12}} a_{n_{22}} \cdots a_{n_{q2}} e^{j(n_{12}+n_{22}+\cdots+n_{q2}) \Delta (t+\tau)} \right\rangle.$$  \hspace{1cm} (3-3)
The indicated time average is performed by causing the $n_{11}$'s to cancel the $n_{j2}$'s thus removing the dependence on $t$. There are $q!$ ways of pairing the $n_{11}$ indices with the $n_{j2}$ indices therefore

$$R_{f(x,q)}(\tau) = q! \sum_{n_1} \sum_{n_2} \cdots \sum_{n_q} |a_{n_1}|^2 |a_{n_2}|^2 \cdots |a_{n_q}|^2 e^{j(n_1+n_2+\cdots+n_q)\Delta \tau}.$$  

If the $(q-1)!$ restrictions on the indices are removed such that $n_1$ can possibly equal $-n_2$, etc., then the autocorrelation function of $f(x(t), q)$ can be written as

$$R_{f(x,q)}(\tau) = q! \left[ \sum_{n_1} |a_{n_1}|^2 e^{j n_1 \Delta \tau} \right] \left[ \sum_{n_2} |a_{n_2}|^2 e^{j n_2 \Delta \tau} \right] \cdots$$

$$= q! \left[ R_x(\tau) \right]^q$$

since

$$R_x(\tau) = \langle x(t) x(t+\tau) \rangle = \left\langle \sum_{n_2} \sum_{n_1} a_{n_2}^* a_{n_1} e^{j(n_2+n_1)\Delta t} e^{j n_1 \Delta \tau} \right\rangle$$

$$= \sum_{n_1=-\infty}^{\infty} |a_{n_1}|^2 e^{j n_1 \Delta t}.$$
It will be shown that the error introduced into Eq. (3-5) by removing the \((q-1)\) restrictions on the indices, that is by allowing the possibility that \(n_1 = -n_2\), \(n_1 = -n_3\), etc., approaches zero as \(\Delta\) approaches zero. To demonstrate this fact the restriction on only \(n_1\) and \(n_2\), in Eq. (3-4), will be removed. The contribution of the terms in the case for which \(n_1 = -n_2\) is

\[
(3-7) \left[ \sum_{n_1=-\infty}^{\infty} \left| a_{n_1}\right|^2 \left| n_{n_1}\right|^2 \right] q! \sum_{n_3} \left| a_{n_3}\right|^2 \cdots \sum_{n_q} \left| a_{n_q}\right|^2 e^{j(n_3+\cdots+n_q)\Delta^2}.
\]

From Eq. (2-23) the quantity in square brackets can be seen to be zero in the limit as \(\Delta\) approaches zero. Since the remaining factor in Eq. (3-7) is finite or zero then the contribution of the terms for which \(n_1 = -n_2\) in Eq. (3-5) is zero. Therefore the contribution to Eq. (3-5) due to terms resulting from the removal of any combination of the \((q-1)\) restrictions can be shown to be zero for \(q\) finite.

Derivation of Orthogonal Functions

At this point sufficient information is available to proceed with the derivation of the form of the orthogonal functions, \(f(x, q)\).

Recalling that the sum of the terms of \(x^q(t)\) at frequencies which are members of set \(S_q\) are defined as \(f(x, q)\) and that there are other terms of \(x^q\) at frequencies which are members of sets \(S_{q-2}\), \(S_{q-4}\), \(\cdots\), then \(f(x, q)\) can be determined by subtracting from \(x^q\) the terms at frequencies which are members of sets other than set \(S_q\).
That is

$$f(x, q) = x^q - \sum_{m} \alpha_{qm} f(x, m)$$

where $m < q$ and $q - m$ is an even integer.

Consider the terms of $x^q$ at frequencies which are members of set $S_p$ where $(q-p)$ is a positive even integer. These terms are of the form

$$a_{n_1} a_{n_2} \cdots a_{n_p} e^{i(n_1 + n_2 + \cdots + n_p) \Delta t}$$

Since $(q-p)$ is an even integer, the average value of the term in brackets is equivalent to the average value of $x^{q-p}$ which is

$$(q-p-1)(q-p-3) \cdots 3 \cdot 1 \sigma_x^{q-p}.$$ 

The amplitude of each of these frequencies is given by the number of ways of choosing a given $(n_1 + n_2 + \cdots + n_p)$ combination from the $q$ indices, multiplied by

$$< x^{q-p} > a_{n_1} a_{n_2} \cdots a_{n_p}.$$ 

Thus the total amplitude of the terms of $x^q$ at each frequency which is a member of set $S_p$ is

$$a_{n_1} a_{n_2} \cdots a_{n_p} \frac{q!}{(q-p)!} (q-p-1)(q-p-3) \cdots 3 \cdot 1 \sigma_x^{q-p}.$$
The function $f(x, p)$ consists of terms which are identical to these terms with the exception that the terms of $f(x, p)$ are smaller by a constant factor. Since there are $p!$ terms of $f(x, p)$ at each frequency, the total amplitude of each frequency term is $p!$ multiplied by the amplitude of each term or

$$(3-11) \quad a_{n_1} a_{n_2} \cdots a_{n_p} p!.$$

The ratio of the expression of Eq. (3-10) to that of Eq. (3-11) is

$$(3-12) \quad \frac{q!}{(q-p)!} \frac{1}{p!} (q-p-1) (q-p-2) \cdots 3 \cdot 1 \sigma_x^{q-p}.$$

Therefore all terms of $x^q$ at frequencies which are members of set $S_p$ can be removed from $x^q$ by subtracting the product of Eq. (3-12) and $f(x, p)$ from $x^q$. Since

$$(3-13) \quad f(x, q) = x^q - \sum_m \alpha_{q-m} f(x, m)$$

it is clear that

$$(3-14) \quad \alpha_{qm} = \frac{q!}{(q-m)!m!} (q-m-1)(q-m-2) \cdots 3 \cdot 1 \sigma_x^{q-m}.$$

The functions, $f(x, q)$ can now be constructed. Since $x^0$ and $x^1$ are equal to 1 and $x$ respectively, then
\[
\begin{align*}
    f(x,0) &= 1 = 1 \\
    f(x,1) &= x = x \\
    f(x,2) &= x^2 - \alpha_{20} f(x,0) = x^2 - \sigma_x^2 \\
    f(x,3) &= x^3 - \alpha_{31} f(x,1) = x^3 - 3 \sigma_x^2 x \\
    f(x,4) &= x^4 - \alpha_{42} f(x,2) - \alpha_{40} f(x,0) = x^4 - 6 \sigma_x^2 x^2 + 3 \sigma_x^4 \\
    f(x,5) &= x^5 - \alpha_{53} f(x,3) - \alpha_{51} f(x,1) = x^5 - 10 \sigma_x^2 x^3 + 15 \sigma_x^4 x \\
    \text{etc.}
\end{align*}
\]

It has been shown previously in Eq. (3-1) that

\[ (3-15) \quad < f(x,p) f(x,q) > = 0 \quad \text{for} \quad p \neq q. \]

The value of \(< f^2(x, q) >\) is simply determined from Eq. (3-5).

Since by definition

\[ R_x(0) = < x^2(t) > = \sigma_x^2 \]

then

\[ (3-16) \quad < f^2(x, q) > = R_f(x, q)f(0) = q! \left[ R_x(0) \right]^q = q! \sigma_x^{2q}. \]

Thus if desired, the orthogonal functions, \( f(x, q) \), could be normalized by dividing each by

\[ (3-17) \quad \sqrt{q!} \sigma_x^{-q}. \]

The functions \( f(x, q) \) can be recognized as having the same form as the Hermite polynomials, \( H_q(x) \). The defining equation for the Hermite polynomials \(^7\) which are modified to be equivalent to the function \( f(x, q) \) is
A recursion relation is helpful for the determination of the polynomials since the successive differentiations become unwieldy. This recursion relation is given by

\[(3-19) \quad H_q(x) = x \frac{d}{dx} H_{q-1}(x) - (q-1) \sigma^2 H_{q-2}(x).\]
The Hermite polynomials are defined in most texts in such a way that the normal distribution function with which they are associated has zero mean and unity variance. This definition is

\begin{equation}
H_n(y) = (-1)^m \left[ \frac{d^m}{dy^m} e^{-y^2/2} \right] e^{y^2/2}
\end{equation}

where the prime on the \( H \) merely indicates the unity variance definition. The orthogonality relation corresponding to this definition is

\begin{equation}
\frac{1}{2\pi} \int_{-\infty}^{\infty} H_m(y) H_p(y) e^{-y^2/2} dy = \begin{cases} 
\frac{m!}{\sigma^m} & \text{if } m = p \\
0 & \text{if } m \neq p
\end{cases}
\end{equation}

By making the change of variable

\begin{equation}
y = x/\sigma
\end{equation}

where \( \sigma^2 \) is the variance of the \( x \) variable, and substituting for \( y \) in these equations, the definition of the Hermite polynomials which matches the definition of \( f(x, m) \) given in Eq. (3-18) can be obtained.

Since

\begin{equation}
\frac{d^m}{dy^m} = \sigma^{-m} \frac{d^m}{dx^m}
\end{equation}
then Eq. (4-1) becomes

\begin{equation}
(4-5) \quad H_m'(\frac{x}{\sigma}) = (-1)^m \sigma^{-m} \left[ \frac{d^m}{dx^m} e^{-x^2/2 \sigma^2} \right] e^{x^2/2 \sigma^2}.
\end{equation}

Similarly Eq. (4-2) can be written as

\begin{equation}
(4-6) \quad \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} H_m'(\frac{x}{\sigma}) H_p'(\frac{x}{\sigma}) e^{-x^2/2 \sigma^2} dx = \begin{cases} m! & m = p \\ 0 & m \neq p \end{cases}.
\end{equation}

Since the highest order term of $H_m'(y)$ is $y^m$ then the corresponding term in $H_m'(\frac{y}{\sigma})$ is $x^m/\sigma^m$. Therefore, the relationship between $f(x, m)$ and $H_m'(x)$ is

\begin{equation}
(4-7) \quad f(x, m) = H_m(x) = \sigma^{-m} H_m'(\frac{x}{\sigma})
\end{equation}

\begin{align*}
&= (-1)^m \sigma^{2m} \left[ \frac{d^m}{dx^m} e^{-x^2/2 \sigma^2} \right] e^{x^2/2 \sigma^2}.
\end{align*}

Likewise, the orthogonality relation becomes

\begin{equation}
(4-8) \quad \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} H_m(x) H_p(x) e^{-x^2/2 \sigma^2} dx = \begin{cases} \sigma^{2m} m! & m = p \\ 0 & m \neq p \end{cases}.
\end{equation}

Since $x(t)$ has been assumed to be ergodic so that time averages and statistical averages can be interchanged then it can be seen that this equation is identical to Eq. (3-1) and (3-5).

