DATA REDUCTION AND COMPUTER GROUP

DOCUMENT 153-71
(Formerly 119-71)

ERROR ANALYSIS AND METHODS FOR ESTIMATING ERRORS IN POSITION, VELOCITY, AND ACCELERATION DATA

WHITE SANDS MISSILE RANGE
KWAJALEIN MISSILE RANGE
YUMA PROVING GROUND
DUGWAY PROVING GROUND
ELECTRONIC PROVING GROUND

ATLANTIC FLEET WEAPONS TRAINING FACILITY
NAVAL AIR WARFARE CENTER WEAPONS DIVISION
NAVAL AIR WARFARE CENTER AIRCRAFT DIVISION
NAVAL UNDERSEA WARFARE CENTER DIVISION NEWPORT

30TH SPACE WING
45TH SPACE WING
AIR FORCE FLIGHT TEST CENTER
AIR FORCE DEVELOPMENT TEST CENTER
AIR FORCE WEAPONS AND TACTICS CENTER
DETACHMENT 2, SPACE AND MISSILE SYSTEMS CENTER

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

Document 153-71

ERROR ANALYSIS AND METHODS FOR ESTIMATING ERRORS IN POSITION, VELOCITY, AND ACCELERATION DATA

Data Reduction and Computing Group
Range Commanders Council

Published May 1971 by the

Secretariat
Range Commanders Council
White Sands Missile Range,
New Mexico 88002

CONTENTS

	Foreword	1
1.0	Preface	3
2.0	Introduction	5
3.0	Error Classification, Accuracy and Precision	7
3.1	Errors in Basic Quantities	13
3.2	Causes of Errors in Instrumentation Tracking Systems	15
3.3	Correlation of Random Error	17
4.0	Methods Used to Estimate Random Errors	21
4.1	Graphical Techniques	22
4.2	Least Squares Point Estimates of Random Error	23
4.3	Multi-Instrument Estimate of Random Error	25
4.4	Variate Difference Method	26
4.5	Least Square Curve Fitting Method	29
4.6	Use of a High Pass Digital Filter	31
4.7	Effects of Autocorrelation on the Estimation of Random Errors	33
5.0	Interval Estimation of Random Errors	39
5.1	The Case of the Normal Distribution	41
5.2	Confidence Intervals on Error Parameter Estimates	43
6.0	Methods Used to Evaluate Systematic Errors	45
6.0.1	Analytical Investigation of Possible Error Sources	45
6.0.2	Comparison of Observations from Different Systems	45
6.0.3	Analysis of Measuring Residuals from Over-Determined Least Squares Solution	46
6.0.4	Analysis of Residuals from a Best Estimate of Trajectory	46
6.0.5	Multilateration	46
6.1	Error Models	47
6.2	Regression Analysis	51
6.3	Analysis of Variance	53
7.0	Theory of Error Propagation	59
7.1	Propagation of Errors into Position	63
7.2	Propagation of Errors into Velocity and Acceleration	69
8.0	Error Tolerance Regions	75
8.1	Tolerance Ellipsoids	77
9.0	Digital Filtering/Smoothing of Data	79
9.1	Effects of Smoothing on Data	83
10.0	Inertial Guidance Data	93
11.1	Error Analysis in the Frequency Domain	99
	Bibliography	117

LIST OF FIGURES

1. Concepts of accuracy, precision and error

1.a Systematic errors

2. Uncorrelated noise

3. Correlogram of AN/FPS-16, R, A and E.

4. Frequency response

5. Digital low pass filter frequency response comparisons

6. Least square regression

7. Least square regression

8. Reduction factor K

9. Standard Normal Frequency Distribution Curve

10. Estimate of Errors

11. Velocity Error for 2nd Degree Polynomial

12. Velocity Error for 3rd Degree Polynomial

13. Velocity Error for 4th Degree Polynomial

14. Acceleration Error for 2nd Degree Polynomial

15. Acceleration Error for 3rd Degree Polynomial

16. Acceleration Error for 4th Degree Polynomial

17. K_a, K_b, and K_c as a function of N

18. Variance reduction curve

19. Effects of folding or aliasing

20. Examples of continuous autocorrelation functions

21. Estimate of autocorrelation function of white noise

22. Estimate of autocorrelation function for first order MARKOV process

23. Estimateof Markov process using 4th order autoregressive scheme

24. Analysis of variance for the autoregression scheme

25. Estimate of spectral density function

26. Regression analysis of data

27. Estimate of spectral distribution function, F

FOREWORD

Mathematical models, methods and techniques which are useful and appropriate for estimating the accuracy of position (in some coordinate systems); velocity, and acceleration data are presented in this document. The development and use of the techniques discussed have evolved through the years and in some cases out of work not related to missile testing. Preparation and coordination of the material contained herein have extended over the past six years with contributions from the various organizations in IRIG. The material is not meant, or expected, to yield complete agreement as to the relative merits or importance of these procedures. However, this document is a step in the direction toward the eventual establishment of IRIG "guide lines" for recommended techniques for determining and presenting the accuracy estimates of data collected from instrumentation systems in support of operational testing. In addition, this volume will provide a source of documentation on current and available procedures as well as definitions concerning error, accuracy, and precision.

1.0 PREFACE

The Inter-Range Instrumentation Group has had as one of its goals the establishment of an IRIG standard for techniques in determining the accuracy estimates for instrumentation systems and methods of presenting these estimates. The IRIG Steering Committee assigned the task of accomplishing this goal to the Data Reduction and Computing Working Group (DR&CWG).

During the past several years DR&CWG has endeavored to investigate the many aspects of this task. Work has proceeded slowly because the formulation of standard and acceptable methods for determining accuracies is not easy. Furthermore, it is difficult to obtain agreement on the means of presenting the accuracy estimates. This work has required extensive communication and coordination on techniques in use at the various member ranges. As a result of this continued exchange of information, a common language and increased understanding have evolved. This evolution has been slow, but it now appears that sufficient agreement between member ranges and adequate technical maturity have developed.

In an effort to get some results completed in print, the DR&CWG decided to break the task into two parts. The first portion, dealing with the positional data accuracies, appeared as IRIG Document Number 104-62. The second portion, scheduled to deal with velocity and acceleration data accuracies, was written in 1963 and coordination was undertaken. However, at the 21st meeting of DR&CWG it was concluded that there was much overlap in the two documents and that it would be desirable to consolidate them since 104-62 was then at least three years old. The resulting document, 103-64, was greatly amplified and contained many new topics.

This new document basically contains the material in 103-64, with amplified material, and includes some of the newer techniques developed and used during the past few years.

2.0 INTRODUCTION

To assess the accuracy of the final result of processing data from range instrumentation is extremely difficult. To obtain agreement among the various member ranges on a standard for techniques is even more difficult. The major reasons for these difficulties stem from the variety and diversity of:

1) Various types and amounts of instrumentation used

2) Application of these instruments

3) Operational techniques employed

4) Data handling procedures

5) Mathematical and computational techniques

6) Types of requirements for reduced data

Although the member ranges have a wide variation in instrumentation and operational techniques, it is generally agreed that the measurement, X_t, collected at an instrument at time, t, is composed of a "true value" or "signal," μ_t, and an error component, ϵ_t. Thus

$$X_t = \mu_t + \epsilon_t$$

or (2.0.1)

$$\epsilon_t = X_t - \mu_t$$

The common problem in data reduction (and error estimation) is to devise mathematical models and corresponding numerical analysis techniques which will effectively separate the signal from the error in some "optimum" manner.

In discussions and conferences pertaining to error estimation, much has been said about the "true value" of a parameter. First, it is emphasized that the true value of a parameter being measured is unknown. If the true value were known there would be no need for further discussion of errors or for taking any measurements in the first place. As Cassius J. Keyser has remarked, "Absolute certainty is a privilege of uneducated minds -- and fanatics." Even though the true value cannot be known exactly, it is still a very useful concept which is used in the construction of mathematical models to represent and estimate the error in a measurement or set of measurements.

Since the true value of the signal is unknown, the approach is to devise mathematical models and experimental design techniques to give "good" or "best" estimates of the signal. In data reduction terminology, these estimates are often referred to as "computed values" or as "standards for comparison" in error analysis. The computed value, $\hat{\mu}_t$, is a numerical function of the measurements X_t or may be a simultaneous measurement from a more accurate instrumentation system or several systems. The estimate of the error is

$$\epsilon_t \approx e_t = X_t - \hat{\mu}_t. \qquad (2.0.2)$$

The quantity e_t is called a residual, and applied error analysis utilizes this quantity as its basic input in estimating error characteristics.

It is important to note that the measurement X_t is a random variable and since $\hat{\mu}_t$ and e_t are numerical functions of X_t, they are also random variables. Each of these quantities are functions of time and thus make up a realization or sample from a stochastic or random process. If the observations are taken at discrete times, they form a sequence of random variables or a stochastic sequence and are commonly referred to as a discrete time series. This concept readily suggests that probability and statistics play an important role in error estimation, as well as in obtaining parameter estimates from the data reduction process.

The concept of accuracy with respect to measured and/or reduced data is closely related to the error therein, but is not identical to it. Certainly "small error" implies high accuracy or "accurate" data, and "large error" corresponds to low accuracy or inaccurate data. In general it is sufficient to state that accuracy is some function of the error distribution, and in the final analysis, accuracy itself must be estimated and based on quantities which are approximated numerically, assumed, or based on statistical estimation of parameters which characterize the error distribution.

3.0 ERROR CLASSIFICATION, ACCURACY, AND PRECISION

In accordance with Section 2.0 the error ϵ_t is estimated by the residual e_t and the accuracy of the measured and/or reduced data is estimated by some function of the residual. This function usually involves the expected value of the error. The expected errors are generally classified into two basic types, the random (noise) errors and the systematic errors. The theoretical error model is

$$\epsilon_t = S_t + N_t \tag{3.0.1}$$

where

S_t = systematic error, and

N_t = random or noise error.

The random error is nondeterministic. This means it is not described by an analytic function but must be characterized in terms of its probability distribution function. The systematic error is generally considered to be deterministic and can be represented with an analytic model. If the error distribution is thought of in terms of variation, then the systematic error comprises the "explained variation" while the random error corresponds to the "unexplained variation."

Another type of error which is not explicitly covered by the model (3.0.1), but which cannot be overlooked is "gross error." This type of error is usually due to some type of malfunction in the measurement process and is not a usable or valid measurement. For this reason data with this type of error is referred to as "bad data," "wild points," or more formally as "outlyer data." This type of error does not belong to the statistical population described by the probability distribution function of N_t and hence must be removed or edited out so it will not bias or corrupt the data reduction process. Many different types of editing techniques are used on the various ranges and are not discussed in this report. However, it is emphasized that if outlyers are ignored in the data reduction process their effect will corrupt reduced data from neighboring valid measurements and render the otherwise valid data unusable.

It is important to note that it may often be difficult to distinguish random from systematic error in a given set of measurements. This may be due to the fact that certain constraint conditions exist (geometry limitations, availability and capability of instrumentation, cost, safety, etc.) and no feasible test can be designed which will allow a reduction process to give adequate residuals; or it may be that S_t is very complicated and an adequate error model is not available; or it may be that the variance of the random error is large and the ratio of systematic error to noise error is small (error signal to noise ratio), which corrupts the estimates made of S_t. Nevertheless, in concept, and often in practice, it is possible to treat random errors statistically while the systematic errors may be estimated in a deterministic manner and corresponding corrections made to the observed data.

The combination of systematic and random error is referred to as total error, ϵ_t, and in this document the accuracy of a measurement is defined as the root mean square of the distribution of the total error. In statistical terms

$$\text{Accuracy} \sim \sqrt{E\left[(\text{total error})^2\right]} = \sqrt{E[\epsilon_t^2]} \tag{3.0.2}$$

where E (·) is the expected value of the quantity in brackets.

If the systematic error and random error are independent, then

$$\text{Accuracy} \sim \sqrt{E\left[(\text{systematic error})^2\right] + E\left[(\text{random error})^2\right]}. \tag{3.0.3}$$

Another important term which should be distinguished from accuracy is "precision." There are probably few words as loosely used by scientific personnel as precision and accuracy. Accuracy can be described as "closeness to the truth" while precision is the "closeness together" or internal consistency of a set of measurements (see Figure 1). From these descriptions we define precision as

$$\text{Precision} \sim \sqrt{E\left[(\text{random error})^2\right]}. \tag{3.0.4}$$

It is seen that accuracy requires precision but precision does not necessarily imply accuracy. If the systematic error $\gg$ random error then in this special case

$$\text{Accuracy} \sim \sqrt{E\left[(\text{systematic error})^2\right]}. \tag{3.0.5}$$

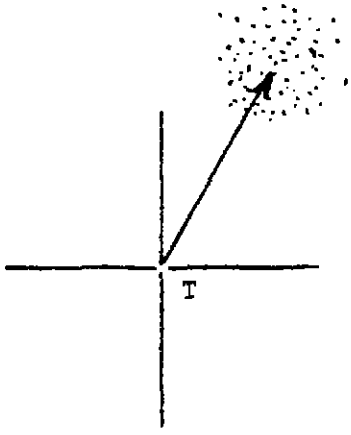
It is recognized that among the ranges there is not necessarily unanimous agreement regarding the classification of errors and the corresponding concepts of precision and accuracy. However, in the interests of practicability the errors are classified as random and systematic with the important distinction that the term "bias error" is considered to be systematic error and "constant bias error" is considered to be the constant component of systematic error.

At times it is convenient to characterize and analyze error in the frequency domain. When this is done for trajectory type data the error classification may be grouped into two categories. The high frequency category may correspond in a sense to random noise errors while the low frequency end of the spectrum may correspond to systematic errors (although not necessarily). The systematic error has a special case corresponding to zero frequency which is constant bias error. The frequency spectrum is a useful tool in the analysis of error and is discussed in more detail in Section 11.0.

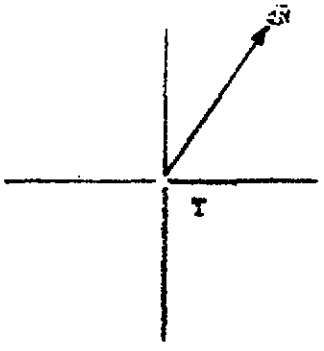
To effectively estimate systematic errors with the idea of predicting what accuracy can be expected on future tests, reliance must be made on experience from past error analyses concerning the variation of systematic errors within a test and from test to test. Viewed in this manner, the systematic errors are random variables which may be characterized by a probability distribution function. An example of how the constant bias estimate from a residual distribution of an instrumentation system such as a radar will vary from test to test is illustrated in Figure 1.a.

The variation of errors within and between tests, and between instrumentation sites leads one to look for techniques which will determine whether or not significant differences exist in the error from variable factors and to estimate what effects changes in these factors have on the error. Techniques relating to this type of error analysis deal with the analysis of variance and are discussed in Section 6.3.

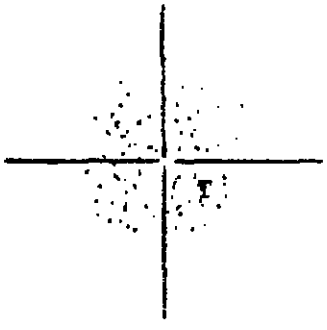
It is recognized that there are other descriptions which may be applied to errors. Correlation of errors in all of its detail may not be fully understood at this time. The various interdependencies of data and the added complications which arise from auto, serial, and/or time series correlation makes it unwise to attempt a full classification at this time; however, the importance of correlation between and within errors should not be ignored and a general discussion of correlation and its effects on error estimation is given in Sections 3.3 and 4.7.



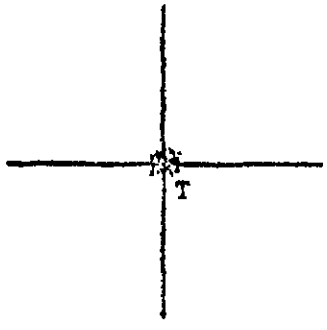
1 - Low Accuracy,
Low Precision



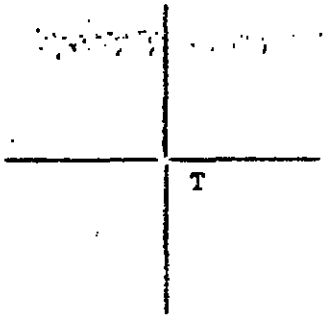
2 - Low Accuracy,
High Precision



3 - Low Precision,
No Systematic Error



4 - High Accuracy,
High Precision



5 - Systematic Error

FIGURE 1. The concepts of accuracy, precision and error are illustrated. Consider T to be a target with the dots representing the results from several firings at the target

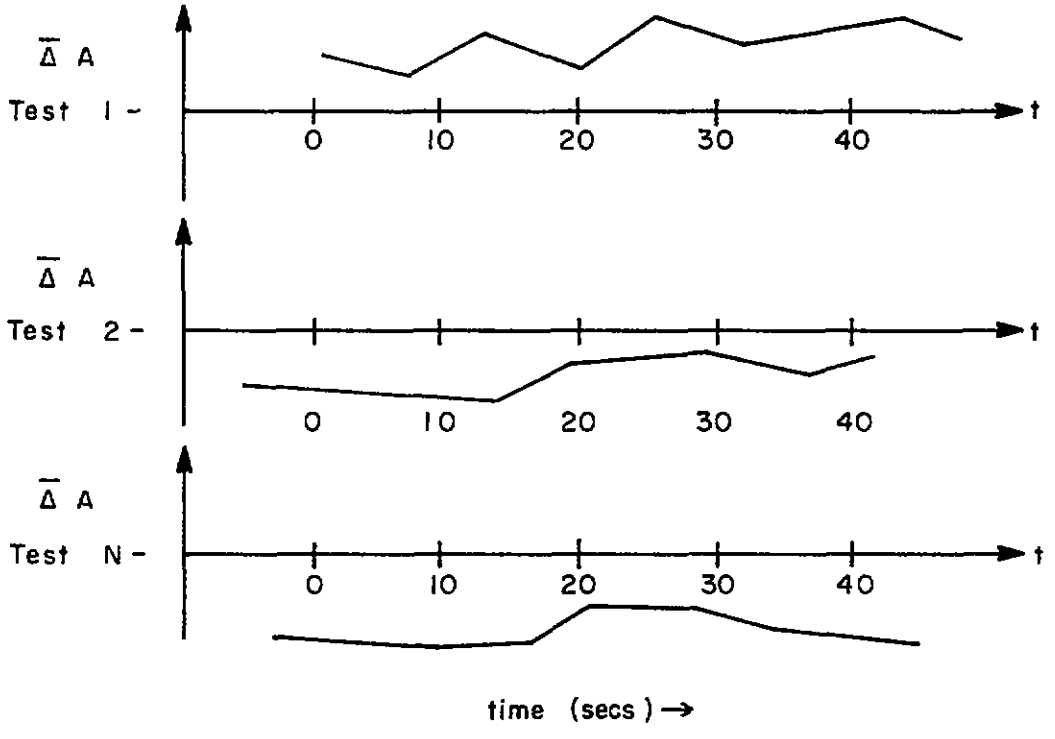


FIGURE I.a The systematic errors $\bar{\Delta}$ in azimuth A, as measured by a radar, have variations as pictured above both from test to test and as a function of time for a particular test.

3.1 ERRORS IN BASIC QUANTITIES

Thus far only errors in measurements or observations have been discussed. However, a great variety of error sources enter into the accuracy considerations for data systems used in tracking missiles. At this time and within the present state-of-the-art, it is not practical or even possible to accurately estimate the error in position, velocity, and acceleration due to each of these error sources. In an attempt to meet demands for more and more stringent accuracy requirements, great effort has been directed toward the improvement of instrumentation. However, it must be realized that even with perfect instrumentation, there are basic theoretical limitations on the accuracy which may be achieved. These limitations may introduce errors which make impossible the accuracy which is being demanded of the tracking systems.

Typical errors associated with some basic quantities are listed below. These exist quite apart from the instrumentation and place limitation or lower bound on the accuracy which may be attained. It is realized that there may be no unanimity regarding these magnitudes, or for that matter, their types, i.e., probable, absolute, etc., but they will be given as points of departure.

a)	First order, Class I surveys (distance)	10 ppm (parts/1000000)
b)	Velocity of light	1-2 ppm
c)	Index of refraction	25 ppm
d)	Changes in refraction due to rapid fluctuations in atmosphere	30-40 ppm
e)	Errors in direction cosines caused by the above fluctuations	10 ppm
f)	Overall ballistic camera accuracy	5-15 ppm
g)	Errors in reading ballistic camera film due to 1°C fluctuation in reading room temperature	5 ppm
h)	Star catalog error	1 ppm
i)	Error in 100 mile base line	1-5 feet
j)	Error in origin of national survey	50-100 feet
k)	Differences between International Spheroid and Clarke 1866 Spheroid	150-300 feet
l)	Undulations of the geoid Δa)	50-100 feet
m)	Capability to determine relative geoid heights at widely separated stations	30 feet
n)	Voltage level (secondary standard)	100 ppm
o)	Mass	.0001-100 ppm
p)	Timing (atomic standard	10^{-12} seconds
q)	Timing (laboratory electronics)	10^{-8} seconds
r)	Timing (instrument crystal	10^{-6} seconds
s)	Timing correlation between separate stations	10^{-5} seconds

Some of the above values are contained in Reference (1).

It is interesting to observe the effects of the quoted survey and refraction errors when propagated into spatial position and velocity at a range of 250 miles. The resulting errors due to these sources alone are:

Position: 15 feet in horizontal coordinates
40 feet in vertical coordinates

Velocity: 0.15 foot/second in horizontal coordinates.
0.40 foot/second in vertical coordinates

Obviously, these are facts which must be considered when imposing tracking accuracy requirements and in designing instrumentation systems to meet them. One is immediately faced with the operation of the law of diminishing returns.

3.2 CAUSES OF ERRORS IN INSTRUMENTATION TRACKING SYSTEMS

Once a set of residuals has been calculated with corresponding estimates of the random and systematic errors, the process of determining the cause of the errors is often extremely difficult. The diffculty encountered in determining the cause of error depends on several factors such as the ability to effectively design and control an experiment, knowledge of the variables which effect the error, correlations and effects between variables, correct assumptions concerning error model and probability distribution functions.

Some causes of random errors in measurements may be tropospheric and ionospheric scintillations, electronic noise, instrumentation wear, mechanical play, granularity, or tolerances in the measurement capabilities of the instrument, errors in computing and timing, multipath, scintillation due to echo or skin track, film reading, and so on.

Systematic errors are caused by physical limitations in achieving a true geodetic survey, electronic, optical and mechanical misalignment, servo lag, phase and frequency drift, frequency and timing bias, encoder nonlinearity, antenna droop, mislevel, lens distortion, beacon delay, miscollimation, dial eccentricity, and nonorthogonality of azimuth and elevation axes. One of the most important types of systematic error is usually considered to be constant bias or zero set error. This error is very important because it is often the most significant component of the systematic error S_t. In addition, it is the easiest component to estimate and to correct for; in fact, under actual flight conditions, the dynamics and geometry are often of such nature that the constant bias error is the only component of the systematic error which can be effectively estimated. Because of its constant nature, attempts to correct the instrumentation for this type of error can be made using static calibration tests before and after the actual flight tracking operation. Although constant bias error in the raw measurements can be considered practically constant for the duration of a single test, its propagated effects on position error may vary considerably during a single test because of sensitivity to tracking geometry. Neighboring spatial points in the time and space domains will in general have nearly constant error due to a fixed bias in misalignment. However, propagated position error would be significantly different at t+10 seconds (close range) from what it would be at t+150 seconds (slant range of several hundred miles). However, position errors due to the misalignment in the period t+140 to t+160 seconds would be expected to remain almost constant depending on geometry changes in that particular interval.

3.3 CORRELATION OF RANDOM ERROR

The random errors $N_1,...,N_n$ corresponding to measurements $X_1,....X_n$ may be interrelated so that the error in the jth measurement may be dependent upon the errors in $X_{j-1},X_{j-2},...X_{j-p}$, $p \leqslant j$. If this relationship exists, the errors are dependent and said to be correlated. In the case of a time series this interdependence is called the autocorrelation and is estimated by means of computing the serial correlation.

From a mathematical standpoint two random variables X and Y are independent if their joint probability density

$$F(x,y) = f(x)\, g(y)$$

where f(x) and g(y) are the probability density functions corresponding to X and Y respectively. Two random variables are uncorrelated if E(XY)=E(X)E(Y). It can be shown that independence implies two random variables are uncorrelated, but not vice versa. The correlation coefficient between the random variable X and Y is defined as

$$\rho_{xy} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{[E(X - \mu_X)^2\, E(Y - \mu_Y)^2]^{\frac{1}{2}}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{3.3.0}$$

From relation (3.3.0) we see the correlation coefficient is the ratio of the covariance to the product of the standard deviations of X and Y. If a sample $X_1,...,X_n$, $Y_1,...,Y_n$ is taken, the correlation coefficient, ρ_{xy}, is estimated by

$$r_{xy} = \frac{(N-1)\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{N\left[\sum_{i=1}^{N} (X_i - \bar{X})^2 \sum_{i=1}^{N} (Y_i - \bar{Y})^2\right]^{\frac{1}{2}}} \tag{3.3.1}$$

If $X_1, X_2,...,X_n$ is a realization the autocorrelation $\rho(k)$ at lag k is estimated by the serial correlation

$$r_k = \frac{\frac{1}{N-k}\sum_{i=1}^{N-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{\left[\frac{1}{N-1}\sum_{i=1}^{N} (X_i - \bar{X})^2 \frac{1}{N-k-1}\sum_{i=1}^{N-k} (X_{i+k} - \bar{X})^2\right]^{\frac{1}{2}}}, \tag{3.3.2}$$

$$k = 0,1,2,...,M<N$$

(3.3.2) is analogous to (3.3.0) in that the autocorrelation function is the ratio of the autocovariance to the standard deviations of X_i and X_{i+k}. If we let the autocovariance estimate at lag k be R(k) and if the stochastic process is stationary, then the correlation coefficient can be replaced by the relation

$$r_k = \frac{R(k)}{R(0)} \tag{3.3.3}$$

(3.3.3) is called the normalized auto-covariance or the circular correlation coefficient. The graph of r_k is sometimes called a correlogram (see Figure 3).

The autocorrelation between errors is especially known to be present if the measurements are from an instrument whose measurements are taken with a small time increment between them. Most types of range instrumentation are of this type and as a result the autocorrelation of the errors occurs frequently.

The effects of high correlation between errors can lead one to trouble in both data reduction and error analysis. If the correlation among errors is high and does not "die out" or become small in a short time the observed data will deceptively appear to be smoother (see Figure 2), or to contain what appears to be low frequency oscillations and trends which could very easily be falsely assumed to be "signal" or valid information, and could lead one to false conclusions about the proper technique for editing and smoothing of the data. In addition, it can be shown that if high correlation exists, then estimates of the variance of the error distribution will be biased (see Section 4.7).

Several methods are used for modeling autocorrelated data. The general idea is to express the random error as a stochastic function which depends on parameters which are random variables. The most common is a pth order stochastic difference equation or pth order Markov process. The first order Markov process gives a probabilistic model of the form

$$N_i = \rho N_{i-1} + \eta_i \tag{3.3.4}$$

where η_i is an independent random variable $0 \leq \rho \leq 1$. The serial correlation coefficient corresponding to this process can be shown to be exponential, i.e., the correlation between the errors in the ith and (i+k)th measurement is ρ^k, where ρ is the correlation coefficient between adjacent measurements. Figure 2 shows independent noise (ρ=0) and correlated noise corresponding to a first order Markov process with ρ=.3, ρ=.6, and ρ=.99, respectively. Inspection of the graphs of the errors in Figure 2 shows that as the correlation becomes high, the data appears smoother and systematic trends appear in the error profile which can be very misleading.

The second order Markov process is

$$N_i = a_1 N_{i-1} + a_2 N_{i-2} + \eta_i$$

where $|a_1| < 1$, $|a_2| < 1$, and η_i is an independent random variable. In this model the autocorrelation damps out exponentially but also contains a sinusoidal oscillation. An example of FPS-16 error estimates which follow this type of autocorrelation is given in Figure 3. Note that on this figure the autocorrelation is low at a lag of 0.5 seconds between points (i.e., the correlation between points separated by 0.5 seconds is low).

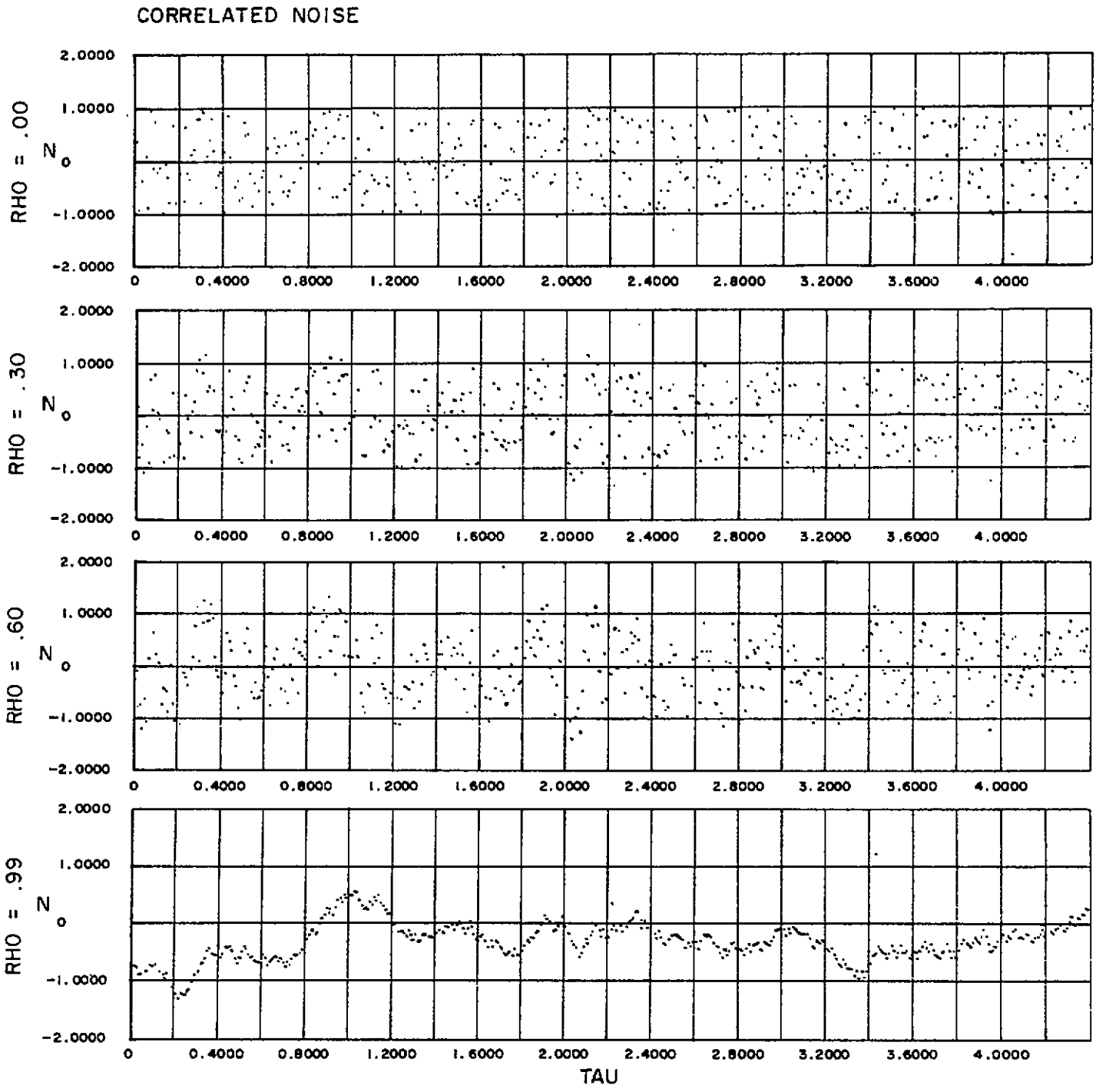


FIGURE 2 Example of uncorrelated noise (RHO = 0) and low correlated noise (RHO = .3) to very high correlated noise (RHO = .99).

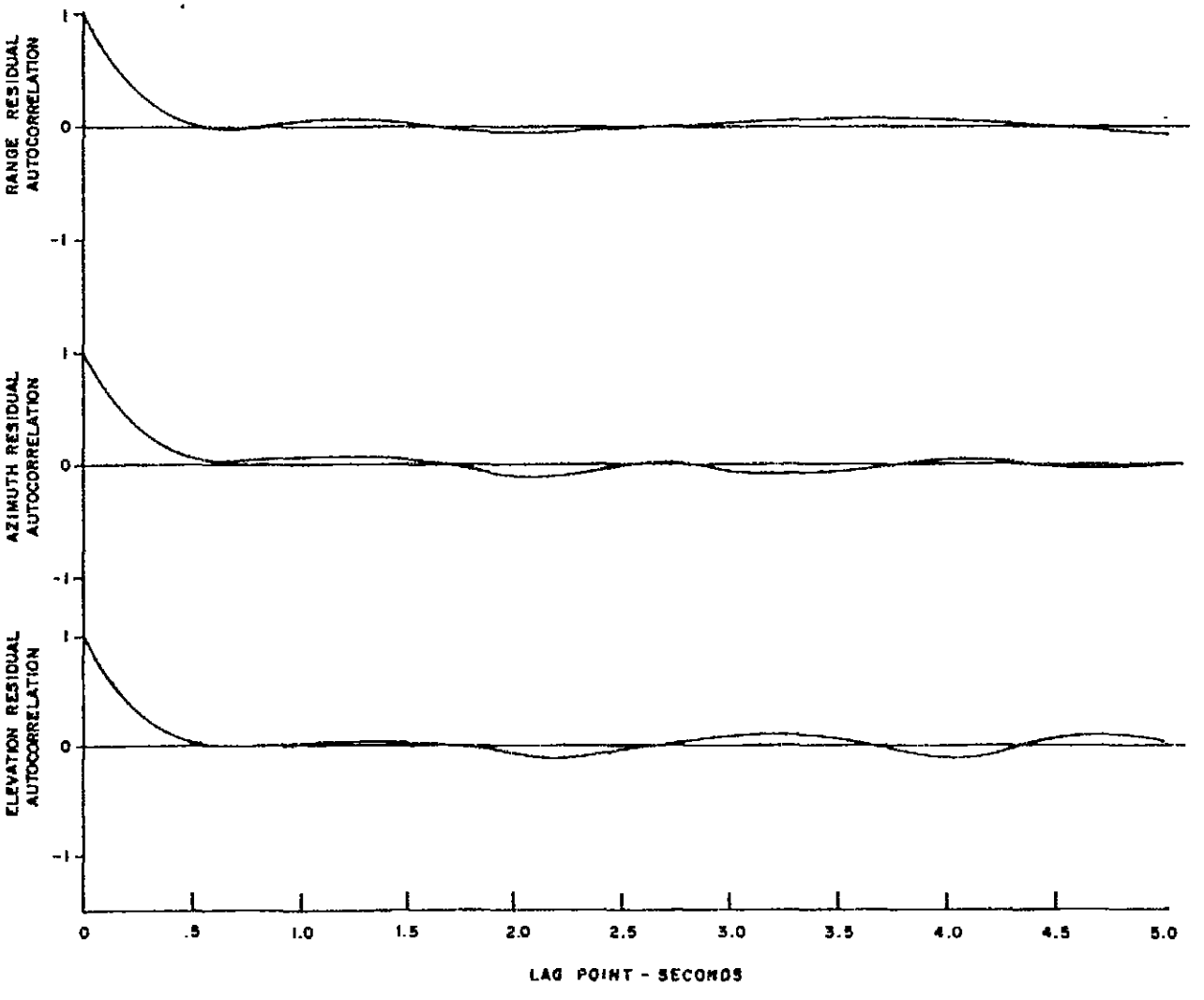


FIGURE 3 CORRELOGRAM OF AN/FPS-16 RANGE, AZIMUTH AND ELEVATION RESIDUAL DATA USING LAG OF .5 SECONDS.

4.0 METHODS USED TO ESTIMATE RANDOM ERRORS

From the previous discussion it is seen that random errors are estimated in terms of parameters which describe or characterize their behavior in terms of a probability distribution function. The parameters estimated are almost always in terms of the first and second sample moments corresponding to the residual distribution. If $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ is a set of errors, then the variance of the random error is expressed as

$$\sigma^2_{\epsilon_i} = E(\epsilon_i^2) - [E(\epsilon_i)]^2, \tag{4.0.1}$$

and if $e_1, e_2, \ldots e_n$ is a set of n residuals from a stationary time series, which estimates this set, then the random error is estimated by

$$\hat{\sigma}_e = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2} \tag{4.0.2}$$

where

$$\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i$$

Relation (4.0.2) is the sample standard deviation of the residual distribution. In general, the trajectory measurements themselves are changing and do not form a stationary process. However, the random error of the measurements is usually assumed to be stationary random process. This means that effective methods for removing nonstationary effects must be devised. Thus, the problem in estimating the random error is to be able to separate the nonstationary signal from the error, obtain a set of residuals, then separate the systematic error from the random error. This process is dependent on many factors including such elements as computational techniques, autocorrelation and cross correlation and assumptions concerning the analytic form of the systematic error model.

It is especially desirable to have a good estimate of the random errors for a particular system on a particular test as a function of time. This information is needed to settle such important questions as whether or not the successive errors in the measured quantity are correlated, what the optimum smoothing functions are, and what the best methods are for estimating velocity and acceleration data from position data.

Assuming that the autocorrelation and the cross correlation are zero, several methods of estimating random errors may be considered, and are given in Sections 4.1 through 4.6.

4.1 GRAPHICAL TECHNIQUES

In some instances graphical profiles of the residual distribution versus time are available for analysis. This is especially true of electronic measuring equipment where the servo feedback voltages are recorded graphically in analog form. In cases such as this, the residual profile can be inspected to see if there are systematic trends, etc. If the residual data is free of trends due to systematic error and correlation, one can select a fixed or variable sample rate and use relation (4.0.2) in estimating the random error.

If the residual is from a population characterized by the normal distribution then a much simpler method is available and is based on peak-to-peak or the sample range over a selected group or block of residuals. In selecting groups, we take k groups each containing n points. The ranges R_j which are peak-to-peak values for each jth group are estimated by

$$R_j = \max e_{ij} - \min e_{ij} \tag{4.1.1}$$

The average range $\bar{R}$ of the sample ranges is then computed, and the standard deviation is estimated from the relation

$$\tilde{\sigma} = a_n \bar{R} \tag{4.1.2}$$

where a_n is the ratio $\sigma/E(R)$ for the standard normal distribution and is available in tables of the distribution of the standardized range for a normal population (see Reference 23 for tables).

Since the peak-to-peak technique is relative to the "envelope" of the residual distribution for each group or subinterval, the size of the subinterval can be selected so that trends can be removed if they exist in the data.

The graphical techniques can be very useful for "quick look" analysis but have a drawback in that they usually involve much manual analysis. When using this technique one must be careful that "wild points" are omitted since the range technique is not as stable as estimating σ by relation (4.0.2). In addition, one must be certain that the serial correlation between sampled points is not high.

4.2 LEAST SQUARES POINT ESTIMATES OF RANDOM ERRORS

This method can be most easily used when a point-by-point least squares fit to a redundant data set is being used for reduction of the data.

Examination of the discrepancies between the solutions and the observed measurements will yield information regarding the distribution of errors in the measurements. By combining many of these residuals one may estimate their standard error. A comparison of this estimate with that predicted by the a priori statistics of the adjustment gives an approximation to the error in this a priori variance data. By repeating this procedure using new estimates for the a priori variance data, an improved figure for the data error may be found.

The sample standard deviation of the space position residual at each ith time point is

$$\theta_i = \sqrt{\frac{\sum_{j=1}^{n} \epsilon_{ij}^2}{kn - 3}} \qquad (4.2.1)$$

where

ϵ_{ij} is the residual distance at the ith time for the jth instrument
k = number of measured parameters (R, A, E, etc.)
n = number of instruments
3 = number of parameters being estimated (X, Y, Z).