An arbitrary function, $g(x)$, can be approximated by a linear combination of Hermite polynomials defined as in Eq. (4-7),
that is

\begin{equation}
(4-9) \quad g(x) \equiv \sum_{m=0}^{M} A_m H_m(x) .
\end{equation}

The coefficients, $A_m$, are derived by multiplying the equation for $g(x)$ by $H_p(x)$ and taking the statistical mean of both sides as follows

\begin{equation}
(4-10) \quad \frac{g(x) H_p(x)}{H_p(x)} = \sum_{m=0}^{M} A_m \frac{H_m(x) H_p(x)}{H_p(x)} .
\end{equation}

Since the set of Hermite polynomials is complete, $g(x)$ cannot be orthogonal to $H_p(x)$ for every value of $p$. Therefore the member on the left side of the equation is not identically equal to zero. Using the orthogonality relation this equation becomes

\begin{equation}
(4-11) \quad g(x) H_p(x) = A_p \rho! \sigma^{2p} .
\end{equation}

Consequently

\begin{equation}
(4-12) \quad A_p = \frac{1}{\rho! \sigma^{2p}} \frac{g(x) H_p(x)}{H_p(x)} .
\end{equation}

By changing the index to $m$ and writing the statistical mean in its integral form the coefficient $A_m$ becomes

\begin{equation}
(4-13) \quad A_m = \frac{1}{\sqrt{2 \pi \sigma^{2m+1}}} \int_{-\infty}^{\infty} g(x) H_m(x) e^{-x^2/2 \sigma^2} dx .
\end{equation}
It will be shown that the above definition of the coefficients of \( H_m(x) \) gives the best mean square approximation of \( g(x) \) in Eq. (4-9). Let the function \( g(x) \) be reexpressed as

\[
(4-14) \quad g(x) \approx \sum_{m=0}^{M} B_m H_m(x)
\]

where the \( B_m \)'s are arbitrary constants. It will be shown that for the mean square value of

\[
(4-15) \quad \left[ g(x) - \sum_{m=0}^{M} B_m H_m(x) \right]
\]

to be minimum, the constants, \( B_m \), must be chosen such that

\[
(4-16) \quad B_m = A_m
\]

where \( A_m \) is given by Eq. (4-13). The mean square value of Eq. (4-15) is always greater than or equal to zero and is given by

\[
(4-17) \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left[ g(x) - \sum_{m=0}^{M} B_m H_m(x) \right]^2 e^{-x^2/2} \, dx \geq 0
\]

After performing the indicated squaring operation and interchanging the order of the sum and integration operations this relation becomes
It can be seen with the aid of Eq. (4-8) and (4-13) that

\[(4-19)\] \[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g^2(x) e^{-x^2/2\sigma^2} dx - 2 \sum_{m=0}^{M} \sigma^{-2m} m! \Lambda_m \hat{\Lambda}_m + \sum_{m=0}^{M} \sigma^{-2m} m! B_m^2 \geq 0.\]

By completing the square the final result is obtained, which is

\[(4-20)\] \[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g^2(x) e^{-x^2/2\sigma^2} dx - \sum_{m=0}^{M} \sigma^{-2m} m! A_m^2 + \sum_{m=0}^{M} \sigma^{-2m} m! [A_m - B_m]^2 \geq 0.\]

It is clear that this function has its smallest value when

\[(4-21)\] \[A_m = B_m.\]

After making this substitution another useful result is obtained. Since

\[(4-22)\] \[\frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} g^2(x) e^{-x^2/2\sigma^2} dx \geq \sum_{m=0}^{M} \sigma^{-2m} m! A_m^2.\]
then the series on the right side of this relation must be convergent
since the left side is independent of $M$. Furthermore since the set
of Hermite polynomials is complete this relation by definition becomes
an equality as $M$ approaches infinity, that is

$$\frac{1}{\sqrt{2\pi}} \sigma^{-2} \int_{-\infty}^{\infty} g(x)^2 e^{-x^2/2\sigma^2} \, dx = \sum_{m=0}^{\infty} \sigma^{-2m} m! A_m^2,$$

which is equivalent to the statement that the mean square error of

$$[g(x) - \sum_{m=0}^{\infty} A_m H_m(x)]$$

is zero. Thus the series for $g(x)$ is said to converge in the mean to
$g(x)$.

Convergence at a Point

The manner in which the Hermite expansion of a function
converges at a point will next be investigated. It will be shown that the
range of $x$ over which this type of convergence is possible is limited
to values of $x$ such that

$$|x| \leq L < \infty.$$

This can be justified in the following manner. Since the series expansion
for $g(x)$ converges in the mean to $g(x)$ then

$$\frac{1}{\sqrt{2\pi}} \sigma^{-2} \int_{-\infty}^{\infty} [g(x) - \sum_{m=0}^{\infty} A_m H_m(x)]^2 e^{-x^2/2\sigma^2} \, dx = 0.$$
The function \( g(x) \) will be assumed to be bounded, continuous and have a finite slope in the vicinity of the point \( x = x_i \). Therefore the integrand of Eq. (4-24) is also continuous and of finite slope in the vicinity of \( x_i \). If the integrand of Eq. (4-24) is greater than zero at \( x_i \), then it must be greater than zero in some finite interval containing \( x_i \). However since the integral of Eq. (4-24) is zero then

\[
(4-25) \quad \frac{1}{\sqrt{2\pi} \sigma} \left[ g(x_i) - \sum_{m=0}^{M} A_m H_m(x_i) \right]^2 e^{-x_i^2 / 2 \sigma^2}
\]

must be zero. If the range of \( x \) is restricted such that

\[
|x| \leq L < \infty
\]

then

\[
e^{-x_i^2 / 2 \sigma^2}
\]

cannot be zero, therefore

\[
(4-26) \quad g(x_i) = \sum_{m=1}^{\infty} A_m H_m(x_i)
\]

The convergence of the Hermite expansion of \( g(x) \) at a point where \( g(x) \) is discontinuous is also such that the mean square error is a minimum at the point. In view of this it can be shown that the expansion converges to the arithmetic mean of the two values of \( g(x) \) immediately adjacent to the point of discontinuity. The argument which will be used to demonstrate this phenomenon is similar to that which
Guillemin uses in the case of the convergence of the Fourier series at a discontinuity. Let the two values of the function $g(x)$ immediately adjacent to the point of discontinuity, $x_0$, be denoted by $g(x_0^+)$ and $g(x_0^-)$ and let the value of the series at $x_0$ be denoted by $s(x_0)$. Then the mean square error at $x_0$ is expressed by

\[
(4-27) \quad \frac{[s(x_0) - g(x_0^+)]^2 + [s(x_0) - g(x_0^-)]^2}{2} \exp^{-\frac{(x_0-2\sigma)^2}{2}}.
\]

If this is to be a minimum, its derivative with respect to $s$ must be zero, that is,

\[
(4-28) \quad s(x_0) - g(x_0^+) + s(x_0) - g(x_0^-) = 0
\]

which yields

\[
(4-29) \quad s(x_0) = \frac{g(x_0^+) + g(x_0^-)}{2}.
\]

These results will be applied in the next chapter.
Preliminary Discussion

The signal quantizer which will be discussed in this chapter can be represented by a two terminal network as is shown in Fig. 1. Even though the quantizer's input and therefore its output are functions of time, its transfer characteristic, a typical example of which is shown in Fig. 2, is independent of time, that is, the output signal, \( y \), is a function of only the magnitude of the input signal, \( x \). Therefore, since the following discussion will be concerned with the quantizer's output with respect to its input, all time dependence indications on \( x \) and \( y \) will be omitted in the remainder of the chapter.

The difference between the quantizer output and input signals will be called quantizing error and will be denoted by

\[
\epsilon = y(x) - x.
\]

A typical example of this error is shown as a function of \( x \) in Fig. 3.

---

**Fig. 1.** --Two terminal network representation of a signal quantizer.
Fig. 2.--Typical quantizer input-output characteristics.

Fig. 3.--Typical quantizing error function.
Quantizing Error Minimization

The development of a method for the determination of quantizer level settings for a general M-level quantizer such that the mean square quantizing error is minimized will be given in this section. This development is based upon the assumption that the quantizer input signal is stationary, has zero mean and a normal probability density function. In addition it has been assumed that the quantizer has as many quantizing levels for positive values of $x$ as for negative values. Consequently the optimum level settings which will result from the method will yield a symmetrical quantizer; however the assymetrical quantizer could also be optimized by using the method which will be described.

References will be made to $a$ - parameters and $b$ - parameters throughout the following discussion. The $a$ - parameters are defined as the magnitudes of the quantizer input signal at which the output assumes a new value and the $b$ - parameters are defined as the magnitude of each output jump. For example, referring to Fig. 2, in the range of $x$ from $a_1$ to $a_2$ the quantizer output is $b_1$. Assuming that the value of $x$ moves in a positive direction, at the instant when $x$ becomes larger than $a_2$ the output jumps by an amount $b_2$ to make the total output $b_1 + b_2$.