The term kn-3 represents the number of degrees of freedom available in the estimate of the variance. In the case of an optical-only solution the degrees of freedom would be 2n-3, and 3n-3 for a solution using n radars. In the case where n_1 radars and n_2 optical instruments were used, we would have $3n_1+2n_2-3$ degrees of freedom.

Position error estimates in the required coordinate system are almost always correlated and as a result, the position error distribution is characterized by the estimate of the variance-covariance or "dispersion" matrix. The variance-covariance matrix estimate for the ith time point is

$$D_i = \begin{bmatrix} \sigma^2_{x_i} & \sigma_{x_i y_i} & \sigma_{x_i z_i} \\ \sigma_{y_i x_i} & \sigma^2_{y_i} & \sigma_{y_i z_i} \\ \sigma_{z_i x_i} & \sigma_{z_i y_i} & \sigma^2_{z_i} \end{bmatrix} = N_i^{-1} \theta_i^2 \qquad (4.2.2)$$

where N^{-1} is the inverse coefficient matrix corresponding to the normal equation obtained by computing the partial derivatives in the least square process.

Care must be taken in applying this method, because it is capable of detecting systematic or bias error as well as random error. An examination of the time-dependent behavior of the residuals should be made to determine whether or not apparent serial correlations or trends among these errors are due to systematic causes.

If corresponding residuals from point to point display "trends," residual systematic errors are indicated which often may be estimated and removed by techniques such as a "best estimate of trajectory" (BET) solution, a multiple regression analysis or numerically filtering over a series of points. After the trends or systematic error estimates have been removed, estimates of random error may be made utilizing the new residuals which correspond to the "unexplained" or random variations about the trend.

In cases where all observations have equal weight, the standard error in the observations may be estimated by (4.2.1). For the weighted case the quadratic form of the dispersion matrix of the residuals may be used to obtain an estimate of the unit variance. That is $\sigma_e^2 = \underline{e}' W^{-1} \underline{e}$ where e is a column vector, W a weight matrix. In this case the dispersion matrix would be

$$D_i = (T_i' \; W_i^{-1} \; T_i)^{-1} \; \hat{\sigma}_i^2 \qquad (4.2.3)$$

In (4.2.3) T is the "design" matrix, T' is its transpose and the coefficient matrix of the normal equations is $N = T'T$. The matrix W_i is the weight matrix. The consistency of this estimate may be established by comparing its diagonal terms with a value of the variances known a priori.

4.3 MULTI-INSTRUMENT ESTIMATE OF RANDOM ERROR

This method, sometimes called Simon-Grubbs, estimates random error using over determination from additional independent systems. In these computations (simplified) the mathematical model for three measurements is

$$y_{i1} = x_{i1} + e_{i1},$$

$$y_{i2} = x_{i2} + e_{i2},$$

$$y_{i3} = x_{i3} + e_{i3}.$$

where y_{ij} is the measured value of the ith characteristic x_{ij} with associated measurement error e_{ij} for the jth instrumentation system. Form the three differences:

$$\Delta_{1,2} = e_{i1} - e_{i2}$$

$$\Delta_{1,3} = e_{i1} - e_{i3}$$

$$\Delta_{2,3} = e_{i2} - e_{i3}.$$

The bias error is tacitly assumed to be some fixed value; the variance estimates for the above difference equations are computed about the mean difference. If the bias is changing, it adds components of variance estimates given below:

$$s^2_{1,2} = s^2_{e1} + s^2_{e2}$$

$$s^2_{1,3} = s^2_{e1} + s^2_{e3}$$

$$s^2_{2,3} = s^2_{e1} + s^2_{e3}$$

The above equations are solved for s^2_{e1}, s^2_{e2}, and s^2_{e3}, the respective error variance estimates for instrumentation systems 1, 2, and 3. The variance of the estimated error variance for instrumentation System One is given by:

$$\text{Var}\ (s^2_{e1}) = \frac{2s^2_{e1}}{(n-1)} + \frac{1}{(n-1)}\left(s^2_{e1}s^2_{e2} + s^2_{e1}s^2_{e3} + s^2_{e2}s^2_{e3}\right).$$

For measurement Systems Two and Three, similar expressions with the subscripts permuted are appropriate.

4.4 VARIATE DIFFERENCE METHOD

The variate difference method is a technique which uses the pth order successive difference to "detrend" data and estimate the variance of the random error about the trend. It can be shown that the coefficients a_i of the linear combinations for the pth difference follow the binomial expansion law and thus the sum of the a_i^2 for the pth difference is

$$\sum_{i=1}^{p+1} a_i^2 = \frac{(2p)!}{p!(2p-p)!} = \binom{2p}{p} \tag{4.4.1}$$

Since successive differences are an analog to the time derivative it is obvious that constant linear and quadratic trends will be removed from the measured data by a first, second and third order differencing scheme. In actual practice the third difference is usually sufficient and in almost all cases the third, fourth and fifth difference estimates give equivalent results.

If $X_1, \ldots X_n$ is a set of measurements or sample from a set of independent random variables with mean μ_x and variance σ_x^2 then the linear combination

$$Y_s = \sum_{i=1}^{n} a_{si} X_i \tag{4.4.2}$$

is a random variable with mean

$$\mu_{y_s} = \left(\sum_{i=1}^{n} a_{s_i} \right) \mu_x \tag{4.4.3}$$

and variance

$$\sigma_{y_s}^2 = \left(\sum_{i=1}^{n} a_{si}^2 \right) \sigma_x^2 \tag{4.4.4}$$

When taking successive differences, the resulting difference turns out to be a linear combination of the original random variable. Each difference gives a new random variable which can be expressed as a sum of the original measurements $\{ X_i \}$. For example, the first difference is

$$Y_{1i} = \Delta^1_{x_i} = X_i - X_{i-1}, \quad a_1 = 1, \; a_2 = -1$$

For the second and third differences we get

$$Y_{2i} = \Delta^2_{x_i} = X_{i+1} - 2X_i + X_{i-1}, \quad a_1 = 1, \; a_2 = -2, \; a_3 = 1,$$

$$Y_{3i} = \Delta^3_{x_i} = X_{i+2} - 3X_{i+1} + 3X_i - X_{i-1}, \; a_1 = 1, \; a_2 = -3, \; a_3 = 3, \; a_4 = -1$$

and so on.

From (4.4.4) and (4.4.1) the equation for the variance of the pth difference is

$$\sigma^2_{\Delta^p_x} = \binom{2p}{p} \sigma^2_x \tag{4.4.5}$$

Now the estimate of $\sigma^2_{\Delta^p_x}$ is

$$\hat{\sigma}^2_{\Delta^p_x} = \frac{1}{n-p} \sum_{i=1}^{n-p} \left(\Delta^p_{x_i}\right)^2, \tag{4.4.6}$$

so that

$$\hat{\sigma}^2_x = \frac{\sum_{i=1}^{n-p} \left(\Delta^p_{x_i}\right)^2}{\binom{2p}{p} (n-p)} \tag{4.4.7}$$

where

$\Delta^p_{x_i}$ = pth successive difference in X_i

n = number of data points in sample

p = order of the successive difference

The variate difference method is widely used since it can be applied to data when no knowledge concerning the trend in the data is available. However, caution must be used because of the assumptions which were made concerning independence of the errors. It should be emphasized that if the error is highly correlated, then the error variance estimates will be biased, and in fact usually underestimated (see Section 4.7). In addition to the correlated effects on the variate difference method the technique is also highly unstable if "wild points" or outlyers are sampled. One method of compensating for autocorrelation in the data is to compute the differences using a larger time spacing Δt between the measured X_i's. However, one must be careful that when large time increments are used, the effects of aliasing do not bias the variance estimates. A more satisfactory technique is to estimate the autocorrelation function and use this to correct the biasing effects on (4.4.7). The amount of biasing caused by autocorrelation is discussed in Section 4.7.

Since the variate difference method is based on differencing it is a useful tool in examining the frequency content of pth order differences obtained. If Δt is the time interval between successive data points, the maximum distinquishable frequency is $(1/2\Delta t)$ cycles per second and is called the Nyquist frequency. Higher order differences around the order of 4, 5 or 6 effectively examines the error contributions of frequencies between $0.82/2\Delta t$ to the Nyquist frequency. This is due to the fact that frequencies below $0.82/2\Delta t$ have been removed by differencing. If we can assume the noise is uniform over all frequencies, we may estimate the component of the variance in the interval $.82/2\Delta t$ to $1/2\Delta t$ by dividing the variance obtained using the variate difference method by four. The frequency range of $0.82/2\Delta t$ to $1/2\Delta t$ is based on the frequency response of the variate difference method. For a fourth order difference the weights are (1, -4, 6, -4, 1). If we take the Fourier transform of these weights we obtain the frequency response as given in Figure 4.

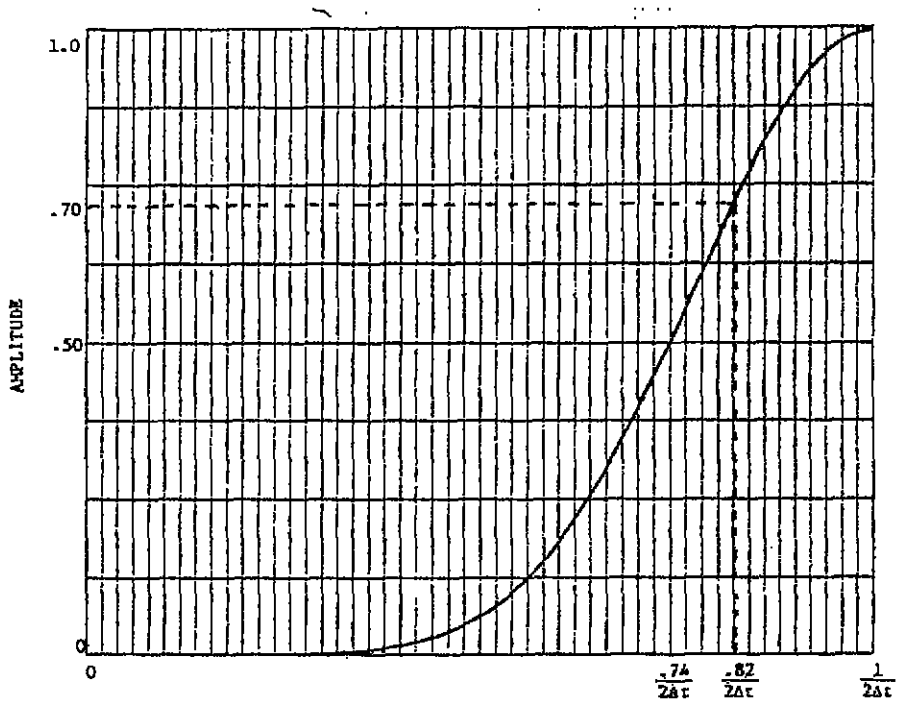


Figure 4 - Frequency response for fourth order variate difference method with lag Δt between successive points.

4.5 THE LEAST SQUARE CURVE FITTING METHOD

In the least square method a polynomial is fitted to the data points by any standard least squares curve fitting technique.

In particular, if a series of measurements, X_1, X_2,...,X_n of a parameter such as a rectangular coordinate are made as a function of time and a polynomial of degree p is fitted to this data and if the corresponding points on the curve are $X'_1, X'_2,...,X'_n$, then:

$$\sigma_x = \sqrt{\frac{1}{(n - p - 1)} \sum_{i=1}^{n} (X_i - X'_i)^2} \tag{4.5.1}$$

is an estimate of the standard deviation of the random error in the measurements $X_1, X_2 ..., X_n$.

If p, the degree of the polynomial, is small compared to n, the number of points, the value of σ_x approximates the usual definition of RMS error, i.e.,

$$\text{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - X'_i)^2} \tag{4.5.2}$$

Estimated standard deviations are frequently plotted against the time of the mid-point of the span, and significant information concerning the variation of random errors with time or geometry can be obtained from these plots.

If the error is serially correlated then the estimates of the variance of the random error will be biased depending on the extent of the correlation. This is discussed in more detail in Section 4.7.

However, in estimating the errors by either the variate difference and/or least square curve fitting methods, it is usually assumed that autocorrelation and cross correlation are zero or very nearly so. In actuality, the following possible conditions may exist in data acquired by range instrumentation:

1. The errors in the measurements X, Y, Z at each point in time are not independent-cross correlation,

2. The errors in successive trajectory points are not mutually independent-serial or autocorrelation,

3. All measurements of a given type do not possess a common variance-heteroscedastic (versus homoscedastic),

4. Time measurements are not error free-time error.

Whenever one or all of the above conditions exist in the acquired data, more advanced and sophisticated techniques are needed to estimate the random errors in the data.

4.6 USE OF A DIGITAL HIGH PASS FILTER

In Section 3.0 random error was characterized in the frequency domain as comprising the higher frequency portion of the power spectrum of the error distribution. If this is the case, then frequency-constraining digital filters can be constructed which will effectively separate the high-frequency components from the low at a designated cut-off frequency. This technique is often more desirable than using a least square polynomial to smooth the data and then subtracting the smooth data from the raw measurement. The reason for the desirability of the digital filter over the least-square polynomial is that the digital filter can be designed so that its frequency response has a sharper roll off with less "leakage" or side lobe effect so that undesirable frequencies do not pass through the filter. Examples of second-degree polynomial frequency responses over 31 and 21 points and digital filters with approximately the same cut-off frequencies are given in Figure 5. It is noted from the graph that the high-frequency "leakage" of the polynomial fits is as high as 25 per cent. The corresponding leakage for a low-pass digital filter over 25 points with a cutoff of 0.05 cycles per unit time is 2 per cent.

When the residual distribution is obtained from a high-pass numerical filter, the standard deviation and RMS of the distribution can be estimated using relations (4.5.1) and (4.5.2) where p is the highest degree of a polynomial which will pass through the corresponding low-pass filter without change.

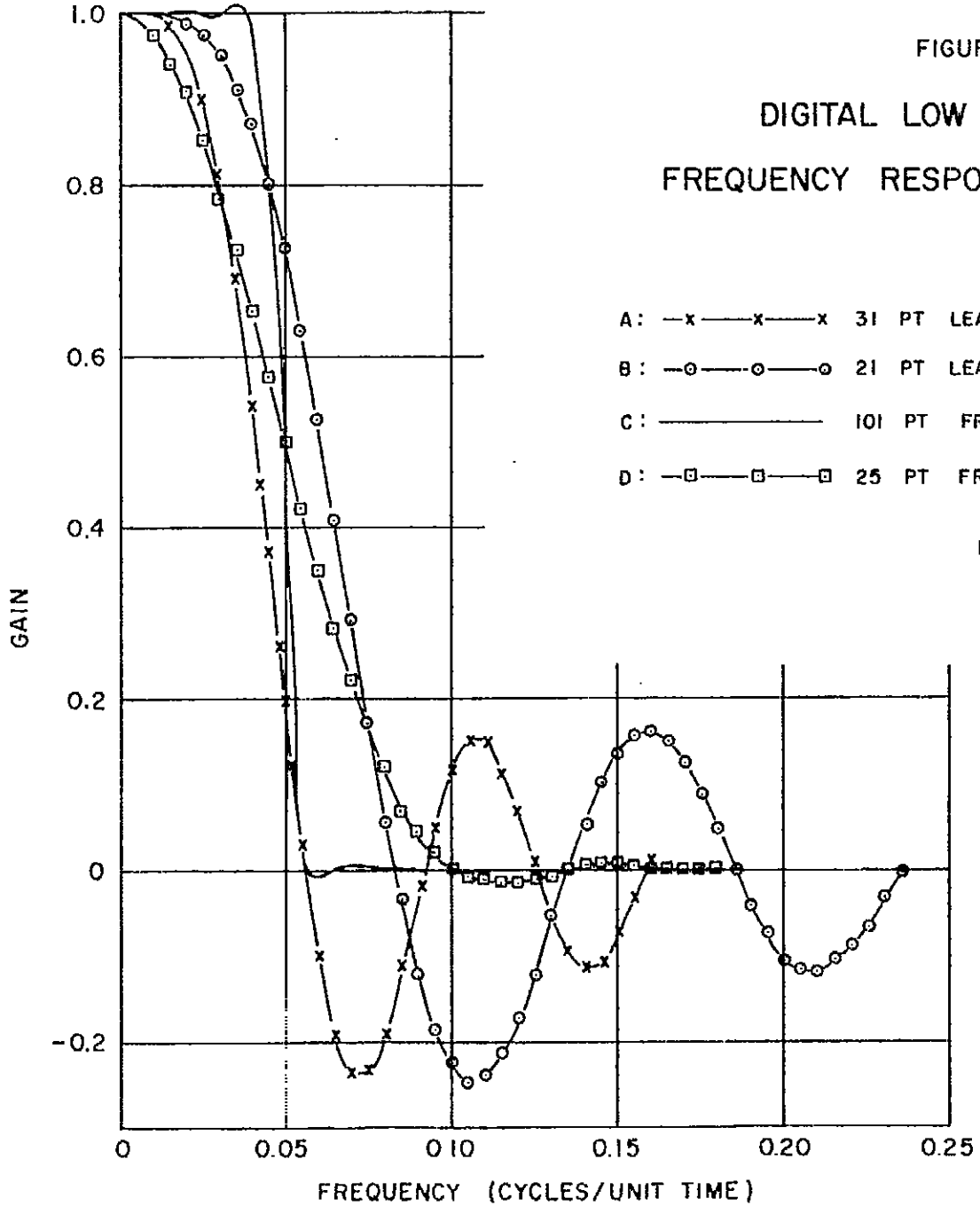


FIGURE 5

DIGITAL LOW PASS FILTER

FREQUENCY RESPONSE COMPARISONS

A: —x——x——x 31 PT LEAST SQUARE QUADRATIC f_c = .04

B: —⊙——⊙——⊙ 21 PT LEAST SQUARE QUADRATIC f_c = .06

C: ———————— 101 PT FREQ CONSTRAINED FILTER f_c = .05

D: —⊡——⊡——⊡ 25 PT FREQ CONSTRAINED FILTER f_c = .05

MAXIMUM LEAKAGE: A — 24 %
B — 25 %
C — 1 %
D — 2 %

4.7 EFFECTS OF AUTOCORRELATION OF THE ESTIMATION OF RANDOM ERRORS

In Section 3.3 it was pointed out that high autocorrelation makes the random error appear smoother and can lead one to false conclusions concerning reduction techniques. This is due to the fact that high correlation may appear as signal information. Because of this trait, autocorrelated random errors will bias the data, and the amount of bias usually depends on how high the correlation is.

To illustrate how autocorrelation affects data reduction and random error estimation, the different random errors illustrated in Figure 2 were superimposed on a known quadratic function, $a_0 + a_1 t + a_2 t^2$, with $a_0 = -1, a_1 = -1, a_2 = 1$, and the variance of each error distribution was $\sigma^2 = 0.01$. The results of approximating the quadratic signal with uncorrelated random error are shown in Figure 6 while the results with high autocorrelated error are given in Figure 7. Each of these figures displays a complete picture of the analysis. The first graph displays the predicted curve (line) with the raw data (signal plus noise) on the same graph as circles. The second graph displays the residuals. The information on the bottom of the figures describes how good the approximation is. A comparision of the two figures is summarized in the following table

TABLE I

Auto Correlation (rho)	Percent Error a_0	a_1	a_2	Standard Error σ_{a_0}	σ_{a_1}	σ_{a_2}	Variance Estimate	Best Degree Fit
0	1.852	3.808	1.474	.07311	.04750	.02310	.0095	2
.99	20.687	29.893	11.406	.05548	.02418	.01176	.0025	10

Table I shows that highly correlated error gives invalid results in every area. The error in the regression coefficients with rho = .99 is much higher and the estimates of the standard errors in the regression coefficients are grossly underestimated and are not consistent with the actual errors. With rho = 0 the variance estimate is very close (.0095 versus 0.01) while the variance estimate for rho = .99 is underestimated. It is noted that the best polynomial fit was correct with uncorrelated noise. However, when rho = .99 was present, the best fit was a polynomial of degree 10. The example given has shown that data containing random error with high autocorrelation can cause the following adverse effects.

1. Often makes data appear smoother.

2. May appear as signal or information.

3. Biases estimates of regression coefficients.

4. Biases estimates of variance of error distribution.

5. Biases estimates of variance of regression coefficients.

6. Biases "goodness-of-fit" tests.

The effects of correlated error on estimates of the variance when the variate difference method is used is also adverse. High correlation causes the variance estimates to be low. If the autocorrelation is a first order variate difference on each point to eliminate a quadratic trend, it can be shown that the variance estimate will be

$$\delta^2 = (1 - \rho)(1 - .5\rho + .1\rho^2)\ \sigma^2 \tag{4.7.0}$$

where

ρ = correlation coefficient of lag 1

σ^2 = variance of error distribution

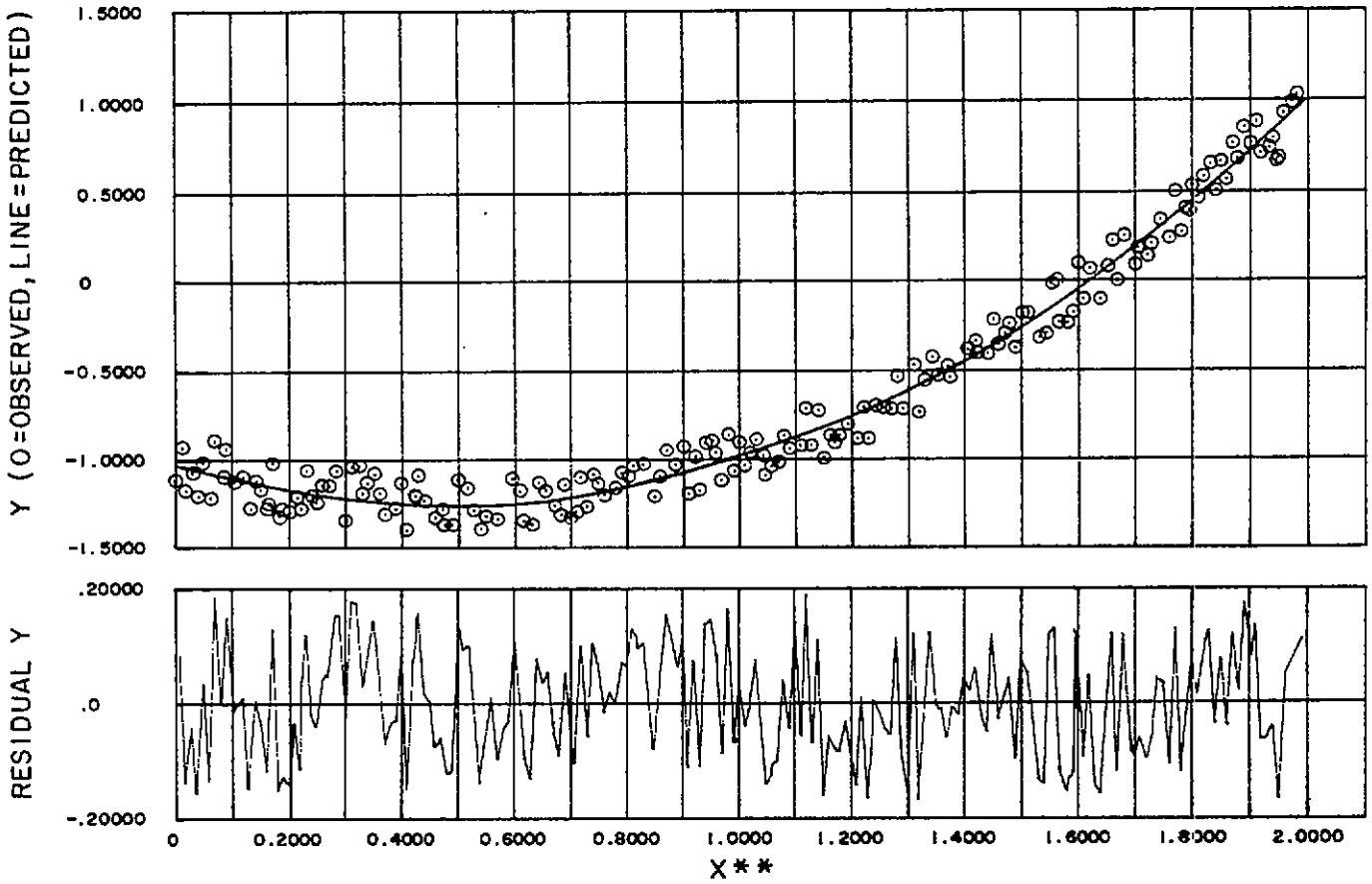
If we use the same technique on every other point (i.e., with a lag of 2) then the variance estimate is

$$\delta^2 = (1 - \sigma^2)(1 - .8\rho + .2\sigma^2)\ \sigma^2 \tag{4.7.1}$$

The results of (4.7.0) and (4.7.1) are illustrated in Figure 8. Note from the figure that with a lag of 1 (every point) and with $\rho = 0.9$, the variance of the error is underestimated by a factor of 0.0631.

It is stressed that there are techniques which may be used to estimate and compensate for autocorrelated error. If the correlation is known a priori, it can be compensated for by "extended" or weighted least-square techniques which utilize the inverse of the autocovariance matrix to obtain a weighted estimate which compensates for correlation. If the autocovariance/correlation is not known, it can be approximated by analysis of residuals from first estimates. An iterative approximation process can be set up to obtain updated estimates of the autocovariance functions which are used for updated estimates of the regression coefficients. Information regarding the autocorrelation function may often be obtained by establishing a history of data analysis from previous operations.

BMD05R OUTPUT OPERATION NO. 999999 RUN NO. 0
CORR. NOISE p1 = 0 DATE JOB NO. 706222



POLYNOMIAL REGRESSION OF DEGREE 2

CASE NO. 1
SUBCASE NO. 0

INTERCEPT (A VALUE) 1.01652-

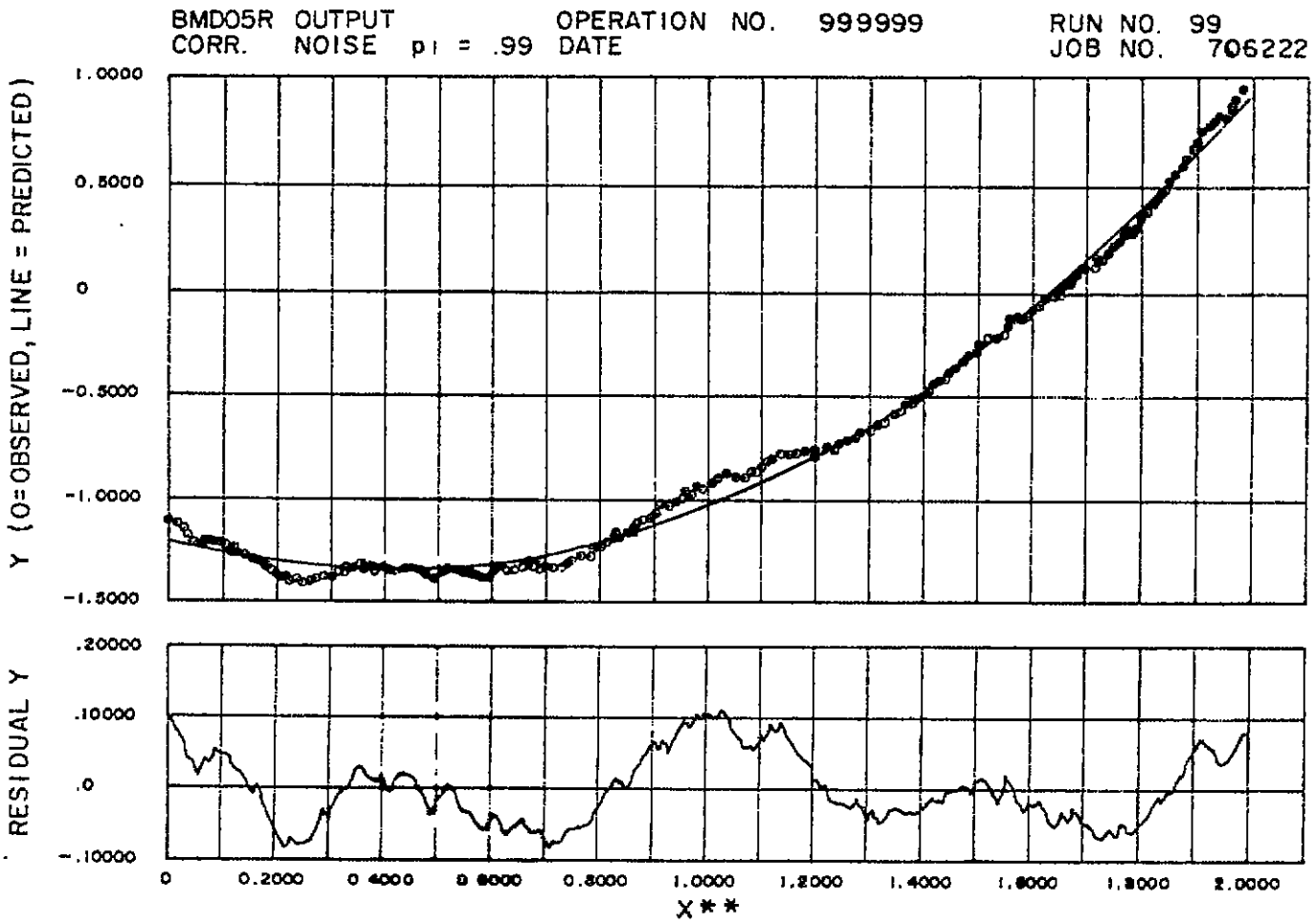
REGRESSION COEFFICIENTS
.96192- .98526

STANDARD ERROR OF REGRESSION COEFFICIENTS
.04750 .02310

ANALYSIS OF VARIANCE FOR POLYNOMIAL OF DEGREE 2

SOURCE OF VARIATION	DEGREE OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F VALUE
DUE TO REGRESSION	2	83.75314	41.87657	4414.24670
DEVIATION ABOUT REGRESSION ...	197	1.86888	.00949	
TOTAL ...	199	85.62202		

FIGURE 6 Least square regression on $y(t) = -1 - t + t^2 + N$ where N is uncorrelated noise with mean zero and variance $\sigma^2 = 0.01$.



POLYNOMIAL REGRESSION OF DEGREE 2

CASE NO 1
SUBCASE NO. 0

INTERCEPT (A VALUE) 1.20687-

REGRESSION COEFFICIENTS

.70107- .80594

STANDARD ERROR OF REGRESSION COEFFICIENTS

.02418 .01176

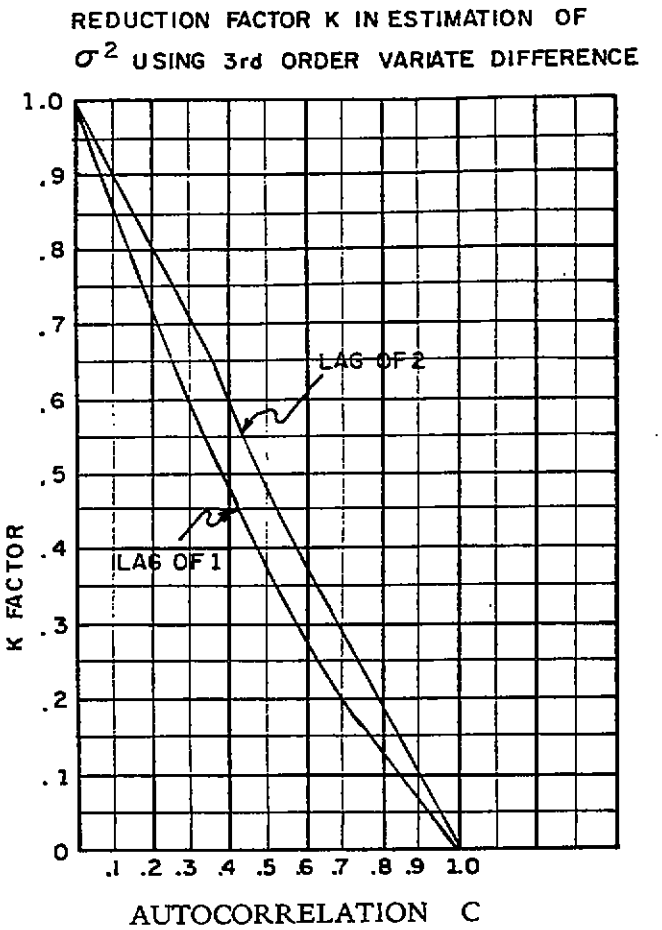
ANALYSIS OF VARIANCE FOR POLYNOMIAL OF DEGREE 2

SOURCE OF VARIATION	DEGREE OF FREEDOM	SUM OF SQUARES	MEAN SQUARE	F VALUE
DUE TO REGRESSION	2	89.13169	44.56584	18127.47094
DEVIATION ABOUT REGRESSION .	197	48432	.00246	
TOTAL ..	199	89.61601		

FIGURE 7 Least square regression on $y(t) = -1 - t + t^2 + N_t$

$N_t = .99\, N_{t-1} + \eta_i$ is highly correlated with $p = .99$, mean = 0 and variance $\sigma^2 = 0.01$

FIGURE 8


REDUCTION FACTOR K IN ESTIMATION OF
σ^2 USING 3rd ORDER VARIATE DIFFERENCE
1.0
.9
.8
.7
.6
.5
.4
.3
.2
.1
0
K FACTOR
LAG OF 2
LAG OF 1
.1 .2 .3 .4 .5 .6 .7 .8 .9 1.0
AUTOCORRELATION C

5.0 INTERVAL ESTIMATION OF RANDOM ERRORS

In the previous section, techniques for point estimation of the variance, standard deviation or root mean square of the random error distribution were discussed. It should be noted that the estimates of these parameters are themselves random variables and have corresponding probability distribution functions. This fact points out that the use of sample random variables (or residuals in the case of error analysis) involves the concept of confidence, tolerance and prediction regions. In the case of one-dimensional random variables, these regions are intervals on the real line. The regions corresponding to higher-dimension random variables will be discussed in Section 8.0 after the covariance matrix has been discussed.

Interval estimation is an important concept and, in general, if we are given a sample of n residuals $e_1 ...,e_n$ from the population $\epsilon_1 ,...,\epsilon_n$, the problem is to find two statistics (L, U) defining an interval which has $100\alpha\%$ probability of containing $100\gamma\%$ of the individual values in the population. Such an interval is called a tolerance interval with lower and upper tolerance limits L and U. The value of α, $0 \leqslant \alpha \leqslant 1$, is called the confidence coefficient. For example, if $\alpha = .90$ and $\gamma = .95$, we are 90% confident that at least 95% of the individual values will be in the interval (L,U).

Besides interval estimation for individual values of the population, we are interested in interval estimates for the parameters, θ, which define the probability density function of the population. We want to find statistics $L(\theta)$ such that

$$\text{Prob}\,(L(\theta) < \theta < U(\theta)) = \gamma \qquad (5.0.1)$$

where θ would be a parameter, such as the mean or variance of the distribution. In this case, the interval $L(\theta)$, $U(\theta)$ is called a confidence interval for the true (but unknown) value of the parameter. Another interval estimate is the prediction interval. In this case, we construct an interval that has the preassigned probability of containing the next observation.

5.1 THE CASE OF THE NORMAL DISTRIBUTION

It is often assumed that the random error ϵ_t is a random variable which has the normal or Gaussian distribution. This assumption may be based on experience or used because it is well known, convenient, and gives approximately correct results. It is pointed out that in the real world many random variables are not normally distributed, and this includes random errors. An example of an error which is uniformly distributed (rather than normally distributed) is round-off error. However, the importance of the normal distribution in error analysis is recognized and some of its basic uses are considered in this document.

A figure of the standard normal distribution (i.e., $\mu = 0$, $\sigma^2 = 1$) is given in Figure 9. If e_t is a sample from the population ϵ_t which corresponds to the normal distribution with mean μ and variance σ^2 (assumed known in this instance), then tolerance intervals of the form $\hat{\mu} \pm k\sigma$ corresponding to $\alpha = 1$ (absolute certainty) can be constructed. The values of k corresponding to the more widely used values of γ are given in the following table

AREAS (100γ%) UNDER THE NORMAL CURVE IN THE INTERVAL $\mu \pm k\sigma$

k	γ
.6745	.5000
.7979	.5751
1.0000	.6827
1.6449	.9000
2.0000	.9545
3.0000	.9973

These particular values of k and γ correspond to the following error quantities

.6745σ: probable error PE
.7979σ: mean absolute error ME
σ: standard deviation
1.6449σ: map accuracy standard MAS)
2σ: significant error
3σ: highly significant or near certainty error

Another important quantity is the standard error of the mean, or standard error of estimate. If the mean and variance of a sample are estimated from n points, then he standard error is $\sigma/\sqrt{n}$.

$$Y = \frac{1}{\sqrt{2\pi}} e^{-X^2/2}, \; u = 0, \; \sigma = 1$$

$X = Z/\sigma$

σ = Standard Deviation

Z = Actual Deviation

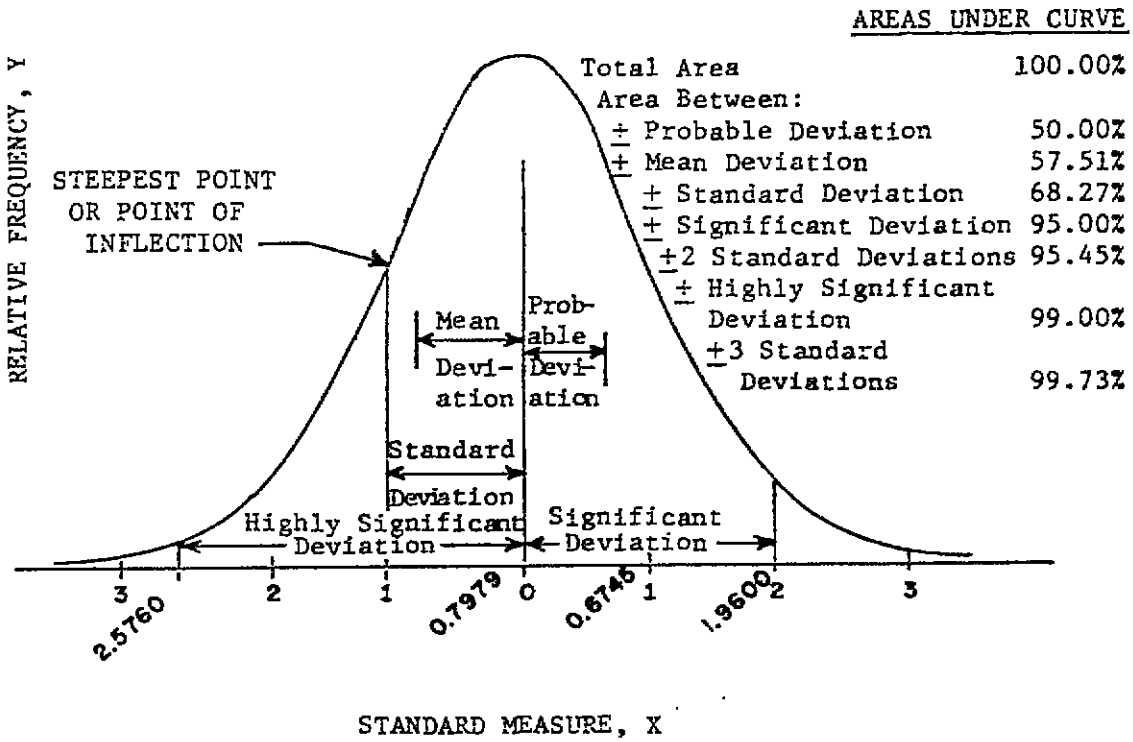


FIGURE 9. Standard Normal Frequency Distribution Curve.