The final equations which yield the optimum quantizer parameters will be obtained by two techniques. The first technique is essentially the same as that used by Max$^{10}$ and will be referred to as the direct
method. It involves the statistical mean of $\mathcal{E}^2$, where $\mathcal{E}$ is the quantizing error, an example of which is shown in Fig. 3. The equations from which the optimum $a$ and $b$ parameter values can be determined are obtained by setting the partial derivatives of the mean square quantizing error with respect to each of the parameters equal to zero.

The second technique, which will be referred to as the Hermite expansion method, yields the same results as the direct method. The Hermite expansion method is convenient in that it suggests a closed form for the unwieldy equations obtained by the direct method. The second technique involves the expansion of the quantizing error function in terms of Hermite polynomials. The equations for determining quantizer parameters are obtained by minimizing this expansion with respect to variations of the $a$ and $b$ parameters.

The Direct Method

The mean square value of the quantizing error, Eq. (5-1), is given by the expression

\begin{equation}
\mathcal{E}^2 = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} (y-x)^2 e^{-x^2/2\sigma^2} \, dx.
\end{equation}

For the case of the symmetrical quantizer the above equation can be simplified to the following expression
Substitution of the equations for individual line segments of $y-x$, indicated in Fig. 3 into the above expression gives

\[
\frac{\varepsilon^2}{\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} (y-x)^2 e^{-x^2/2\sigma^2} \, dx + \ldots
\]

The general form of the partial derivatives of $\varepsilon^2$ with respect to the $a_m$'s can be obtained by performing the differentiation with respect to an arbitrary value of $a_m$. Thus let

\[m = p\]

and

\[
\frac{\partial \varepsilon^2}{\partial a_p} = \frac{2}{\sqrt{2\pi}} \left\{ \left[ \left( \sum_{m=1}^{p} b_m \right) - a_p \right]^2 - \left[ \left( \sum_{m=1}^{p-1} b_m \right) - a_p \right]^2 \right\} e^{-a_p^2/2\sigma^2} = 0
\]
The terms which remain after the squaring and subtraction operations are performed are

\[ b_p \left( b_p + 2 \sum_{m=1}^{p-1} b_m - 2 \sigma_p \right) \]

however since \( b_p \) is not zero then

\[ b_p = 2 \left[ \sigma_p - \sum_{m=1}^{p-1} b_m \right] \quad (5-6) \]

For the case where \( p = 1 \)

the summation term does not exist, therefore it appears that

\[ b_1 = 2 \sigma_1 + 0 \]

This relation can be verified by simply observing

\[ \frac{\partial^2 \mathcal{E}}{\partial \sigma_i^2} = \frac{2}{\sqrt{2\pi} \sigma} \left[ a_i^2 - (b_i - a_i)^2 \right] e^{-\frac{a_i^2}{2\sigma^2}} = 0 \quad (5-7) \]

Since \( b_1 \) is not zero then

\[ b_1 = 2 \sigma_1 \quad (5-8) \]

Fig. 4 has been drawn in such a way that Eq. (5-6) is satisfied which can be seen more clearly if the equation is rewritten as

\[ \frac{b_p}{2} = \sigma_p - \sum_{m=1}^{p-1} b_m \quad (5-9) \]
Fig. 4.--Quantizing error with restriction imposed.
where

\[ p = 1, 2, 3, \ldots, M. \]

It is apparent from Fig. 4 that the effect of the requirements on the quantizing error set by this equation has been to cause equal absolute magnitudes of the error function, \( \mathcal{E}(x) \), at the points immediately to the left and to the right of each \( a_p \). Thus it is clear that the extremum obtained when each partial derivative is set to zero is a minimum.

The \( M \) equations obtained in the preceding discussion are not sufficient to solve for the \( 2M \) unknowns. The remaining equations are obtained by setting the partial derivatives of the mean square quantizing error, \( \mathcal{E}^2 \), with respect to the \( b_m \)'s equal to zero. The equations thus obtained are at first glance rather unwieldy; however the Hermite expansion method will suggest a closed form for these equations. Before proceeding with the second method, the partial derivative of \( \mathcal{E}^2 \) with respect to one of the \( b_m \)'s, namely \( b_p \), is determined as follows:

\[
(5-10) \quad \frac{\partial \mathcal{E}^2}{\partial b_p} = \frac{1}{\sqrt{2\pi} \sigma} \left\{ 2 \int_{a_p}^{a_{p+1}} \left[ \sum_{m=1}^{p} b_m \right] e^{-x^2/2\sigma^2} \, dx \right. \\
+ 2 \int_{a_{p+1}}^{a_{p+2}} \left[ \sum_{m=1}^{p+1} b_m \right] e^{-x^2/2\sigma^2} \, dx + \ldots \\
+ 2 \int_{a_M}^{\infty} \left[ \sum_{m=1}^{M} b_m \right] e^{-x^2/2\sigma^2} \, dx \right\} = 0.
\]
This can be reduced to the form

\[
(5-11) \quad \sum_{m=1}^{p} b_m \int_{a_p}^{a_{p+1}} e^{-x^2/2\sigma^2} \, dx + \sum_{m=1}^{p+1} b_m \int_{a_{p+1}}^{a_{p+2}} e^{-x^2/2\sigma^2} \, dx + \ldots
\]

\[
+ \sum_{m=1}^{M-1} b_m \int_{a_{M-1}}^{a_M} e^{-x^2/2\sigma^2} \, dx + \sum_{m=1}^{M} b_m \int_{a_M}^{\infty} e^{-x^2/2\sigma^2} \, dx = \int_{a_p}^{\infty} e^{-x^2/2\sigma^2} \, dx = \sigma^{-2} \exp{-a_p^2/2\sigma^2}.
\]

This, then, is the general form of the second half of the \(2M\) equations required to solve for the optimum quantizer levels. The discussion of these equations will be continued in the section on the Hermite expansion method.

The Hermite Expansion Method

The exact value of the quantizing error function at points of discontinuity has not yet been defined. Therefore, let the value of \(E(x)\) at each point of discontinuity be equal to the arithmetic mean of the values of \(E(x)\) immediately to the left and to the right of the point of discontinuity. This, as was explained in Chapter IV, insures that the Hermite expansion of \(E(x)\) at every point of the range of \(x\) such that \(|x| \leq L < \infty\).

It is obvious that the value of \(L\) must be larger than the largest value of \(a_m\) which is \(a_M\). This introduces no problem, however, since for a practical quantizer the largest value of \(a_m\) is seldom larger than
three times the standard deviation of the $x$ signal, $\sigma$.

As a preliminary step, in the expansion of the quantizing error function the function $y_m(x)$ which is shown in Fig. 5 will be expanded.

![Diagram](image)

Fig. 5. --Component function of quantizer output function.

It can be seen that the quantizer output function and subsequently the quantizing error function can be synthesized by adding $M$ of those functions with appropriate values of $a_m$ and $b_m$. The defining equations for $y_m(x)$ are

\begin{align}
(5-12) \quad y_m(x) &= -b_m \quad ; x < -a_m \\
&= -b_m/2 \quad ; x = -a_m \\
&= 0 \quad ; -a_m < x < a_m \\
&= b_m/2 \quad ; x = a_m \\
&= b_m \quad ; x > a_m .
\end{align}

The form of the coefficients for the Hermite expansion

\begin{align}
(5-13) \quad y_m(x) &= \sum_{p=0}^{\infty} c_{mp} H_p(x) ; \quad |x| \leq L < \infty
\end{align}
is given by Eq. (4-13). Thus the expression for $C_{mp}$ is

$$
(5-14) \quad C_{mp} = \frac{1}{\sqrt{2\pi} \sigma^{2p+1} p!} \int_{-\infty}^{\infty} y_m(x) H_p(x) e^{-x^2/2\sigma^2} dx
$$

where $\sigma^2$ is the variance of the variable $x$. By taking advantage of the odd symmetry of $y_m(x)$ the expression for $C_{mp}$ may be written as

$$
(5-15) \quad C_{mp} = \frac{2^{b_m}}{\sqrt{2\pi} \sigma^{2p+1} p!} \int_{a_m}^{\infty} H_p(x) e^{-x^2/2\sigma^2} dx.
$$

This integral can be simply evaluated with the aid of the defining expression for the Hermite polynomials. Thus since

$$
(5-16) \quad H_p(x) e^{-x^2/2\sigma^2} = \sigma^{-2p} (-1)^p \frac{\partial^p}{\partial x^p} (e^{-x^2/2\sigma^2})
$$

then

$$
(5-17) \quad \int_{a_m}^{\infty} H_p(x) e^{-x^2/2\sigma^2} dx = \sigma^{-2} H_{p-1}(a_m) e^{-a_m^2/2\sigma^2}.
$$

Finally

$$
(5-18) \quad C_{mp} = \frac{2^{b_m}}{\sqrt{2\pi} \sigma^{2p-1} p!} H_{p-1}(a_m) e^{-a_m^2/2\sigma^2}.
$$

The expansion for $y(x)$ is given by...
(5-19) \[ y(x) = \sum_{m=1}^{M} \sum_{p=-1}^{\infty} c_{mp} H_p(x) = \sum_{m=1}^{\infty} \sum_{p=-1}^{\infty} c_{mp} H_p(x) \quad \text{for} \quad |x| \leq L < \infty, \]

Only odd values of \( p \) are involved because of the symmetry of \( y(x) \).

Therefore,

(5-20) \[ y(x) = \sum_{p=1}^{\infty} D_p H_p(x) \quad |x| \leq L < \infty \]

where

(5-21) \[ D_p = \frac{2}{\sqrt{2\pi} \sigma^2 p!} \sum_{m=1}^{M} b_m H_{p-1}(a_m) e^{-q_m^2/2\sigma^2}. \]

In the remainder of the discussion the indication that the index \( p \) assumes only odd values will be dropped.

Finally the expansion for the quantizing error is

(5-22) \[ E(x) = y(x) - x = y(x) - H_l(x) = \sum_{p=1}^{\infty} D_p H_p(x) - H_l(x) \quad \text{for} \quad |x| \leq L < \infty. \]

As a consequence of Eq. (5-22), the mean square quantizing error can be written as

(5-23) \[ \overline{E^2} = \left[ \left( \sum_{p=1}^{\infty} D_p H_p(x) \right) - H_l(x) \right] \left[ \left( \sum_{q=1}^{\infty} D_q H_q(x) \right) - H_l(x) \right], \]

which after carrying out the indicated multiplication of terms becomes

(5-24) \[ \sum_{p=1}^{\infty} \sum_{q=1}^{\infty} D_p D_q H_p(x) H_q(x) - 2 \sum_{p=1}^{\infty} D_p H_p(x) H_l(x) + H_l(x)^2. \]
The orthogonal property of the Hermite polynomials permits a considerable simplification of Eq. (5-24), thus

\[ \sum_{p=1}^{\infty} \sigma^{-2p} p! \left[ \frac{2}{2\pi} \right] \left[ \sum_{m=1}^{M} b_m H_p(a_m) e^{-q_m^2/2 \sigma^2} \right]^2 \]

which after substitution of the expression for \( D_p \) becomes

\[ \sum_{p=1}^{\infty} \sigma^{-2p} p! \left[ \frac{2}{2\pi} \sigma^{-2p-1} p! \right] \left[ \sum_{m=1}^{M} b_m H_p(a_m) e^{-q_m^2/2 \sigma^2} \right]^2 \]

By taking the partial derivative of \( \overline{\mathcal{E}^2} \) with respect to an arbitrary \( a_k \) and setting the result equal to zero, an extremum with respect to \( a_k \) is obtained; thus

\[ \frac{\partial \overline{\mathcal{E}^2}}{\partial a_k} = 2 \sum_{p=1}^{\infty} \sigma^{-2p} p! \left[ \frac{2}{2\pi} \sigma^{-2p-1} p! \right] \left[ \sum_{m=1}^{M} b_m H_p(a_m) e^{-q_m^2/2 \sigma^2} \right] \]

Recalling that

\[ H_p(x) e^{-x^2/2 \sigma^2} = \sigma^{-2p} (-1)^p \frac{\partial^p}{\partial x^p} \left( e^{-x^2/2 \sigma^2} \right) \]
then it can be seen that

\begin{equation}
(5-29) \quad \frac{d}{dx} (H_p(x) e^{-x^2/2\sigma^2}) = -\frac{1}{\sigma^2} H_{p+1}(x) e^{-x^2/2\sigma^2}.
\end{equation}

Thus Eq. (5-27) can be written as

\begin{equation}
(5-30) \quad 2 \sum_{p=1}^{\infty} \sigma^2 p! D_p \left( \frac{-2b_k}{\sqrt{2\pi}} \sigma^{-2p+1} H_p(q_k) e^{-q_k^2/2\sigma^2} \right)
- 2 \sigma^2 \left( \frac{-2b_k}{\sqrt{2\pi}} \sigma^{-3} H_l(q_k) e^{-q_k^2/2\sigma^2} \right) = 0
\end{equation}

which after simplification becomes

\begin{equation}
(5-31) \quad b_k \sum_{p=1}^{\infty} D_p H_p(q_k) - b_k H_l(q_k) = 0.
\end{equation}

Since \( b_k \) is not equal to zero then

\begin{equation}
(5-32) \quad \sum_{p=1}^{\infty} D_p H_p(q_k) = H_l(q_k) = q_k
\end{equation}

but

\begin{equation}
(5-33) \quad y(x) = \sum_{p=1}^{\infty} D_p H_p(x) \quad \text{for} \quad |x| \leq L < \infty
\end{equation}

therefore

\[ y(a_k) = a_k \quad \text{for} \quad 1 \leq k \leq M. \]

Since the point \( a_k \) is a point of discontinuity of \( y(x) \), the value to which the Hermite expansion converged at this point is actually the arithmetic mean of the values of \( y(x) \) on either side of \( a_k \) as was
shown in Eq. (4-29). In other words, the midpoints of discontinuities of the function $y(x)$ must lie along the line $y(x) = x$

as is shown in Fig. 6.

![Diagram showing discontinuity midpoints along 45° line.](image)

Fig. 6.--Discontinuity midpoints lie along 45° line.

Thus it can be seen that

\[
y(a_1) = b/2 = a_1
\]

\[
y(a_2) = b_1 + b_2/2 = a_2
\]

\[
\vdots
\]

(5-35) \[
y(a_k) = \sum_{m=1}^{k-1} b_m + \frac{b_k}{2} = a_k
\]
Returning to the comparable general expression as obtained by the direct method, Eq. (5-36), and solving for \( a_p \), the relation

\[
a_p = \frac{b_p}{2} + \sum_{m=1}^{p-1} b_m
\]

is obtained which can be seen to be identical to Eq. (5-35).

Next by taking the partial derivative of \( \overline{E^2} \) with respect to an arbitrary \( b_k \) and setting the result equal to zero, an extremum with respect to \( b_k \) is obtained, thus

\[
\frac{\partial \overline{E^2}}{\partial b_k} = 2 \sum_{p=1}^{\infty} \sigma^{2p} \frac{2}{\sqrt{2\pi} \sigma^{2p-1} p!} \sum_{m=1}^{M} b_m H_{p-1}(q_m) e^{-q_m/2\sigma^2} \cdot \frac{2}{\sqrt{2\pi} \sigma^{2p-1} p!} \frac{2}{\sigma} H_{p-1}(q_k) e^{-q_k/2\sigma^2} - 2 \sigma^2 \left[ \frac{2}{\sqrt{2\pi} \sigma} H_0(q_k) e^{-q_k/2\sigma^2} \right] = 0
\]

which can be simplified to

\[
\sum_{p=1}^{\infty} D_p H_{p-1}(q_k) = H_0(q_k) = 1
\]

for \( 1 \leq k \leq M \). This rather interesting expression will be seen to be useful in the evaluation of the integral of

\[
y(x) e^{-x^2/2\sigma^2}
\]
Since

\[ y(x) = \sum_{p=1}^{\infty} D_p H_p(x) \quad ; \quad |x| \leq L < \infty , \]

then

\[ \int_{a_k}^{L} y(x)e^{-x^2/2\sigma^2} \, dx = \int_{a_k}^{L} \sum_{p=1}^{\infty} D_p H_p(x) e^{-x^2/2\sigma^2} \, dx . \]

The defining equation for the Hermite polynomials can once again be used to evaluate this integral after the order of the sum and integration are interchanged. Since

\[ H_p(x) e^{-x^2/2\sigma^2} = \sigma^2 P_{(-)}^{(-1)} \frac{\partial P}{\partial x} \left( e^{-x^2/2\sigma^2} \right) \]

then

\[ \int_{a_k}^{L} H_p(x)e^{-x^2/2\sigma^2} \, dx = \sigma^2 H_{p-1}(a_k) e^{-a_k^2/2\sigma^2} - \sigma^2 H_{p-1}(L) e^{-L^2/2\sigma^2} . \]

But since the last term of this expression approaches zero as \( L \) gets large, then Eq. (5-39) can be written

\[ \int_{a_k}^{L} y(x)e^{-x^2/2\sigma^2} \, dx = \sum_{p=1}^{\infty} \sigma^2 D_p H_{p-1}(a_k) e^{-a_k^2/2\sigma^2} . \]
The final result is obtained by substituting Eq. (5-37) into this expression which yields

\[(5-42) \int_{a_k}^{L} y(x) e^{-x^2/2\sigma^2} \, dx = \sigma^2 e^{-a_k^2/2\sigma^2} \]

The error introduced by substituting infinity for \( L \) in the upper limit of the integral approaches zero as \( L \) is made large, for example

\[\int_{L}^{\infty} e^{-x^2/2\sigma^2} \, dx < 0.0001\]

for \( L = 4 \).

Thus

\[(5-43) \int_{a_k}^{\infty} y(x) e^{-x^2/2\sigma^2} \, dx = \sigma^2 e^{-a_k^2/2\sigma^2} \; ; \quad x \leq L < \infty \]

\[1 \leq k \leq M\]

This equation can be shown to be identical to the corresponding equation, Eq. (5-11), obtained by the direct method.

In summary, the \( 2M \) equations necessary for the evaluation of the \( 2M \) coefficients, \( a_m \) and \( b_m \), for the symmetrical quantizer are

\[(5-44) \int_{a_k}^{\infty} y(x) e^{-x^2/2\sigma^2} \, dx = \sigma^2 e^{-a_k^2/2\sigma^2}\]

\[(5-45) b_k = 2 \left[ a_k - \sum_{m=1}^{k-1} b_m \right]\]
for \( 1 \leq k \leq M \),

where \( \sigma^2 \) is the variance of the \( x \) variable and where \( y(x) \) is

of the form shown in Fig. 2.

These simultaneous equations are solvable by using the digital computer. The computer is programmed to adjust the values of the \( a_k \)'s, beginning with the initial values which are included as input data, until Eq. (5-44) is satisfied to some specified degree of accuracy for all values of \( k \) such that

\[ 1 \leq k \leq M \]

The values of \( b_k \) which are required in this computation are of course obtained from Eq. (5-45). For the program used, the only restriction on the initial values of the \( a_k \)'s is that they be chosen in such a way that the corresponding values of \( b_k \) are all greater than zero.

That the Eq. (5-44) and (5-45) represents solutions for minimum mean square quantizing error has been verified experimentally by calculating and observing larger mean square quantizing errors for values of the \( a \) and \( b \) coefficients which are different from the values given by the above equations.

The values of the \( a \) and \( b \) parameters which were obtained by the digital computer for two different symmetrical quantizers are shown in Tables 1 and 2. An indication of the degree of accuracy to which the final values of \( a_k \) and \( b_k \) satisfy Eq. (5-44) can be obtained by using the expression
The lower limit of $\eta$ is a function of only the amount of computer time allotted to the job.