5.2 CONFIDENCE INTERVALS ON ERROR PARAMETER ESTIMATES

Suppose $e_1,\ldots,e_n$ is a sample observed from the population $\epsilon_t \sim$ normal (μ,σ^2) where the population parameters μ and σ^2 are unknown. Assuming that the observations are independent, then μ and σ are estimated by the sample mean $\bar{e}$ and sample standard deviation s_e, where

$$\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i,$$

(5.2.1)

$$s_e = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (e_i - \bar{e})^2}$$

$\bar{e}$ and s_e are random variables with $\bar{e} \sim$ normal $(\mu,\sigma^2/n)$ and the quantity

$$Z = \frac{(n-1)s_e^2}{\sigma^2} \tag{5.2.2}$$

has the Chi-square distribution with n - 1 degrees of freedom (i.e., Z has X^2_{n-1} distribution). It can further be shown that the statistic

$$t = \frac{\bar{e} - \mu}{s_e/\sqrt{n}} \tag{5.2.3}$$

has the "student" T-distribution with n - 1 degrees of freedom. From relation (5.2.3) we can obtain a confidence interval for μ by

$$\text{Prob}\left(t_{n-1,\left(\frac{1-\gamma}{2}\right)} < \frac{\bar{e} - \mu}{s_e/\sqrt{n}} < t_{n-1,\left(\frac{1+\gamma}{2}\right)} \right) = \gamma$$

or (5.2.4)

$$\bar{e} \pm t_{n-1,\left(\frac{1+\gamma}{2}\right)} \frac{s_e}{\sqrt{n}}$$

is a $100\gamma\%$ confidence interval on μ, where t depends on the confidence coefficient α and the degrees of freedom υ which in the case of n observations correspond to (n - 1). When $t_{\upsilon,\alpha} = 1$, then the confidence interval is $\bar{e} \pm s_e/\sqrt{n}$ and the quantity $s_e/\sqrt{n}$ is called standard error of the mean or just "standard error." Values for $t_{\upsilon,\alpha}$ can be obtained from tables in most statistical texts.

In a similar manner we can get a confidence interval on σ utilizing (5.2.2). From (5.2.2) we get

$$\text{Prob}\left[\chi^2_{\left(\frac{1-\gamma}{2}\right)} < \frac{(n-1)s_e^2}{\sigma^2} < \chi^2_{\left(\frac{1+\gamma}{2}\right)}\right] = \gamma \tag{5.2.5}$$

which leads to a $100\gamma\%$ confidence interval on σ^2 which is

$$\frac{(n-1)s_e^2}{\chi^2_{\left(\frac{1+\gamma}{2}\right)}} < \sigma^2 < \frac{(n-1)s_e^2}{\chi^2_{\left(\frac{1-\gamma}{2}\right)}} \tag{5.2.6}$$

Values of χ^2_α can be obtained from Chi-square tables in statistical texts.

6.0 METHODS USED TO EVALUATE SYSTEMATIC ERRORS

The problem of evaluating systematic errors in a given instrumentation system is considerably more difficult than that of evaluating random errors, unless a suitable comparison standard is available. Unfortunately, the systematic errors can be several orders of magnitude greater than the random errors and no absolute comparison standard exists. There are, however, several approaches which can be used to evaluate systematic errors.

6.0.1 ANALYTICAL INVESTIGATION OF POSSIBLE ERROR SOURCES

This involves representing the system by as comprehensive a mathematical model as possible. Reasonable perturbation of each of the model parameters is introduced and the resultant effects of this on the data are calculated. From this analysis may be obtained an estimate of the likely range of systematic errors. This approach is limited by the store of scientific knowledge, which makes it extremely difficult to construct a sufficiently comprehensive mathematical model of the system. This would appear to be especially true of propagation anomalies which particularly affect electronic systems.

6.0.2 COMPARISON OF OBSERVATIONS FROM DIFFERENT SYSTEMS

If it is known that System I is significantly more accurate in an absolute sense than System II, then an estimate of the systematic error in II can be made by noting the discrepancies between observations of the two systems. The estimate of the total error in System II is made by computing the difference

$$\Delta = (\text{System II} - \text{System I})$$

at each time point. The difficulty with this approach lies in selecting a suitable standard. Emphasis must be placed upon the fact that upper bounds on possible systematic errors in the standard must be known, and these bounds must be narrow compared with reasonably lower bounds for the systematic error in the system being investigated. Thus, for the calibration standard it is absolute accuracy, not relative or internal accuracy, which is of prime importance. This fact makes it unprofitable to compare systems indiscriminately. Also, it is necessary to employ a mathematical model of the system being treated in order to transform the standard data and propagate any error in the standard into the tested system.

The observations collected and used for error analysis may be from actual operational tests or from tests which are specifically designed for the purpose of error analysis. Tests of the latter type are called calibration tests. The most simple, economical, and frequently used calibration technique is the "static" calibration test. This utilizes calibration equipment and objects as "line-ups," boresight towers, and stars. In most instances, however, the static calibration test is not effective in estimating the many errors encountered during an actual operation. The reason for this is that most of the systematic error components are dynamic in nature and depend upon variables such as distance, tracking rates, atmospheric conditions (temperature pressure, humidity, etc.), and the direction in which the instrumentation system is pointing. As a result, dynamic calibration tests utilizing a good test design and a standard with sufficiently small error provides a better estimate of the systematic error.

6.0.3 ANALYSIS OF MEASURING RESIDUALS FROM OVER–DETERMINED LEAST SQUARES SOLUTIONS

When a given instrumentation system provides a redundancy of data, a least square solution may be performed. The adjustment leads to a set of measuring residuals (or corrections) for each trajectory point. The criterion for adjustment is the principle of maximum likelihood. If these residuals indicate, from point-to-point, a randomness about zero, there is good indication that the systematic error is small. If the residuals do not display this randomness about zero, one must conclude that the original observations are biased or have systematic error. A total error computed from the residuals gives an estimate of the upper limit of the systematic error in the observations.

6.0.4 ANALYSIS OF RESIDUALS FROM A BEST ESTIMATE OF TRAJECTORY

When there is simultaneous track with more than one instrumentation system, a Best Estimate of Trajectory (BET) can be computed. This trajectory and the residuals of the individual systems referenced to it can then be analyzed to determine estimates of the systematic errors. The BET is a powerful method of estimating systematic errors in both the reduced data and the measured parameters. There are several methods of obtaining a BET, depending upon the criteria for test. The most powerful of the methods developed to date involves the use of mathematical models of the errors of the individual systems. In all the applications of BET to error studies, the basic idea is that, according to the criterion agreed upon, a best estimation of what the observations should have been is made and then compared with the observed data.

6.0.5 MULTILATERATION

As in the case of BET, when redundant data from several instruments is available a multilateration technique which uses inputs from several different sources such as doppler, pulse radar data, and inertial guidance data can be used. This technique has the error model design as part of the solution and attempts to "weight out" the effects of error sources by associating each source with an uncertainty which is estimated from a priori testing and is updated by the multilateration solution. If error sources are correlated, they adversely affect the solution and must be taken out, using estimates from calibration performed prior to the test. Once this is done, estimates of the magnitudes of all the error model coefficients are made simultaneously employing a maximum likelihood technique, which is mechanized using the Kalman filter approach.

6.1 ERROR MODELS

In Section 3.0 it was stated that systematic errors are deterministic. This means the systematic error S(t) can be represented analytically in terms of a mathematical model. Since the random error is nondeterministic, it must be represented by some probability model. In general, one may set up error models which fall into the following categories:

I. Deterministic

Parametric estimation—an analytic model which describes the error in terms of parameters which are the components of the error.

Nonparametric estimation—a numerical model which approximates the error numerically, i.e., as a sequence of numbers such as the output from a digital filter.

II. Nondeterministic

Parametric estimation—a stochastic equation depending upon parameters which are random variables, e.g., the expression of a random process by a pth order Markov Process. If the process is stationary it can be characterized in terms of parameters which index its probability density function.

The most simple systematic error model would be

$$S(t) = \mu \tag{6.1.1}$$

In this case the systematic error is a constant bias, "offset" or "zero set" error. If S(t) is estimated by a sequence of n Δj's, then

$$\mu \triangleq \bar{\Delta} = \frac{1}{n} \sum_{j=1}^{n} \Delta j \tag{6.1.2}$$

The Δj's also contain random error and the averaging in (6.1.2) separates μ from N(t). The estimate of the random error parameter would be

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (\Delta j - \bar{\Delta})^2}. \tag{6.1.3}$$

If the error components are independent, then the error model can be expressed as a linear combination of independent variables, i.e.,

$$\varepsilon_i = \sum_{j=1}^{k} \alpha_j x_i + \eta_i \tag{6.1.4}$$

The error model given in (6.1.1) can be expanded to include terms which describe various error components. A good example is a typical error model frequently used for the AN/FPS-16 radars. The error models for range, azimuth, and elevation are given below.

$$\Delta R_j = a_0 + a_1\dot{R}_j + a_2\ddot{R}_j + a_3\dddot{R}_j + a_4 f(S/N) + a_5 R_j + a_6 R_j\dot{R}_j \qquad (6.1.5)$$

$$+ a_7 \csc E_j + a_8 \sin \frac{2\pi}{6000}\dot{R}_j + a_9 \cos \frac{2\pi}{6000}\dot{R}_j + a_{10}\frac{X_j}{R_j}$$

$$+ a_{11}\frac{y_j}{R_j}\, a_{12}\frac{z_j}{R_j} + N_j$$

where

a_0 = constant bias
a_1 = timing delay
a_2 = acceleration servo lag
a_3 = jerk servo lag
a_4 = beacon delay
a_5 = oscillator drift or scale factor
a_6 = time dilation
a_7 = residual refraction
a_8 = resolver nonlinearity
a_9 = resolver nonlinearity
a_{10} = x survey error
a_{11} = y survey error
a_{12} = z survey error
N_j = random error

$$\Delta A_j = b_0 + b_1\dot{A}_j + b_2\ddot{A}_j + b_3\dddot{A}_j + b_4 \sin A_j \tan E_j + b_5 \cos A_j \tan E_j \qquad (6.1.6)$$

$$+ b_6 R_j\dot{A}_j + b_7 \sec E_j + b_8 \tan E_j + b_9 \sin A_j + b_{10} \cos A_j$$

$$+ b_{11}\frac{1}{R_j}\cos A_j \sec E_j + B_{12}\frac{1}{R_j}\sin A_j \sec E_j + N_j$$

where

b_0 = constant bias
b_1 = timing bias
b_2 = acceleration servo lag
b_3 = jerk servo lag
b_4 = mislevel
b_5 = mislevel
b_6 = time dilation
b_7 = collimation
b_8 = nonorthogonality
b_9 = encoder nonlinearity
b_{10} = encoder nonlinearity
b_{11} = x survey
b_{12} = y survey
N_j = random error

$$\Delta E_j = c_0 + c_1 \dot{E}_j + c_2 \ddot{E}_j + c_3 \dddot{E}_j + c_4 \sin A_j + c_5 \cos A_j + c\, R_j \dot{E}_j \qquad (6.1.7)$$

$$+ c_7 \cot E_j + c_8 \sin E_j + c_9 \cos E_j + c_{11} \frac{1}{R_j} \sin A_j \sin E_j$$

$$+ c_{12} \frac{1}{R_t} \cos A_j \sin E_j + c_{13} \frac{1}{R(t)} \cos E_j + N_j$$

where

c_0 = constant bias
c_1 = timing delay
c_2 = acceleration servo lag
c_3 = jerk servo lag
c_4 = mislevel
c_5 = mislevel
c_6 = time dilation
c_7 = residual refraction
c_8 = encoder nonlinearity
c_9 = encoder nonlinearity
c_{10} = x survey
c_{11} = y survey
c_{12} = z survey
N_j = random error

In the linear error models the dependent variable is the total error residual, and the independent variables are usually assumed to be uncorrelated and are determined by the physics or geometry of the situation. The independent variables are assumed to be known without (or at least with negligible) error. Since the independent variables depend on the physics and geometry of a particular test for a certain instrumentation system, it is important to stress that the error model depends not only on analysis of the physics involved, but on the design of the test itself. For example, if one had a near-perfect standard for comparison for obtaining good error estimates, one could not obtain a valid estimate of the mislevel error coefficients if the test design failed to enable the instrumentation system to traverse less than 180° in azimuth. Also, one could not estimate scale factor error if the range remained approximately constant. It is obvious that in order to estimate dymanic errors one must have a corresponding dynamic test design.

One proposed test for dynamic error estimation involves the use of a calibration satellite. This type of vehicle would allow tracking in virtually every quadrant over a wide variety of rates, look angles, and positions. Further, its trajectory can be predicted very accurately using the equations of motion as a model. The orbital parameters can be determined by BET techniques, using world-wide networks of instrumentation systems. Many problems would still exist in a program of this type, such as unmodeled parameters in the orbital model, nonuniformity in the earth's gravitational field, etc. However, during certain satellite passes and combinations of passes, the effects of such errors can be minimized. For further detailed information on the use of a calibration satellite, see Reference 24.

6.2 REGRESSION ANALYSIS

The linear error models discussed in the previous section are generally approximated utilizing multiple linear regression techniques. The techniques for regression analysis are discussed in this section and the techniques for evaluating the validity of the analysis are discussed in Section 6.3, Analysis of Variance.

When an acceptable standard is available and $e_1, e_2, \ldots, e_n$ obtained they are assumed to be a realization from the process

$$\varepsilon_i = \alpha_o + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \ldots + \alpha_k X_{ik} + \eta_i \tag{6.2.1}$$

$i = 1, 2, \ldots, n$. The α's are unknown parameters and η_i is the random error. η is assumed to be a stochastic process whose elements are homoscedastic, independent, with mean $\mu_\eta = 0$ and variance σ_η^2. In matrix form (6.2.1) is

$$\underline{\epsilon} = X\underline{\alpha} + \underline{\eta} \tag{6.2.2}$$

where $\underline{\epsilon}$ and $\underline{\eta}$ are n x 1 column vectors. $\underline{\alpha}$ is a (k+1) x 1 column vector and X is an n x (k+1) matrix of known elements

$$X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \tag{6.2.3}$$

The matrix X consists of the n values of the k independent variables and is called the design matrix. In the analysis and design of experiments, the independent variables are called factors, the dependent variable (in this case the error) is called the response variable or yield. The n values of the independent variables are called levels.

In the standard least squares analysis of (6.2.2), estimates $\underline{a}$ are made of the parameters α such that the sum

$$S = \sum_{i=1}^{n} [e_i - a_o - a_1 X_{i1} - \ldots - a_k X_{ik}]^2 \tag{6.2.4}$$

is minimized. This done by taking

$$\frac{\partial S}{\partial a_o} = \frac{\partial S}{\partial a_1} = \ldots = \frac{\partial S}{a_k} = 0. \tag{6.2.5}$$

The system (6.2.5) gives the normal equations for $\underline{a}$. These are

$$(X^T X)\underline{a} = X^T \underline{e} \qquad (6.2.6)$$

and it can be shown that $(X^T X)$ is a positive definite matrix and has an inverse. Thus

$$\underline{a} = (X^T X)^{-1} X^T \underline{\epsilon} \qquad (6.2.7)$$

The a's are called regression coefficients and are least square estimators of the parameters α. Since the a's are dependent upon the e_i's they themselves are random variables and are characterized by a kth dimensional probability density function. Since the system of normal equations was linear, the variance of the regression coefficients is estimated by the sample covariance matrix:

$$\mathrm{Var}(\underline{a}) \doteq \hat{e}^2_{\underline{a}} = [X^T X]^{-1} S_N^2 \qquad (6.2.8)$$

where

$$S_N^2 = \sum_{i=1}^{n} \frac{\hat{N}_i^2}{n - k - 1}. \qquad (6.2.9)$$

S_N^2 is the estimate of the variance of η, computed from the residuals $\hat{N}_i$, where

$$\hat{N}_i = e_i - [a_o + \sum_{j=1}^{k} a_j X_{ij}] \qquad (6.2.10)$$

i.e., the $\hat{N}$'s are the residuals about the regression estimate of ϵ.

If the η's are assumed to be characterized by the normal distribution, then since the a's are linear combinations, they are also normal variates with expected value estimated by (6.2.7) and variances estimated by (6.2.8). Confidence intervals on the parameters $\sigma^2_{\alpha_i}$ and α_i can be obtained using the Chi-square and normal distributions as discussed in Section 5.2.

If the errors are not independent, then it can be shown that if this is not accounted for, the estimate of α will be biased. The techniques for handling correlated error come under regression analysis with "extended" or weighted least squares approximation. A detailed discussion of this topic is found in Reference 26.

6.3 ANALYSIS OF VARIANCE

Many of the previous sections show that error analysis is actually the analysis of the variation of measurements about some accepted standard. The problem involved in error analysis is to evaluate it in terms of explained variation (systematic error) and unexplained variation (random error). If one should obtain error estimates from several sources such as various tests, different instrumentation sites/systems, several runs, flights, times, aircraft, missiles, etc., he might desire to consolidate the results and design a technique to break the total variation into components which will explain the effects these varied sources have on the total variation. For example, the error models in Section 6.1 give several sources which will contribute to the total variance of the error. The questions are "how much do they contribute" and "is the contribution due to any one source significant?" If the answers to questions such as these can be obtained, we would then know which variables to use or delete in our regression analysis. The technique used to break variation in data into source components is called Analysis of Variance, which is highly dependent upon the design of experiments briefly mentioned in Section 6.2.

The sources of variation considered in analysis of variance are called variables or factors. The factors may be quantitative (such as a dynamic error model) or qualitative (such as different missiles or radars). Analysis of variance can correspond to several test designs. The most simple test design is the "completely randomized" or between-group and within-group design. In this case there is only one source of variation. For example, suppose that total error estimates y are made for a particular radar for k different tests from the same type of missile. The data would be grouped according to tests and the problem is to estimate the amount of variation (out of the total variation) between tests; the amount of variation within tests, and to test whether or not such variation is "significant." Such a test is called completely randomized because the grouping of the tests is accomplished in a random manner to eliminate any systematic trends which might exist from test to test. The mathematical model for such a design is

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \; j = 1, \ldots, n_i; \; i = 1, \ldots, k \qquad (6.3.1)$$

where

y_{ij} = ith measurement from the jth group

n_i = number of observations in the ith group

μ = mean of population from which groups are sampled

α_i = deviation of mean of ith group from μ

ϵ_{ij} = unexplained variation in y_{ij}.

The α's are regarded as random variables with mean zero and variance σ_α^2. The term ϵ_{ij} is often called experimental error, and in the between-within experiment $\sigma_{\epsilon ij}^2$ would be the variance within the groups. Table 6.3.0 gives a typical between-within analysis of variance design (one-way analysis of variance) and illustrates how the parameters are estimated.

TABLE 6.3.0

ONE WAY ANALYSIS OF VARIANCE

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	Expected Mean Square
Between groups	k-1	$\sum_i n_i(y_i - \bar{\bar{y}})^2 = A$	A/k-1	$\sigma^2 + k\sigma_\alpha^2$
Within groups	N-k	$\sum_{i,j} (y_{ij} - \bar{y}_i)^2 = B$	B/N-k	σ^2
TOTAL	N-1	$\sum_{ij} (y_{ij} - \bar{y}_i)^2 = C$	$\frac{C}{N-1}$	

$$\bar{y}_i = \frac{1}{n_i} \sum_j y_{ij}, \quad N = \sum_i n_i$$

$$\bar{\bar{y}} = \sum_i \frac{n_i \bar{y}_i)}{N}$$

$$k = (N^2 - \sum_i n_i^2)/[N(k-1)]$$

If the errors are not independent, the analysis of variance is not valid. Errors from different groups may reasonably be assumed uncorrelated, but errors within groups are often correlated, especially if the measurements within a group come from a time series. However, this can be offset by randomizing within each group Thus, in order to be effective, the one-way analysis of variance should be completelyrandomized.

The analysis of variance usually has a "test of significance" associated with it. That is, we hypothesize that the means of each group are equal, then (under the assumption the groups have been completely randomized and are independent normal variates), test for the rejection of this hypothesis at a prescribed probability level The most widely used test is the ratio of the variance from the source (between) to the variance of the experimental error. This forms a statistic which has the F probability distribution with (k-1) and N-k degrees of freedom. This is the same as saying

$$\text{Prob}\left\{\frac{S_1^2}{S_2^2} = \frac{\text{variance between groups}}{\text{variance within groups}} < K\right\} = \int_0^K F_{\substack{(N-k) \\ (k-1)}} dF. \qquad (6.3.2)$$

Suppose we select K such that $\text{Prob}(S_1^2/S_2^2 < k) = .95$, and then set up the hypothesis $H_0: \mu_1 = \mu_2 = \ldots = \mu_k = \mu$ (i.e., the means of each group are equal). From the analysis of variance we estimate F

$$\tilde{F} = \frac{S_1^2}{S_2^2}. \qquad (6.3.3)$$

If $\tilde{F}$ is then compared to the theoretical value of F from the F distribution, and $\tilde{F} > F$ we reject H_0 and say the group means are significantly different. If $\tilde{F} < F$ we cannot reject H_0 and there is no significant difference.

When certain factors which are known are placed in the experimental design, additional constraints are added to the design. For example, we could select the p radars and k groups from the same type of test. We block the k groups and block the p radars in each of the k groups. This type is a randomized block design as shown below

TESTS

RADARS	1	2	...	k
1	y_{11}	y_{12}	⋮	y_{1k}
2	y_{21}	y_{22}	⋮	y_{2k}
⋮	⋮	⋮	⋮	⋮
p	y_{p1}	y_{p2}	...	y_{pk}

The analysis for the randomized block design is called a two-way analysis of variance, and the component variances measure the effects between tests and between radars with respect to the variance of the experimental error. If the cells in the design of Figure 6.3.0 contain r replicates each, then it becomes possible to estimate the effects of interaction between radars and tests.

The completely randomized design and the randomized block design are special cases of a more generalized design technique called factorial design. This technique handles several variables, each with several levels and each having replicates. This type of analysis could be used in conjunction with the error models (6.1.5) and (6.1.7). The error model for radar range error (Relation 6.1.5) would consist of 12 factors, each having n levels (corresponding to n observations), with one replicate. An example of an analysis of variance on a similar radar range error model is given in Table 6.3.1.

Table 6.3.1 is an example of an analysis which summarizes Sections 6.0, 6.1, 6.2, and 6.3, and needs some explanation. The coefficient of determination estimates the percentage of variation out of the total variation, which is explained by the regression analysis. The computed T value is the ratio of the regression coefficient to its standard error estimate. The partial correlation coefficient is an estimate of correlation between the two variables, keeping the effects of the other variables fixed. The multiple correlation coefficient is the correlation between the observed dependent variable and the regression estimate. The "proportion of variance cumulated" estimates the percent of the total variation due to each individual variable. Examination of the table shows that the fit explains only 49% of the total variations with 31.4% and 17.2% of this due to scale factor and timing. The T value

shows these two regression coefficients are significantly different from zero at the 95% confidence level. The total error estimate is RMS (ΔR) = 9.7 feet. The bias estimate is -7.54 feet, while the noise estimate is 4.63 feet. Since the ratio [(mean ΔR)/$\bar{\sigma}_{\Delta R}$)] is small it is difficult to obtain a large coefficient of determination (i.e., a large proportion of the total error is random or "unexplained"). However, the test for significance and T values indicates the regression fit is significant.

TABLE 6.3.1

ANALYSIS OF VARIANCE OF RADAR ERROR MODEL (EXAMPLE)

1. Error Model: $\Delta R_j = \alpha_0 + \alpha_1 \dot{R}_j + \alpha_3 \ddot{R}_j + N_j$
2. Coefficient of Determination: 0.4906
3. Multiple Correlation Coefficient: 0.7004
4. Standard Error of Estimate ($\hat{\sigma}_n$) = 4.63336
5. Intercept Estimate (a_0): -7.54538

ANALYSIS OF VARIANCE

Source of Variation	D.F.	Sum of Squares	Mean Squares	F Value
Due to regression	3	5396.28357	1798.76118	83.7879
Deviation about regression . .	261	5603.15576	21.46803	
TOTAL . .	264	10999.43933		

Variable	Mean	Std. Deviation	Reg. Coeff.	Std. Error of Reg.Coe.	Computed T Value	Partial Corr. Coe	Sum of Sq. Added	Prop. Var. Cum.
R	11829.42249	2329.27798	0.00132	0.00013	10.21585	0.53446	3455.36624	0.31414
$\dot{R}$	-0.34323	38.72200	0.07054	0.00746	9.45630	0.50516	1898.11476	0.17256
$\ddot{R}$	0.53988	0.27293	-1.54071	1.09114	-1.41202	-0.08707	42.80281	0.00389
ΔR	7.23463	6.45481						

7.0 THEORY OF ERROR PROPAGATION

In addition to estimating the error in the measurements from instrumentation systems, it is necessary to estimate the error in the final reduced data. This data usually consists of position, velocity, and acceleration information in some coordinate (usually rectangular) system. Thus, errors in range, azimuth, and elevation or direction cosines, film coordinates, range sums or differences must be propagated into rectangular coordinates. Furthermore, if velocity and acceleration are to be inferred from this positional data, the manner in which the errors propagate must be known. At least three factors are of importance here: (1) the manner in which observing sites are located with respect to each other (assuming that several are involved), (2) the errors in the measurement systems, and (3) the position of the object in space whose coordinates are to be determined.

Let x be a random variable with mean μ and variance σ_x^2. If $y = a_0 + a_1 x$, it can be shown that

$$\sigma_y^2 = a_1^2 \sigma_x^2 \tag{7.0.1}$$

Further, if $x_1, x_2, x_3, \ldots, x_n$ are independent random variables with variance σ_x^2, then the sum

$$y_j = \sum_{i=1}^{n} a_{ji} x_i, \quad j = 1,2,\ldots,N \tag{7.0.2}$$

has mean

$$\mu_{y_j} = \left[\sum_{i=1}^{n} a_{j_i} \right] \mu_x \tag{7.0.3}$$

and variance

$$\sigma_{y_j}^2 = \sum_{i=1}^{n} a_{j_i}^2 \sigma_x^2. \tag{7.0.4}$$

Further, the covariance of y_j, y_k, $j, k = 1,2,\ldots, N$ is

$$\sigma_{y_j y_k} = \left(\sum_{i=1}^{n} a_{j_i} a_{k_i} \right) \sigma_x^2 \tag{7.0.5}$$

If $y = f(x)$ is some function, we can approximate $f(x)$ in the neighborhood of $x = x_o$ by a first order Taylor series expansion. That is

$$y \doteq f(x_o) + \frac{df}{dx_o} (x - x_o) \tag{7.0.6}$$

In the same manner, using (7.0.1) we approximate the variance in y by the relation

$$\sigma_y^2 \doteq \left(\frac{df}{dx_o} \right)^2 \sigma_x^2 \tag{7.0.7}$$

Now suppose $y = f(x_1,x_2)$, then the first order Taylor Series expansion in the neighborhood of $(x_1,x_2) = (x_1^o,x_2^o)$ is

$$y \doteq f(x_1^o,x_2^o) + \frac{\partial f}{\partial x_1} \Big|_{x_1^o} (x_1 - x_1^o) + \frac{\partial f}{\partial x_2} \Big|_{x_2^o} (x_2 - x_2^o), \tag{7.0.8}$$

and from (7.0.4) and (7.0.5) the form of σ_y^2 would be

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_{x_2} + 2 \frac{\partial f}{\partial x_1} \frac{\partial f}{\partial x_2} \sigma_{x_1 x_2} . \tag{7.0.9}$$

In general if $y_i = y_i(x_1,x_2,\ldots,x_n)$, $i = 1,2,\ldots,m$, then the Taylor Series expansion estimate of the covariance matrix of y is

$$\theta_y^2 = R S_x R^T \tag{7.0.10}$$

where

$$\hat{\sigma}_y^2 = \begin{bmatrix} \hat{\sigma}_{y_1}^2 & \hat{\sigma}_{y_1y_2} & \cdots & \hat{\sigma}_{y_1y_m} \\ \hat{\sigma}_{y_2y_1} & \hat{\sigma}_{y_2}^2 & \cdots & \hat{\sigma}_{y_2y_m} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\sigma}_{y_ny_1} & \hat{\sigma}_{y_ny_2} & \cdots & \hat{\sigma}_{y_m}^2 \end{bmatrix} \qquad (7.0.11)$$

$$S_x = \begin{bmatrix} \hat{\sigma}_{x_1}^2 & \hat{\sigma}_{x_1x_2} & \cdots & \hat{\sigma}_{x_1x_n} \\ \hat{\sigma}_{x_2x_1} & \hat{\sigma}_{x_2}^2 & \cdots & \hat{\sigma}_{x_2x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\sigma}_{x_nx_1} & \hat{\sigma}_{x_nx_2} & \cdots & \hat{\sigma}_{x_n}^2 \end{bmatrix} \qquad (7.0.12)$$

$$R = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_2}{\partial x} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \frac{\partial y_m}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \qquad (7.0.13)$$

and

R^T is the transpose of R

If the x_i's are independent random variables then S_x is a diagonal matrix, i.e., the covariances are all zero. In this case

$$\sigma_{y_iy_j} = \frac{\partial y_i}{\partial x_1} \frac{\partial y_j}{\partial x_1} \sigma_x^2, + \left[\frac{\partial y_i}{\partial x_2}\right] \sigma_{x_2}^2 + \ldots + \left[\frac{\partial y_i}{\partial x_n}\right] \sigma_{x_n}^2 \qquad (7.0.14)$$

The relation $\sigma_y^2 = RS_xR^T$ is usually called the covariance equation and is classically referred to as the Gaussian law of error propagation. This technique is based directly upon approximating the change in the dependent variable by the differential, i.e., the linear portion of a Taylor Series expansion. This leads one to wonder how much error is involved in the variance estimates. Actual examples using simulation techniques have shown the method to be very good and, as John Tukey is quoted as saying, "The most important conclusion is that the classical propagation formula is much better than seems to be realized. Examples indicate that it is quite likely to suffice for most work."

7.1 PROPAGATION OF ERRORS INTO POSITION

Since most of the measured quantities are angle, direction cosines, film coordinates, range sums, and range differences, the effects of errors in these observations are present in rectangular coordinates derived from these quantities. While it is most convenient to consider the errors in the actual measurements of the instrumentation system when discussing its accuracy, these errors do not necessarily give the desired information. The relation between errors in range, azimuth, and elevation and errors in derived position data depends upon three factors: (1) the location and geometry of the instrumentation sites, (2) the errors in the measurements of the system, and (3) the position of the point in space that is to be determined.

Methods can be developed and programmed which give estimates of the errors in rectangular coordinates when items (1), (2), (3) in the above paragraph are taken into consideration. The end product of these programs tests estimates of errors in rectangular coordinates, and is referred to as geometric dilution of precision (GDOP). For examining the capabilities of given instrumentation on a particular test, significant estimates of the rectangular coordinate accuracies can be obtained by using the nominal trajectory for the test, the actual instrument sites to be used, and the best estimates of the errors of measurement for the particular systems. Further information on GDOP may be obtained from the Bibliography.

Although GDOP is widely used, the more sophisticated concept of an ellipsoid of error is gaining acceptance as spatial measurements are made with higher orders of precision. Whereas GDOP utilizes only information from the diagonal terms ($\sigma_X^2, \sigma_Y^2, \sigma_Z^2$) of the variance-covariance matrix, the ellipsoid of error utilizes all of the information which can be extracted from the entire matrix. If an ellipsoid of error is used, confidence regions can be set up about each spatial position point and the orientation of the axes of the ellipsoid can be obtained. This is discussed in more detail in Section 8.0.

The basic problem in error propagation is to determine the effect of the random errors of measurement in the particular instruments on the rectangular coordinate data. Suppose, for example, that the quantities, α, β, γ, are measured at time, t, with errors $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$. If there is a known mathematical relation between X, Y, Z and α, β, γ:

$$\begin{aligned} X &= f_1(\alpha,\beta,\gamma), \\ Y &= f_2(\alpha,\beta,\gamma), \\ Z &= f_3(\alpha,\beta,\gamma), \end{aligned} \tag{7.1.0}$$

then the errors in X, Y, and Z produced by $\Delta\alpha$, $\Delta\beta$, $\Delta\gamma$ can be approximated by

$$\Delta X = \frac{\partial X}{\partial \alpha}\Delta\alpha + \frac{\partial X}{\partial \beta}\Delta\beta + \frac{\partial X}{\partial \gamma}\Delta\gamma, \tag{7.1.1}$$

$$\Delta Y = \frac{\partial Y}{\partial \alpha}\Delta\alpha + \frac{\partial Y}{\partial \xi}\Delta\beta + \frac{\partial Y}{\partial \gamma}\Delta\gamma,$$

$$\Delta Z = \frac{\partial Z}{\partial \alpha}\Delta\alpha + \frac{\partial Z}{\partial \beta}\Delta\beta + \frac{\partial Z}{\partial \gamma}\Delta\gamma,$$

where the partial derivatives are evaluated using f_1, f_2, f_3 and the measured values of α, β, γ. This type of estimation of the errors is also satisfactory when discussing the effects of systematic errors if these errors are relatively small with respect to the magnitudes of X, Y and Z.

A discussion for other systems involving multiple stations and least square solutions is based upon the same ideas. However, they are complicated by the fact that more variables are involved and more complex mathematical manipulation must be carried out. For simplicity, methods are illustrated here for radar.

The relations between the measured quantities, azimuth (A), elevation (E), and range (R) and the space position (X,Y,Z) in left handed coordinates are:

$$X = R \cos A \cos E$$

$$Y = R \sin A \cos E \tag{7.1.2}$$

$$Z = R \sin E.$$

Errors may exist singly in one of the measurement R, A, E or simultaneously in R, A, E. Using the above relations and simultaneous errors in R, A, E, the first order effects of these errors upon X, Y, Z are given by:

$$\begin{aligned} \Delta X &\doteq \Delta R \cos E \cos A - R\,\Delta A \cos E \sin A - R\,\Delta E \sin E \cos A \\ \Delta Y &\doteq \Delta R \cos E \sin A + R\,\Delta A \cos E \cos A - R\,\Delta E \sin E \sin A \\ \Delta Z &\doteq \Delta R \sin E + R\,\Delta E \cos E \end{aligned} \tag{7.1.3}$$

Using the covariance equation (7.0.10), we estimate the elements of the variance-covariance matrix σ for the radar assuming the errors in R, A, E are uncorrelated. The estimates are

$$\sigma_X^2 = \left(\frac{\partial X}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial X}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial X}{\partial E}\right)^2 \sigma_E^2$$

$$= \sigma_R^2 \cos^2 A \cos^2 E + \sigma_E^2 R^2 \cos^2 A \sin^2 E + \sigma_A^2 R^2 \sin^2 A \cos^2 E,$$

$$\sigma_Y^2 = \left(\frac{\partial Y}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial Y}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial Y}{\partial E}\right)^2 \sigma_E^2$$

$$= \sigma_R^2 \sin^2 A \cos^2 E + \sigma_A^2 R^2 \cos^2 A \cos^2 E + \sigma_E^2 R^2 \sin^2 A \sin^2 E,$$

$$\sigma_Z^2 = \left(\frac{\partial Z}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial Z}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial Z}{\partial E}\right)^2 \sigma_E^2$$

$$= \sigma^2 \sin^2 E + \sigma_E^2 R^2 \cos^2 E,$$

$$\sigma_{xy} = \frac{\partial X}{\partial R}\frac{\partial Y}{\partial R}\sigma_R^2 + \frac{\partial X}{\partial A}\frac{\partial Y}{\partial A}\sigma_A^2 + \frac{\partial X}{\partial E}\frac{\partial Y}{\partial E}\sigma_E^2 \qquad (7.1.4)$$

$$= (\cos A \sin A \cos^2 E)\,\sigma_R^2 - (R^2 \cos E^2 \sin A \cos A)\,\sigma_A^2$$

$$+ (R^2 \cos A \sin^2 E \sin A)\,\sigma_E^2,$$

$$\sigma_{xz} = \frac{\partial X}{\partial R}\frac{\partial Z}{\partial R}\sigma_R^2 + \frac{\partial X}{\partial A}\frac{\partial Z}{\partial A}\sigma_A^2 + \frac{\partial X}{\partial E}\frac{\partial Z}{\partial E}\sigma_E^2$$

$$= (\cos A \cos E \sin E)\,\sigma_R^2 - (R^2 \cos A \sin E \cos E)\,\sigma_E^2,$$

$$\sigma_{yz} = \frac{\partial Y}{\partial R}\frac{\partial Z}{\partial R}\sigma_R^2 + \frac{\partial Y}{\partial A}\frac{\partial Z}{\partial A}\sigma_A^2 + \frac{\partial Y}{\partial E}\frac{\partial Z}{\partial E}\sigma_E^2$$

$$= (\sin A \cos E \sin E)\,\sigma_R^2 - (R^2 \sin A \sin E \cos E)\,\sigma_E^2.$$

The variances and covariances correspond to the error distribution which has been propagated into a left handed coordinate system tangent to the spheroid with the radar at the origin of the system.

In many instances, it is desired to propagate the error into a coordinate system with a different origin and orientation. For example, it is often desired to propagate the error into a coordinate system whose origin is a launching pad and whose +X axis is down the launch azimuth. In this case the R A E data is first transformed into a local tangent plane system, (X_{TP},$Y_{TP}$$Z_{TP}$). Based on the geodetic position (ϕ,λ) of the radar, the tangent plane coordinates are transformed so that each axis is parallel to a geocentric (earth centered) system, translated to geocentric origin, translated to new origin (launch pad), then the axes are rotated into the tangent plane system with the +X axis parallel to the desired orientation. This process is accomplished by the following matrix equation

$$[P]_{\alpha}^{T} \left[[R] \begin{bmatrix} X_{TP} \\ Y_{TP} \\ Z_{TP} \end{bmatrix} + [G_R] - [G_P] \right] = \begin{bmatrix} X'_{TP} \\ Y'_{TP} \\ Z'_{TP} \end{bmatrix} \tag{7.1.5}$$

where

[R] = matrix which rotates radar tangent plane into geocentric system

$[G_R]$ and $[G_P]$ are geocentric coordinates of radar and launch pad

$[P]^T$ = Matrix which rotates axes parallel to local tangent plane so translation from earth's center can be obtained. [P] corresponds to [R] only is for the launch pad.

α = orientation azimuth of transformed system

$\begin{bmatrix} X'_{TP} \\ Y'_{TP} \\ Z'_{TP} \end{bmatrix}_{\alpha}$ = local tangent plane coordinates with origin at launch pad P with +X axis oriented parallel to α degrees

The matrices [R] and [P] are linear transformations whose elements are constants which are functions of α, ϕ and λ. Since the transformations are linear, the propagated estimate for the covariance matrix in the new coordinate system is

$$\theta_P = A\, \theta_{TP}\, A^T \tag{7.1.6}$$

where θ_{TP} = covariance matrix with respect to radar tangent plane

$$A = [P]^T [R]$$

It is pointed out that the error propagation techniques can propagate from X,Y,Z back to the original observations of R,A,E. In order to do this the geometry (X,Y,Z) and the covariance matrix must be known. For example, for

$$R = f(X,Y,Z)$$

then

$$\sigma_R^2 = \left(\frac{\partial R}{\partial X}\right)^2 \sigma_X^2 + \left(\frac{\partial R}{\partial Y}\right)^2 \sigma_Y^2 + \left(\frac{\partial R}{\partial Z}\right) \sigma_Z^2$$

$$+ 2 \left(\frac{\partial R}{\partial X}\right)\left(\frac{\partial R}{\partial Y}\right) \sigma_{XY} + 2 \left(\frac{\partial R}{\partial X}\right)\left(\frac{\partial R}{\partial Z}\right) \sigma_{XZ} + 2 \left(\frac{\partial R}{\partial Y}\right)\left(\frac{\partial R}{\partial Z}\right) \sigma_{YZ}. \qquad (7.1.7)$$

Should it be necessary to differentiate the position to acquire velocity, it is assumed that any low-frequency systematic error is virtually eliminated in the process. It is noted. however, that cyclic errors may not be disposed of in this manner.

It is noted that, under certain conditions, the effects of total errors can be propagated into spatial position error. If the variance of the systematic error corresponding to the same point in space is estimated over several tests, then the effects of this can be propagated into spatial position error. In this case the variance of the total error input for propagation would be

$$\sigma^2_{TOTAL} = \sigma^2_{SYSTEMATIC} + \sigma^2_{RANDOM} \qquad (7.1.8)$$

and the propagated error would have a variance-covariance matrix corresponding to total spatial position error.

7.2 PROPAGATION OF ERRORS INTO VELOCITY AND ACCELERATION

In general, position and range rate data are the basic measurements received by instrumentation systems which are external to a vehicle in flight. The basic trajectory-related measurements from on board systems are acceleration components parallel to the missile coordinate system. In the first case, velocity and acceleration data are obtained by numerically estimating the first and second derivatives of the position data. In the latter, velocity and position are obtained by successive integration with respect to time.

Very low frequency systematic errors in position data may be regarded as constant over short spans of time. In this instance a measured variable can be represented as a function of time by

$$V = F(t) + B \tag{7.2.1}$$

where B is the bias error. Upon numerical differentiation with respect to time we obtain

$$\dot{V} = F'(t) \tag{7.2.2}$$

which is independent of B. The significance of this fact is that a differentiation derived from a measured function with low frequency bias error is virtually free of this bias. However, higher frequency cyclic error and random error will persist and must be considered. The amount of random error for a particular instant of time is, of course, unknown and cannot be predicted as a function of time, but under the assumption that it is a stationary process (at least during short spans of data), its statistical properties remain constant.

Inertial guidance systems currently in use in many missile programs provide examples of acceleration measuring devices. The functions these instruments perform are varied and complicated. Inertial guidance systems are discussed in more detail in Section 10.0. In this section we are concerned primarily with errors propagated due to numerical differentiation of position data with respect to time.

If "raw" velocities in space position data were computed by successive differences, the "average" velocity estimate over the time increment would be

$$\dot{X} \doteq \Delta X_{\frac{2i+1}{2}} = \frac{1}{\Delta t}(X_i - X_{i-1}). \tag{7.2.3}$$

The X_i's are random variables and assumed to be independent with mean μ and variance σ_X^2. Then, using the propagation techniques of Section 7.0, the variance in ΔX is

$$\text{Var}\left(\frac{\Delta X}{\Delta t}\right) = \frac{\sigma_{X_i}^2 + \sigma_{X_{i-1}}^2}{\Delta t^2} = \frac{2\sigma_X^2}{\Delta t^2}. \tag{7.2.4}$$

From (7.2.4) it is obvious that the random variable ΔX contains more random error than X. For example, if σ_X= 2 feet and Δt = 0.1 second, then the error in raw velocity would be $\pm 20\sqrt{2}$ or ± 28.28 feet per second. In the case of "raw" acceleration it is much worse, and for σ_X = 2 feet, $\sigma_{\ddot{X}} = \pm 500$ feet per second^2! These examples were given to show that filtering/smoothing techniques are needed to achieve reasonable $\dot{X}$ and $\ddot{X}$ estimates when derived from numerically differentiating position data.

Most methods presently used filter the position data as a function of time and then numerically differentiate the filtered data. The most common type of filter used is a kth degree polynomial which is fitted by some criterion (usually least squares) to a span of m points. The first and second derivatives of the polynomial are evaluated at an appropriate time, usually the midpoint, since this gives velocity and acceleration data with smaller random errors than end point estimates. The point of evaluation which gives minimum random error is between the end and center points. The moving arc technique of fitting the polynomial of kth degree of the data, X_2 to X_{m+1} and t_2 to t_{m+1}, and evaluating first and second derivatives is used. This is continued until all desired velocity and acceleration components are computed. In the so-called simple differencing method, the degree of the polynomial is 2. If the data is equally spaced in time, i.e.,

$$t_j = t_0 + (\Delta t)\, j, \tag{7.2.5}$$

then the polynomial filters for velocity and acceleration will consist of weights b_j and c_j and the velocity and acceleration estimates are made from the following linear combinations:

$$\frac{dX_i}{dt} \doteq \frac{1}{\Delta t} \sum_{j=1}^{m} b_j\, X_{i-j+d} \tag{7.2.6}$$

and

$$\frac{d^2X_i}{dt^2} \doteq \frac{1}{(\Delta t)^2} \sum_{j=1}^{m} c_j\, X_{i-j+d} \tag{7.2.7}$$

respectively, where d is related to the delay (i.e., if m = 2n+1 and evaluation is at the end point then d = m). The b_j and c_j depend on the degree k of the polynomial, the span of points, m, used in the moving average, and the number d. The b_j and c_j are fixed constants and are precomputed. Using formulas of the type above and assuming that a polynomial of kth degree approximates the data to a sufficient degree of accuracy over the span of m points, and that the errors in successive values of X_j are uncorrelated, the relation between the random errors in position data and the random errors in velocity and acceleration data can be expressed in the form

$$\theta_{\dot{X}} = \frac{\theta_X}{\Delta t} \left(\sum_{j=1}^{m} b_j^2 \right)^{\frac{1}{2}} \tag{7.2.8}$$

and

$$\hat{\sigma}_{\ddot{X}} = \frac{\hat{\sigma}_X}{(\Delta t)^2} \left(\sum_{j=1}^{m} c_j^2 \right)^{\frac{1}{2}} \tag{7.2.9}$$

Similar relations can be obtained if the errors are correlated and the form of the correlation is known. In the case of midpoint smoothing the equations for $\sigma_{\dot{X}}$ and $\sigma_{\ddot{X}}$ become

$$\hat{\sigma}_{\dot{X}} = \frac{\hat{\sigma}_X}{\Delta t} \sqrt{\frac{12}{m(m^2 - 1)}} \simeq \frac{\sigma_X}{\Delta tm} \frac{\sqrt{12}}{\sqrt{m}}, \tag{7.2.10}$$

for second degree polynomial, and

$$\hat{\sigma}_{\ddot{X}} = \frac{\sigma_X}{(\Delta t)^2} \sqrt{\frac{720}{m(m^2 - 1)(m^2 - 4)}} \simeq \frac{\sigma_X}{(\Delta tm)^2} \frac{\sqrt{720}}{\sqrt{m}} \tag{7.2.11}$$

for a third degree polynomial. It is evident from the approximations that if the sampling rate is stepped up and the smoothing time span (Δtm) held constant, the velocity and acceleration errors are reduced approximately by a factor of $m^{-\frac{1}{2}}$. Truncation errors limit the extent to which this can be applied. If smoothing is performed at some point other than the midpoint the same general argument applies.

In the discussion of the effects of random errors on velocity and acceleration, it was pointed out that the method presented was valid only if the errors in successive values of a given coordinate are uncorrelated. This assumption is not true for a systematic error by definition. Hence, different techniques are required for determination of the effects of systematic errors.

In the case of single station systems, the errors in velocity and acceleration in Cartesian coordinates may be found by differentiating the XYZ equations with respect to time and utilizing the Gaussian approximation.

$$\sigma_{f_i}^2 = \sum \left(\frac{\partial f_i}{\partial \alpha_j} \right)^2 \sigma_{\alpha_j}^2 \tag{7.2.12}$$

where f_i = X, Y, Z, $\dot{X}$, $\dot{Y}$, $\dot{Z}$ and α_j = R, A, E, $\dot{R}$, $\dot{A}$, $\dot{E}$ for radar. For example, the variances of the velocity errors for radar become

$$\sigma^2_{\dot{X}} = (-\dot{E} \sin E \cos A - \dot{A} \cos E \sin A)^2 \sigma^2_R$$

$$+ (-\dot{R} \sin E \cos A - \dot{E} R \cos E \cos A + \dot{A} R \sin E \sin A)^2 \sigma^2_E \quad (7.2.13)$$

$$+ (-\dot{R} \cos E \sin A + \dot{E} R \sin E \sin A - \dot{A} R \cos A \cos E)^2 \sigma^2_A$$

$$+ (\cos A \cos E)^2 \sigma^2_{\dot{R}} + (-R \cos A \sin E)^2 \sigma^2_{\dot{E}} + (-R \sin A \cos E)^2 \sigma^2_{\dot{A}}$$

$$\sigma^2_{\dot{Y}} = (-\dot{E} \sin E \sin A + \dot{A} \cos E \cos A)^2 \sigma^2_R$$

$$+ (\dot{R} \cos E \cos A - \dot{E} R \sin E \cos A - \dot{A} R \cos E \sin A)^2 \sigma^2_A \quad (7.2.14)$$

$$+ (-\dot{R} \sin E \sin A - \dot{E} R \cos E \sin A - \dot{A} R \sin E \cos A)^2 \sigma^2_E$$

$$+ (\sin A \cos E)^2 \sigma^2_{\dot{R}} + (-R \sin A \sin E)^2 \sigma^2_{\dot{E}} + (R \cos E \cos A)^2 \sigma^2_{\dot{A}}$$

$$\sigma^2_{\dot{Z}} = (\dot{E} \cos E)^2 \sigma^2_R + (\dot{R} \cos E - \dot{E} R \sin E)^2 \sigma^2_E \quad (7.2.15)$$

$$+ (\sin E)^2 \sigma^2_{\dot{R}} + (R \cos E)^2 \sigma^2_{\dot{E}}$$

The expressions for the covariance terms $\sigma_{\dot{x}\ddot{y}}$, $\sigma_{\dot{x}\ddot{z}}$, $\sigma_{\dot{y}\ddot{z}}$ may be written according to the following expression for $\sigma_{\dot{x}\ddot{y}}$.

$$\sigma_{\dot{x}\ddot{y}} = \sigma_{\dot{y}\ddot{x}} = \frac{\partial \dot{X}}{\partial R}\frac{\partial \dot{Y}}{\partial R}\sigma_R^2 + \frac{\partial \dot{X}}{\partial A}\frac{\partial \dot{Y}}{\partial A}\sigma_A^2 + \frac{\partial \dot{X}}{\partial E}\frac{\partial \dot{Y}}{\partial E}\sigma_E^2 + \frac{\partial \dot{X}}{\partial \dot{R}}\frac{\partial \dot{Y}}{\partial \dot{R}}\sigma_{\dot{R}}^2 + \frac{\partial \dot{X}}{\partial \dot{A}}\frac{\partial \dot{Y}}{\partial \dot{A}}\sigma_{\dot{A}}^2 + \frac{\partial \dot{X}}{\partial \dot{E}}\frac{\partial \dot{Y}}{\partial \dot{E}}\sigma_{\dot{E}}^2 \cdot \tag{7.2.16}$$

In a similar manner, a partitioned variance-covariance matrix can be estimated giving all position, velocity, acceleration error variances and covariances. The matrix would be

$$\bar{\sigma} = \left|\begin{array}{ccc:ccc:ccc}
\sigma_X^2 & \sigma_{xy} & \sigma_{xz} & \sigma_{x\dot{x}} & \sigma_{x\dot{y}} & \sigma_{x\dot{z}} & \sigma_{x\ddot{x}} & \sigma_{x\ddot{y}} & \sigma_{x\ddot{z}} \\
\sigma_{xy} & \sigma_Y^2 & \sigma_{yz} & \sigma_{y\dot{x}} & \sigma_{y\dot{y}} & \sigma_{y\dot{z}} & \sigma_{y\ddot{x}} & \sigma_{y\ddot{y}} & \sigma_{y\ddot{z}} \\
\sigma_{xz} & \sigma_{yz} & \sigma_Z^2 & \sigma_{z\dot{x}} & \sigma_{z\dot{y}} & \sigma_{z\dot{z}} & \sigma_{z\ddot{x}} & \sigma_{z\ddot{y}} & \sigma_{z\ddot{z}} \\
\hdashline
\sigma_{\dot{x}x} & \sigma_{\dot{x}y} & \sigma_{\dot{x}z} & \sigma_{\dot{X}}^2 & \sigma_{\dot{x}\dot{y}} & \sigma_{\dot{x}\dot{z}} & \sigma_{\dot{x}\ddot{x}} & \sigma_{\dot{x}\ddot{y}} & \sigma_{\dot{x}\ddot{z}} \\
\sigma_{\dot{y}x} & \sigma_{\dot{y}y} & \sigma_{\dot{y}z} & \sigma_{\dot{x}\dot{y}} & \sigma_{\dot{Y}}^2 & \sigma_{\dot{y}\dot{z}} & \sigma_{\dot{y}\ddot{x}} & \sigma_{\dot{y}\ddot{y}} & \sigma_{\dot{y}\ddot{z}} \\
\sigma_{\dot{z}x} & \sigma_{\dot{z}y} & \sigma_{\dot{z}z} & \sigma_{\dot{x}\dot{z}} & \sigma_{\dot{y}\dot{z}} & \sigma_{\dot{Z}}^2 & \sigma_{\dot{z}\ddot{x}} & \sigma_{\dot{z}\ddot{y}} & \sigma_{\dot{z}\ddot{z}} \\
\hdashline
\sigma_{\ddot{x}x} & \sigma_{\ddot{x}y} & \sigma_{\ddot{x}z} & \sigma_{\ddot{x}\dot{x}} & \sigma_{\ddot{x}\dot{y}} & \sigma_{\ddot{x}\dot{z}} & \sigma_{\ddot{X}}^2 & \sigma_{\ddot{x}\ddot{y}} & \sigma_{\ddot{x}\ddot{z}} \\
\sigma_{\ddot{y}x} & \sigma_{\ddot{y}y} & \sigma_{\ddot{y}z} & \sigma_{\ddot{y}\dot{x}} & \sigma_{\ddot{y}\dot{y}} & \sigma_{\ddot{y}\dot{z}} & \sigma_{\ddot{x}\ddot{y}} & \sigma_{\ddot{Y}}^2 & \sigma_{\ddot{y}\ddot{z}} \\
\sigma_{\ddot{z}x} & \sigma_{\ddot{z}y} & \sigma_{\ddot{z}z} & \sigma_{\ddot{z}\dot{x}} & \sigma_{\ddot{z}\dot{y}} & \sigma_{\ddot{z}\dot{z}} & \sigma_{\ddot{x}\ddot{z}} & \sigma_{\ddot{y}\ddot{z}} & \sigma_{\ddot{Z}}^2
\end{array}\right| \tag{7.2.17}$$

In the case of multiple station systems requiring least squares solutions for data reduction, the variances cannot be readily determined analytically as for radar. The same reasonable assumption is made, however, that the variances of the instrument errors represents a normal distribution at the same point in the trajectory over many tests. With this the Gaussian approximation is again utilized, except that the partial derivatives are replaced by increments, i.e.,

$$\frac{\partial f_i}{\partial \alpha_j} \approx \frac{\Delta f_i}{\Delta \alpha_j}$$

The ratio of $\Delta f_i/\Delta \alpha_j$ is determined by numerically computing Δf_i with a predetermined value of $\Delta \alpha_j$ with all other j values of the instrument variables held constant.

8.0 ERROR TOLERANCE REGIONS

Confidence and tolerance intervals on the point estimates of error parameters were discussed in Section 5.2. The next points in question are confidence regions on two- and three-dimensional spatial position data. In general, the type of tolerance regions used consist of ellipses in the two-dimensional case and ellipsoids in the three-dimensional cases. (Special cases of these are circles and spheres.)

In obtaining any type of confidence/tolerance region one must first make an assumption concerning the underlying probability distribution function, and then know or have a good estimate on the parameters which characterize the distribution. The usual assumption is that the underlying distribution is approximately Gaussian which means that all nth dimension distributions can be characterized in terms of the first and second moments of the distribution (i.e., the means and variances). Thus, the basic tool for the confidence region is the variance or variance-covariance matrix. The probability density function for the Gaussian distribution in three dimensions is

$$f(X,Y,Z) = \frac{1}{(2\pi)^{3/2}|\sigma|^{1/2}} e^{-1/2\ Q^{-1}(X,Y,Z)}$$

where

$|\sigma|$ = determinant of covariance matrix σ

$Q^{-1}(X,Y,Z)$ = quadratic form of the inverse covariance matrix σ^{-1}

The quadratic form of the matrix is

$$(X, Y, Z)\,[\sigma^{-1}]\begin{pmatrix}X\\Y\\Z\end{pmatrix} = X^2\sigma_x^{2\prime} + Y^2\sigma_y^{2\prime} + Z^2\sigma_z^{2\prime} + 2XY\sigma_{xy}^{\prime} + 2XZ\sigma_{xz}^{\prime} + 2YZ\sigma_{yz}^{\prime} \qquad (8.0.1)$$

where the prime designates the elements of σ^{-1}. The above relation is the general equation of an ellipsoid and it can be transformed into standard form

$$\frac{\hat{X}^2}{\sigma_{\hat{x}}^{2\prime}} + \frac{\hat{X}^2}{\sigma_{\hat{y}}^{2\prime}} + \frac{\hat{Z}^2}{\sigma_{\hat{z}}^{2\prime}} = 1 \qquad (8.0.2)$$

by rotating the axes of the general form equation parallel to the $(\hat{X},\hat{Y},\hat{Z})$ coordinate system. It can be shown that this process is equivalent to diagonalizing the inverse covariance matrix which, in turn, is equivalent to finding the eigenvalues of σ^{-1}.

The diagonal matrix D of σ^{-1} is the matrix of eigenvalues

$$D = \begin{bmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_1 & 0\\ 0 & 0 & \lambda_3\end{bmatrix} \qquad (8.0.3)$$

and its quadratic form is

$$[\hat{X},\ \hat{Y},\ \hat{Z}]\ D \begin{pmatrix} \hat{X} \\ \hat{Y} \\ \hat{Z} \end{pmatrix} = \lambda_1 \hat{X}^2 + \lambda^2\ \hat{Y}^2 + \lambda_3 \hat{Z}^2 = C^2$$

or

$$\frac{\hat{X}^2}{C^2/\lambda_1} + \frac{\hat{Y}^2}{C^2/\lambda_2} + \frac{\hat{Z}^2}{C^2/\lambda_3} = 1. \tag{8.0.4}$$

The variances of the standard trivariate or ellipsoidal normal distribution are

$$\sigma^2_{\hat{x}} = C^2/\lambda_1,\ \sigma^2_{\hat{y}} = C^2/\lambda_2,\ \sigma^2_{\hat{z}} = C^2/\lambda_3 \tag{8.0.5}$$

It is noted that if the errors in X, Y and Z are independent, then the general equation of the ellipsoid has no cross product terms and corresponds to a matrix σ^{-1}, which is already a diagonal matrix. If the errors are independent and $\sigma^2_{\hat{x}} = \sigma^2_{\hat{y}} = \sigma^2_{\hat{z}}$, the distribution is the spherical normal distribution (3 dimensions) or the circular normal when two axes are equal and the third is zero.

When the errors in X, Y and Z are correlated, then σ has off-diagonal terms and the corresponding error ellipsoid is not in standard form, but is skewed in space. The eigenvalues of σ^{-1} provide the length of the axes and the eigenvectors of σ^{-1} will tell the orientation of each of the axes with respect to the coordinate system $(\bar{X},\bar{Y},\bar{Z})$.

8.1 TOLERANCE ELLIPSOIDS

Once an error ellipsoid has been obtained, the question of how much of the spatial error distribution is contained in the ellipsoid is encountered. Since the errors are assumed to be normal, it can be shown that the axes of the standard ellipsoid equation have the Chi-square distribution with three degrees of freedom. With three degrees of freedom, the probability, p. that a point will be in an ellipsoid with constant term C^2 is

$$p = \int_0^{C^2} \chi_3^2 \, d\chi^2 \tag{8.1.0}$$

The value of C^2 corresponding to p can be obtained from any Chi-square table. The following table gives values of C versus p.

p	C
.95	2.79
.90	2.50
.80	2.15
.70	1.91
.50	1.54
.20	1.00

TABLE 1

Thus, for a 90% error tolerance ellipse, C = 2.50 so that the axes would be

$$\sigma_{\hat{x}} = \frac{2.5}{\sqrt{\lambda_1}}, \quad \sigma_{\hat{y}} = \frac{2.5}{\sqrt{\lambda_2}}, \quad \sigma_{\hat{z}} = \frac{2.5}{\sqrt{\lambda_3}} \tag{8.1.1}$$

The values of C = 1.538 and C = 4.0 are the values of C corresponding to "probable" ellipsoid of error and ellipsoidal near-certainty error, respectively.

In the two-dimensional case, the same procedure is used except that σ^{-1} is a 2x2 matrix and the standard equation of the ellipse is Chi-square with two degrees of freedom. A table of p versus C is given below.

p	C
.99	3.37
0.95	2.45
0.90	2.146
0.865	2.00
0.50	1.177
0.394	1.00

TABLE 2

When C = 1.774 we have a tolerance ellipse called the probable ellipse of error. When the axes are equal, the tolerance region corresponding to C = 1.774 is called "Circular Probable Error" (CPE).

In actual practice $\sigma_x^2 \neq \sigma_y^2 \neq \sigma_z^2$ and the errors will not be independent, so it is usually more valid to work with ellipsoids and ellipses rather than circles or spheres. There are certain instances, however, when it is desired to obtain a circular or spherical tolerance region from an ellipse or ellipsoid. That is, it is desired to find the radius of a circle (or sphere) which contains the same number of points as the tolerance ellipse (ellipsoid) with axes $C\sqrt{\lambda_1}$, $C\sqrt{\lambda_2}$. To accomplish this, we make the change of variables letting $\hat{x} = r \cos\theta$, $\hat{y} = r \sin\theta$ then integrating the bivariate elliptical normal distribution with respect to r and θ we find p as a function of

$$A = \text{Max}\left[\frac{\sigma_{\hat{x}}}{\sigma_{\hat{y}}}, \frac{\sigma_{\hat{y}}}{\sigma_{\hat{x}}}\right] \quad \text{and} \quad B = \frac{R_p}{\sqrt{\sigma_{\hat{x}}\sigma_{\hat{y}}}}$$

where R_p is the radius corresponding to p. The following table gives A and B versus p = 0.90.

A	B
1.0	2.1460
1.5	2.2501
2.0	2.4565
2.5	2.6865
3.0	2.9118
4.0	3.3290
5.0	3.7058
10.0	5.2111
15.0	6.3756
20.0	7.3594
50.0	11.631
100.0	16.449

TABLE 3

For example, if the maximum ratio of $\sigma_{\hat{x}}$ to $\sigma_{\hat{y}}$ is A = 2.0 then the radius of a circle containing 90% of the error distribution would be $R_{.90} = 2.4565\sqrt{\sigma_{\hat{x}}\sigma_{\hat{y}}}$.

The radius R_p of a sphere containing p% of an error distribution can be obtained from a probability ellipsoid in a similar but more complicated manner, and will not be discussed in this report (See Reference 29 for details).

9.0 DIGITAL FILTERING/SMOOTHING OF DATA

As stated in Section 2.0, the measured or raw data contains signal with error superimposed on it. The basic reason for using a digital filter is to separate or suppress the noise error from the signal. There are many different filtering techniques used at the various ranges but, in general, a digital filter is evaluated by the following criteria:

1) How effectively the noise components of the input data are attenuated or removed.

2) How much distortion in the signal.

3) How effectively the filter recovers and smooths the first, second, and higher order derivatives.

4) Amount of serial correlation in the data.

5) Amount of computing time required.

It would be desirable to select a filter which would give the best of each item above, but unfortunately this is not usually the case. Often a filter may be designed to accomplish 1, 2, or 3 in some "best" manner, but usually Item 5 is then sacrificed.

In general, most filters used are linear operators on the raw data. If $X_1 \ldots, X_n$ is a set of raw data, the filtered data is usually expressed as a linear combination of subsets or data spans of the set. For example, if the filter gives a point estimate corresponding to the center time point of a span consisting of 2N+1 points then the expression for the output of the filter would be

$$\hat{X}_t = \sum_{i=-N}^{i=+N} W_i X_{t+i} \tag{9.0.1}$$

where the W_i are filter weights. These weights are constrained so that

$$\sum W_i = 1 \tag{9.0.2}$$

This is necessary since the signals occuring in missile trajectory work usually have a very large low-frequency trend, and the response of the filter at these low frequencies should be exactly one (i.e., no distortion or biasing at zero frequency).

If the filter is an end point estimation or "predictor" filter, then the weighted sum would consist of k previous points.

Many of the digital filters being used are of the general-purpose type. The parameters which define the weights of such a filter are adjusted according to the sampling rate of the input data and some estimate as to the frequency composition of the signal. Some types of general-purpose filters are

1) Straight moving average

2) Classical least squares polynomial fit

3) Constrained least squares polynomial fit

4) Variable span

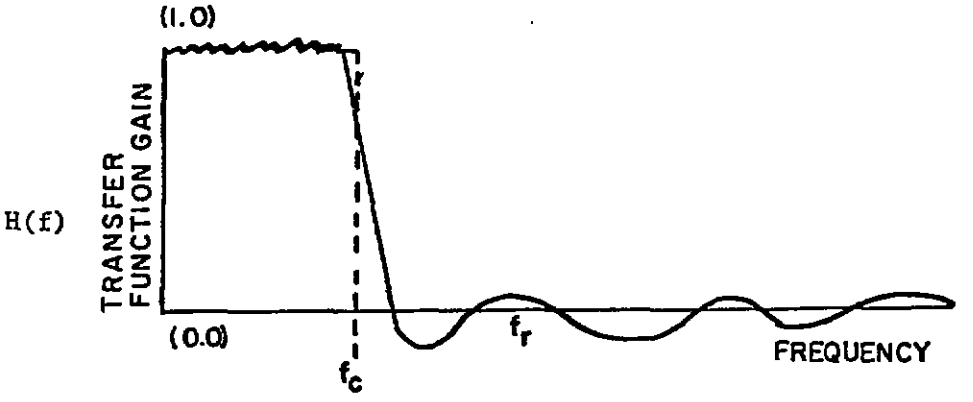
5) Frequency constraining

The straight moving average is comparable to a center point estimate from a first degree polynomial in item (2). The polynomial curve fitting technique gives a single point estimate from a polynomial curve fit to n consecutive data points. The estimate may be end point, center point, or any other point on the curve. The constrained least squares uses the classical approach but adds constraint conditions on the filter parameters based on some known information concerning the data preceding the current filter span. The variable span filter is any of the filter types with the added characteristic that the filter span is adjusted ("opened up" or "closed out" and "reinitialized") based on some characteristics of the input data.

All linear digital filters are frequency-constraining in that they suppress certain frequencies, pass other frequencies without distorting, and may distort or amplify other frequencies.

Because of the low frequency trend characteristics in trajectory data, the filters used to smooth this data are of the "low pass" type. That is, the filter weights are designed to suppress high frequencies and to output the low frequency components with a minimum amount of distortion. In many filters such as the least square polynomial type, the cut-off frequency is determined as a function of the number of points in the filter span. As a number of points in the filter span becomes larger, the cut off frequency becomes lower, i.e., more smoothing is applied to the data. There are other filters, however, which construct the weights for a specific cut-off frequency, and the increase in the filter span does not change the cut-off frequency of the filter, but rather improves the filter's frequency response corresponding to the desired cut-off frequency.

In addition to low-pass filters, weights can be constructed for high-pass, band-pass, and band-reject filters. All filters mentioned thus far can be evaluated by estimating their frequency response function H(f). The frequency response is estimated by taking the Fourier Transform of the weights W_i as a function of i. A typical frequency response versus the ideal frequency response or step function is given on the following page:



An example of some frequency response functions was discussed in Section 4.6 and illustrated in Figure 5. The frequency response function is very important when it comes to error analysis, because knowledge concerning the cut-off frequency, the high-frequency "leakage" (side lobe effect), the sharpness of filter roll-off at the cut-off frequency, and the distortion or amplification of the low frequencies, can indicate whether or not the appropriate filter was used for a given set of data. The frequency response for a finite number of weights will never follow the ideal response curve. The frequency response oscillates or has side lobes at the higher frequencies because the filter is truncated in the time domain, and significant weights W_i have been discarded, causing discontinuities in the time response of the filter. These time discontinuities in the time domain result in the oscillations in the frequency domain at the higher frequencies. To avoid this result, a function must be chosen which decays very quickly in the time domain, so that truncation at a reasonable span length will discard data multipliers of much smaller magnitudes. Slowness of decay in the time domain is caused by discontinuities in the derivatives, particularly the zeroth and other low order derivatives, in the frequency domain. This suggests that, where possible, functions having all continuous derivatives be chosen for use.

The error introduced into the filtering process by the oscillations in the transfer function (curve) is inversely proportional to the number of points used in the filter and the roll-off frequency specified. However, there are means whereby this response can be controlled and a fixed cut off specified, and at the same time have a rapid roll off and very low passage in high frequency areas. Such filters are described in Reference 3.

A discussion of differentiation and prediction filters (see Reference 4) would be too lengthy to pursue in detail. Nevertheless, a few guiding comments will be added here. Two frequency filters with different cut-off frequencies can be "cascaded" in the data reduction process to achieve the digital equivalent of either a band-pass or a band-elimination filter, depending upon the arrangement of the data multipliers and of the filter outputs and residuals. To preserve the high rate of change in slope

and avoid filter lags at the acceleration discontinuities at staging, it becomes necessary to have a variable span filter, and collapse the filter weighting function into the discontinuities and build up the weighting function again on the other side. Unfortunately, such a process does not, in general, preserve the desired qualities of the function. Therefore, if an integration step accompanies the smoothing process, such as occurs in deriving velocity from acceleration, it is necessary either to filter through the discontinuity, thereby sacrificing the high-frequency response of the data, or to perform the integration process first and then filter, collapsing the filter when desired. However, this does require additional computations since the raw data is generally at a higher sample rate than the smoothed data, and in common with the prefiltering approach, careful editing of the raw data is required to prevent integration errors due to large noise spikes in the raw data (see Reference 2).

Another type of frequency digital filtering has become feasible and desirable due to a recently derived algorithm for spectral analysis called the Fast Fourier Transform (FFT) (Reference 27). Besides making digital spectrum analysis more attractive from an economic standpoint, the FFT has enabled many to change their concepts of digital filtering, in that the intellectually appealing approach of filtering in the frequency domain is now often simpler, faster, and just as effective as filtering in the time domain, although two transforms between the time and frequency domain are employed in the process. Suppose, for example, we desire to filter the data $X_1, X_2, \ldots, X_n$ with a low-pass filter at some cut-off frequency f_c. Instead of computing the weights for a time domain convolution we use the FFT to estimate the real and imaginary parts of the spectrum of the time series. Once we obtain the spectrum we truncate it at the appropriate cut-off frequency then take the inverse FFT. The resulting data in the time domain corresponds to the truncated spectrum and, as a result, has been "low pass" filtered with cut-off f_c.

In addition to the so-called general-purpose filters, there are special-purpose digital filters which obtain each filtered data point from the preceding data point using the recursive relationships and information concerning the variance of preceding data contained in the "state vector." The state vector describes the mathematical model of the total system generating the data which is to be filtered. For this filter to function in an optimum manner, the following must be given: (1) system parameters and their configuration, (2) state vector initial values and variance-covariance matrix of state variables, (3) time and effects of special events. The special-purpose filter is much more sophisticated from a mathematical standpoint than the general-purpose or simple digital filter. The special-purpose filter can be a very effective analysis tool especially when used in conjunction with multi-instrument solutions. One of the more well known special-purpose filters is the Kalman filter (see Reference 28).

9.1 EFFECTS OF SMOOTHING ON DATA

There is a diversity of opinion among people who process and analyze data concerning the many numerical methods used in smoothing measured quantities and in the retrieval of information from erroneous observations. The literature on this subject is voluminous. Of the many methods, all use assumptions on the functional form of the basic data trend or the statistical properties or origin of errors in an effort to obtain a numerical process which will, within the imposed constraints, improve the data by minimizing errors. The choice of a technique must depend upon the objectives sought. Some investigation of a given technique's potential in describing the data must be made. For example, statistical smoothing methods depend upon the character of errors, their distribution, their variances, and their dependence or independence. One might choose a second-degree moving arc polynomial over successive 51 point spans of 10 samples per second on position data to achieve a certain reduction in error variance.

Least squares polynomial smoothing has long been used as a general-purpose filter. How well the objectives may be achieved in choosing a smoothing technique may be determined by the knowledge of the functional equations of motion and the error characteristics. Figures 10 thru 17 contain important facts relative to polynomial smoothing and Figure 18 corresponds to a particular set of frequency-constraining digital filters. A description of each follows:

Figure 10 shows estimates of position and velocity error using a polynomial of degree two and various (N_s) smoothing intervals. Evaluation is at midpoint of smoothing interval.

Figure 11 gives the relative smoothing in velocity determinations obtained by using a span of 2n+1 points and evaluating at points other than the midpoint for the case where X can be expressed as a quadratic function of time and the errors in X are uncorrelated.

Figure 12 give curves for higher degree polynomial fits for both velocity and accleration errors.

Figure 17 gives the values

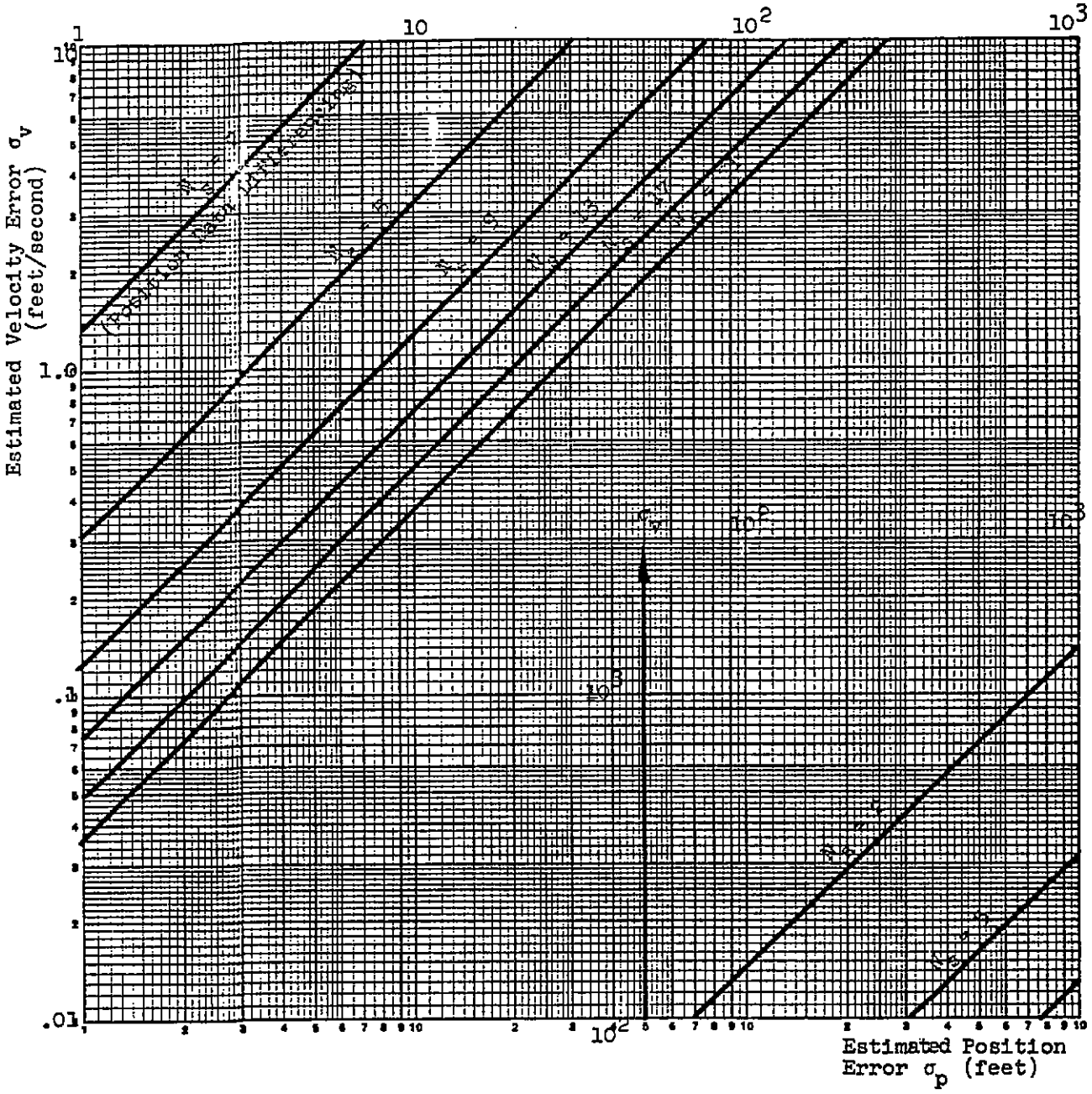
$$K_a = \left(\sum_{j=1}^{m} a_j^2 \right)^{\frac{1}{2}}, \qquad K_b = \left(\sum_{j=1}^{m} b_j^2 \right)^{\frac{1}{2}}, \qquad K_c = \left(\sum_{j=1}^{m} c_j^2 \right)^{\frac{1}{2}}$$

for center point smoothing as a function of the number N of points in the smoothing span.

Figure 18 gives the reduction factor in the standard deviation of the random error of position data. The family of curves represents low-pass filters with spans from 3 to 400 points with cut-off frequencies from 0.0200 hertz to 0.500 hertz.