The mean square quantizing error, as calculated from Eq. (5-25) and (4-23), is given in each case.

**TABLE 1**

**OPTIMUM PARAMETERS FOR A FOUR LEVEL SYMMETRICAL QUANTIZER**

<table>
<thead>
<tr>
<th>$k$</th>
<th>$a_k$</th>
<th>$b_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.220</td>
<td>.4400</td>
</tr>
<tr>
<td>2</td>
<td>.6800</td>
<td>.4800</td>
</tr>
<tr>
<td>3</td>
<td>1.2000</td>
<td>.5600</td>
</tr>
<tr>
<td>4</td>
<td>1.8700</td>
<td>.7800</td>
</tr>
</tbody>
</table>

$\eta = .0006$

Mean Square Quantizing Error $\xi^2 = .02786$
TABLE 2

OPTIMUM PARAMETERS FOR A SEVEN LEVEL SYMMETRICAL QUANTIZER

\( M = 7 \) \hspace{1cm} \( \sigma = 1 \)

<table>
<thead>
<tr>
<th>k</th>
<th>( a_k )</th>
<th>( b_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.13778</td>
<td>.27556</td>
</tr>
<tr>
<td>2</td>
<td>.41500</td>
<td>.27888</td>
</tr>
<tr>
<td>3</td>
<td>.70236</td>
<td>.29585</td>
</tr>
<tr>
<td>4</td>
<td>1.01471</td>
<td>.32884</td>
</tr>
<tr>
<td>5</td>
<td>1.36889</td>
<td>.37953</td>
</tr>
<tr>
<td>6</td>
<td>1.78893</td>
<td>.46054</td>
</tr>
<tr>
<td>7</td>
<td>2.38672</td>
<td>.73505</td>
</tr>
</tbody>
</table>

\( \eta = .00088 \)

Mean Square Quantizing Error \( \overline{e^2} = .0108 \)
CHAPTER VI

SPECIFICATION OF A QUANTIZER FILTER

The addition of a filter to the output of the quantizer can reduce certain undesirable components of the quantized signal such as the quantizing noise and also the effects of the noise component of the input signal applied to the quantizer. Prior to the specification of such a filter, the autocorrelation function of the quantizer output signal will be derived. As will be shown, the method used yields a particularly convenient form.

Quantizer Output Autocorrelation Function

In the previous chapter the quantizer output function was expanded in terms of Hermite polynomials. Due to the fact that the terms resulting from this expansion are uncorrelated functions of the input signal, this representation will be shown to be useful in the computation of the autocorrelation function of the quantizer output signal.

From Eq. (5-20)

\[ y(x(t)) = \sum_{p=1}^{\infty} D_p H_p(x(t)) \quad ; \quad |x| \leq L < \infty \]
where

\[(6-2) \quad D_p = \frac{2}{\sqrt{2\pi} \sigma^2} \sum_{m=1}^{M} b_m H_{p-1}(a_m)e^{-a_m^2/2\sigma^2}.\]

Hence the autocorrelation function of \( y(x(t)) \) is

\[(6-3) \quad R_y(\tau) = \frac{y(x(t))y(x(t+\tau))}{\sum_{p=1}^{\infty} \sum_{q=1}^{\infty} D_p D_q H_p(x(t))H_q(x(t+\tau))}.\]

In Chapter III the time average of

\( f(x(t), p) f(x(t+\tau), q) \)

was determined. Therefore since

\( f(x, p) = H_p(x) \)

and since \( x(t) \) is ergodic, then from Eq. (3-1) and (3-5)

\[(6-4) \quad H_p(x(t))H_q(x(t+\tau)) = \begin{cases} p! [R_x(\tau)]^p & ; p = q \\ 0 & ; p \neq q \end{cases}.\]

The final form of the autocorrelation of the quantized output function in the case where \( x(t) \) is normally distributed is, therefore

\[(6-5) \quad R_y(\tau) = \sum_{p=1}^{\infty} p^2 D_p^2 [R_x(\tau)]^p \quad \text{where} \quad R_x(\tau) = x(t)x(t+\tau).\]
Some of the implications of this method of obtaining the autocorrelation function of the output of a nonlinear device will be given.

The portion of \( y(x(t)) \) which is correlated with the input signal, \( x(t) \), is simply the first term of Eq. (6-1) which is

\[
D_1 H_1(x(t)) = D_1 x(t).
\]

By the nature of the Hermite polynomials the remainder of the terms of \( y(x(t)) \) are uncorrelated with each other and with \( x(t) \). Therefore the autocorrelation function of the component of the output which is correlated with the input signal is

\[
\left[ D_1 x(t) \right] \left[ D_1 x(t + \tau) \right] = D_1^2 R_x(\tau).
\]

The portion of \( y(x(t)) \) which is introduced by the nonlinear nature of the quantizer is given by all terms of Eq. (6-1) excluding the first term, that is

\[
\sum_{p = 3}^{\infty} D_p H_p(x(t)).
\]

The autocorrelation function of those terms is

\[
\sum_{p = 3}^{\infty} p! D_p^2 \left[ R_x(\tau) \right]^p.
\]

The autocorrelation function of the total output is the sum of these two components, Eq. (6-7) and (6-9), which is consistent with Eq. (6-5).
Optimum Linear Smoothing Filter

In this section a system which consists of a quantizer feeding a linear filter, as is shown in Fig. 7, will be considered.

![Filter diagram](image)

**Fig. 7.** - Filter optimizing configuration.

The input signal of the quantizer, which as before is denoted by \( x(t) \), now has a signal component, \( x'(t) \), and a noise component, \( n(t) \). The relation for the autocorrelation function of the quantizer output derived in the previous section is still valid if the autocorrelation of the input signal is expressed as

\[
R_X(\tau) = E[x(t) + n(t)]x(t + \tau) + n(t + \tau) = R_{X'}(\tau) + R_n(\tau) .
\]

The signal and noise will be assumed to be statistically independent therefore

\[
(6-10) \quad R_X(\tau) = x'(t) x'(t + \tau) + n(t) n(t + \tau) = R_{X'}(\tau) + R_n(\tau) .
\]

The filter is to be chosen in such a way that the mean square value of the difference between the filter output and the signal portion of the quantizer input is minimum thus the Wiener Hopf theory will be
used for the specification of the filter. The expression for the realizable optimum linear smoothing filter\textsuperscript{11} according to this theory

\[
H_{opt}(s) = \begin{bmatrix}
S_{x'y'}(s) \\
S_{yy'}(s)
\end{bmatrix}^+ \begin{bmatrix}
S_{yy}(s)
\end{bmatrix}
\]

(6-11)

where \( S_{yy'}(s) \) and \( S_{yy'}(s) \) are the portions of the Laplace transform of \( R(\tau) \) with singularities in the right half and left half planes respectively. Thus

\[
\mathcal{L} \begin{bmatrix} R_y(\tau) \end{bmatrix} = S_{yy'}(s) S_{yy'}(s).
\]

The cross power spectrum of \( x'(t) \) and \( y(t) \), \( S_{x'y}(s) \), is equal to the Laplace transform of the cross correlation of these signals which is

\[
R_{x'y}(\tau) = x'(t) y(x(t+\tau)) = x'(t) \sum_{p=1}^{\omega} D_p H_p(x(t+\tau)). \]

Since the only component of \( y(x(t)) \) which is correlated with \( x'(t) \) is contained in the first term of the series then

\[
R_{x'y}(\tau) = D_1 x'(t) \left[ x'(t+\tau) + n(t+\tau) \right]
\]

(6-12)

\[
= D_1 R_x'(\tau)
\]

since \( x'(t) \) and \( n(t) \) are independent.

Finally the symbol

\[
\begin{bmatrix}
\end{bmatrix}^+
\]
denotes the partial fraction expansion of the function enclosed with respect to the poles in the left half plane.

If the requirement that the filter be realizable is dropped, that is if \( h(t) \) is not identically zero for negative values of \( t \), then the resulting filter is called an infinite-lag smoothing filter and \( H(s) \) is given by

\[
(6-13) \quad H(s) = \frac{S_{xy}(s)}{S_{yy}(s)}.
\]

The filter specified in this way is valuable when long records of \( x(t) \) and \( y(t) \) are available before filtering is begun. For example a long segment of \( y(t) \) might be stored in a computer or on magnetic tape in such a form that it could be convolved with the impulse response of the filter to yield the filtered output.

As an example, the infinite-lag smoothing filter characteristic will be calculated for the optimized, seven level, symmetrical quantizer, the parameters of which are tabulated in Table 2. The form of the auto-correlation functions of the signal and noise components to be used is

\[
(6-14) \quad R_x(\tau) = a e^{-|\tau|}
\]

and

\[
(6-15) \quad R_n(\tau) = b e^{-c|\tau|} \quad \text{where} \quad c > 1
\]

respectively. These signals will be assumed to have zero mean, be normally distributed and be statistically independent. The variance of
the quantizer input signal which must be used in the calculation of the Hermite expansion of the quantizer output signal is

\[ (6-16) \quad \sigma_x^2 = R_x(0) = R_x(0) + R_n(0) = a + b. \]

Some of the coefficients of the Hermite expansion of the optimized, seven level quantizer transfer function for the case where \( \sigma_x^2 = a + b = 1 \) are given in Table 3. In addition the corresponding values of \( (\sigma_x^2)^p p! D_p^2 \) which are needed in the expression for the autocorrelation function of \( y(x) \), Eq. (6-5), are included.