Estimates of Errors in Velocity Data (σ_v) vs. Errors in Position Data (σ_p) at Midpoints of Dif. Quadratic Smoothing Intervals(N_s)

Figure 10



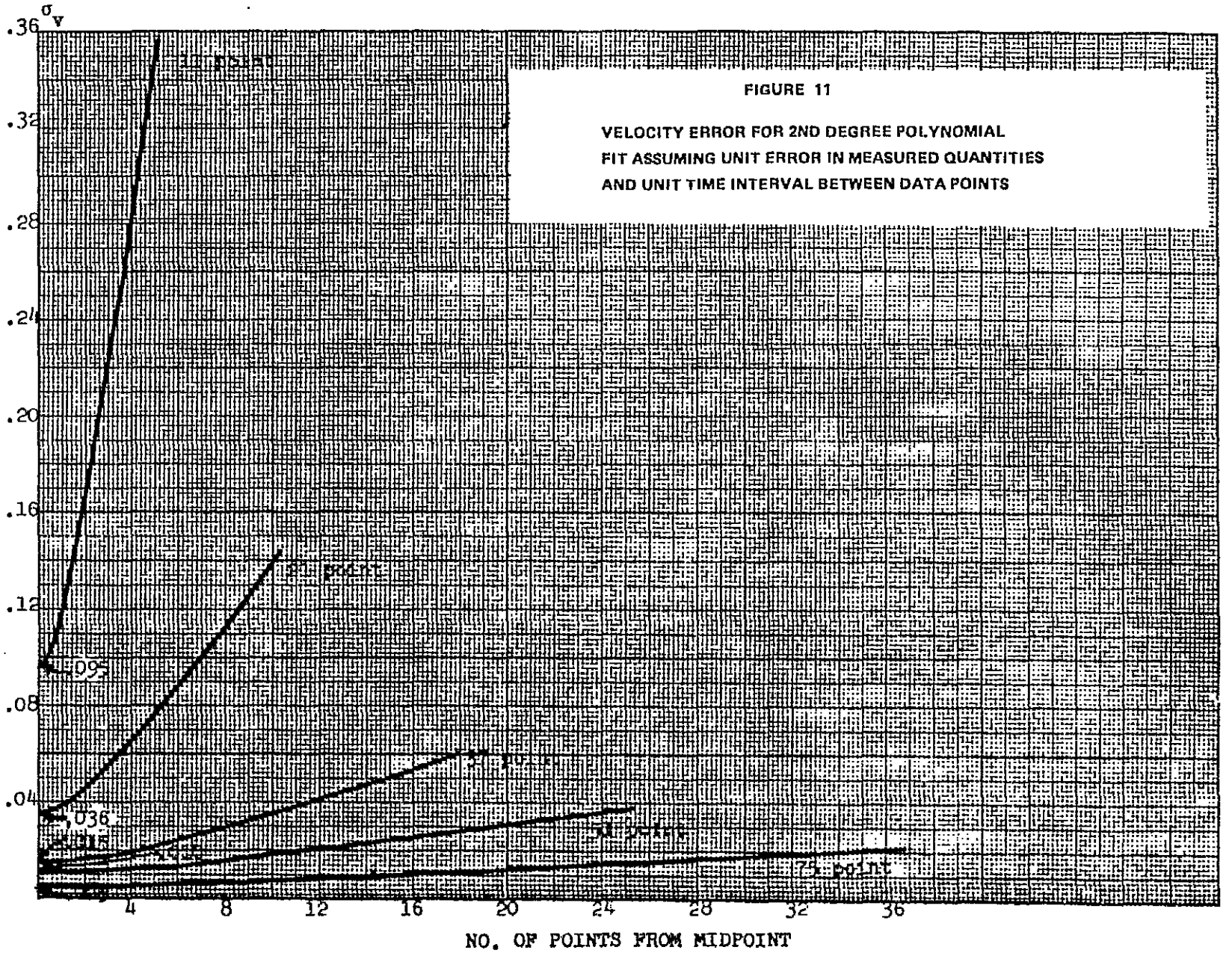


FIGURE 11

VELOCITY ERROR FOR 2ND DEGREE POLYNOMIAL FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES AND UNIT TIME INTERVAL BETWEEN DATA POINTS

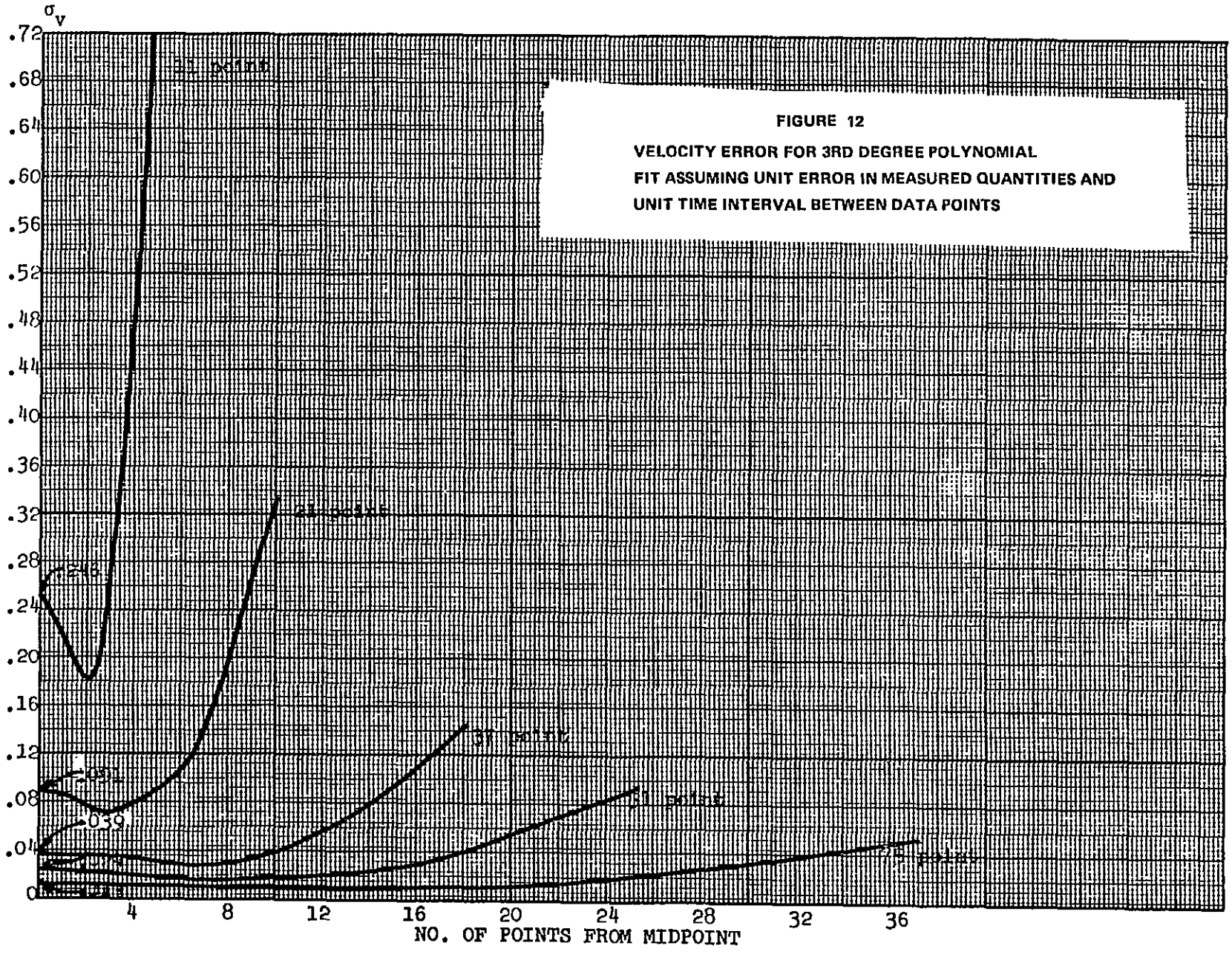


FIGURE 12

VELOCITY ERROR FOR 3RD DEGREE POLYNOMIAL FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES AND UNIT TIME INTERVAL BETWEEN DATA POINTS

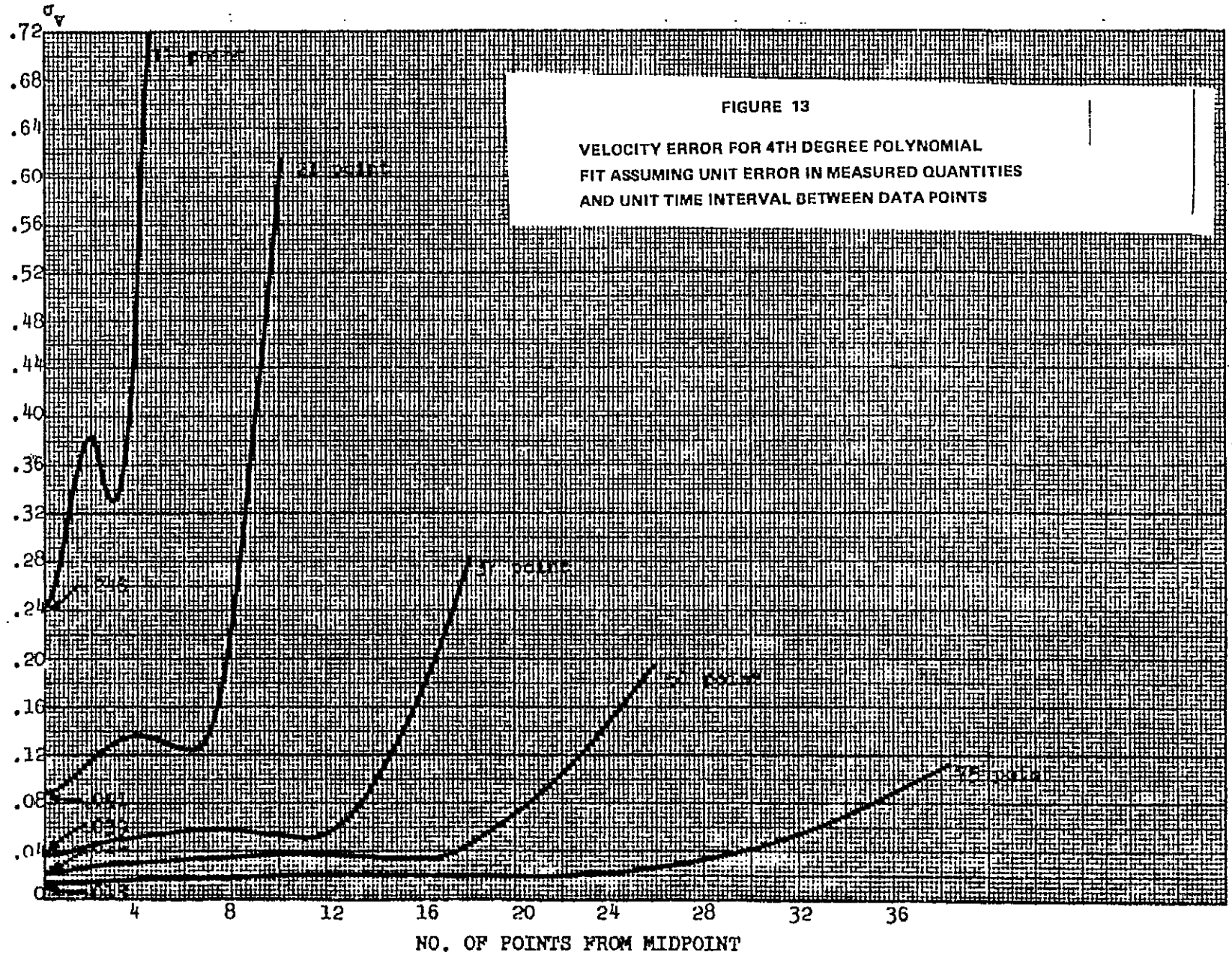


FIGURE 13

VELOCITY ERROR FOR 4TH DEGREE POLYNOMIAL
FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES
AND UNIT TIME INTERVAL BETWEEN DATA POINTS

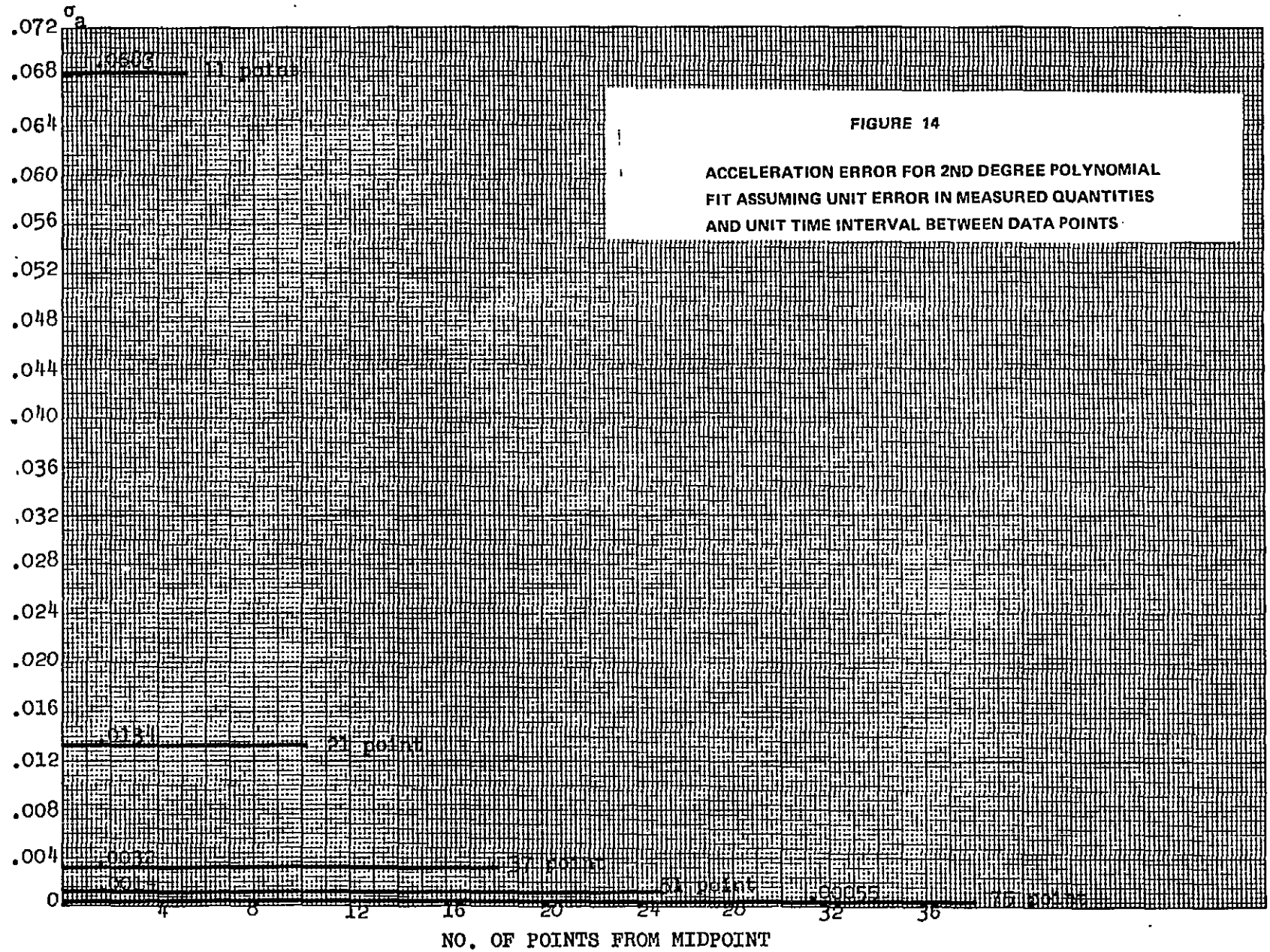


FIGURE 14

ACCELERATION ERROR FOR 2ND DEGREE POLYNOMIAL FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES AND UNIT TIME INTERVAL BETWEEN DATA POINTS

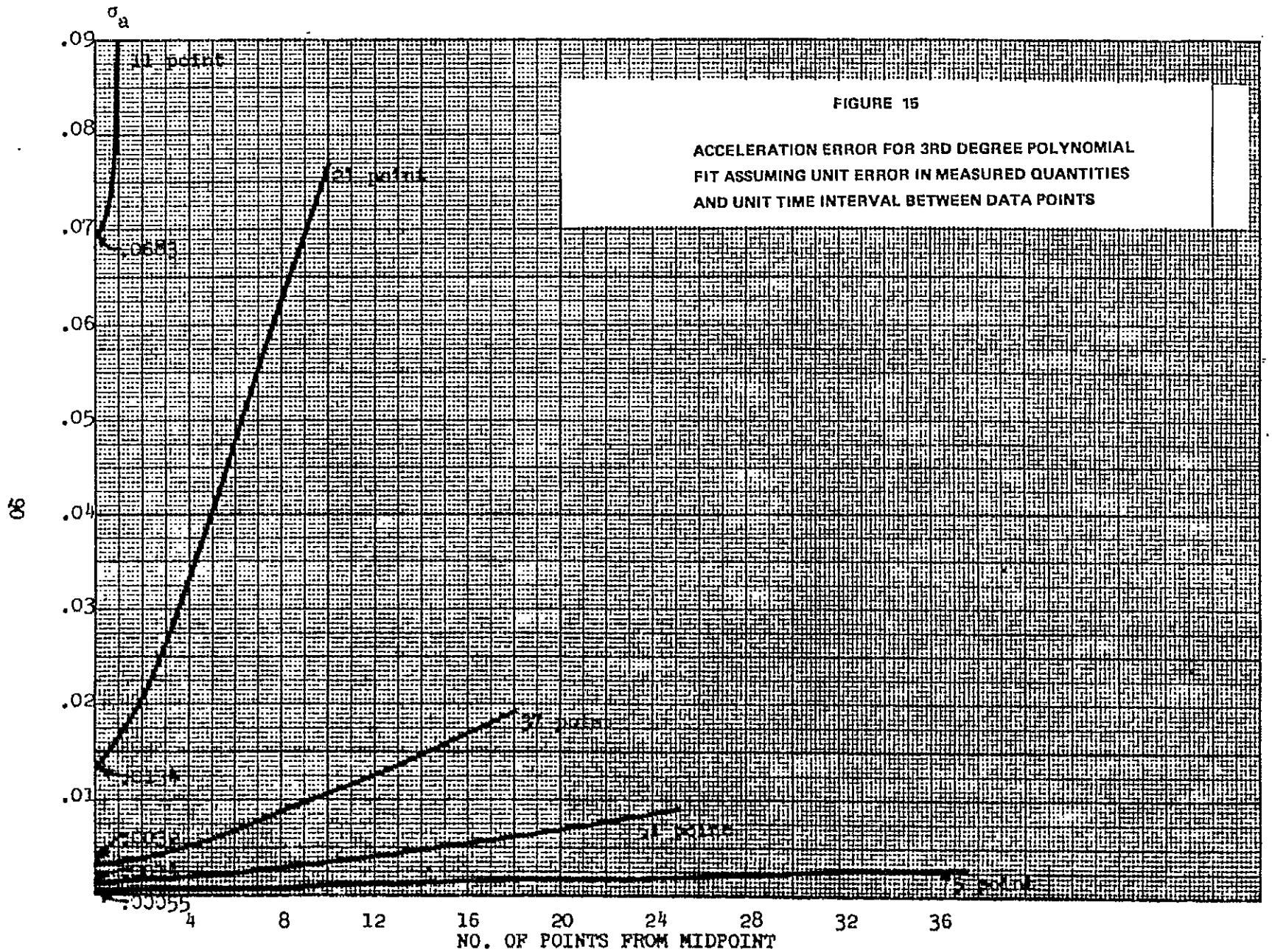


FIGURE 15

ACCELERATION ERROR FOR 3RD DEGREE POLYNOMIAL FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES AND UNIT TIME INTERVAL BETWEEN DATA POINTS

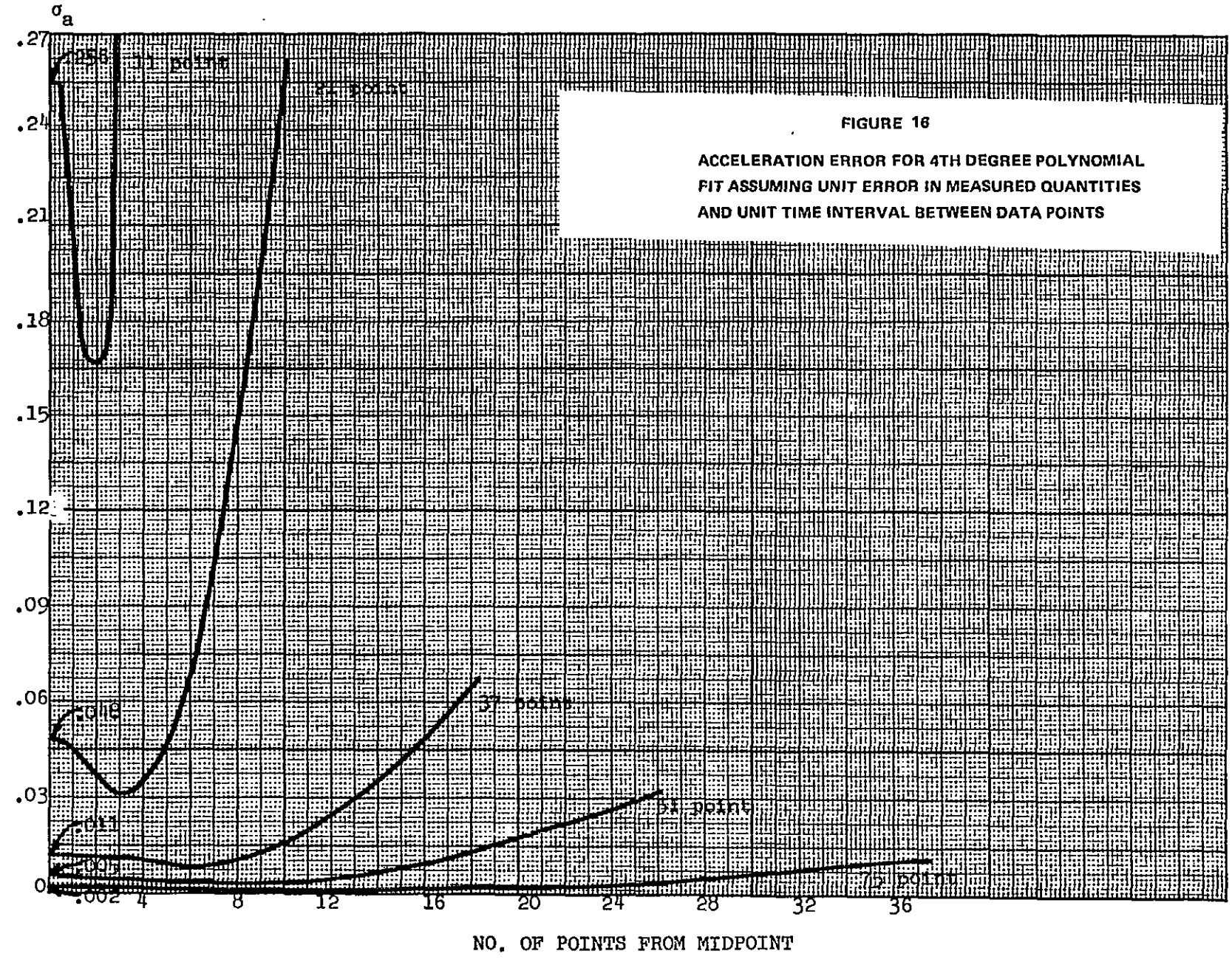


FIGURE 16

ACCELERATION ERROR FOR 4TH DEGREE POLYNOMIAL FIT ASSUMING UNIT ERROR IN MEASURED QUANTITIES AND UNIT TIME INTERVAL BETWEEN DATA POINTS

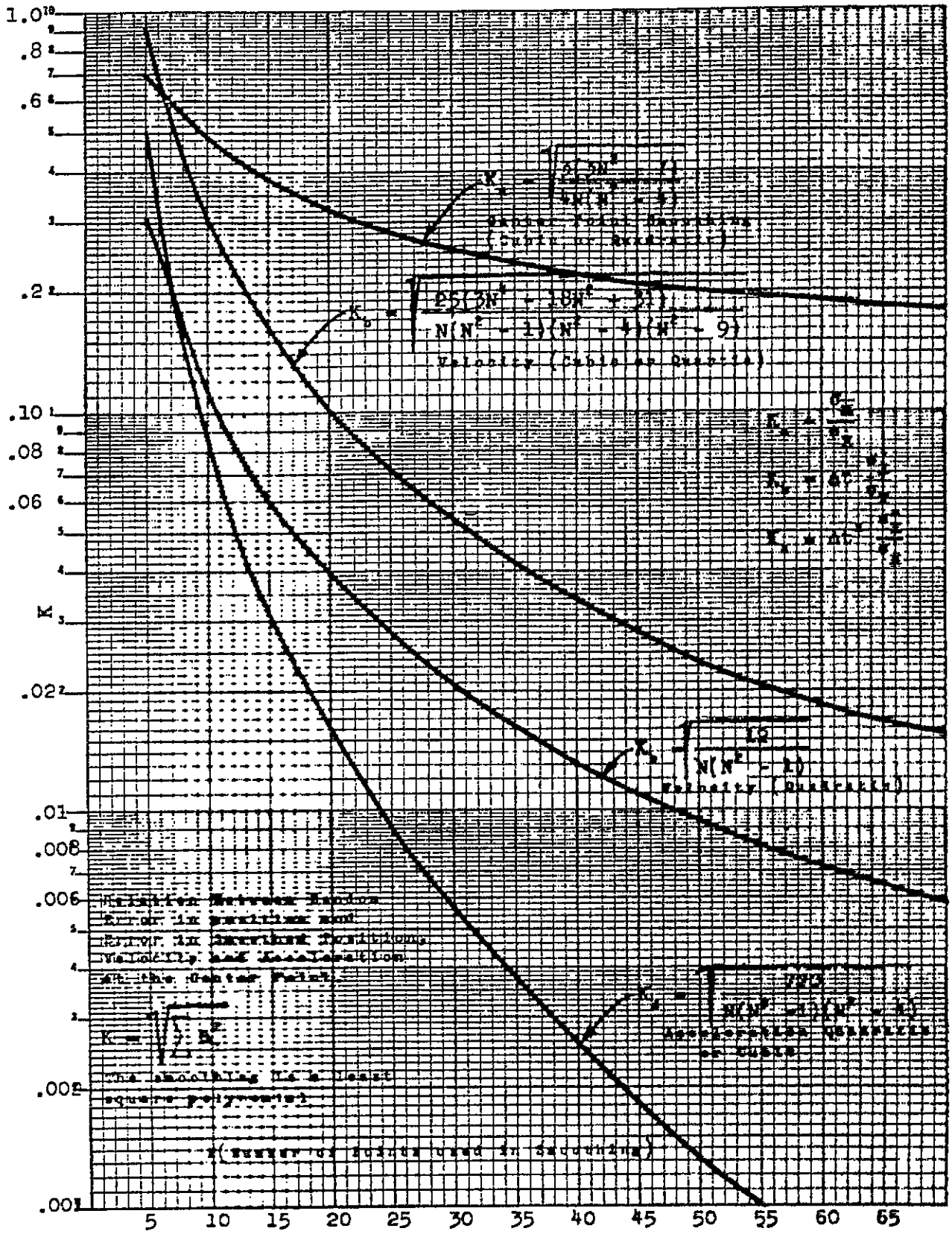


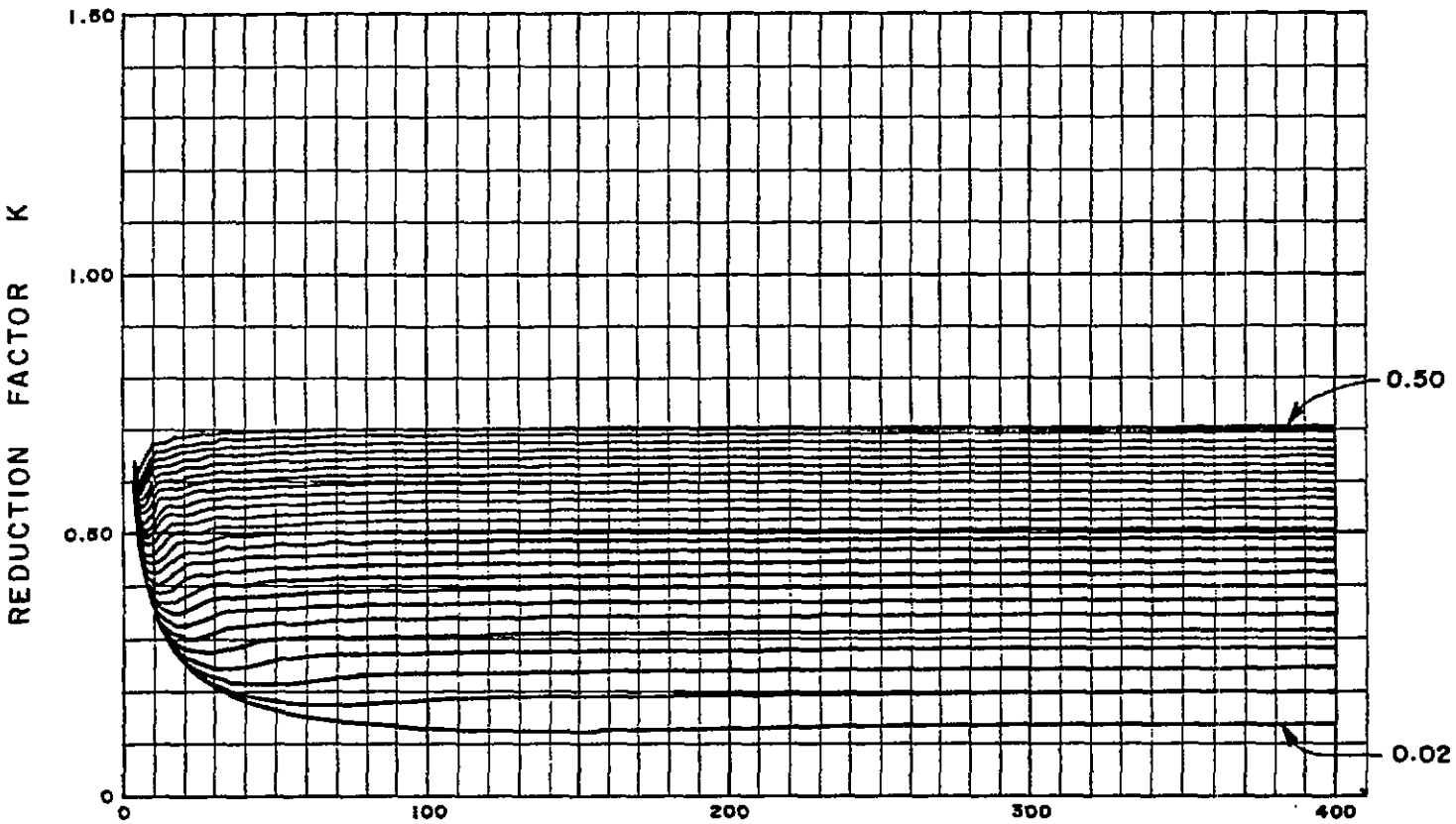
Figure 17 K_a, K_b and K_c as a function of N

FIGURE 18

VARIANCE REDUCTION CURVE FOR WEIGHTING FACTOR & NUMBER

FILTER TYPE = LOW

CUT-OFF FREQ = .0200 .1400 .2600 .3800 .5000
.0400 .1600 .2800 .4000
.0600 .1800 .3000 .4200
.0800 .2000 .3200 .4400
.1000 .2200 .3400 .4600
.1200 .2400 .3600 .4800



10.0 INERTIAL GUIDANCE DATA

Traditionally, attention and effort have been centered upon attaining velocity and accleration data from position data. Techniques for determining the accuracies of the derived quantities have been discussed. Recently, there has been a development of strong interest in utilizing data from accelerometers of guidance systems to assess velocity accuracies. Historically, this may appear to be a paradox. Development of the many tracking systems on the various ranges had, as a primary objective, the evaluation of guidance systems. Hence, it seems odd now to consider the use of guidance systems to evaluate tracking systems. However, the weight of accumulated evidence and experience has forced the conclusion that the quality, performance, and reliability of these systems are high. Nevertheless, the simple integration of acceleration data to produce velocities has an inherent disadvantage in that additional unknowns are introduced.

In view of the above, techniques have been developed for combining this data with that of other sources to establish trajectories. It has been suggested that errors in velocity arise from accelerometer errors such as bias, scale factor, quadratic nonlinearity, platform misalignment, gyro mass unbalance, and gyro constant drift. Errors from these sources are relatively small during the initial portions of flight, but after 50 seconds or so tend to exceed those present in data from the better CW Doppler interferometers. For a typical system, these errors have been identified as follows, in the order listed:

$$
\begin{aligned}
\Delta\dot{X}_0 &= K_0 t \quad , \quad K_0 = 30 \times 10^{-6} g \\
\Delta\dot{X}_1 &= K_1 \dot{X} \quad , \quad K_1 = 10 \text{ ppm} \\
\Delta\dot{X}_2 &= K_2 / a_x^2 dt \quad , \quad K_2 = 3 \text{ ppm/g} \\
\Delta X_3 &= K_3 \dot{Z} \quad , \quad K_3 = 2 \text{ sec.} \\
\Delta\dot{X}_4 &= K_4 \dot{Z}^2 \quad , \quad K_4 = 0.01 \text{ deg/hr/g} \\
\Delta\dot{X}_5 &= K_5 a_z t dt \quad , \quad K_5 = 0.05 \text{ deg/hr.}
\end{aligned}
\qquad (10.0.1)
$$

It may be observed that few of these are identifiable in functional form and in certain cases a high degree of correlation is present. Furthermore, some of these are time dependent and tend to grow with time. A factor not appearing above but one exerting considerable influence on the analysis of velocity is the propagation of tracking noise into the velocity vector. This also tends to grow with time, and hence, tends to obliterate the data towards the end of powered flight. It should be noted that on conventional solid-propellant vehicles, loss of guidance data is usual following missile burnout because of the dynamics of the vehicle attendant to thrust termination. Consequently, no free-fall flight data are ordinarily available.

For powered flight data analysis, velocity error estimates may be obtained directly by comparing the velocity data derived from the tracking systems positional data with inertial guidance velocity data. Inertial guidance data are subject to long term drift errors, but exhibit good characteristics in the

very low frequency errors (errors below, .05 cycles per second); whereas, the data from external systems are subject to errors in the high frequency and middle frequency regions (errors above .05 cycles per second), but exhibit good characteristics in the low frequency or near zero frequency region. It is this very fact, i.e., that these differences exist in the frequency characteristics of the two systems, that the guidance data may be used as a standard for noise estimation in the external data, and the external data, in turn, may be used as a standard for the estimation of long term drifts or biases in the guidance data.

The analysis of free-fall data may proceed along two lines. It may be assumed that the observed X, Y, Z data are serially correlated and processed through smoothing filters in order to obtain velocity data ($\dot{X}$, $\dot{Y}$, $\dot{Z}$) and acceleration data ($\ddot{X}$, $\ddot{Y}$, $\ddot{Z}$). Then estimates of the error in the derivative data may be obtained as a function of the variance estimate of the random + cyclic error in position. The position error variance which propagates into velocity and acceleration is a composite of the high and low frequency noise in the data. This error is also referred to as the random + correlated error, or perhaps most appropriately described as the total error minus the constant bias error. Constant bias errors in the position data introduce negligible effects in the derivative velocity and acceleration data, since by their very nature they tend to disappear in the process of differentiation. It is (cyclic) low frequency error which has the most deleterious effect on the quality of the derivative data. Smoothing over spans appreciably shorter than the typical low frequency does little to reduce its influence. The rigorous propagation of correlated errors through any specified smoothing filter requires a knowledge of the autocorrelation function of the errors. If the autocorrelation coefficients for successive values of X and 1, P_1, P_2, P_3..., the variance of the value of X obtained from the smoothing filter

$$X_i = \sum_{j=-m}^{m} b_j X_{i+j} \qquad (10.0.2)$$

is given by

$$\sigma^2_{x_i} = \sigma^2_x \, b \, R \, b^T \qquad (10.0.3)$$

in which b is the row vector of data multipliers

$$b = (b_{-m} \; b_{-m+1} \cdots b_{-1} b_0 \; b_1 \cdots b_{m-1} \; b_m) \qquad (10.0.4)$$

R is the auto-correlation matrix:

$$R = \begin{bmatrix} 1 & P_1 & P_2 & P_3 & \cdots & P_{2m} \\ P_1 & 1 & P_1 & P_2 & \cdots & P_{2m-1} \\ P_2 & P_1 & 1 & P_1 & \cdots & P_{2m-2} \\ P_3 & P_2 & P_1 & 1 & \cdots & P_{2m-3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ P_{2m} & P_{2m-1} & P_{2m-2} & P_{2m-3} & & 1 \end{bmatrix} \qquad (10.0.5)$$

and σ_x^2 is the variance of X_i.

Similarly, the variance of $\ddot{X}$ obtained from the smoothing filter

$$\ddot{X}_i = \sum_{j=-m}^{m} c_j X_{i+j} \qquad (10.0.6)$$

is given by

$$\sigma_{\ddot{x}_i}^2 = \sigma_x^2 \, c \, R \, c^T \qquad (10.0.7)$$

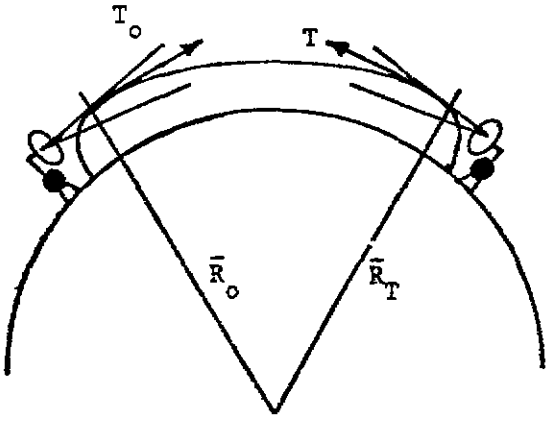
in which c is the row vector of data multipliers

$$c = (c_{-m} c_{-m+1} \cdots c_{m-1} c_m), \qquad (10.0.8)$$

R is the autocorrelation matrix.

An estimate of the autocorrelation function is best obtained from the residuals about a free fall ellipse (Keplerian ellipse) fitted to a long span (preferably 100 to 200 seconds) of post burnout data. The autocorrelation function may be obtained from the power spectrum analysis computer program.

There have been many difficulties associated with the use of aircraft and ballistic camera coverage on missile tests to calibrate external tracking system velocity errors to acceptable accuracies, which have led to the use of a second technique. This method consists of fitting the Keplerian equations of motion to data taken at both ends of the free flight trajectory, i.e., data taken by an up-range tracking system of the missile position after missile burnout and data taken by a down-range tracking system of missile position prior to atmospheric re-entry. Position parameters are fitted to a Taylor series expansion of the equations of motion truncated at the fourth power. Because the time between the two data spans is accurate, velocities obtained from the trajectory reconstructed in this manner will also be accurate (typically on the order of 0.1 fps) although there may be sizeable position uncertainties (typically 200-300 feet) in the data taken at each end.



Center of Earth

At $t = t_o$

$$\bar{R}_T = \bar{R}_o + \bar{V}_o(T - t_o) + \int_{t_o}^{T} \int_{t_o}^{t} \bar{g}\, dtdt. \qquad (10.0.9)$$

Hence,

$$V_o = \frac{\bar{R}_T - \bar{R}_o}{T - t_o} - \frac{1}{T - t_o} \int_{t_o}^{T} \int_{t_o}^{t} \bar{g}\, dtdt. \qquad (10.0.10)$$

The derivation of this result is not difficult. From the Figure, R_0 is the radius at injection time t_0 and T is time of re-entry where radius is $\bar{R}_T$. Since this portion of the trajectory is substantially outside the atmosphere the use of no-drag formulation with only gravity force acting is justified. Thus, acceleration is:

$$a = \bar{g}$$

or

$$\frac{dv}{dt} = \bar{g}$$

$$dv = \bar{g}\, dt \qquad (10.0.11)$$

$$V = \int_{t_0}^{T} \bar{g}\, dt + C_1$$

and

$$R = V_0(T - t_0) + \int_{t_0}^{T} \int_{t_0}^{T} \bar{g}\, dtdt + C_2$$

or (10.0.12)

$$\bar{R}_T = R_0 + V_0(T - t_0) + \int_{t_0}^{T} \int_{t_0}^{T} \bar{g}\, dtdt.$$

This may then be solved for V_0.

11.0 ERROR ANALYSIS IN THE FREQUENCY DOMAIN

A very effective tool to use in the evaluation of error is spectral analysis. It is assumed that the errors e_t are a realization from a stochastic process $\{\epsilon_t\}$ which is stationary in the wide sense. That is, the process has a constant mean and its autocovariance is independent of time, but is dependent on time lag. That is

$$E(\epsilon_t) = \mu(t) = \text{CONSTANT}$$

and (11.0.1)

$$R_{\epsilon_t}(k) = E[(\epsilon_t - \mu)(\epsilon_{t+k} - \mu)], \quad k = 0,1,2,\ldots,m.$$

The assumption of stationarity is generally true over subsets of the data, and if ϵ_t is not stationary it can usually be "prewhitened" or "detrended" so that nonstationary effects can be removed from the data. The variance of ϵ_t is obtained from R(k) when k = 0, i.e., $\sigma_{\epsilon_t}^2 = R(O)$. The autocorrelation function of ϵ_t is given by

$$\rho(k) = \frac{R(k)}{R(0)}.$$

It can be shown that if ϵ_t is a stationary process then the autocovariance has the following properties

$$\begin{aligned} &R(0) > 0 \\ &R(-k) = R(k) \\ &|R(k)| \leq R(0) \\ &\lim_{k\to\infty} R(k) = 0 \end{aligned} \qquad (11.0.2)$$

Further, it can be shown that the autocorrelation matrix

$$[\rho] = \begin{bmatrix} \rho(o) & \rho(1) & \rho(2) & \ldots & \rho(p) \\ \rho(-1) & \rho(0) & \rho(1) & \ldots & \rho(p-1) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(-p) & \rho(-p+1) & \ldots & & \rho(0) \end{bmatrix} \qquad (11.0.3)$$

is positive definite

The spectral density function for continuous ε_t is

$$f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} R(\tau) d\tau \tag{11.0.4}$$

and if ε_t is discrete (stochastic sequence) then

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\tau=\infty} e^{-i\omega\tau} R(\tau). \tag{11.0.5}$$

Conversely, if we know the spectral density $f(\omega)$ then

$$R(\tau) = \int_{-\infty}^{\infty} e^{i\omega\tau} f(\omega) d\omega \tag{11.0.6}$$

and

$$R(\tau) = \int_{-\pi}^{\pi} e^{i\omega\tau} f(\omega) d\omega \tag{11.0.7}$$

for the continuous and discrete cases respectively. In digital analysis it is clear we are interested in using the discrete case to approximate the continuous function, and from (11.0.7) we can see that the quantities

(11.0.8)

$$\frac{1}{2\pi} R(-\tau), \quad (\tau = 0, \pm 1, \pm 2, \ldots.),$$

are the Fourier coefficients of the function $f(\omega)$. In actuality ϵ_t is generally continuous but a realization e_t is usually discrete. In addition to the spectral density function, another important function in the frequency domain is the spectral distribution function

$$F(\lambda) = \int_{-\pi}^{\lambda} f(\omega)\,d\omega \tag{11.0.9}$$

which distributes the variance or power as a function of frequency increments. If the Fourier Transform of the autocorrelation function is taken, then the total area under the spectral density curve is unity. If the Fourier Transform of the autocovariance is taken, then the total area under the curve is $R(o) = \sigma^2$ (in this case the spectral density is often called the power spectrum).

In actual practice the autocovariance is approximated from n error estimates $e_1,\ldots,\epsilon_n$ over a total of $m < n$ lags.

$$\hat{R}_e^{(p)} = \frac{1}{n-p} \sum_{i=1}^{n-p} (e_i - \bar{e})(e_{i+p} - \bar{e}),\ p = 0,1,2,\ldots,M \tag{11.0.10}$$

where

$$R_e^{(p)} = R_e(p\Delta t),\ \Delta t = \text{time increment.}$$

The "raw" spectral density is estimated by numerically estimating the Fourier transform of $R_e^{(p)}$ using the trapezoidal rule

$$\hat{f}(s) = \frac{2\Delta t}{\pi} \sum_{p=0}^{M} \varepsilon_p R_e^{(p)} \cos \frac{sp\pi}{M},\ s=0,1,\ldots,M$$

where

$$\varepsilon_p = \begin{cases} 1, & 0 < p < M \\ 1/2, & p = 0, M \end{cases} \tag{11.0.11}$$

$$\hat{f}(s) = \hat{f}(\omega_s) = \hat{f}\left(\frac{s\pi}{M\Delta t}\right)$$

The raw spectral estimates are smoothed using a Hamming spectral window. In the frequency domain the Hamming weights give a three-point moving average and the filtered spectrum estimates are

$$\bar{f}(0) = .54\ \hat{f}(0) + .46\ \hat{f}(1)$$

$$\bar{f}(s) = .23\ \hat{f}(s-1) + .54\ \hat{f}(s) + .23\ \hat{f}(s+1) \qquad (11.0.12)$$

$$\bar{f}(M) = .54\ \hat{f}(M) + .46\ \hat{f}(M-1)$$

The Hamming weights are used to reduce the side band effect caused by discrete approximation when frequencies are not exactly on the discrete spectral estimate point. The frequency response of the Hamming weights and others are well known and there is much literature on the subject. Since

$$\omega_s = \frac{s\pi}{M\Delta t} = 2\pi f_s,$$

then the spectral density estimates will be in increments of $\Delta f = \frac{1}{2M\Delta t}$ for $0 \leq f \leq \frac{1}{2\Delta t}$. The frequency $\frac{1}{2\Delta t}$ is called the Nyquist or "folding frequency," f_N, and is the maximum frequency which can be resolved for a given sampling rate Δt. The Nyquist frequency is called the folding frequency because it can be shown that if significant frequencies exist in the data which are higher than $\frac{1}{2\Delta t}$ hertz, they will "fold" about the Nyquist frequency and appear on the spectrum as a lower frequency between 0 and $\frac{1}{2\Delta t}$ hertz. This is illustrated in Figure 19 below. For example, suppose ϵ_t is a continuous (analog) function which contains a 60 cycle "hum" component from a power source and it is digitized at $\Delta t = .05$ seconds. $\frac{1}{2\Delta t} = 10$ cycles $= f_N$ and the 60 cycle component would falsely appear as a zero frequency component or bias in ϵ_t. The folding of higher frequencies is called aliasing.

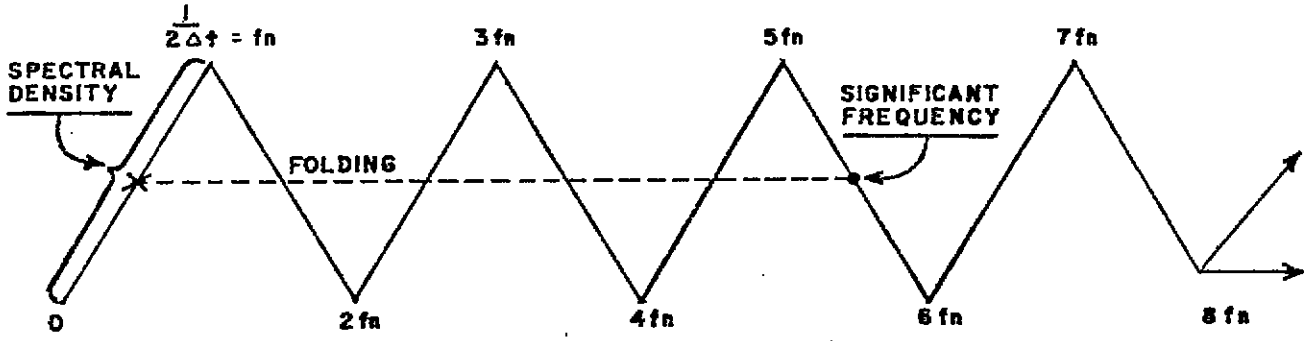


FIGURE 19 EFFECTS OF FOLDING OR ALIASING

The concept of aliasing immediately points to the fact that if the data are sampled with Δt such that $\frac{1}{2\Delta t}$ <significant frequencies, erroneous frequencies will appear in the data. The effects of these frequencies on the error in position, velocity and acceleration data can be highly significant. If the data are such that a large Δt (low sampling rate) is required, then one must be assured that significant frequencies do not exist above the Nyquist frequency. This can be assured by prefiltering the data using a higher sampling rate and a low pass filter with a cut-off frequency equal to the Nyquist frequency corresponding to the desired low sampling rate. This is sometimes necessary when spectral analysis on low frequency error components is to be accomplished.

Since the relation between the autocovariance and the spectral density function is known, we can utilize these two functions in the analysis of many types of data and for error analysis specifically. For this, several examples of autocovariance and spectral density functions corresponding to different error models are given (see Figure 20).

Suppose $e_1,\ldots,e_n$ is a realization from a random process which is independent (uncorrelated). Since the e_i are discrete, the realization will have a low pass frequency band and since the realization is uncorrelated, it is called white noise. In this case

$$R(k) = \begin{cases} \sigma^2, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{11.0.13}$$

Then, according to (11.0.2) the corresponding spectral density is

$f(\omega)=\frac{\sigma^2}{2\pi}$. Thus stationary sequences of uncorrelated random errors are characterized by the fact that their spectral densities are constant over the interval $\omega = 0$ to $\omega = \pi/\Delta t$. An example of the estimated correlation of white noise from radar data is given in Figure 21.

Suppose now that the realization $e_1, e_2, \ldots, e_n$ is correlated with autocovariance

$$R(k) = \sigma^2 \rho^k, \quad |e| < 1. \tag{11.0.14}$$

Autocorrelation is $\frac{R(k)}{R(0)} = \rho^k$ and this form of correlation corresponds to a first order Markov process, i.e., $\epsilon_t = \rho\, \epsilon_{t-1} + \eta_t$. The autocorrelation decays exponentially and from (11.0.5) it can be shown that

$$f(\omega) = \frac{\sigma^2}{2\pi} \frac{1 - e^2}{|e^{i\omega} - e|^2} \tag{11.0.15}$$

which is an exponential type spectral density function (see Figure 20).

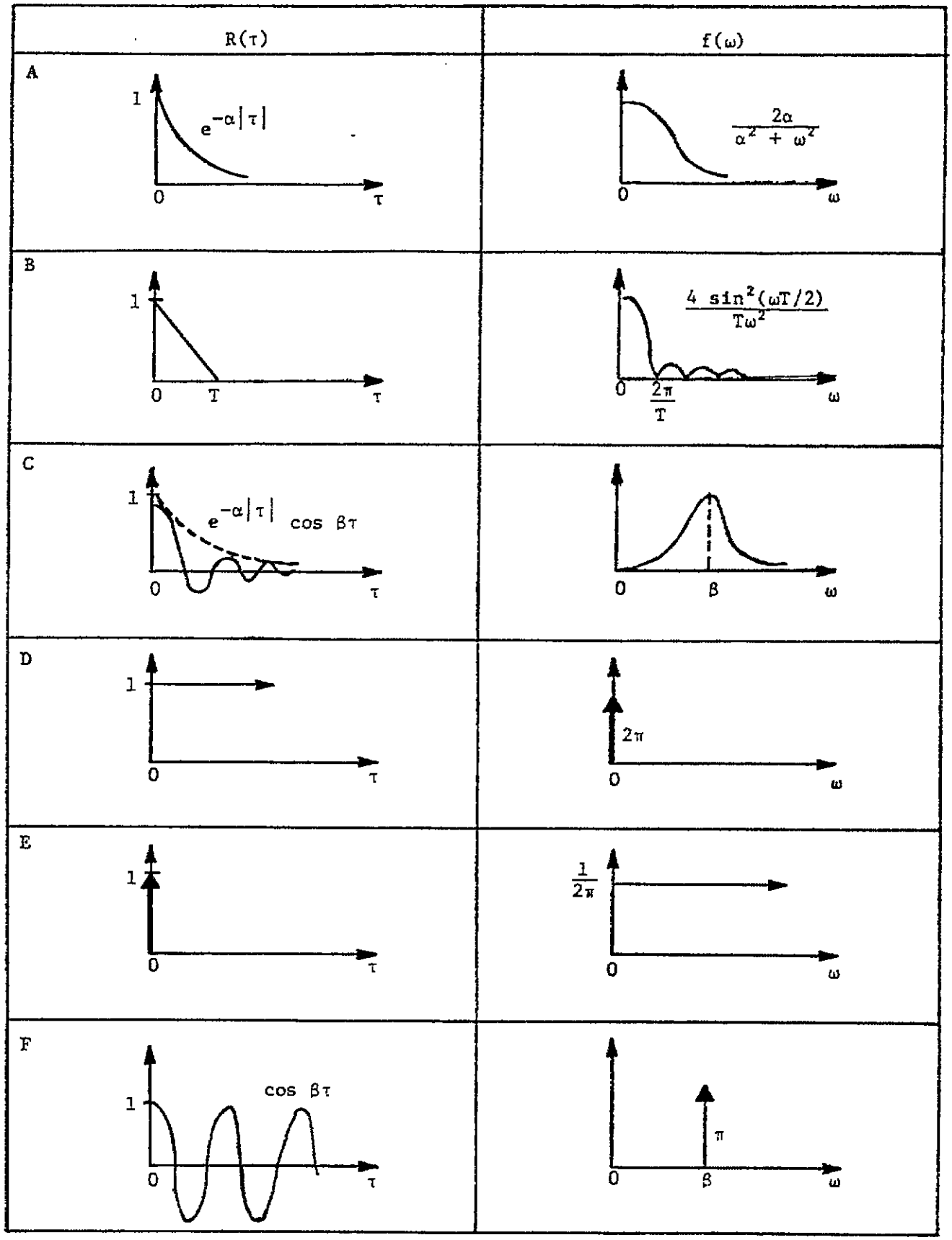


Figure 20 Examples of some continuous autocorrelation functions and corresponding spectral density functions. Types A and C are the continuous analogs to a first and second order Markov Process respectively.

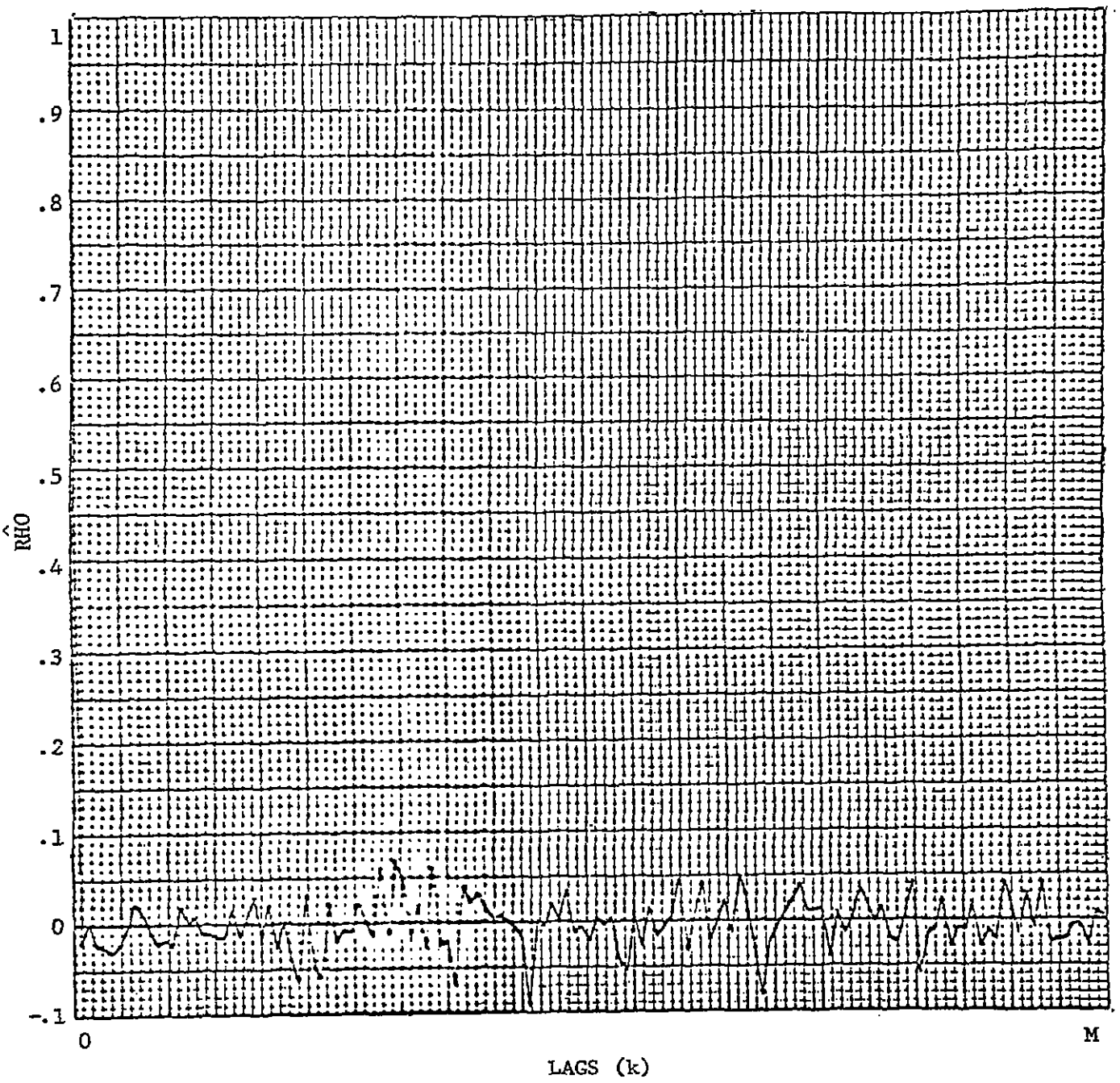


FIGURE 21 Estimate of autocorrelation function of independent random error (white noise).

Markov processes can be estimated by the so-called autoregressive scheme. For example, the autocorrelation and the power spectrum can be estimated for a particular set of error estimates. If the shape of the autocorrelation and power spectrum indicate that the underlying stochastic process is Markov, it can be approximated by an autoregressive scheme. To accomplish this, the error estimates are staggered in the following manner

e_5	e_4	e_3	e_2	e_1
e_6	e_5	e_4	e_3	e_2
e_7	e_6	e_5	e_4	e_3
⋮	⋮	⋮	⋮	⋮
e_n	e_{n-1}	e_{n-2}	e_{n-3}	e_{n-4}

The first column is considered to be the dependent stochastic variable. The relation $e_i = \alpha_1 e_{i-1} + \alpha_2 e_{i-2} + \alpha_3 e_{i-3} + \alpha_4 e_{i-4} + \eta_i$ is then estimated by regression techniques. An analysis of variance can indicate which coefficients are significant, and then the process can be repeated using the appropriate order for the Markov process. An example of such a technique is illustrated in the autoregressive fit in Figure 23. The corresponding autocorrelation estimate is in Figure 22. An analysis of variance corresponding to the autoregressive scheme is tabled in Figure 24. The analysis of variance indicates the process is of the first order because the regression coefficients a_2, a_3, a_4, are not significant. Figure 23 gives the graph of the approximation of η_i. If $e_1,\ldots,e_n$ is a realization from the process $\epsilon_t = a_1 \epsilon_{t-1} + a_2 \epsilon_{t+2} + \eta_t$, η_t is independent random variable, then it can be shown the autocorrelation function is an exponentially damped harmonic of the form.

$$\rho_k = \left[a_2^{k/2} \sin (kV + W)\right] / \sin W \qquad (11.0.16)$$

where

$$V = \cos^{-1}\left[\frac{-a_1}{2a_2^{\frac{1}{2}}}\right]$$

$$W = \tan^{-1}\left[\frac{(1 + a_2)\tan V}{1 - a_2}\right]$$

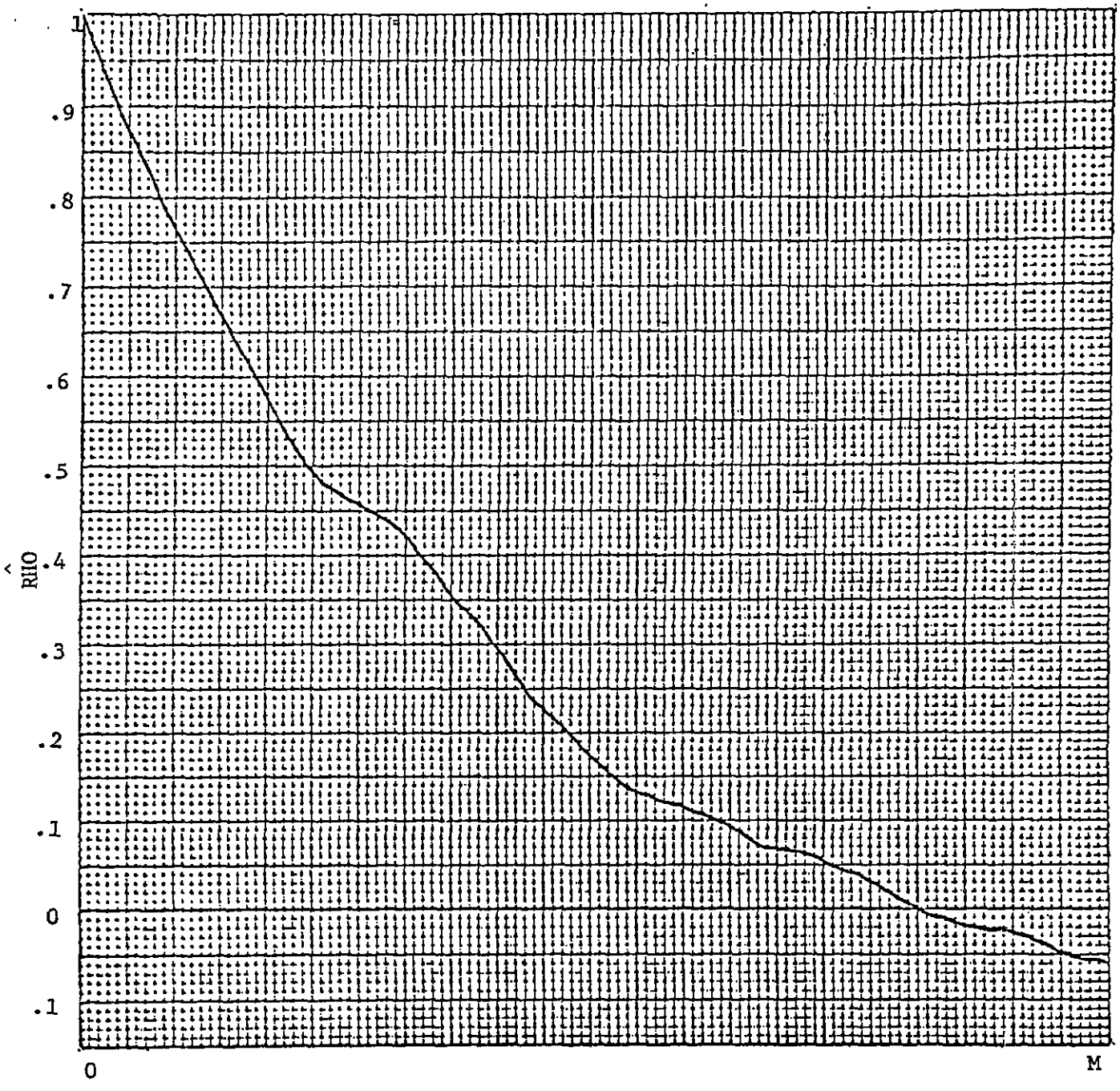
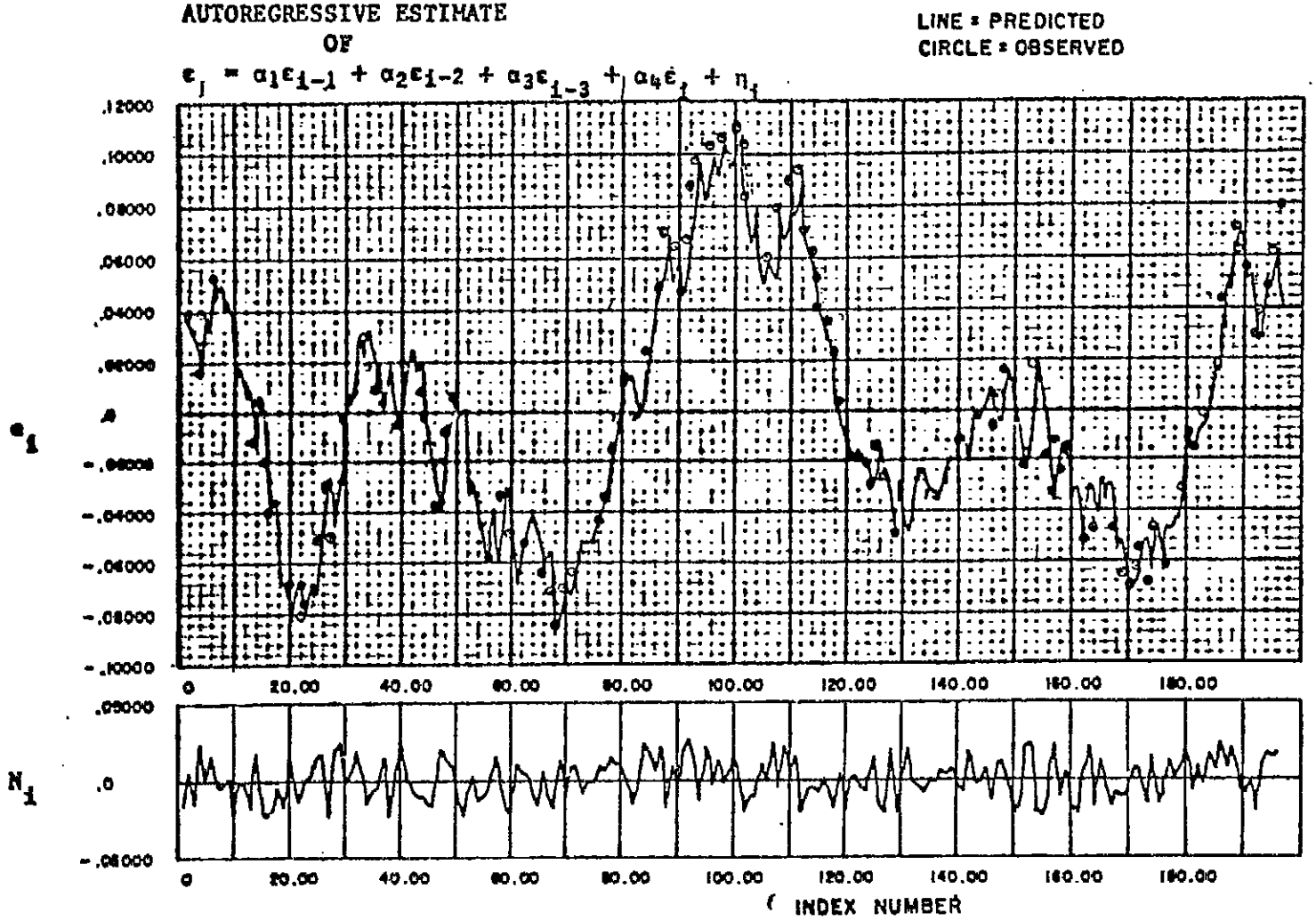


FIGURE 22 Estimate of autocorrelation function for first order MARKOV process with $\rho = \alpha_1 = 0.99$. The estimate at LAG k=1 is 0.97. Note the error in the estimate for large lags. RHO is not supposed to become negative for this particular process.



COEFFICIENT ESTIMATES:

A(0) =	.00005
A(1) =	.96846
A(2) =	.15390
A(3) =	.15523-
A(4) =	.01164-

FIGURE 23 Estimate of Markov process using 4th order autoregressive scheme. The regression coefficients indicate that the process is first order with $\rho = .96848$. See Figure 24 for analysis of variance.

ANALYSIS OF VARIANCE FOR THE AUTOREGRESSION SCHEME

SOURCE OF VARIATION	D.F.	SUM OF SQUARES	MEAN SQUARES	F VALUE
DUE TO REGRESSION	4	0.42199	0.10550	570.1688
DEVIATION ABOUT REGRESSION . .	191	0.03534	0.00019	
TOTAL . .	195	0.45733		

VAR. NO.	MEAN	STD. DEVIATION	REG. COEFF.	STD. ERROR REG. COE.	COMP. T	PARTIAL CORR. COE.	S.S.	PROP. VAR. CUM.
1	-0.00172	0.04830	0.96848	0.07262	13.33615	0.69439	0.42096	0.92048
2	-0.00167	0.04838	0.15390	0.10022	1.53557	0.11043	0.00000	0.00001'
3	-0.00143	0.04875	-0.15523	0.09948	-1.56043	-0.11220	0.00102	0.00223
4	-0.00117	0.04909	-0.01164	0.07200	-0.16164	-0.01170	0.00000	0.00001
5	-0.00164	0.04843						

OCOMP. CHEDK ON FINAL COEFF. -0.01164

FIGURE 24 The F value indicates autoregressive scheme is a significant fit. The computed T value shows that the only significant regression coefficient is a_1. The amount of explained variation due to $a_1 e_i$ is 90% of the total variation.

The spectral density corresponding to ρ_k will be a band of frequencies about ω_o =W and the general form of the spectral density will be

$$f(\omega) = \frac{A(\omega^2 + b^2)}{W^2 + 2a\omega^2 + b^4}$$

where

$$A = \frac{k}{\pi}$$

$$b = \sqrt{k^2 + W^2} \qquad (11.0.17)$$

$$a = k^2 - W^2$$

The continuous case of damped oscillatory covariance function is

$$R(\tau) = \sigma^2 e^{-\alpha|\tau|} \cos \omega_o \tau \qquad (11.0.18)$$

Another example of random error would be the so-called moving average of a sequence of uncorrelated random variable. The stochastic model is

$$\varepsilon_j = \sum_{k=1}^{n} a^k \eta_{j-i}, \ |a| < 1, \qquad (11.0.19)$$

where the η's are independently distributed with zero mean. In this case it can be shown that in the limit as $n \to \infty$ the autocovariance function is of the form

$$R(k) = \frac{a^k}{1 - a^2} \tag{11.0.20}$$

which corresponds to a spectral density similar to (11.0.15) if $a = \rho$ and $\sigma^2 = \frac{1}{1 - a^2}$

The most general type of correlated noise is one where the correlation is varying and its graph has varying amplitude and periods. In this case the stochastic error model is

$$\varepsilon_t = \sum_{i=1}^{n} (a_i \cos \omega_i t + b_i \sin \omega_i t) \tag{11.0.21}$$

where the a_i and b_j are independent random variables with mean zero and variance $\sigma^2_{a_i} + \sigma^2_{b_i} = \sigma^2_i$. In this case the autocovariance function is

$$R(k) = \sum_{i=1}^{n} \sigma_i^2 \cos \omega_i k \tag{11.0.22}$$

and the spectral density function is a sequence of frequency bands made up of "impulses" corresponding to $\sigma_i^2/\Delta\omega$ at ω_i. The analytic form of $f(\omega)$ is

$$f(\omega) = \pi \sum_{i=1}^{n} \sigma_i^2 [\delta (\omega - \omega_i) + \delta(\omega + \omega_i)] \tag{11.0.23}$$

where δ is the Dirac delta function.

(11.0.21) is often the type of spectral density encountered in error analysis. An example of (11.0.23) is given in Figure 25.

Analysis of error in the frequency domain can be made more efficient from a computational standpoint if the fast Fourier transform is used (Reference 27). The autocovariance function can be obtained with less computation time by taking the fast Fourier transform, computing the modulus, smoothing, then taking the inverse fast Fourier transform.

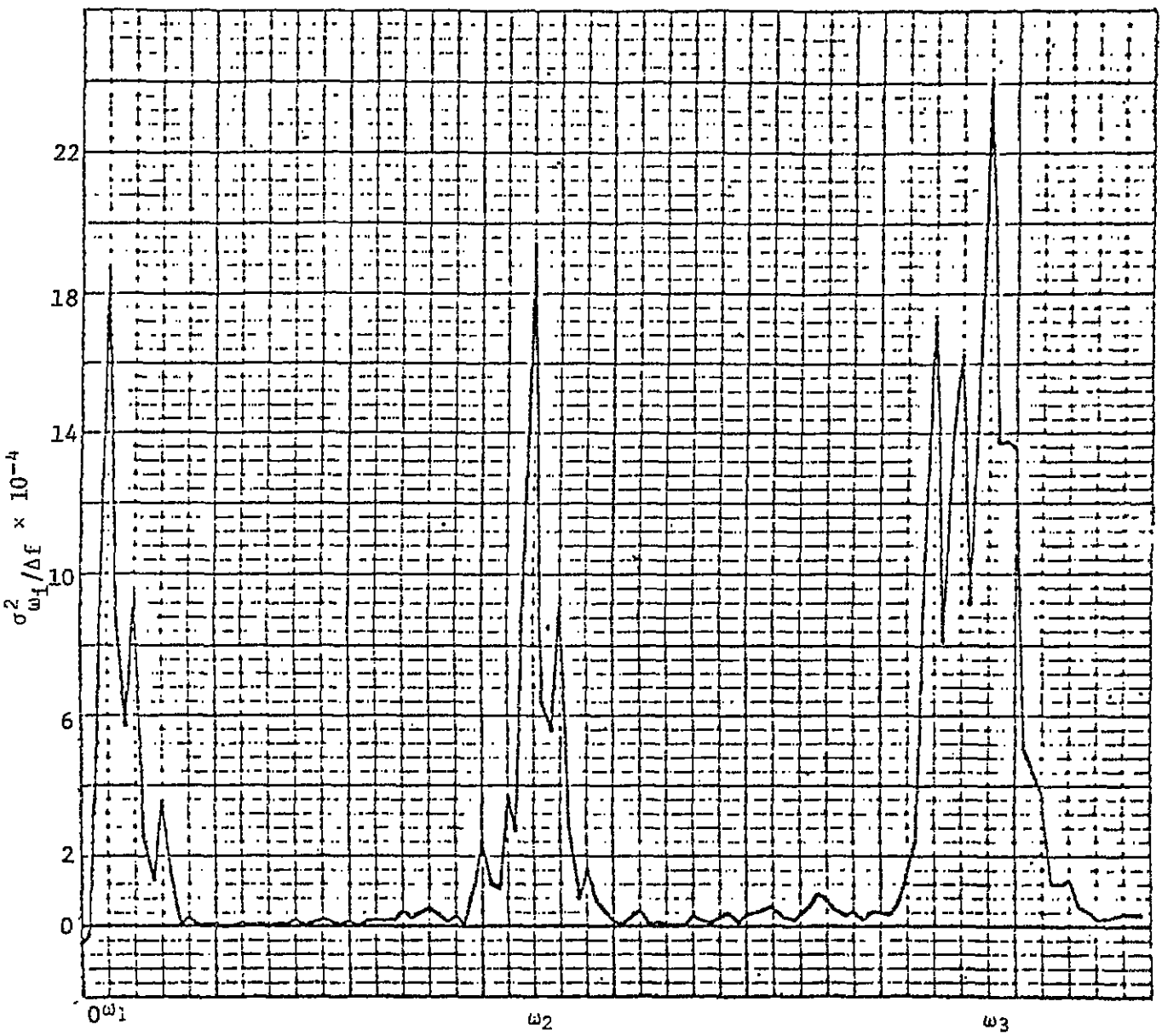


FIGURE 25 Estimate of the spectral density function for a stochastic process with random amplitudes and phases. The ordinate estimates the variance per frequency increment $\Delta f = \frac{1}{2M\Delta t}$.

Once the frequency content and autocovariance of the error ϵ_t have been estimated, one can filter the data to remove both high frequency random error and oscillatory systematic errors of moderately low frequency (say 0.05 hertz). In addition, the autocovariance function can be used to remove biased estimates of the variance of the error by extended least squares approximation. The analysis of data in the frequency domain can be utilized in conjunction with multiple regression and analysis of variance. For example, suppose the error is a stationary process of the Form (11.0.21) However, the significant frequencies ω_i are not necessarily harmonics and are not known. A spectral analysis would show significant frequency components and the corresponding squared amplitudes (see Figure 25). With k significant ω_i's selected from the spectrum, (11.0.21) could be approximated using multiple regression and solving for the a_i's and b_i's. An analysis of variance would then enable one to determine how good an approximation would be made on the expected values of a_i, b_j and their standard deviations σ_{a_i}, σ_{b_i}.

As an example a regression analysis was done in conjunction with error data corresponding to the spectrum estimates in Figure 25. The results of this analysis are given in graphical form in Figure 26. An analysis of variance indicated the regression analysis explained 82% of the total variation in the data. Finally, an estimate of the spectral distribution function corresponding to the spectral density function is given in Figure 27. The spectral distribution gives the estimate of variance as a function of frequency bands. From the graph we see that the total variance estimate corresponds to the intercept: $\hat{F}(0) - \hat{F}(\frac{1}{2\Delta t}) = 0.954$. The variance estimate over the band $\omega_A < \omega_3 < \omega_B$ is $F(\omega_B) - F(\omega_A)$ or $\sigma^2_{\omega_A-\omega_B} = 0.0365$.

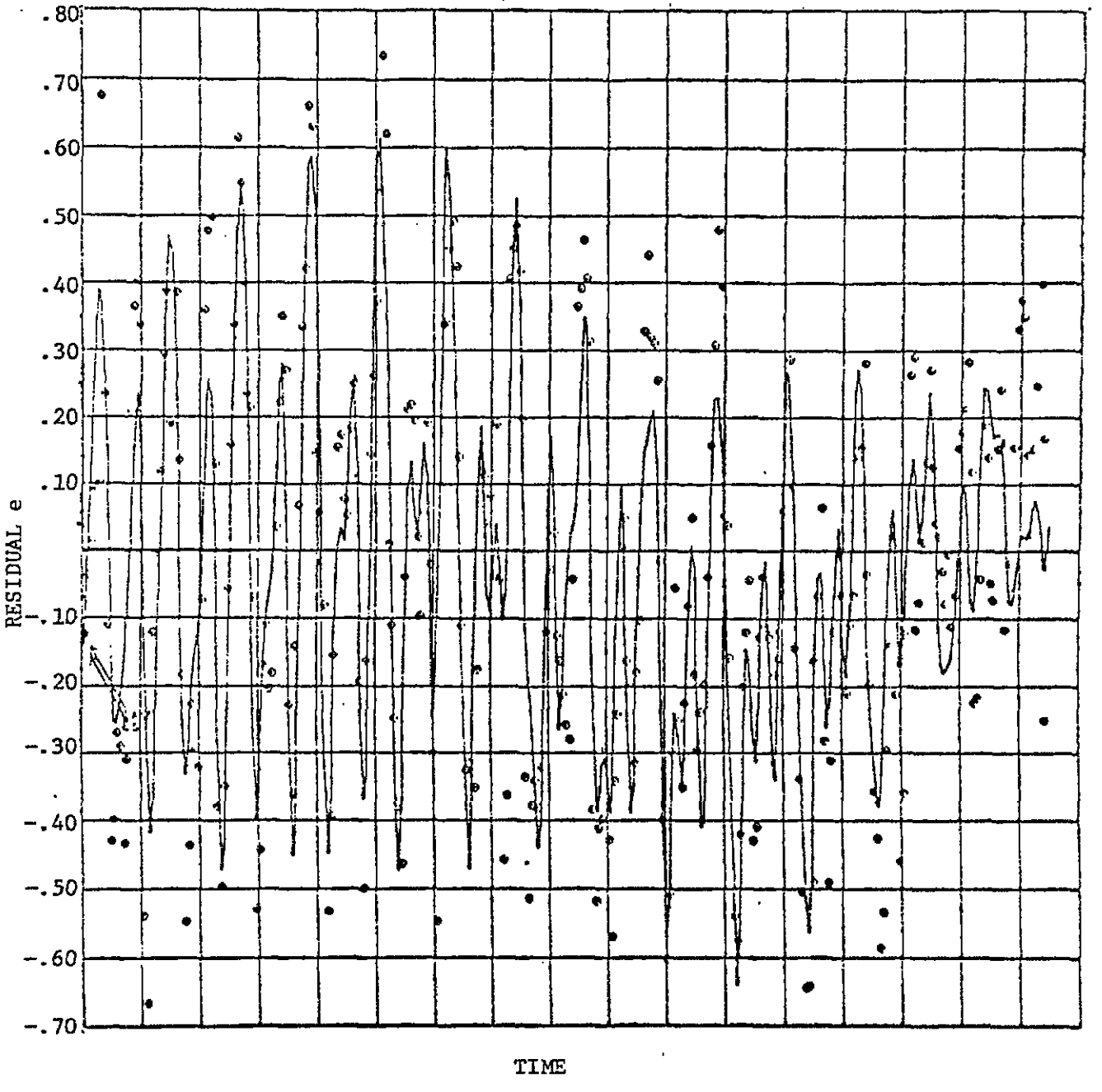


FIGURE 26 Regression analysis of data based on frequency information obtained from spectral density estimate (FIGURE 25). Line is predicted error while dots represent error measurements.