<table>
<thead>
<tr>
<th>( p )</th>
<th>( D_p )</th>
<th>( p! D_p^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( 9.90 \times 10^{-1} )</td>
<td>( 9.80 \times 10^{-1} )</td>
</tr>
<tr>
<td>3</td>
<td>( -7.49 \times 10^{-3} )</td>
<td>( 3.37 \times 10^{-4} )</td>
</tr>
<tr>
<td>5</td>
<td>( -1.70 \times 10^{-3} )</td>
<td>( 3.47 \times 10^{-4} )</td>
</tr>
<tr>
<td>7</td>
<td>( -1.58 \times 10^{-5} )</td>
<td>( 1.27 \times 10^{-6} )</td>
</tr>
<tr>
<td>9</td>
<td>( 1.44 \times 10^{-6} )</td>
<td>( 7.49 \times 10^{-5} )</td>
</tr>
<tr>
<td>101</td>
<td>( 4.35 \times 10^{-68} )</td>
<td>( 1.78 \times 10^{-5} )</td>
</tr>
<tr>
<td>103</td>
<td>( -5.22 \times 10^{-70} )</td>
<td>( 2.70 \times 10^{-5} )</td>
</tr>
<tr>
<td>105</td>
<td>( 5.68 \times 10^{-72} )</td>
<td>( 3.49 \times 10^{-5} )</td>
</tr>
</tbody>
</table>
As can be seen from the table, the series

\[(6-17) \quad R_y(\tau) = \sum_{p=1}^{\infty} \sigma_x^{2p} p! D_p^2 \left[ R_x(\tau) \right]^p \]

converges very slowly. However, it will be shown that this is not a serious deficiency especially in light of the fact that the sum of the series

\[\sum_{p=1}^{\infty} \sigma_x^{2p} p! D_p^2 \]

\[p \text{ odd} \]

is known from Eq. (4-23), that is

\[(6-18) \quad \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y(x)^2 e^{-x^2/2\sigma_x^2} \, dx = \sum_{p=1}^{\infty} \sigma_x^{2p} p! D_p^2 \]

Thus by introducing the binomial expansion for

\[\left[ R_x(\tau) \right]^p \]

the final form of \( R_y(\tau) \) for the example used is

\[(6-19) \quad R_y(\tau) = \sum_{p=1}^{\infty} \sigma_x^{2p} p! D_p^2 \sum_{k=0}^{p} a^{p-k} b^k (p) \exp \left[ \frac{p + (c-1)k}{|\tau|} \right] \]

where

\[\binom{p}{k} \]

is the symbol for the binomial coefficients.
The power spectrum of $y(x)$, which is needed for the specification of the optimum filter, is obtained by taking the Laplace transform of $R_y(\tau)$. This relation in its exact form is

\[
(6-20) \quad S_{yy}(s) = \sum_{p=1, \text{p odd}}^{\infty} \sigma^{2p} p! D_p^2 \sum_{k=0}^{p} a^{p-k} k^k \binom{p}{k} \frac{2[p+(c-1)k]}{[p+(c-1)k]^2 - s^2}.
\]

The power spectrum of the quantizer output given by this equation could obviously be calculated to any accuracy desired by including a sufficient number of terms; however in view of the slow convergence of the coefficients,

\[
\sigma^{2p} p! D_p^2
\]

an approximate form of this equation will be described.

Since the filter to be specified is expected to be low pass, an upper limit, $s_1$, can be placed on the range of interest of $s$ in the expression for the power spectrum of the quantizer output signal and therefore a value of $p$, $p_1$, can be chosen such that for values of $p$ larger than $p_1$

\[
[p+(c-1)k]^2 >> s_1^2,
\]

thus Eq. (6-20) can be rewritten as
(6-21) \[ S_{yy}(s) = \sum_{p=1}^{p_l} \sigma^{2p} p! D_p^2 \sum_{k=0}^{p} a^{p-k} b^k (p) \frac{2 \left[ p + (c-1)k \right]}{\left[ p + (c-1)k \right]^2 - s^2} \]

\[ + \sum_{p=p_l+2}^{p_2} \sigma^{2p} p! D_p^2 \sum_{k=0}^{p} a^{p-k} b^k (p) \frac{2}{p + (c-1)k} \]

+ \( K \)

where \( p_2 \) is some value of \( p \) beyond which it is no longer desirable to calculate the terms as indicated and where \( K \) is the sum of the remainder of the terms. An upper bound for \( K \) can be obtained in the following manner. Since the function

\[ \sum_{k=0}^{p} a^{p-k} b^k (p) \frac{2}{p + (c-1)k} \]

monotonically decreases as \( p \) increases and since

(6-22) \[ \sum_{p=p_2+2}^{\infty} \sigma^{2p} p! D_p^2 = \frac{1}{\sqrt{2\pi} \sigma} \int_{-\infty}^{\infty} \left[ y^2(x) e^{-x^2/2\sigma^2} \right] dx \]

then an upper bound for \( K \) is given by
In the examples to follow the power spectrum of the optimized seven-level, symmetrical quantizer, the parameters of which are given in Table 2, will be calculated by using Eq. (6-21) for four different combinations of values of \( p_1 \), \( p_2 \) and \( K \). In addition the optimum infinite-lag smoothing filter system function will be calculated for each case. In this way the effect of different approximations of \( S_{yy}(s) \) on the filter system function can be demonstrated. The autocorrelation functions used in the examples for the signal and noise components of the quantizer are

\[
0.8 \exp(-|\tau|)
\]

and

\[
0.2 \exp(-5|\tau|)
\]

respectively.

For Cases 1 and 2 the values of \( p_1 \) and \( p_2 \) will be 9 and 299 respectively while for Cases 3 and 4 these values will be 19 and 299. The values of \( K \) used for Cases 1 and 3 will be obtained by the upper bound representation given by Eq. (6-23) while for Cases 2 and 4 the value of \( K \) used will be zero, representing the lower bound on \( K \). The equation for \( S_{xy}(s) \), Eq. (6-12), involves no approximations and thus the same value will be used for all four cases.

\[
\sum_{p=p_2+2}^{\infty} \sigma^2 p! D_p \left[ \sum_{k=0}^{p_2+2} \frac{p_2+2-k}{k} \left( \frac{p_2+2}{p_2+2+(c-1)k} \right)^2 \right]
\]
Case 1.

Eq. (6-21) after the substitution of the appropriate parameter values is

\[
S_{yy}(s) = \left[ \sum_{p=1}^{9} p! D_p^2 \sum_{k=0}^{p} (0.8)^{p-k} (0.2)^k \binom{p}{k} \frac{2(p+4k)}{(p+4k)^2 - s^2} \right]
+ \left[ \sum_{p=11}^{299} p! D_p^2 \sum_{k=0}^{p} (0.8)^{p-k} (0.2)^k \binom{p}{k} \frac{2}{p+4k} \right]
+ \left[ \sum_{p=301}^{\infty} p! D_p^2 \sum_{k=0}^{301-k} (0.8)^{301-k} (0.2)^k \binom{301}{k} \frac{2}{301+4k} \right].
\]

As an aid to evaluating this equation some of the terms of the expression within the first set of square brackets, that is the terms for which

\[(p + 4k) > 9,\]

will be simplified by assuming that the range of interest of \(s^2\) is less than \(9^2\), therefore this group of terms will actually be calculated in the manner indicated by the expression within the second set of square brackets of Eq. (6-24). Thus the power spectrum of the quantizer output is

\[
S_{yy}(s) = \left[ \frac{1.57}{1-s^2} + \frac{1.035 \times 10^{-3}}{3^2 - s^2} + \frac{1.96}{5^2 - s^2} + \frac{1.819 \times 10^{-3}}{7^2 - s^2} \right]
+ \left[ \frac{2.741 \times 10^{-3}}{9^2 - s^2} + 2.752 \times 10^{-5} \right] + \left[ 3.404 \times 10^{-5} \right] + \left[ 2.82 \times 10^{-5} \right]
\]

where the constants in the second and third brackets indicate the contributions of the second and third brackets of Eq. (6-24) respectively.
This equation expressed in terms of a common denominator becomes

\[
(6-26) \quad S_{yy}(s) = \frac{8.976 \times 10^{-5} (s^2 - (2.993)^2)(s^2 - (3.418)^2)}{(s^2 - 1)(s^2 - 3^2)(s^2 - 5^2)} \times \\
\frac{(s^2 - (6.999)^2)(s^2 - (8.997)^2)(s^2 - (198.5)^2)}{(s^2 - 7^2)(s^2 - 9^2)}
\]

which after the cancellation of nearly identical terms can be written in final form as

\[
(6-27) \quad S_{yy}(s) = \frac{8.976 \times 10^{-5} (s^2 - (3.418)^2)(s^2 - (198.5)^2)}{(s^2 - 1)(s^2 - 5^2)}
\]

The value of the cross power spectrum which is given by the Laplace transform of the cross correlation function of Eq. (6-12) is

\[
S_{x'y'}(s) = 0.8 \cdot \frac{2}{1 - s^2} = \frac{1.585}{1 - s^2}
\]

From these two relations the system function for the optimum infinite-lag smoothing filter can be obtained in the following manner

\[
(6-28) \quad H(s) = \frac{S_{x'y'}(s)}{S_{yy}(s)} = -1.768 \times 10^4 \frac{(s^2 - 5^2)}{(s^2 - (3.418)^2)(s^2 - (198.5)^2)}
\]
and finally the inverse Laplace transform of this function, that is the filter impulse response, is

\[
(6-29) \quad h(t) = 0.875 e^{-3.418|t|} + 44.5 e^{-198.5|t|}.
\]

The system functions of the filters for the remaining cases to be considered are calculated in the same manner therefore some of the steps will be omitted.

Case 2.

In this example the lower bound representation of \( K \) is used and the equation for \( S_{yy}(s) \) becomes

\[
(6-30) \quad S_{yy}(s) = \left[ \sum_{p=1 \text{ even}}^{2} p! D_p^2 \sum_{k=0}^{p} (0.8)^{p-k} \left( \frac{0.2}{k!} \right) \frac{2(p+4k)}{(p+4k)^2 - s^2} \right] + \left[ 0 \right].
\]

Thus

\[
(6-31) \quad S_{yy}(s) = 6.156 \times 10^{-5} \frac{(s^2 - (3.418)^2)(s^2 - (2.397)^2)}{(s^2 - 1)(s^2 - 5^2)}.
\]
Consequently

\[ (6-32) \quad H_\omega(s) = \frac{-2.575 \times 10^4 (s^2 - 5^2)}{(s^2 - (3.418)^2)(s^2 - (239.7)^2)} \]

and

\[ (6-33) \quad h(t) = 0.875 e^{-3.418|t|} + 53.7 e^{-239.7|t|} \]

Cases 3 and 4 represent an improvement in accuracy of the approximation since \( p_1 \) is nineteen rather than nine.

Case 3.

\[ (6-34) \quad S_{yy}(s) = \left[ \sum_{p=1}^{19} p! D_p \sum_{k=0}^{p} \frac{(0.8)^{p-k} (0.2)^k}{k!} \frac{2(p+4k)}{(p+4k)^2 - s^2} \right] \]

\[ + \left[ \sum_{p=21}^{299} p! D_p \sum_{k=0}^{p} \frac{(0.8)^{p-k} (0.2)^k}{k!} \frac{2}{p+4k} \right] \]

\[ + \left[ \sum_{p=301}^{\infty} p! D_p \sum_{k=0}^{301} \frac{(0.8)^{301-k} (0.2)^k}{k!} \frac{301}{301+4k} \right] \]

and therefore
\[ S_{yy}(s) = \frac{1.57}{1-s^2} + \frac{1.035 \times 10^{-3}}{3^2 - s^2} + \frac{1.96}{5^2 - s^2} + \frac{1.819 \times 10^{-3}}{7^2 - s^2} + \frac{2.74 \times 10^{-3}}{9^2 - s^2} + \frac{7.724 \times 10^{-4}}{11^2 - s^2} + \frac{2.438 \times 10^{-3}}{13^2 - s^2} + \frac{3.17 \times 10^{-4}}{15^2 - s^2} + \frac{1.409 \times 10^{-3}}{17^2 - s^2} + \frac{4.639 \times 10^{-4}}{19^2 - s^2} + 7.991 \times 10^{-6} \left[ 2.514 \times 10^{-5} + 2.82 \times 10^{-5} \right]. \]

The final form of this equation is

\[ S_{yy}(s) = \frac{6.133 \times 10^{-5} (s^2 - (3.418)^2) (s^2 - (240.3)^2)}{(s^2 - 1)(s^2 - 5^2)} \]

and then

\[ H(s)_{-\infty} = \frac{-2.59 \times 10^{4} (s^2 - 5^2)}{(s^2 - (3.418)^2)(s^2 - (240.3)^2)} \]

hence

\[ h(t) = 0.875 e^{-3.418|t|} + 53.9 e^{-240.3|t|}. \]
The form of the equation for $S_{yy}(s)$ for Case 4 is identical to that for Case 3 except that the expression within the third square bracket which represents $K$ is set to zero.

**Case 4.**

\[
S_{yy}(s) = \frac{3.313 \times 10^{-5} (s^2 - (3.418)^2)(s^2 - (326.95)^2)}{(s^2 - 1)(s^2 - 5^2)}
\]

Hence

\[
H(s) = \frac{-4.79 \times 10^4 (s^2 - 5^2)}{(s^2 - (3.418)^2)(s^2 - (326.95)^2)}
\]

and

\[
h(t) = 0.874 e^{-3.418 |t|} + 73.25 e^{-326.95 |t|}
\]

Each of the filters obtained by using the four approximations for the power spectrum of the quantized signal has a similar low frequency response as evidenced by the first term of each impulse response function. In each case the second term of these functions describes a relatively tall and narrow function, thus it is the second term which determines the high frequency response of the filter. The variations in
the four impulse response functions appear in the second term of these functions which is to be expected since the approximations of the power spectrum of the quantized signal involved only high frequency terms. Cases 3 and 4 can be expected to give the most accurate filter since more terms were included before the approximations were made.

The quantized signal can be considered to be composed of three types of terms. The terms of the first type are correlated with only the information bearing portion of the quantizer input signal, \( x'(t) \). The terms of the second type are correlated with the noise portion of the quantizer input signal, and finally those terms which are not correlated with either part of the input signal are of the third type. The terms of the third type are introduced by the nonlinear nature of the quantizer and are referred to as quantizer noise. The power of each of these three groups of terms at the input and at the output of the filter will be given in order to indicate the relative effectiveness of the filter.

These results were found to be essentially the same for the filters obtained in all four cases therefore the results for only Case 3 are given.

The total power of the first two types of terms can be obtained by using Eq. (6-7) and setting \( \tau \) to zero. Thus

\[
(6-42) \quad R_y(0) = D_1^2 \left[ R_x(0) + R_n(0) \right] = 0.98 \left[ 0.8 + 0.2 \right].
\]
The first of these two terms is the power of the terms correlated with \(x'(t)\) which is

\[ \sigma_{y,x'}^2 = 0.784 \]

and the second term is the power of terms correlated with \(n(t)\)

\[ \sigma_{y,n}^2 = 0.196 \]

The power of the terms of the quantized signal of the third type is found by subtracting these two components of power from the total power as obtained from Eq. (6-5). Thus

\[
\begin{align*}
\sigma_{y,\text{quant. noise}}^2 &= \sigma_{y,\text{total}}^2 - (\sigma_{y,x'}^2 + \sigma_{y,n}^2) \\
&= 0.985 - 0.980 = 0.005.
\end{align*}
\]

Since the filter is linear, the power of each of the types of terms at the filter output can be treated separately. The power of the terms of the first type is given by

\[
\begin{align*}
\sigma_{g,x}^2 &= R_{g,x}(0) = \frac{1}{2\pi j} \int_{-\infty}^{\infty} S_{y,x}(s) |H_s(s)|^2 ds = 0.578
\end{align*}
\]

where \(S_{y,x}(s)\) is the Laplace transform of

\[ R_{y,x}(\tau) = D_{\tau}^2 R_{x}(\tau). \]
Similarly the power of the terms of the second type at the filter output is

\[(6-45) \quad \sigma^2_{g,n} = R_{g,n}(0) = \frac{1}{2\pi j} \int_{-\infty}^{\infty} S_{y,n}(s)|H_\omega(s)|^2 ds = 0.0895.\]

And finally the power of the quantizing noise at the filter output is

\[(6-46) \quad \sigma^2_{g,\text{quant. noise}} = R_{g,\text{quant. noise}}(0) = \frac{1}{2\pi j} \int_{-\infty}^{\infty} S_{yy}(s)|H_\omega(s)|^2 ds \]

\[- (\sigma^2_{g,x} + \sigma^2_{g,n}) = 0.6700 - 0.6675 = 0.0025.\]

These values will be tabulated in Table 4.

<table>
<thead>
<tr>
<th>TABLE 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIGNAL POWER COMPONENTS</td>
</tr>
<tr>
<td>Filter Input</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Signal Variance</td>
</tr>
<tr>
<td>Noise Variance</td>
</tr>
<tr>
<td>Quantizing Noise Variance</td>
</tr>
</tbody>
</table>

It must be recognized that the criterion for the optimum linear filter was that the mean square value of the quantity

\[g(t) - x(t)\]

be a minimum and that the best that the linear filter can do is to strike a compromise between passing those frequencies associated with the
desired signal and treating unfavorably the remainder of the frequencies.

The mean square value of the above quantity can be shown to be

\[(6-47) \quad [g(t) - x(t)]^2 = \sigma_x^2 - \frac{1}{2\pi} \int_{-\infty}^{\infty} S_{yy}(\omega) |H_\omega(s)|^2 ds\]

\[= 0.80 - 0.67 = 0.13\]

As a consequence of this comparatively large error, the degree of reduction of the noise components of the filtered signal seems reasonable.

A portion of the noise present at the input of the quantizer could be removed by filtering of the input of the quantizer, which would effectively reduce the amplitude of the noise. A better optimum value for the filter on the output of the quantizer could then be found which would attenuate the desired signal to a lesser degree and still remove a portion of the undesired components.

The techniques used in the preceding examples were based on the decaying exponential form of the autocorrelation function for the quantizer input signals. The same techniques with some variations could be used in the case of other forms of the autocorrelation functions.
CHAPTER VII

EXPERIMENTAL FILTERING RESULTS

In order to obtain experimental results of the ability of the optimum infinite-lag linear filter to recover a desired signal from a quantized desired signal plus noise mixture, a digital computer simulation of the components involved has been used. The filter which was simulated was specified in Case 3 of the previous chapter. Thus the quantizer parameters to be used are those of the optimized seven level symmetrical quantizer of Table 2 and the power spectra of the signal and noise components of the quantizer input signal are

\[ S_X'(s) = \frac{1.6}{1 - s^2} \]

and

\[ S_n(s) = \frac{2}{25 - s^2} \]

respectively.

The first portion of this chapter will be devoted to a discussion of the manner in which two uncorrelated time signals which approximate random signals with normal distributions and with appropriate power spectrum can be specified. The final section of the chapter contains the
computer simulation results, that is strip charts of the signal and noise components of the quantizer input, the quantizer output and the filtered output.