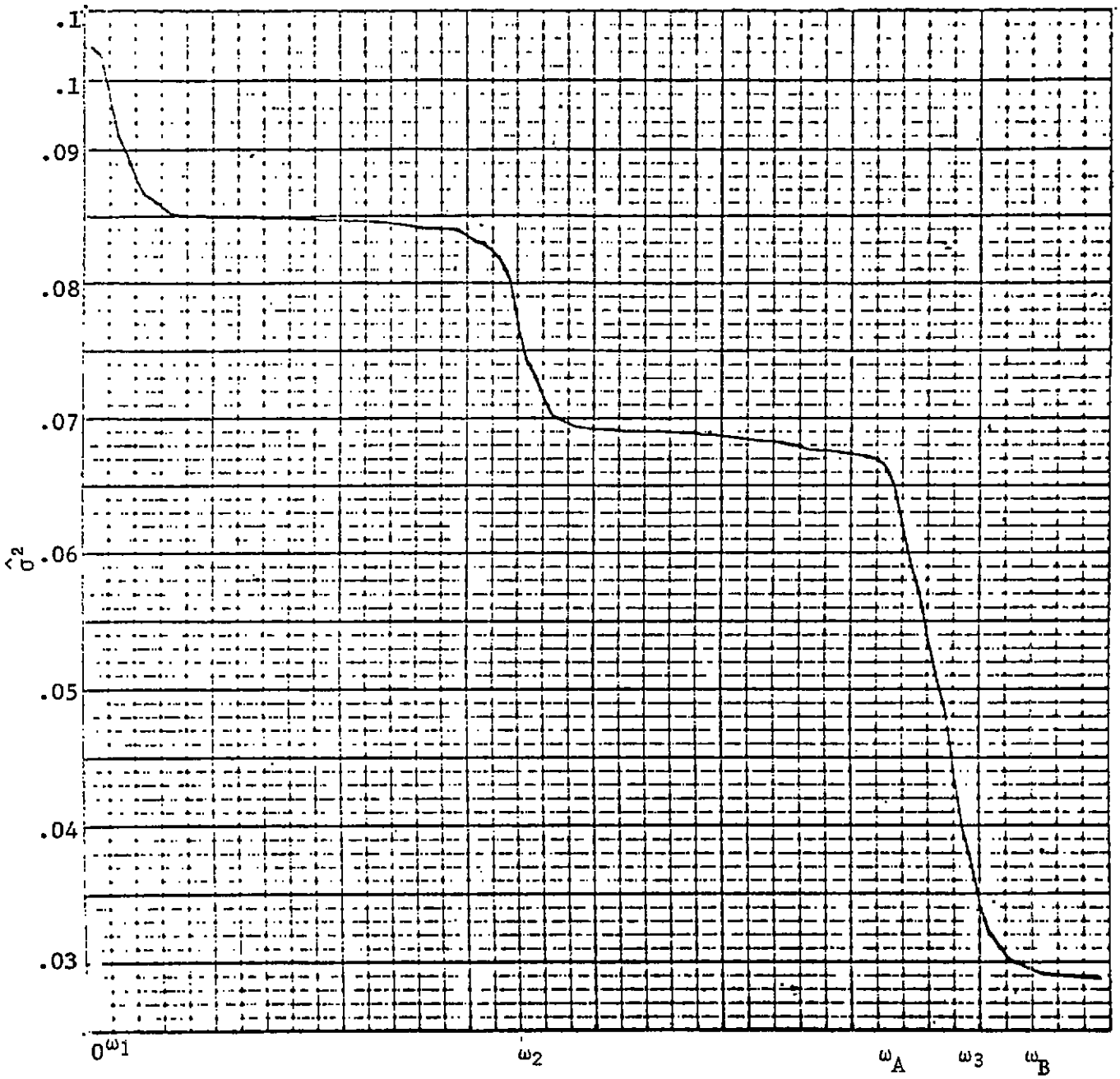


FIGURE 27 Estimate of spectral distribution function, F, corresponding to spectral density function in FIGURE 25.

BIBLIOGRAPHY

References

1. National Academy of Science -- Report of Ad Hoc Panel on Basic Measurements dated December 8, 1961, Washington, D. C.

2. Inertial Guidance -- University of California English and Physical Sciences Extension Series, edited by G. R. Petman, Jr., John Wiley and Sons, New York, 1962.

3. Gennery, D. G. and Garrett, J. R., Paper Fourth Joint AFMTC-Range Users Conference, Orlando AFB, Florida, February 1963.

4. Trimble, G. D., RCA/MTP Mathematical Services TM-63-3, Polynomial Filters with Continuity Constraints, AFMTC, October 1963.

5. Grubbs, F. E., "On Estimating Precision of Measuring Instruments and Product Variability," J.A.S.A., Volume 43, (1948) pp 243-264.

6. Simon, L. E., "On the Relation of Instrumentation to Quality Control," Instruments, Vol. 19, November 1946.

7. Moore, J. A., Thompson, W. A., "On the Problem of Negative Estimates of Variance," Paper at Sixth Conference on the Design of Experiments in Army Research, Development and Testing, BRL, APG, Maryland (October 1960).

8. Kingsley, O. L., "Analysis of Some Trajectory Measuring Instrumentation Systems," Paper at Sixth Conference on the Design of Experiments, BRL, APG, Maryland (October 1960).

9. Kendall, M. G., The Advanced Theory of Statistics, Vol II. Third Edition, C. Griffin & Company, LTD., London (1951).

10. Davis, R. C. "Techniques for the Statistical Analysis of Cinetheodolite Data." NAVORD Rpt 1299, China Lake, California (March 22, 1951).

11. Anderson, R. L., and Bancroft, T. A., Statistical Theory in Research, Maple Press, York, Pennsylvania (1952).

12. Morse, A. P., and Grubbs, F. E., "The Estimation of Dispersion from Differences," Annals of Math. Stat., Vol. 18, (1947), pp 194-214.

13. Deming, W. Edward, Statistical Adjustment of Data, John Wiley and Sons, New York.

14. Sterrett, John K., Manual for Moving Polynomial Arc Smoothing, BRL Report No. 840, dated November 1952, Aberdeen Proving Ground, Maryland.

15. Sterrett, John K., A Bibliography of Techniques for Smoothing or Adjustment of Data, BRL Technical Note No. 571, dated December 1951, Aberdeen Proving Ground, Maryland.

16. Brown, D., RCA Data Processing Technical Memo No. 54, "A General Solution to the Problem of Adjustment of Trajectory Observations," March 6, 1959.

17. Drucker, A., Space Technology Laboratories, STL/TN-60-000-19180, "Rocket Guidance Systems Test Analysis, Part I, Flight Test Analysis.

18. Eisenhart, C., Journal of Research, National Bureau of Standards, "Realistic Evaluation of the Precision and Accuracy of Instrument Calibration Systems." 67C, pp. 161-187, 1963.

19. Grenander, H., Statistical Analysis of Stationary Time Series, John Wiley and Sons, New York, 1957.

20. Mood, A. M. and Graybill, F. A., Introduction to the Theory of Statistics, McGraw-Hill Book Company, Inc., New York, 1953.

21. Parzen, E., Technometrics, 3, pp. 167-190, "Mathematical Considerations in the Estimation of Spectra," 1961.

22. Tukey, J. W., Probability and Statistics, "An Introduction to the Measurement of Spectra," John Wiley and Sons, Inc., New York, 1959.

23. Owen, D. B., Handbook of Statistical Tables, Addison-Wesley, Inc., Reading, Mass., pp. 138-140, 1962.

24. Technical Proposal for the Experimental Calibration of Range Instrumentation by the use of Geodetic Satellites, edited by M. A. Bondelid, Tech Note 3274-3-65, Pacific Missile Range, October 1965.

25. Morris, G. W., Polynomial Smoothing, Differentiation and Integration by Least Squares, Tech Note 3280-17, PMR, 1963.

26. Chew, V., Regression Analysis with Correlated Errors, RCA TM No. 64-14, AFETR, January 1964.

27. Cooley J., Lewis, P., Welch, P., The Fast Fourier Transform Algorithm and its Applications, IBM Research Paper RC-1743, February 9, 1967, Revised July 14, 1967, Yorktown Heights, New York.

28. Advances in Control Systems, Theory and Applications, edited by C. T. Leondes, Academic Press, New York, Vol. 3, pp. 219-292, 1966.

29. Thomas C. F., Sphere of Equal Probability for a Three Dimensional Normal Distribution, Tech. Note 3420-1-67, Pacific Missile Range, March 1967.

RCCS DOCUMENT TARGET BOX 01

DOCUMENT #: 154-71
(Formerly 120-71)

NAME: IRIG Standards for Distributing Raw Radar Antenna Data

TYPE: DOCUMENT

DATA REDUCTION AND COMPUTER GROUP

IRIG STANDARD 154-71
(Formerly 120-71)

IRIG STANDARDS FOR DISTRIBUTING RAW RADAR ANTENNA DATA

WHITE SANDS MISSILE RANGE
KWAJALEIN MISSILE RANGE
YUMA PROVING GROUND
DUGWAY PROVING GROUND
ELECTRONIC PROVING GROUND

ATLANTIC FLEET WEAPONS TRAINING FACILITY
NAVAL AIR WARFARE CENTER WEAPONS DIVISION
NAVAL AIR WARFARE CENTER AIRCRAFT DIVISION
NAVAL UNDERSEA WARFARE CENTER DIVISION NEWPORT

30TH SPACE WING
45TH SPACE WING
AIR FORCE FLIGHT TEST CENTER
AIR FORCE DEVELOPMENT TEST CENTER
AIR FORCE WEAPONS AND TACTICS CENTER
DETACHMENT 2, SPACE AND MISSILE SYSTEMS CENTER

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

IRIG STANDARD 154-71

IRIG STANDARD FOR DISTRIBUTING RAW RADAR ANTENNA DATA

Prepared by

Data Reduction and Computing Group
Range Commanders Council

September 1971

Published by

Secretariat
Range Commanders Council
White Sands Missile Range
New Mexico 88002

CONTENTS

PURPOSE - 1

APPLICATION - 1

LIMITATIONS - 1

STANDARD REFERENCES AND PROCEDURES - - - - - - - - - - - - - - 1

FIGURE 1, IRIG Raw Radar Antenna Data Format - - - - - - - - - - - - 2

FIGURE 2, IRIG Raw Radar Antenna Data Format Details - - - - - - - - - 3

FIGURE 3, Remote Tracking Constants Data - - - - - - - - - - - - - - - 7

IRIG STANDARD FOR DISTRIBUTING RAW RADAR ANTENNA DATA

PURPOSE

This document defines a standard procedure for distributing raw radar antenna data from remote tracking sites via teletype.

APPLICATION

The standard procedure set forth herein is recommended, where applicable, for the interchange of raw radar antenna data among the National and Service Ranges and other ranges associated with IRIG.

LIMITATIONS

This standard does not apply to the distribution of vector acquisition data to remote tracking sites.

STANDARD REFERENCES AND PROCEDURES

1. Coordinate System—Raw radar antenna data are referred to a polar earth-fixed, non-inertial, site tangent coordinate system.

2. Teletype Message Format—The IRIG Raw Radar Antenna Data Format in Figure 1 is used for low speed teletype transmission (100 words per minute) of digital data. One message consists of 47 characters and is transmitted at the rate of one message every six seconds. Each message will be preceded by and ended by special characters representing "Start of Message" and "End of Message." Full details associated with each field of the format are shown in Figure 2.

3. Remote Tracking Site Constants Data—Figure 3 contains detailed constants data relative to the conversion of raw radar antenna data into engineering units and geodetic site coordinates information.

Figure 1

IRIG RAW RADAR ANTENNA DATA FORMAT

Line 1	JJ
Line 2-n	NSSXQHHMMSSAAAAAAAEEEEEERRRRRRRRRRRRRRRR
Line n+1	#

Key:

JJ	=	Start of Message
N	=	Data Type
S	=	Remote Tracking Site Address
X	=	Radar Type
Q	=	On/Off Target Code
H	=	Hours, Greenwich
M	=	Minutes, Greenwich
S	=	Seconds, Greenwich
A	=	Azimuth
E	=	Elevation
R	=	Range
$\dot{R}$	=	Range Rate
#	=	End of Message

Figure 2

IRIG RAW RADAR ANTENNA DATA FORMAT DETAILS

Line No.	Character No.	Character	Note	Description
				Start of Message
1	1	J		Always the Same
	2	J		Always the Same
	3	↑		Figure Shift
	4	<		Carriage Return
	5	⫟		Line Feed
	6	↓		Character Shift
				47 Character Message
2-(N)	1	⫟		Line Feed
	2	↑		Figure Shift
	3	↑		Figure Shift
	4	N	1	Indicates Data Type
	5	S	2	Remote Tracking Site Address (Tens)
	6	S	2	Remote Tracking Site Address (Units)
	7	X	3	Indicates Radar Type
	8	Q		On/Off Track Code — On = 2; Off = 0
	9	H		Tens of Hours
	10	H		Units of Hours
	11	M		Tens of Minutes

Line No.	Character No.	Character	Note		
	12	M		Units of Minutes	
	13	S		Tens of Seconds	
	14	S		Units of Seconds	
	15	A	4	Azimuth Bits	21, 20, 19
	16	A		Azimuth Bits	18, 17, 16
	17	A		Azimuth Bits	15, 14, 13
	18	A		Azimuth Bits	12, 11, 10
	19	A		Azimuth Bits	9, 8, 7
	20	A		Azimuth Bits	6, 5, 4
	21	A		Azimuth Bits	3, 2, 1
	22	E	4	Elevation Bits	21, 20, 19
	23	E		Elevation Bits	18, 17, 16
	24	E		Elevation Bits	15, 14, 13
	25	E		Elevation Bits	12, 11, 10
	26	E		Elevation Bits	9, 8, 7
	27	E		Elevation Bits	6, 5, 4
	28	E		Elevation Bits	3, 2, 1
	29	R	4	Range Bits	27, 26, 25
	30	R		Range Bits	24, 23, 22
	31	R		Range Bits	21, 20, 19
	32	R		Range Bits	18, 17, 16
	33	R		Range Bits	15, 14, 13
	34	R		Range Bits	12, 11, 10
	35	R		Range Bits	9, 8, 7
	36	R		Range Bits	6, 5, 4
	37	R		Range Bits	3, 2, 1

Line No.	Character No.	Character	Note		
	38	R	4	Range Rate Bits	27, 26, 25
	39	R		Range Rate Bits	24, 23, 22
	40	R		Range Rate Bits	21, 20, 19
	41	R		Range Rate Bits	18, 17, 16
	42	R		Range Rate Bits	15, 14, 13
	43	R		Range Rate Bits	12, 11, 10
	44	R		Range Rate Bits	9, 8, 7
	45	R		Range Rate Bits	6, 5, 4
	46	R		Range Rate Bits	3, 2, 1
	47			Carriage Return	
				End of Message	
(n+1)	1	=		Line Feed	
	2	↑		Figure Shift	
	3	#		# Symbol	
	4	↓		Character Shift	

Note 1: This character may be varied through any of the figure characters at the option of the computer operator to meet specific documented requirements. Normally, however, this character is defined as follows:

Character 4	Data Type
2	Real Time Data
3	Simulated Data
7	Last Sample

Note 2: These characters are used to identify the remote tracking site from which the raw radar antenna data was derived. The remote tracking site addresses are assigned and controlled by SAMTEC (DOTS) for this IRIG Standard.

Note 3: This character may be varied through any of the figure characters at the option of the computer operator to meet specific documented requirements.

Note 4: See Figure 3 for Azimuth, Elevation, Range, and Range Rate Least Significant Bit weights. Range Rate Bit #27 represents Sign: 0 = increasing and 1 = decreasing.

FIGURE 3

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Fischer, 1960)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.XXXXXXXXXXXX	X.XXXXXX	.XXXXX	XXX°XX'XX.XXXX"	XXX°XX'XX.XXXX"	XXXX.XX	XXXX.XX
AIR FORCE EASTERN TEST RANGE									
Patrick AFB, FL 32925									
Cape Kennedy	AN/FPS-16 (S/N-11)		0.048828125	1.953125	N/A	28-28-54-0694	80-34-36.2102	-80.95	44.17
Ascension Island	AN/FPS-16 (S/N-14)		0.048828125	1.953125	N/A	(-)7-57-6.5477	14-24-44.6124	213.65	302.97
Range Measurement Lab	AN/FPS-16 (S/N-19)		0.048828125	1.953125	N/A	28.13-38.6320	80-36-22.7691	-79.22	46.85
Range Measurement Lab	AN/FPS-16 (S/N-43)		0.048828125	1.953125	N/A	26-36-56.4886	78-20-52.9075	-80.18	48.92
Pretoria, South Africa	AN/MPS-25 (S/N-5)		0.048821125	1.953125	N/A	(-)25-56-36.6257	(-)28-21-29.5952	5292.24	5196.86
Patrick AFB	AN/FPQ-6 (S/N-1)		0.006103515625	1.953125	N/A	28-13-35.2993	80-35-58.2106	-77.02	49.02
Antigua	AN/FPQ-6 (S/N-2)		0.006103515625	1.953125	N/A	17-8-37.7717	61-47-33.9426	13.13	138.78
Ascension	AN/TPQ-18 (S/N-1)		0.006103515625	1.953125	N/A	(-)7-58-23.0373	14-24.5.3384	322.04	411.34
Grand Turk	AN/TPQ-18 (S/N-2)		0.006103515625	1.953125	N/A	21-27-45.7842	71-7-56.1089	-44.34	118.11
Merritt Island	AN/TPQ-18 (S/N-3)		0.006103515625	1.953125	N/A	28-25-29.2132	80-39-52.6127	-88.68	36.91
T-AGM 20 USNS Redstone	AN/FPS-16 (S/N-49)		0.048828125	1.953125	N/A	As Assigned	As Assigned		

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES Based on Clarke 1866		SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			ELEV & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)		
			.xxxxxxxxxxx	x.xxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx° xx' xx.xxxx	xxxx.xx	xxxx.xx
AIR FORCE FLIGHT TEST CENTER									
Edwards AFB, CA 93523									
Edwards AFB	AN/FPS-16 (S/N-38)		0.048828125	1.000000	N/A	34-58-10.5952	117-55-45.8582		2703.08
Edwards AFB	AN/FPS-16 (S/N-41)		0.048828125	1.000000	N/A	34-58-13.6315	117-55-50.0201		2721.58
Edwards AFB	NIKE		0.097656250	1.953125	N/A	34-56-47.2421	117-54-37.6828		2561.10
Edwards AFB	AN/MPS-19C		0.097656244	1.220703	N/A	34-57-38.6066	117-54-38.3760		2663.95
Edwards AFB	AN/MPS-19S		0.097656244	1.220703	N/A	34-56-43.3456	117-54-44.9942		2570.30
Edwards AFB	NASA-1					34-56-59.313	117-53-11.524		2339.67
Ely, NV	Ely-C					39-18-31.231	115-05-08.168		9264.275
Ely, NV	Ely - CAPRI					39-18-31.130	115-05-09.802		9269.037
Beatty, NV	NASA-2					37-04-54.036	116-49-02.938		4964.3
Ely, NV	NASA-3					39-18-31.047	115-05-08.299		9257.7

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Clarke 1866)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.XXXXXXXXXXX	X.XXXXXX	XXXXX	XXX°XX'XX.XXXX"	XXX°XX'XX.XXXX"	XXXX.XX	XXXX.XX
AIR PROVING GROUND CENTER									
Eglin AFB, FL 32542									
Site A-20	AN/FPS-16 (S/N-17)		0.048828125	1.953125	N/A	30-25-17.063	86-47-52.912		93.23
Site A-20	AN/FPS-16 (S/N-20)		0.048828125	1.953125	N/A	30-25-17.064	86-47-53.558		93.18
Site A-20	AN/FPS-16 (S/N-31)		0.048828125	1.953125	N/A	30-25-17.062	86-47-52.272		91.38
Site A-20	AN/FPS-16 (S/N-42)		0.048828125	1.953125	N/A	30-25-16.570	86-47-52.434		77.41
Site D-3	AN/FPS-16 (S/N-23)		0.048828125	1.953125	N/A	29-40-31.400	85-20-48.516		43.06
Site D-3	AN/FPS-16 (S/N-27)		0.048828125	1.953125	N/A	29-40-30.991	85-20-48.058		44.86
Site D-4	AN/FPS-16 (S/N-32)		0.048828125	1.953125	N/A	28-11-19.217	82-47-38.693		42.88
Site D-4	AN/FPS-16 (S/N-39)		0.048828125	1.953125	N/A	28-11-19.683	82-47-39.065		42.90

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Clarke 1866)			
			ELEV. & AZ (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xxxxxxxxxxxx	x.xxxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx°xx'xx.xxxx"	xxxx.xx	xxxx.xx
ELECTRONIC ENVIRONMENT TEST RANGE									
Ft Huachuca, AZ 85613									
Ft Huachuca	AN/FPS-16 (S/N-29)		0.0488266	0.500		31-34-15.246	110-22-12.551		4961.253
Ft Huachuca	Mod. II.V		0.0976551	0.488281		31-35-10.5841	110-21-48.3813		4719.752
Ft Huachuca	Mod. II.V		0.0976551	0.488281		Not Available	Not Available		
Oatman Mountain	AN/MPS-26 (S/N-6)		0.976551	1.00000		33-03-31.893	113-08-38.418		1737.146

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES		SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (± N)	LONGITUDE (± W)		
			.xxxxxxxxxxxx	x.xxxxxx	xxxxx	xxx°xx'xx.xxxx"	xxx° xx' xx.xxxx"	xxxx.xx	xxxx.xx
FLEET COMPUTER PROGRAMMING CENTER									
San Diego, CA 92147									
FCPC	AN/SPS-30		Not Available	Not Available	N/A	32-42-1.25	117-14-55.58		360.0
FCPC	AN/SPS-43		Not Available	Not Available	N/A	32-42-0.75	117-14-56.00		360.0

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES *		SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)		
			.xxxxxxxxxxxx	x.xxxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx° xx' xx.xxxx"	xxxx.xx	xxxx.xx
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION									
Kauai	AN/FPS-16 (S/N-30)		0.048828125	1.953125	N/A	22-07-35.825	200-19-53.962		3789.55
Bermuda	AN/FPS-16 (S/N-34)		0.048828125	1.953125	N/A	32-20-52.208	295-20-44.694		65.13
Bermuda	AN/FPQ-6 (S/N-5)		0.006103515625	1.953125	N/A	32-20-51.705	295-20-44.905		69.23
Wallops Island	AN/FPS-16 (S/N-8)		0.048828125	1.953125	N/A	37-50-28.393	284-30-52.378		40.66
Wallops Island	AN/FPQ-6 (S/N-4)		0.006103515625	1.953125	0.03125	37-51-36.509	284-29-25.236		49.06
Carnaron	AN/FPQ-6 (S/N-4)		0.006103515625	1.953125	N/A	(-)24-53-50.755	113-42-57.764		160.77
Tananarive	Capri		0.048828125	1.953125	N/A	(-)19-00-00.991	47-18-54.191		4390.96
Wallops	AN/MPS-26		0.048828125	1.220703125	N/A	As Assigned	As Assigned		
T-ACM-19 USNS Vanguard	AN/FPS-16 (S/N-47)		0.048828125	1.953125	N/A	As Assigned	As Assigned		

*Based on Clarke 1866, exceptions - Carnaron is based on Hayford International 1910
- Tananarive is based on Hayford International 1910

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	◆ RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Clarke 1866)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xxxxxxxxxxxx	x.xxxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx° xx' xx.xxxx"	xxxx.xx	xxxx.xx
NAVAL AIR TEST CENTER									
Patuxent River, MD									
Patuxent River	AN/M33		.04883	0.24414	600	38-17-48.887	76-22-34.449		35
Patuxent River	AN/M33		.04883	0.24414	600	38-17-50.407	76-22-34.079		26.20
Patuxent River	AN/MPS-19		.04883	0.24414	2800	38-17-49.187	76-22-35.054		35
Patuxent River	AN/MPS-19		.04883	0.24414	2800	38-17-49.877	76-22-35.581		26.01
Patuxent River	AN/MPS-26		.09766	1.000000	8000	38-17-49.926	76-22-35.328		35
Patuxent River	AN/MPS-26		.09766	1.000000	8000	38-17-51.116	76-22-33.940		25.89
Patuxent River	SCR-584		.04883	0.24414	Manual	38-17-52.141	76-22-34.103		25.66

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Fischer 1960)			
			ELEV. & AZ (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xx+xxxxxxxxx	x.xxxxxx	x*xxx	xxx°xx'xx.xxxx"	xxx°xx'xx.xxxx"	xxxx.xx	xxxx.xx
PACIFIC MISSILE RANGE									
Pt Mugu, CA 93041									
Pt Mugu	AN/FPS-16 (S/N-2)		0.00274658203125	1.0000000	N/A	34-07-17.7627	119-09-02.6212		41.32
Pt Mugu	AN/FPS-16 (S/N-3)		0.00274658203125	1.0000000	N/A	34-07-20.2121	119-09-08.7900		41.28
Pt Mugu	AN/FPS-16 (S/N-10)		0.00274658203125	1.0000000	N/A	34-07-21.5967	119-09-12.2764		41.17
Pt Mugu	AN/FPS-16 (S/N-24)		0.00274658203125	1.0000000	0.03125	34-07-22.9818	119-09-15.7641		41.18
Pt Mugu	AN/MPS-16 (S/N-3)		0.0054931640625	1.220703125	N/A	34-06-47.8925	119-07-19.7333		21.56
Pt Mugu	AN/MPS-16 (S/N-1)		0.0054931640625	1.220703125	N/A	34-06-48.1558	119-07-20.0448		20.72
Pt Mugu	AN/MPS-16 (S/N-4)		0.00274658203125	1.220703125	N/A	34-06-47.7777	119-07-19.3963		21.52
Pt Mugu	MOD II (S/N-23)		0.00274658203125	1.0000000	N/A	34-06-47.5631	119-07-16.0773		72.22
Pt Mugu	AN/FPQ-10 (S/N-5)		0.00274658203125	1.953125	N/A	34-06-46.7788	119-07-14.9957		72.60
			0.00140625	1.953125	N/A	34-06-46.7788	119-07-14.9957		72.60
Pt Mugu	AN/FPQ-10 (S/N-6)		0.00274658203125	1.953125	N/A	34-06-46.6527	119-07-16.4969		72.55
SNI	AN/FPS-16 (S/N-7)		0.00274658203125	1.0000000	N/A	33-14-49.7394	119-31-10.9605		930.05
SNI	AN/FPS-16 (S/N-13)		0.00274658203125	1.0000000	N/A	33-14-52.2873	119-31-13.2946		931.90
SNI	AN/FPS-16 (S/N-15)		0.00274658203125	1.953125	N/A	33-14-54.8352	119-31-15.6286		930.00
SNI	MOD II (S/N-12)		0.00274658203125	1.0000000	N/A	Not Available	Not Available		
SNI	AN/FPQ-10 (S/N-1)		0.00274658203125	1.953125	N/A	33-14-46.5444	119-31-08.1217		935.05
SNI	AN/FPQ-10 (S/N-4)		0.00274658203125	1.953125	N/A	33-14-47.1911	119-31-08.6250		935.98
Barsan	AN/MPS-26 (S/N-2)		0.00274658203125	1.220703125	N/A	22-01-32.0560	159-46-59.0867		29.60
Barsan	M-33 (S/N-689)		0.0439453125	2.0000000	N/A	22-01-30.4434	159-46-59.9275		29.40
Makaha	AN/FPQ-10 (S/N-2)		0.00274658203125	1.953125	N/A	22-08-04.3293	159-43-41.6469		1559.37
Makaha	AN/FPQ-10 (S/N-3)		0.00274658203125	1.953125	N/A	22-08-05.1624	159-43-41.5137		1544.85
Makaha	AN/FPQ-12		-XYZ OUTPUT ONLY-		N/A	22-08-06.7431	159-43-42.9511		1516.80

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS./SEC)	GEODETIC COORDINATES (Based on Fischer 1960)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xxxxxxxxxxxx	x.xxxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx°xx' xx.xxxx"	xxxx.xx	xxxx.xx
PACIFIC MISSILE RANGE (Cont)									
Pt Mugu, California									
Barsan	AN/MPS-26 (S/N-5)		0.00274658203125	1.220703125	N/A	22-01-31.1279	159-46-59.6072		29.52
Akau, JA	AN/MPS-25 (S/N-4)		0.00274658203125	1.0000000	N/A	16-45-36.0610	169-30-52.5275		22.45
Akau, JA	AN/MPS-25 (S/N-3)		0.00274658203125	1.0000000	N/A	16-45-36.6852	169-30-54.7195		25.34
Wheeling	AN/FPS-16 (S/N-46)		0.00274658203125	1.953125	N/A	As assigned	As Assigned		
Wheeling	AN/MPS-25 (S/N-2)		0.00274658203125	1.0000000	N/A	As Assigned	As assigned		

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Clarke 1866)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xxxxxxxxxxxx	x xxxxxx	.xxxxx	xxx°xx'xx.xxxx"	xxx°xx'xx.xxxx"	xxxx xx	xxxx.xx
SANDIA CORPORATION									
Albuquerque, NM 87115									
Tonopah Test Range, NV	AN/MPS-25 (S/N-1)		0.0027465820	1.000000	N/A	37-44-26.1871	116-46-41.5499		6032.78
Tonopah Test Range, NV	M-33		0.0054931641	3.051758	N/A	37-47-09.6070	116-45-15.9866		5532.93
Barking Sands, Kauai	Nike Ajax		0.1953125	3.051758	N/A	16-43-53.9860	169-32-05.4040		17.05
Barking Sands, Kauai	Nike Ajax		0.1953125	3.051758	N/A	16-43-53.7570	169-32-07.0370		17.05
Johnston Atoll	Nike Ajax		0.1953125	3.051758	N/A	22-03-41.0507	159-46-53.4250		17.85
Johnston Atoll	Nike Ajax		0.1953125	3.051758	N/A	22-03-39.5483	159-46-53.9670		16.70

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Fischer 1960)			
			ELEV. & AZ. (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)	SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			.xxxxxxxxxxxx	x.xxxxxx	-xxxxx	xxx°xx'xx-xxxx"	xxx°xx'xx-xxxx"	xxxx.xx	xxxx.xx
SPACE AND MISSILE TEST CENTER									
Vandenberg AFB, CA 93437									
Tranquillon	AN/FPS-16 (S/N-18)		0.048828125	1.953125	N/A	34-34-58.4506	120-33-40.1517	2095.80	2170.11
Tranquillon	AN/FPS-16 (S/N-21)		0.048828125	1.953125	N/A	34-34-59.4945	120-33-38.5037	2095.90	2170.21
Point Pillar	AN/FPS-16 (S/N-5)		0.048828125	1.953125	N/A	37-29051.9825	122-29-53.8848	136.88	198.06
Point Pillar	AN/FPQ-6 (S/N-6)		0.006103515625	1.953125	0.03125	37-20-52.4985	122-20-57.3631	104.77	165.95
Vandenberg AFB	AN/TPQ-18 (S/N-5)		0.006103515625	1.953125	0.03125	34-39-57.6303	120-34-51.6978	329.11	403.51
T-AGM-6 USNS Watertown	CAPRI (S/N-1)		0.048828125	1.953125	N/A	As Assigned	As Assigned		
T-AGM-7 USNS Huntsville	CAPRI (S/N-2)		0.048828125	1.953125	N/A	As Assigned	As Assigned		

REMOTE TRACKING SITE CONSTANTS DATA

LOCATION	RADAR	ADDRESS	LEAST SIGNIFICANT BIT (LSB)		RANGE RATE (YDS/SEC)	GEODETIC COORDINATES (Based on Clarke 1866)		SPHEROID HEIGHT (FT)	MEAN SEA LEVEL (FT)
			ELEV & AZ (MILS)	RANGE (YDS)		LATITUDE (+ N)	LONGITUDE (+ W)		
			.xxxxxxxxxxx	x.xxxxxx	xxxxx	xxx°xx'xx.xxxx	xxx°xx'xx.xxxx	xxxx.xx	xxxx.xx
WHITE SANDS MISSILE RANGE									
White Sands Missile Range, NM 88002									
R112 C Station	AN/FPS-16 (S/N-1)		0.048828125	1.000000	N/A	32-21-28.6149	106-22-12.8363		4048.40
R113 C Station	AN/FPS-16 (S/N-6)		0.048828125	1.953125	0.02	32-21-28.6233	106-22-09.3409		4049.20
R114 C Station	AN/FPS-16 (S/N-28)		0.048828125	1.000000	N/A	32-21-28.6308	106-22-05.8461		4047.17
R124 Phillips Hill	AN/FPS-16 (S/N-16)		0.048828125	1.000000	N/A	33-26-42.4099	106-07-53.5662		5326.81
R329 Talos	AN/FPS-16 (XN-3)		0.048828125	1.000000	N/A	32-24-20.5830	106-20-56.5646		4048.03
R122 Holloman	AN/FPS-16 (S/N-4)		0.048828125	1.000000	N/A	32-54-07.9523	106-05-55.0272		4148.49
R123 Holloman	AN/FPS-16 (S/N-9)		0.048828125	1.000000	0.02	32-54-04.9767	106-05-55.0398		4149.40
R125 Tularosa	AN/FPS-16 (S/N-40)		0.048828125	1.000000	N/A	33-05-45.7916	106-09-30.9405		4175.66
R127 Stallion	AN/FPS-16 (S/N-12)		0.048828125	1.000000	0.02	33-48-49.8277	106-39-30.3878		5025.01
R128 Stallion	AN/FPS-16 (S/N-22)		0.048828125	1.000000	N/A	33-48-46.8622	106-39-30.3754		5017.28
R280 Blanding	AN/MPS-25 (XN-1)		0.048828125	1.000000	N/A	37-31-28.668	109-35-25.565		5771.31
R281 Blanding	AN/MPS-25 (XN-2)		0.048828125	1.000000	N/A	37-31-27.288	109-35-28.803		5785.09
R218 Green River	AN/FPS-16 (S/N-26)		0.048828125	1.000000	N/A	38-58-44.452	110-06-42.076		4338.89
R219 Green River	AN/FPS-16 (S/N-33)		0.048828125	1.000000	N/A	38-58-44.008	110-06-37.146		4338.96

AFSC Form 1848, Oct 80

RCCS DOCUMENT TARGET BOX 01

DOCUMENT #: 155-91

NAME: General Principles of Digital Filtering and a Survey of Filters in Current Range Use

TYPE: DOCUMENT

DOCUMENT 155-91

DATA REDUCTION AND COMPUTER GROUP

GENERAL PRINCIPLES OF DIGITAL FILTERING AND A SURVEY OF FILTERS IN CURRENT RANGE USE

WHITE SANDS MISSILE RANGE
KWAJALEIN MISSILE RANGE
YUMA PROVING GROUND
DUGWAY PROVING GROUND
ELECTRONIC PROVING GROUND

ATLANTIC FLEET WEAPONS TRAINING FACILITY
NAVAL AIR WARFARE CENTER WEAPONS DIVISION
NAVAL AIR WARFARE CENTER AIRCRAFT DIVISION
NAVAL UNDERSEA WARFARE CENTER DIVISION NEWPORT

30TH SPACE WING
45TH SPACE WING
AIR FORCE FLIGHT TEST CENTER
AIR FORCE DEVELOPMENT TEST CENTER
AIR FORCE WEAPONS AND TACTICS CENTER
DETACHMENT 2, SPACE AND MISSILE SYSTEMS CENTER

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

DOCUMENT 155-91

GENERAL PRINCIPLES OF DIGITAL FILTERING AND A SURVEY OF FILTERS IN CURRENT RANGE USE

AUGUST 1991

Prepared by

Data Reduction and Computer Group
Range Commanders Council

Published by

Secretariat
Range Commanders Council
U.S. Army White Sands Missile Range
New Mexico 88002-5110

TABLE OF CONTENTS

		Page
CHAPTER 1 - GENERAL PRINCIPLES OF DIGITAL FILTERING		1-1
1.1	Time Domain Measurement Data	1-3
1.2	Error Assumptions	1-5
1.3	Basic Purpose and Filter Use	1-7
1.4	Linear Filters	1-8
1.5	Filter Weight Functions	1-9
1.6	General Purpose Filters	1-10
1.7	Special Purpose Filters	1-10
1.8	Nonrecursive Filters	1-10
1.9	Recursive Filters	1-11
1.10	Filter Weights Construction	1-12
CHAPTER 2 -FILTERS ANALYSIS IN THE TIME DOMAIN		2-1
2.1	Filter Characteristics Analysis	2-1
2.2	Autocorrelation	2-1
2.3	Unit Impulse Response	2-2
2.4	Variance Reduction Factor	2-5
2.5	Simulation	2-7
2.6	Monte Carlo Methods	2-8
2.6.1	Generation of Random Numbers From a General Distribution Function	2-9
2.6.1.1	Definition of Discrete Case	2-9
2.6.1.2	Method of Random Number Generation	2-9
2.6.1.3	Method of Random Number Generation	2-10
2.6.2	Random Numbers From a Uniform Distribution	2-10
2.6.3	Random Normal Deviates	2-11
2.7	Residuals	2-12
CHAPTER 3 - FILTER ANALYSIS AND DESIGN IN THE FREQUENCY DOMAIN		3-1
3.1	Analysis of Filter Characteristics in the Time Domain	3-1
3.2	Discrete Fourier Transform	3-1
3.3	Transfer Function of a Filter	3-5
3.4	Cutoff Frequency, Roll-Off, and Quality of Filter	3-8
3.5	Phase Shift of a Filter	3-11
3.6	Construction of Filter Weights	3-11
3.7	Aliasing	3-13
3.7.1	Aliasing in the Time Domain	3-14
3.7.2	Aliasing in the Frequency Domain	3-16
3.7.3	Sampling Theorem and the Nyquist Frequency	3-17
3.8	Stability of Filter	3-18

Page

CHAPTER 4 - CATEGORIES OF DIGITAL FILTERS..... 4-1

4.1 Low-Pass Digital Filters..... 4-1
4.2 High-Pass Digital Filters..... 4-2
4.3 Band Pass Filters..... 4-3
4.4 Band-Rejection Filters..... 4-5
4.5 Phase-Shift Filters..... 4-7
4.6 Filter Combinations..... 4-7

CHAPTER 5 - FREQUENCY DOMAIN FILTER DESIGN..... 5-1

5.1 Infinite Impulse Response (IIR) Filter Design..... 5-1
5.1.1 Modes of IIR Filter Expression..... 5-2
5.1.1.1 Direct Form..... 5-2
5.1.1.2 Cascade Form..... 5-3
5.1.1.3 Parallel Form..... 5-5
5.1.2 Converting From Analog to Digital..... 5-6
5.1.2.1 Impulse Invariance..... 5-7
5.1.2.2 Bilinear Transformation..... 5-9
5.1.2.3 Direct Mapping of Differentials..... 5-11
5.1.2.4 Matched Z-Transformation..... 5-15
5.1.3 Low-Pass IIR Filters..... 5-16
5.1.3.1 Butterworth Filters..... 5-16
5.1.3.2 Chebyshev Filters..... 5-17
5.1.3.3 Elliptic Filters..... 5-18
5.2 Finite Impulse Response (FIR) Filter Design..... 5-20
5.2.1 Modes of FIR Filter Expression..... 5-20
5.2.1.1 Direct Form..... 5-21
5.2.1.2 Cascade Form..... 5-21
5.2.1.3 Frequency Sampling Form..... 5-21
5.2.1.4 Linear-Phase Form..... 5-22
5.2.2 FIR Filter Design Techniques..... 5-22
5.2.2.1 Windowing..... 5-23
5.2.2.2 Frequency Sampling Design..... 5-28
5.2.2.3 Equiripple Design..... 5-29

CHAPTER 6 - KALMAN FILTERS..... 6-1

6.1 Linear Discrete Kalman Filter..... 6-3
6.1.1 Definitions..... 6-3
6.2 The Linear Continuous Kalman Filter..... 6-7
6.3 Extended Kalman Filters..... 6-9
6.4 Example..... 6-12

Page

CHAPTER 7 - FILTERS IN CURRENT RANGE USE..... 7-1

7.1 Least Squares Filters..... 7-1
7.1.1 Simple Moving Average Filters..... 7-1
7.1.2 Least Squares Polynomial Moving Arc Filters..... 7-2
7.2 Position and Velocity Constrained Least Square Filters..... 7-5
7.2.1 Position Constraint..... 7-6
7.2.2 Velocity Constraint..... 7-6
7.3 Orthogonal Polynomials..... 7-7
7.4 Least Squares Polynomial Moving Arc Filters Using Recursive Sums..... 7-10
7.5 Derivative Information Recovery by a Selective Integration Technique (DIRSIT)..... 7-12
7.6 Digital Filter X (DFX)..... 7-16
7.7 Quadratic Digital (QD) Filter..... 7-19

APPENDIX A

APEP..... A-2
FAST LEAST SQUARES..... A-4
WEIGHED LEAST SQUARES..... A-5
WLSRE IN MIT026..... A-6
M-STATION..... A-8
PBS PROGRAM, M-STATION..... A-9
PBS PROGRAM, DERIVE..... A-10
PBS FILTER..... A-11
PBS PROGRAM, OPFILT..... A-12
ATAGAS KALMAN FILTER..... A-13
MOTION TRACKING KALMAN FILTER..... A-14
TEC TRACKER..... A-15
DRIFT TRACKING KALMAN FILTER..... A-16
EATS KALMAN FILTER..... A-17
TRIDENT KALMAN FILTER..... A-18
QD - WSMR..... A-19
QD - EGLIN..... A-20
QD - APO SAN FRANCISCO..... A-21
QD - YUMA..... A-22
QUADRATIC POLYNOMIAL FILTER..... A-23
QD - NATC..... A-24
AVRAGE..... A-25
DUZ2..... A-26
SDIFFR..... A-27
DIRSIT..... A-29
UFTAS FILTER OVERLAY..... A-30
BET..... A-31
RTUF..... A-33
RFILTR..... A-34
RCHECK..... A-35

APPENDIX A (CONT'D) Page

EYBALL..... A-36
ON-AXIS..... A-37
DYNO43..... A-38
GNFL..... A-39
GNSM..... A-40
FIRFILT..... A-41
RECFLT..... A-42
FILMAX..... A-43
MDPT51..... A-44
BSMW..... A-45
SMW..... A-46
FREQUENCY RESPONSE ANALYSIS..... A-47

REFERENCES

LIST OF FIGURES

CHAPTER 1 - GENERAL PRINCIPLES OF DIGITAL FILTERING

Figure No.		Page
1-1	Graph of power spectrum $G(\omega)$ versus frequency ω, comparing true frequency signal {Y(n)} with noisy frequency signal {e(n)}.....	1-6
1-2	Power spectrum of Gaussian white noise.....	1-7

CHAPTER 2 - FILTER ANALYSIS IN THE TIME DOMAIN

Figure No.		
2-1	Smooth input data. Sampling period - 0.05 seconds.....	2-6
2-2	Real-time response. Point span - 20; sampling period - 0.05 seconds.....	2-6
2-3	A closed-loop simulation logical diagram.....	2-8
2-4	Graphic description of how the random number x(k) is determined. It is obtained with a cumulative distribution function F(x(n)) (in top graph) or F(x) (in bottom graph) and a random variable y(k) uniformly distributed over [0,1].....	2-11

CHAPTER 3 - FILTER ANALYSIS AND DESIGN IN THE FREQUENCY DOMAIN

Figure No.		
3-1	Comparison of multiplications required by direct calculation and FFT algorithm.....	3-3
3-2	Power spectrum of output, where the input is white noise and where the filter passes only frequencies between $f_1$and f_2.....	3-4
3-3	Ideal low-pass filter.....	3-6
3-4	Ideal high-pass filter.....	3-7
3-5	Ideal band-pass filter.....	3-7
3-6	Tolerance limits for approximation of ideal low-pass filter.....	3-9
3-7	Comparison of roll-offs for filters A and B.....	3-10
3-8	Comparison of the quality of filter A with the quality of filter B.....	3-10
3-9	Foldover or aliasing resulting from observing a wheel's increasing rotational speed.....	3-14
3-10	Aliasing in the time domain.....	3-15

Figure No. Page

3-11 Bottom sinusoid shown as the sume of three frequency components, C2, C1, and S1..... 3-15
3-12 Sampling relations for analog and digital systems for properly sampled inputs..... 3-16
3-13 The effects of undersampling on the digital frequency response..... 3-17

CHAPTER 4 - CATEGORIES OF DIGITAL FILTERS

Figure No.