Random Signal Development

Papoulis\textsuperscript{12} shows that it is possible to construct a signal, $g(t)$, having a power spectrum a positive function, $S(\omega)$. First the function,

$$w(t) = \sum_{k=-\infty}^{\infty} \alpha_k \delta(t-kc)$$

(7-3)

is formed, where $\alpha_k$ is a member of a sequence of numbers to be determined. The time autocorrelation $R_w(t)$ of this function is given by

$$R_w(t) = \lim_{r \to \infty} \frac{1}{2rc} \int_{-rc}^{rc} \sum_{m=-\infty}^{\infty} \alpha_m \int_{-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \alpha_p \delta(t+\tau-pc) \, d\tau$$

(7-4)

$$= \lim_{r \to \infty} \frac{1}{2rc} \sum_{m=-r}^{r} \sum_{p=-\infty}^{\infty} \alpha_m \alpha_p \int \delta(t-(p-m)c) \, d\tau.$$
By letting \( k = p - m \)
then

\[
(7-5) \quad R_w(t) = \sum_{k=-\infty}^{\infty} A_k \delta(t-kc)
\]

where

\[
(7-6) \quad A_k = \lim_{r \to \infty} \frac{1}{2rc} \sum_{m=-r}^{r} \alpha_m \alpha_{k+m}.
\]

The sequence of numbers \( \alpha_m \) will be chosen in such a way that

\[
(7-7) \quad \frac{1}{\pi} \lim_{r \to \infty} \frac{1}{2rc} \sum_{m=-r}^{r} \alpha_m \alpha_{m+k} = \begin{cases} B & ; k = 0 \\ 0 & ; k \neq 0 \end{cases}
\]

The autocorrelation \( R_w(t) \) of the sequence of uncorrelated pulses \( w(t) \) is then given by

\[
(7-8) \quad R_w(t) = B \delta(t)
\]

Therefore its power spectrum \( S_w(s) \) is flat and

\[
(7-9) \quad S_w(s) = B.
\]

Given the desired power spectrum, \( S(s) \), a realizable function, \( Y(s) \), which exhibits the minimum-phase-shift property, can be found such that
The minimum-phase-shift property requires that the poles and zeros of \( Y(s) \) appear only in the left-half plane and thereby insures that \( Y(s) \) can be determined from its amplitude.

The response of the system \( Y(s) \) to the input \( w(t) \) is then

\[
(7-11) \quad g(t) = \int_{-\infty}^{\infty} w(\tau) y(t-\tau) \, d\tau = \sum_{k=-\infty}^{\infty} \alpha_k y(t-kc)
\]

where \( y(t) \) is the inverse Laplace transform of \( Y(s) \). The power spectrum of \( g(t) \) is given by

\[
(7-12) \quad S_g(s) = S_w(s) \left| Y(s) \right|^2 = B \left| Y(s) \right|^2 = S(s)
\]

which is the desired power spectrum.

The members of the sequence of numbers, \( \{\alpha_k\} \), are chosen at random from a finite set of numbers, \( S_r \), of \( (Q + 1) \) different magnitudes, \( a_q \); \( 0 \leq q \leq Q \). Since positive and negative values are allowed for each non-zero magnitude there are \( (2Q + 1) \) different members in \( S_r \). The probability of the selection of a given one of these members for a given \( \alpha_k \) is \( \frac{1}{2Q + 1} \). The magnitudes of the members of \( S_r \) are selected in such a way that they approximate a normal distribution of zero mean and unity variance.
It can be shown that the sequence \( \{ \alpha_k \} \) constructed of members of the set \( S_R \) chosen at random satisfies the condition

\[
\lim_{r \to \infty} \frac{1}{2c_r} \sum_{m=-r}^{r} \alpha_m \alpha_{m+k} = 0 \quad ; \quad k \neq 0 .
\]

Given the magnitudes of the members of set \( S_R \), the product \( c_B \) in Eq. (7-7) can be determined in the following manner:

\[
(7-13) \quad \lim_{r \to \infty} \frac{1}{2r} \sum_{m=-r}^{r} \alpha_m^2 = \lim_{r \to \infty} \frac{1}{2r} \left[ \frac{2r}{2Q+1} (a_q)^2 + \frac{2r}{2Q+1} (a_{q-1})^2 + \ldots + \frac{2r}{2Q+1} (a_1)^2 \right] = c_B .
\]

It can be seen that this expression is not dependent on \( r \) and thus

\[
(7-14) \quad c_B = \frac{1}{2Q+1} \left[ 2 (a_q)^2 + 2 (a_{q-1})^2 + \ldots + 2 (a_1)^2 + (a_0)^2 \right] .
\]

For example let \( Q = 9 \) and \( c = .02 \). The cumulative normal distribution function for the variable \( x \) which has a zero mean and unity variance is

\[
(7-15) \quad F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du .
\]

The ten values of \( x \) which are obtained by dividing the half of the range of \( F(x) \) which is greater than .5 into ten equal pieces are given in Table 5. These values of \( x \) are set equal to \( a_q \). The values of \( a_q^2 \)
are also included in the table.

TABLE 5.--Magnitudes of Coefficients for Random Function Expression

<table>
<thead>
<tr>
<th>q</th>
<th>F(x)</th>
<th>x = a_q</th>
<th>a_q^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.50</td>
<td>.000</td>
<td>.0000</td>
</tr>
<tr>
<td>1</td>
<td>.55</td>
<td>.126</td>
<td>.0159</td>
</tr>
<tr>
<td>2</td>
<td>.60</td>
<td>.253</td>
<td>.0640</td>
</tr>
<tr>
<td>3</td>
<td>.65</td>
<td>.385</td>
<td>.1482</td>
</tr>
<tr>
<td>4</td>
<td>.70</td>
<td>.524</td>
<td>.2746</td>
</tr>
<tr>
<td>5</td>
<td>.75</td>
<td>.675</td>
<td>.4556</td>
</tr>
<tr>
<td>6</td>
<td>.80</td>
<td>.842</td>
<td>.7090</td>
</tr>
<tr>
<td>7</td>
<td>.85</td>
<td>1.037</td>
<td>1.0754</td>
</tr>
<tr>
<td>8</td>
<td>.90</td>
<td>1.282</td>
<td>1.6435</td>
</tr>
<tr>
<td>9</td>
<td>.95</td>
<td>1.645</td>
<td>2.7060</td>
</tr>
</tbody>
</table>

From Eq. (7-14) the value of the constant, B, is

\[(7-16)\]

\[B = \frac{1}{0.02} \cdot \frac{2 \times 7.0922}{19} = 37.327.\]

Since the desired power spectrum for the signal portion of the quantizer input signal is

\[(7-17)\]

\[S_{x'}(s) = \frac{1.6}{1 - s^2}\]

then the required value of \(Y(s)\) is determined as follows

\[(7-18)\]

\[
\left|Y_{x'}(s)\right|^2 = \frac{S_{x'}(s)}{B} = \frac{0.04286}{1 - s^2} = 0.207 \cdot \frac{0.207}{1 + s}.
\]
Therefore for \( Y_x(s) \) to be realizable, that is for \( y(t) \) to be identically zero for \( t \) less than zero, the portion with the pole in the left half plane is chosen and

\[
(7-19) \quad Y_x'(s) = \frac{0.207}{1 + s}.
\]

The inverse Laplace transform of this function is

\[
(7-20) \quad y_x'(t) = 0.207 U(t) e^{-t}
\]

where \( U(t) \) is the unit step function. Therefore the desired time function \( x'(t) \) is

\[
(7-21) \quad x'(t) = \sum_{k=-\infty}^{\infty} 0.207 \alpha_k U(t - 0.02k) e^{-(t - 0.02k)}
\]

where the \( \alpha_k \)'s are chosen at random from the \((2Q + 1) = 19\) members of \( S_r \), the magnitudes of which are given in the \( a_q \) column of Table 5.

The desired power spectrum of the noise component of the quantizer input is

\[
(7-22) \quad S_n(s) = \frac{2}{25 - s^2}
\]
therefore

(7-23) \[ Y_n(s) = \frac{0.231}{s + 5} \]

and

(7-24) \[ y_n(t) = 0.231 \ U(t) e^{-5t} \]

and finally

(7-25) \[ n(t) = \sum_{k=-\infty}^{\infty} 0.231 \beta_k \ U(t - 0.02k) e^{-5(t - 0.02k)} \]

The \( \beta_k \)'s are also chosen at random from the nineteen members of the set \( S_r \); however, they are chosen independently of the \( \alpha_k \)'s.

Discussion of Simulation Results

The desired signal component of the quantizer input, as given by Eq. (7-21), is shown in Fig. 8 as a solid line. The noise component of the quantizer input, Eq. (7-25), is shown in Fig. 9. The effect of the higher frequency content of the noise portion of these two quantizer input signals caused, as expected, a greater number of zero crossings of the curve of Fig. 9. The input signal to the simulated quantizer is expressed as the sum of Eq. (7-21) and (7-25). This waveform is shown in Fig. 10. Figure 11 shows the response of the optimized seven-level
Fig. 8.--Desired signal component of quantizer input indicated by the solid line. Filter output signal indicated by the broken line.

Fig. 9.--Noise component of quantizer input.

Fig. 10.--Quantizer input signal.

Fig. 11.--Quantizer output signal.
Fig. 8

$x'(t)$ (solid line)

$g(t)$ (broken line)

Fig. 9

$n(t)$

Fig. 10

$x(t) = x'(t) + n(t)$

Fig. 11

$y(t)$
symmetrical quantizer to the input signal of Fig. 10. The parameters for the quantizer, which was simulated, are given in Table 2. The final result of the simulation, the filter output signal, $g(t)$, was obtained by convolving the filter impulse response function of Eq. (6-28) with the quantizer output signal. The filter output signal is shown as a broken line superimposed upon the desired signal component of the quantizer input in Fig. 8. As can be seen, the filter output is an approximate representation of $x(t)$. Recalling from Eq. (6-47), the theoretical mean square filtering error was

$$\left[ g(t) - x'(t) \right]^2 = 0.13$$

Due to the comparatively large value of this function the compliance of the curve for $g(t)$ to the curve for $x(t)$ in Fig. 8 can be expected to be good but not outstanding.
CHAPTER VIII

CONCLUSIONS

The technique of representing the output of a quantizer as a series of Hermite polynomials can be applied to most nonlinear devices if the input signal is normally distributed. To fall into this category, the transfer function of the nonlinear device, that is the relation expressing the output of the device as a function of the input, must be single valued, contain no impulse-type terms, and have a finite number of discontinuities.

The requirement that the input to these devices be normally distributed can often be relaxed provided that the input signal to the nonlinear device can be derived from a signal with a normal distribution which is often the case.

The determination of an optimum combination of two filters, the first to remove undesired noise from the quantizer input signal and the second to filter the quantized signal is another area for further work. It appears likely that the filters should not be restricted to linear filters since it appears that certain nonlinear filters are quite effective in suppressing quantizer noise.
BIBLIOGRAPHY