4-1 Typical frequency response of low-pass filter (represented by the curved line)..... 4-2
4-2 Frequency response of wide band pass filter..... 4-4
4-3 Method of converting a low-pass filter with shallow roll-off into a band pass filter..... 4-5
4-4 Frequency response of band-rejection filter..... 4-6

CHAPTER 5 - FREQUENCY DOMAIN FILTER DESIGN

Figure No.

5-1 Direct form realization of IIR filter..... 5-3
5-2 Cascade form realization of the IIR filter..... 5-4
5-3 Parallel-form realization with the real and complex poles grouped in pairs..... 5-6
5-4 The mapping of a horizontal strip of width $2\pi/T$ in the s plane to the unit circle of the z plane, through the impulse invariance method.. 5-8
5-5 The mapping of the left half of the s plane to the unit circle of the z plane, using the bilinear transformation..... 5-10
5-6 The relation between analog and digital frequence scales for the bilinear transformation..... 5-10
5-7 s plane to z plane mapping of $j\Omega$ axis for method of backward differences..... 5-13
5-8 s plane to z plane mapping of $j\Omega$ axis for method of forward differences..... 5-14
5-9 Dependence of Butterworth magnitude characteristic on the order N..... 5-17
5-10 Type I and II Chebyshev filters of odd and even orders..... 5-18
5-11 Equiripple approximation in both passband and stop band..... 5-19

LIST OF FIGURES (Cont'd)

Figure No.		Page
5-12	(a) Ideal low-pass filter characteristic with cutoff at ω_c and (b) impulse response corresponding to ideal low-pass filter.....	5-23
5-13	(a) Truncated version of impulse response in figure 5-12, (b) Truncated response shifted so as to make system causal, and (c) Filter frequency responce resulting from truncation of impulse response.....	5-24
5-14a	Common window functions and their transforms: (a) rectangular, (b) triangular (Fejer-Bartlett), (c) Hanning, (d) Hamming, and (e) Kaiser.....	5-26
5-14b	Common window functions and their transforms: (a) rectangular, (b) triangular (Fejer-Bartlett), (c) Hanning, (d) Hamming, and (e) Kaiser.....	5-27
5-15	Commonly used windows for FIR filter design.....	5-28
5-16	Equiripple approximation of a low-pass filter...	5-30

CHAPTER 7 - FILTERS IN CURRENT RANGE USE

7-1	Example acceleration curve.....	7-14
7-2	Corresponding velocity curve.....	7-14
7-3	Example acceleration curve for $t > t_1$.....	7-15

LIST OF TABLES

CHAPTER 3 - FILTER ANALYSIS AND DESIGN IN THE FREQUENCY DOMAIN

Table No.		
3-1	COMPARISON OF REQUIRED MULTIPLICATION OPERATIONS USING THE FFT AND THE DIRECT DFT.....	3-3

CHAPTER 5 - FREQUENCY DOMAIN FILTER DESIGN

5-1	COMMON WINDOWING FUNCTIONS.....	5-25

CHAPTER 6 - KALMAN FILTERS

6-1	SUMMARY OF DISCRETE KALMAN FILTER EQUATIONS.....	6-6

LIST OF TABLES (Cont'd)

Table No. — Page

6-2 SUMMARY OF CONTINUOUS KALMAN FILTER EQUATIONS (WHITE MEASUREMENT NOISE)..... 6-8

CHAPTER 7 - FILTERS IN CURRENT RANGE USE

7-1 A STARTING CONDITION FOR A DIRSIT FILTER WHOSE SPAN IS 2m+1 POINTS..... 7-12

CHAPTER 1

GENERAL PRINCIPLES OF DIGITAL FILTERING

Data reduction includes an inherent requirement to filter measured data. The part of the measured values which is noise, that is, which does not characterize the phenomena being measured, must be removed so that the data expresses the characteristics of the system being tested. (A system is a device which interrelates the excitation and the response.) Often it is necessary to perform integration and differentiation based on measured data. The digital filter is a means by which these operations can be performed. Without data filtering, the measurements and the derived values from these measurements often depict physical phenomena that are virtually impossible. Aircraft and missiles are predictable, so it is possible to design filters which model their physical performance.

The application of digital filters is continually growing as more computers are placed in line in the measuring systems, as classical measurement systems are being upgraded, and as new measuring data concepts are being introduced. Video signal image enhancement, video signal transmission, the Global Positioning System, and aircraft onboard systems will necessitate the development of new digital filtering theories. Filters have many different functions; for example, filters are used in water purification to remove those particles which are undesirable. Filters are also used in coffee pots to keep the grounds separate from the liquid. Digital filters serve much the same purpose; they separate true data from errors which are added during the measurement process.

Signal processing filters were originally circuits with frequency selection characteristics. Some of the filters which developed as a result of this concept were the Butterworth, the Chebyshev, and the elliptic. These filters can be designed to meet some specific criteria regarding amplitude and phase response, but because filters of this type work on data frequency characteristics, they are not very useful in their original realizations on data recorded in digital form. There are time domain realizations which exhibit similar characteristics on time domain data. When digital realizations exist in software, a design error can be easily corrected; when they do not exist, however, roundoff error, storage requirements, and delays become important design factors.

Digital filters are not always derived from a frequency domain realization. Filters can be used directly with digital data, based on the statistical characteristics of the useful signal and noise. The Kalman filters belong to this class of

filters. Some of the principles used in filters of this type were developed by Gauss in his concept of weighted least squares. Although he used this concept in data analysis, it was not until the advent of digital computers that the full potential of this concept was realized. A few extensions have been added to Gauss' work, and many techniques have been developed based on his principles.

Related to digital filters are the terms smoother, filter, predictor, and wild point. These terms are included because the range surveys in appendix A include references to these terms. Given measurements in the time interval [0,t], a smoother produces estimates for times $t_s < t$, a filter produces estimates for times $t_f = t$, and a predictor produces estimates for times $t_p > t$. Thus, optimal estimates are obtained postflight by smoothers (using all recorded data) and are obtained in real time by filters (using all data up to the present). A wild point is a measured data value that fails to meet some predetermined statistical criteria. A "wild point application" of a digital filter is the replacement of a wild point.

A few of the current digital filtering applications by the Range Commanders Council (RCC) member ranges are

(1) in rocket firing to give position, velocity, and acceleration in real time for range safety information;

(2) in inertial navigation where navigation errors have to be estimated in real time for corrections to be applied;

(3) in radar and theodolite data processing to give trajectory estimates for position, velocity, and acceleration;

(4) in satellite trajectory determination;

(5) in processing airplane flight test data; and

(6) in reconstructing photographs or voices which have been transmitted digitally.

This document catalogues some types of digital filters widely used by the RCC member ranges, explores the principles behind these filters, and shows the design techniques for specific applications. Obviously, a rigorous, theoretical treatment of the subject cannot be given in this document, nor can all the details of design techniques. The intent is to give the reader a feel for the subject of digital filtering as defined by the members of the RCC Data Reduction and Computer Group. Noted are the appropriate bibliography references for any of the subject areas where more detail might be helpful to the reader.

1.1 Time Domain Measurement Data

Two representations can be used for discrete data in the time domain. The most commonly used definition for a sampled function is the sequence {f(n)} whose elements are

$$f(n)=f(t(n))=f(t)\Big|_{t=t(n)} \qquad (1\text{-}1)$$

where n is an index for the times at which samples are taken and f(t) is the continuous function from which the samples are taken. In further discussion in this paper, this representation will be used for measurement data. Another mathematical representation for a sampled function in the time domain, f*(t), is

$$f^*(t) = \sum_k f(t(k))\ \delta(t-t(k)) \qquad (1\text{-}2)$$

where the t(k)'s are the instants in time at which the function is sampled and $\delta(t)$ is the Dirac delta function or impulse function having the properties of

$$\delta(t-a) = \begin{cases} 0 & t \neq a \\ \infty & t = a \end{cases} \qquad (1\text{-}3)$$

$$\int_{a-\epsilon}^{a+\epsilon} \delta(t-a)\ dt = 1 \qquad \epsilon > 0 \qquad (1\text{-}4)$$

The above definition of $\delta(t)$ is from reference 47, page 15. Another common definition is

$$\delta(t) = 0,\ t \neq 0,$$
$$\delta(0) = \infty,\ \text{and}$$

$$\int_{-\epsilon}^{\epsilon} \delta(t)dt = 1. \qquad \epsilon > 0 \qquad (1\text{-}5)$$

As explained in reference 48, page 70, the function represents an imaginary rectangle whose base is 0, height is ∞, and area is 1.

The first representation allows filters to be analyzed using some of the calculus techniques of continuous functions and allows bridging of the gap between discrete and continuous analysis. In many current papers on this topic, the mathematical model for the measurements and the state of a system are given in terms of vectors. In this notation, a group of variables is

treated as one vector variable; hence, a compactness of notation is obtained, and very general filters can be discussed without becoming immersed in details. For the sake of completeness, this representation is mentioned here.

Consider a sequence of measurements which are represented in a form similar to equation (1-2). Instead of considering a sequence of real numbers, consider a sequence of vectors, {X(n)}. A particular element X(n) of this sequence consists of measurements made at time t(n) and should relate to the true state of the process from which the measurements were taken. For instance, X(n) could consist of measurements of azimuth, elevation, and range for some object in space. It is known that the sequence {X(n)} can be related, for example, to the position, velocity, and acceleration in the x, y, and z directions. Now assume that the true state of the process Y(t) can be represented by the differential equation

$$\frac{dY(t)}{dt} = F(Y(t),t) \tag{1-6}$$

where Y(t) is a vector composed of scalar functions y(t,1),..., y(t,N) and all the derivatives of each of these scalars; F(Y(t),t) is a vector function that is, in general, nonlinear and has as each element a function of all the elements in Y. Y(t) is known as the state vector and could consist of x, y, and z coordinate positions along with the first and second derivatives of these positions, for instance. Further assume that X(n), the measurement vector, is related to the state vector as

$$X(n) = G(Y(t(n)),t(n)) + N(n) \tag{1-7}$$

where G is the measurement function of the state variable and N(n) is the noise vector or the measurement error vector. In many applications, equation (1-6) constitutes a linear and homogeneous system, and the relationship between the measurement vector and the state vector is assumed linear. In this case, equations (1-6) and (1-7) reduce to

$$\frac{dY}{dt} = A(t)Y(t) \tag{1-8}$$

$$X(n) = M(n)Y(n)+N(n) \tag{1-9}$$

where A(t) is a matrix function of time and M(n) is a matrix known as the measurement matrix. In digital filtering, equations (1-6) and (1-7) are used to give some type of best estimate for Y(t).

In many cases, the measurements are made on certain elements of the state vector itself; that is, the measurement and state vector frequently represent the same test item. For example, an airplane's azimuth and elevation may be measured from several different stations to obtain its "true" azimuth and elevation. In this case, the matrix M(n) is diagonal and contains only 1s and 0s on the diagonal. Most future discussions will operate under this assumption. For ease of understanding, this document will consider the case where an attempt is made to estimate one variable from measurements made on that variable. The state differential equation will be implied rather than given explicitly.

1.2 Error Assumptions

In digital filter analysis, three types of errors are considered: (1) errors in the mathematical model, (2) random errors in the observations, and (3) unwanted discrete frequencies in the observations. The first type of error is caused by a lack of understanding of the physical system for which the estimate is to be made. The second error type is due to inaccurate measurements and is assumed to be normally distributed and unbiased. The third error type is caused by either the reception of noise along with the desired frequency signal or the intention to block out one or more bands of frequencies to suit a specific application. (If a digital filter is used for the second reason, then the third type of "error" is not really an error.)

The random errors are discussed next. Assume that {x(n)} is a sequence of measurements corrupted by noise {e(n)} and that the true value of the process is given by {Y(n)}. In all applications the noise is considered to be additive so that the equation relating the state variable to the measured variable at time t(n) is given by

$$x(n) = Y(n)+e(n). \tag{1-10}$$

It is often assumed that the noise is Gaussian with zero mean and uncorrelated with itself. The quantity which is most often used to characterize the noise is the variance given by equation

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} e(i)^2 \tag{1-11}$$

where N is the total number of measurements being considered. How much this variance is reduced by the filter may be used as the criterion for determining the filter characteristics. Caution must be exercised when using this criterion, because {e(i)} is a time sequence and thus, in most cases, correlated with itself.

Now consider errors of unwanted discrete frequencies in the observations. Again consider the relationship between the true value Y(n), the measurement x(n), and the error e(n) to be of the form in equation (1-10) with {e(n)} being a signal containing unwanted frequencies. In this case, it is assumed that {Y(n)} and {e(n)} are not in the same frequency band. This property is best illustrated using the power spectral density function, the quantity which indicates the extent that a particular range of frequencies present in the data. In the case when the frequency range of the signal and error are nonoverlapping, then the power spectral density function $G(\omega)$ would appear as shown in figure 1-1.

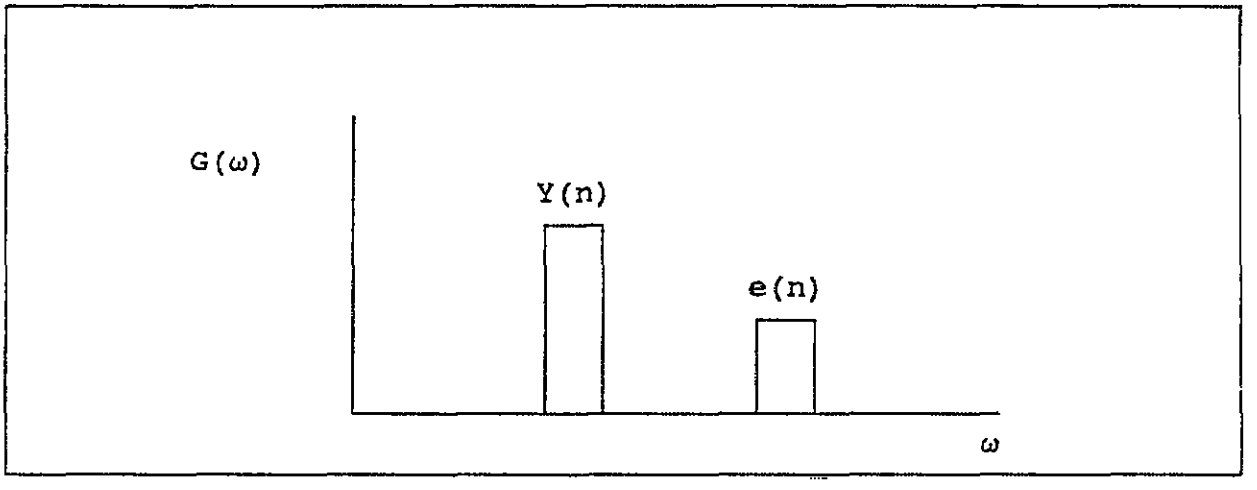


Figure 1-1. Graph of power spectrum $G(\omega)$ versus frequency ω, comparing true frequency signal {Y(n)} with noisy frequency signal {e(n)}.

In this case, a filter would be designed having a frequency response such that {Y(n)} would pass through unchanged, whereas {e(n)} would be suppressed as much as possible.

The power spectral density function for Gaussian white noise is a constant as shown in figure 1-2. When {e(n)} is of this form, the signal and noise are in overlapping frequency ranges; hence, part of the noise is treated as signal and is not suppressed. However, the desired signal is often of a low enough frequency and of a sufficient amplitude that most of the noise in this frequency range can be neglected.

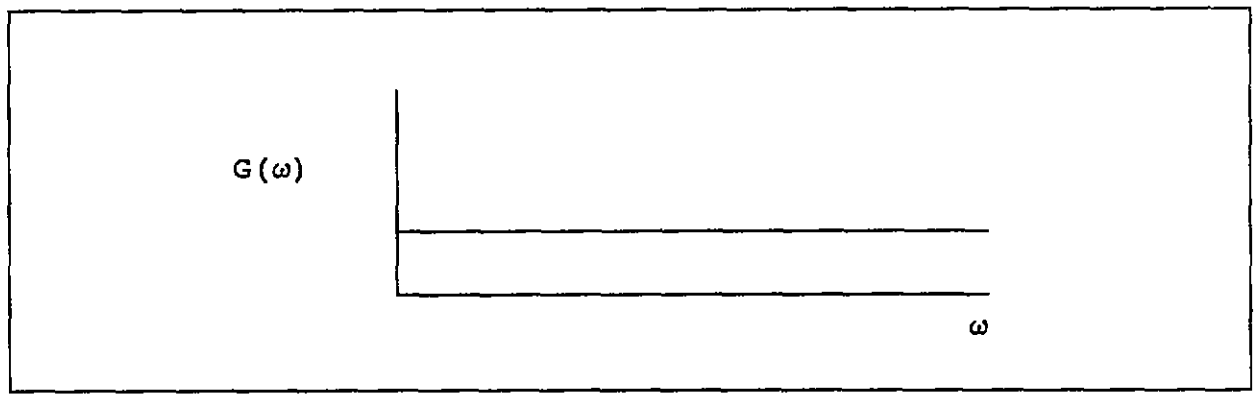


Figure 1-2. Power spectrum of Gaussian white noise.

1.3 Basic Purpose and Filter Use

Filters are used to ascertain information about a process from measurements made on that process when these measurements are corrupted by noise. Any of the following data should be incorporated into the filter if known:

(1) the differential equations describing the process,

(2) the statistics describing the noise, and

(3) the frequency range of the signals describing the process.

In many applications the differential equation describing the system is not known. In these cases, it is often assumed that over small enough time intervals the differential equation of the process is given by

$$\frac{d^k y}{dt^k} = 0 \qquad (1\text{-}12)$$

for some positive integer k.

The noise most often encountered in physical applications is Gaussian; however, often the variance is not known. When the variance of the noise is not known, a least squares criterion is often used to design a filter. The least squares criterion requires that

$$\sum_{n=0}^{N} (\hat{y}(n)-y(n))^2 = \text{a minimum} \qquad (1\text{-}13)$$

where $\{\hat{y}(n)\}$ is the sequence obtained by some type of fit to the data, which is used to give a filtered output. Combining equations (1-12) and (1-13) gives birth to the many filters which are based on least squares polynomial approximation. There are many variations of these filters used in a variety of applications in both real and nonreal time. Probably the widest application of these filters is found in the processing of tracking data. For example, in processing radar data, azimuth, elevation, and range are measured. From these measurements a raw x, y, and z are computed. The digital filter is used to obtain smoothed values for x, y, and z as well as the first and second derivatives for these quantities. Other uses for filters, based on least squares polynomials, include processing of theodolite tracking data and obtaining rates of climb and accelerations from altitude and airspeed data measured onboard an airplane. When the differential equation describing the process and the statistics of the noise are both known, then a Kalman filter can be used to estimate the process. The Kalman filter has found wide application in the fields of guidance and navigation. (The Kalman filter is discussed in chapter 6.)

When the signal describing the process lies in a certain frequency band, a filter is designed which will allow only the frequencies in this band to pass. (Passing frequencies of a certain band are discussed in chapter 4.) These filters are used for square-law detection, frequency-selective smoothing, phase and amplitude determination, and smoothing differentiation. Specific applications include processing of digitally transmitted photographs, medical records, and tracking data. The determination of the best filter for any particular application requires a careful analysis. The filter which is ultimately selected will depend on the physical system under analysis, the type of measurements being taken, and the constraints in computation time.

1.4 Linear Filters

Linear filters are filters which conform to

$$F(x+ay) = F(x)+aF(y), \qquad (1\text{-}14)$$

where F denotes the filtering operation, x and y denote inputs to the filter, and a is a constant. For such filters, the output at a particular time equals the input multiplied by a set of weights which are not a function of the output as given by the equation

$$y(k) = \sum_j h(k,j)x(k-j) \qquad (1\text{-}15)$$

The set $\{h(k,j)\}$ denotes the weight function described in paragraph 1.5. The advantage of linear filters is that they operate on signal and noise independently, so it is relatively easy to determine how they will treat both signal and noise.

Most of the filters in current use are linear; least squares moving-arc polynomials and most applications of frequency-constrained and Kalman filters fall in this category. (The least squares moving-arc polynomials filter is discussed in chapter 7.) If the system under consideration is nonlinear, then some type of linearizing scheme is used so that some of the nice properties of linearity can be retained.

1.5 Filter Weight Functions

To illustrate the idea of filter weight functions, consider a continuous linear filter having one input variable and one output variable. Such a filter is defined by

$$y(t) = \int_{-\infty}^{\infty} h(t,T)x(t-T)dT \qquad (1\text{-}16)$$

The integral in equation (1-16) is the well-known convolution integral. Here x(t) is the input, y(t) is the output, and h(t,T) is the filter weight function. The discrete analog to equation (1-16) is

$$y(j) = \Sigma_k h(j,k)x(j-k) \qquad (1\text{-}17)$$

where {x(j)} is the input sequence, {y(j)} is the output sequence, and {h(j,k)} is a set of weights corresponding to the weight function previously defined. The last equation characterizes the nonrecursive digital filter discussed in paragraph 1.8. The summation in equation (1-17) is known as the convolution sum. The filters, as defined, can be time varying; that is, the weight function can change with time. In the continuous case, if h(t,T) is a function of T only, then the filter is time invariant. Similarly, in the discrete case, if {h(j,k)} is a function of k only, then the digital filter is time invariant. The equation for a digital filter then becomes

$$y(j) = \Sigma_k h(k)x(j-k) \qquad (1\text{-}18)$$

A digital filter is time invariant when the system is being modeled by a constant coefficient differential equation, and the relationship between the state variables and the measurements does not vary with time. When measurements are made at equal increments in time, the set of weights for a time-invariant filter remains the same for each output point. Thus, time invariance is a desirable feature from a computational point of view.

1.6 General Purpose Filters

Many filters in use today are general purpose; that is, they are not restricted to a particular application. The most commonly used of these filters are the least squares polynomial filters. These filters are usually designed so that the degree and number of points to be used can be specified. Within this category there are many variations. For instance, a filter can be constrained so that it passes through the last filtered point and has continuous slope through that point. In addition, the specific implementation can vary; for example, the recursive sums or orthogonal polynomials are different implementations of the polynomial filters.

Most of the frequency-constrained filters mentioned in paragraph 1.4 are general-purpose filters. Any one of these filters can be used in a number of applications. Some uses of general-purpose filters were mentioned in paragraph 1.3. In many instances, these filters can be moved from one application to another without modification.

1.7 Special Purpose Filters

Certain filters are designed for a special mission. Once designed and checked out, these filters then ideally remain the same for the duration of the mission. Two important categories of special purpose filters are range safety and navigation system. An example of a special purpose range-safety filter is the QD filter, (see paragraph 7.7). This filter is suited for its mission because of the speed with which it can deliver smoothed position, velocity, and acceleration in real time. It also has the ability to do spike editing in real time. (Spike editing is defined as editing out the wild points or spikes.) For navigation filtering, special implementations of the Kalman filter are usually used. The Kalman filter is well suited for this application because of its ability to give optimal estimates of a system's state variables in time.

1.8 Nonrecursive Filters

In paragraph 1.5, a digital filter was defined in equation (1-17). For a filter defined in this way, each output point is a function only of the input and the weight function. Any filter for which the current output is not a function of previous outputs such as the one defined by equation (1-17) is said to be nonrecursive. Nonrecursive filters are analogous to open-loop filters in servomechanisms. Many implementations of the least squares polynomial filters and of the frequency-constrained filters fall into this category. Such filters are not subject to problems of instability; that is, an error occurring in a computed output point does not propagate into future output points.

If the right-hand side of equation (1-18) represents a finite sum, then it is called a finite impulse response (FIR) system or FIR filter. In FIR filters, if the right-hand side of equation (1-18) contains q terms, then the filter can have an impulse response at most q samples long. If the last impulse occurred more than q samples ago, the entire right-hand side of equation (1-18) is zero. A presentation of design methods using FIR filters, along with a summary of advantages and disadvantages of using this filter type, is given in chapter 5.

1.9 Recursive Filters

Recursive filters use past output values in computing the current output value. These filters take advantage of past computations in such a way that output values are used to yield information about previous input values to the filter. In this way, storage and computation time are saved. The general form of a recursive filter having one input and one output is

$$y(k) = \sum_{0}^{N} a(j)x(k-j) + \sum_{1}^{M} b(j)y(k-j) \qquad (1\text{-}19)$$

where $\{x(j)\}$ is the input, $\{y(j)\}$ is the output, and $\{a(j),b(j):j=1,\ldots\}$ is the set of weights. Some recursive filters (for example, Kalman and QD) use the previous output to predict a value for the current output and use the current measured input to correct the current output. Recursive filters are analogous to closed-loop filters in servomechanisms. Another commonly used name for such filters is infinite impulse response (IIR) filters or IIR systems. In IIR filters, all outputs y(k) will be influenced by all previous impulse responses a(j) in equation (1-19) regardless of how large k is. Each impulse response has an influence on an infinite number of terms y(k).

The savings in computation time and storage achieved by using recursive filters do not come without a price. If not properly designed, the filters can suffer from instability. Because each filter is computed from previous values $y(j), j<k$, an error induced in any one of these y(k)'s will have an effect in each future y(k). If such errors do not die out as k increases, then the filter is unstable and is not usable. Ways of analyzing the stability characteristics of filters are discussed in chapter 2. Another factor to consider in using recursive filters is the fact that they are not self starting. For the first few output points there will be no previous outputs to use in equation (1-19). Hence, some other method must be devised through which the initial y(k)'s are determined. Methods of choosing starting values depend, of course, on the particular application in

question and are discussed in references 1 and 42 and in chapter 7. A presentation of design methods using IIR filters, along with a summary of advantages and disadvantages, is given in chapter 5.

1.10 Filter Weights Construction

Again consider a nonrecursive, nontime-varying linear filter of the form

$$y(k) = \sum_j h(j,k)x(k-j) \tag{1-20}$$

The problem in designing a filter is finding the weights $\{h(j,k)\}$ in such a way that some design criterion is satisfied. Some of the examples discussed in succeeding chapters are

(1) choosing weights by requiring that the span of $\{x(j)\}$ under consideration be fitted by a least squares polynomial,

(2) choosing the weights using Fourier methods by requiring that the frequency response of the filter fit an idealized frequency response function, and

(3) using the z-transform to derive filter weights from the transfer function of a given analog filter (see chapter 3).

When the filter is recursive, and thus takes the form

$$y(k) = \sum_{0}^{N} a(j)x(k-j) + \sum_{1}^{M} b(j)y(k-j) \tag{1-21}$$

then there is a corresponding nonrecursive filter which is theoretically the same. In this case, the weights are either derived in terms of the corresponding weights for the nonrecursive case or are derived directly using z-transforms.

CHAPTER 2

FILTER ANALYSIS IN THE TIME DOMAIN

2.1 Filter Characteristics Analysis

Analysis of a filter in the time domain consists of determining the characteristics. These characteristics are

(1) Stability. With a unit impulse input to a recursive filter, the filter is unstable if the output oscillates or never dies down to zero amplitude or attenuates.

(2) Attenuation. With a nonzero input parameter, attenuation is the amount the filter has reduced its amplitude in the output.

(3) Time Lag. The amount of time it takes for a filter to supply a best estimate of an input parameter.

(4) Distortion. Attenuating or amplifying amplitudes of different frequencies by different amounts in an undesirable manner is called distortion. If the filter causes minimal distortion, it is a good fit or well-modeled.

To analyze these characteristics, the following techniques are frequently used:

(1) Autocorrelation

(2) Unit Impulse Response

(3) Variance Reduction Factor

(4) Simulation

(5) Monte Carlo Methods

(6) Residuals

2.2 Autocorrelation

When data are fed into a filter, the output data may be more (linearly) correlated in time than the input data; that is, series of output data values may be more correlated than their corresponding series of input data values. The type and amount of this additional correlation will depend on the filter weights which are used as data multipliers. For example, if the filter

frequency response has large side lobe effects and passes frequencies which were supposed to be suppressed, the resultant output will show these effects and they can be measured if the serial autocorrelation in the data is estimated. The serial autocorrelation, used in analyzing time lag, is computed by

$$r(k) = \frac{\sum_{i=1}^{N-k} (x(i)-\bar{x}(1))(x(i+k)-\bar{x}(2))}{\left(\sum_{i=1}^{N-k} (x(i)-\bar{x}(1))^2 \sum_{i=1}^{N-k} (x(i+k)-\bar{x}(2))^2 \right)^{1/2}} \qquad (2\text{-}1)$$

where x(1) and x(2) are the means of the first N-k and the last N-k data points and

N = number of points in the sample,
k = 0,1,2,..., M<N, the number of lags.

The value r(k) represents the (linear) correlation between the first and last N-k data values from a series of N points. This technique can be used in conjunction with simulation techniques by entering white noise into the filter and then measuring the autocorrelation of the output. The amount of correlation imposed on the data by the filter is thus estimated.

Caution must be used if an attempt is made to "smooth" the same data more than once in a sequential manner because of the correlation which may be imposed on the data by the first smoothing process. In cases where it is desirable to smooth position and then again smooth computing velocity, a thorough analysis should be made to ensure that autocorrelation imposed by the first smoothing process does not bias the estimates made by the second process.

2.3 Unit Impulse Response

For a continuous, linear, time-invariant system, a unit impulse response h(t) can be defined as the response of the system at time t because of a unit impulse at time 0, that is, because of $\delta(0)$. (Unit impulse is another term for impulse function (see page 1-3.)) A system is linear if the input $c_1f_1(t)+c_2f_2(t)$ produces an output $c_1g_1(t)+c_2g_2(t)$ for all $f_1(t)$ and $f_2(t)$, where inputs $f_1(t)$ and $f_2(t)$ produce outputs $g_1(t)$ and $g_2(t)$.

The input f(t) and output g(t) of the system are related by

$$g(t) = \int_{-\infty}^{\infty} h(\tau)f(t-\tau)d\tau \qquad (2\text{-}2)$$

where $h(\tau)$ is the unit impulse response. Thus, the output equals the convolution of the unit impulse response and the input.

For a discrete, linear, time-invariant system, the unit impulse response is the sequence $\{h_n\}$ where h_n is the response of the system at state n caused by a unit impulse at state 0. For the discrete case, the input-output relationship can be described as

$$y_n = \sum_{m=-\infty}^{\infty} h_m x_{n-m} \qquad (2\text{-}3)$$

where $\{x_n\}$ and $\{y_n\}$ are the input and output signals, and $\{h_n\}$ is the unit impulse response.

For either case, the output of a linear, time-invariant system is thus the convolution of its input and its unit impulse response. For this purpose, equation (2-3) will represent the system of interest, since digital filters are defined to be discrete, linear, time-invariant systems. As seen from equation (2-3), the unit impulse response can be regarded as a set of weights. Thus, from what was discussed in paragraph 1.5, the unit impulse response completely determines the characteristics of a digital filter. Reference can then be made either to the weights, given past data, in the averaging process or to the unit impulse response.

In reality, it is important to know how much a filter distorts a deterministic input. One way to determine the distortion is to analyze unit impulse response. The z-transform is a convenient tool for the analysis of the response of digital filters when the unit impulse response is known. The z-transform for a sequence $\{x_n\}$ is defined as

$$X(z) = \sum_{n=-\infty}^{\infty} x_n z^{-n} \qquad (2\text{-}4)$$

Stated earlier in this section is that the output of the system can be found by convolving the input and unit impulse response. The output can be found easier through the use of the z-transform. An example taken from reference 50 demonstrating the use of the z-transform in determining the output signal or the unit impulse response is described next.

Let the input signal be {2 1} and the unit impulse response be {8 4 2 1}. Referring to equation (2-2), the Convolution Theorem, which states that the z-transform of the convolution of two functions is equal to the product of the z-transforms of those functions, is applied (see reference 23 or 50). The z-transforms of the input signal and unit impulse response are

$$2 + z^{-1} \quad \text{and} \quad 8 + 4z^{-1} + 2z^{-2} + z^{-3} \tag{2-5}$$

The product of those z-transforms is

$$16 + 16z^{-1} + 8z^{-2} + 4z^{-3} + z^{-4}. \tag{2-6}$$

This polynomial is the z-transform of the output signal. Hence, our output signal is {16 16 8 4 1}.

To find the impulse response, the input and output must be given. Then, the Convolution Theorem can be applied by dividing the z-transform of the output by the z-transform of the input and expanding the result as a polynomial in z^{-1}. The coefficients of this polynomial represent the impulse response. To find the inpulse response, the process described in the previous example is reversed. For any given filter, the unit impulse response, which returns to and remains at zero, is an indication of the filter's stability, time-lag characteristic, and amplitude attenuation.

Another means of specifying a filter is by its step response, which is, the response to a constant signal. A constant signal is represented by a step function, which is described as

$$u(t-a) = \begin{cases} 0 & t < a \\ 1 & t \geq a \end{cases} \tag{2-7}$$

The unit step function is the integral of the unit impulse function. Likewise, the step response is the integral of the impulse response. Quite often, the function used is equal to 1 for a finite period; that is, $u(t) = 1$ for $a \leq t \geq b$, for some $b>a$, and $u(t) = 0$ for $t>b$.

Sometimes in actual practice, rather than using the impulse function, a smooth input step function is generated. A smooth input step function is a function that does not switch values abruptly, as does the step function just defined, but rather that

changes values in a continuous manner. There are many examples in which a smooth step function simulates real-life situations more accurately than does the ideal step function. For example, the acceleration of a rocket at motor burnout does not vanish instantaneously, but tapers off to zero gradually.

As an example of using the step function, the smooth input step function shown in figure 2-1 can be used as a model to evaluate a typical general-purpose, recursive, second-order filter. (An nth order filter is a filter that can be designed with an nth order differential equation. The acceleration at rocket motor burnout discussed in the previous paragraph can be simulated using a second-order filter). The step response is shown in figure 2-2. A recursive filter must be given initial values, since each output value depends on the previous output value. The first three points in figure 2-2 represent the initial values used by the prospective filter.

2.4 Variance Reduction Factor

Let F be an arbitrary digital filter. Assume that the input and output of F are stationary random processes. A process is called stationary if its statistics such as mean and variance are not affected by a shift in the time origin. Theoretically, assume that F has been filtering the same process for an infinitely long time for the output to be stationary. Further, assume that each of the noise processes for the input and output data are uncorrelated with themselves. Let v(1) and v(2) be the variances of the input and output random processes. Then the variance reduction factor is defined as

$$R = \frac{v(2)}{v(1)} \tag{2-8}$$

The variance reduction factor must be interpreted with caution, because the assumptions underlying the estimation of variances are, in most cases, not realistic. The parameter R is not necessarily less than one, although R is generally less than one in a practical case.

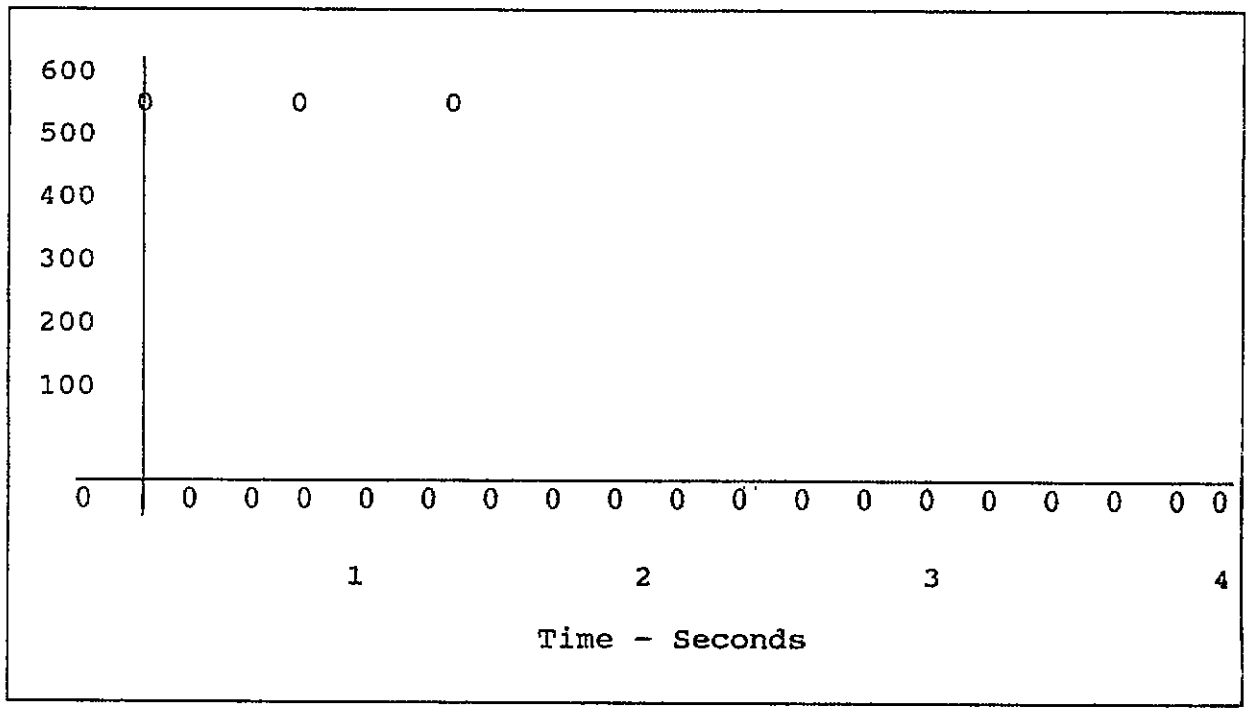


Figure 2-1. Smooth input data. Sampling period - 0.05 seconds.

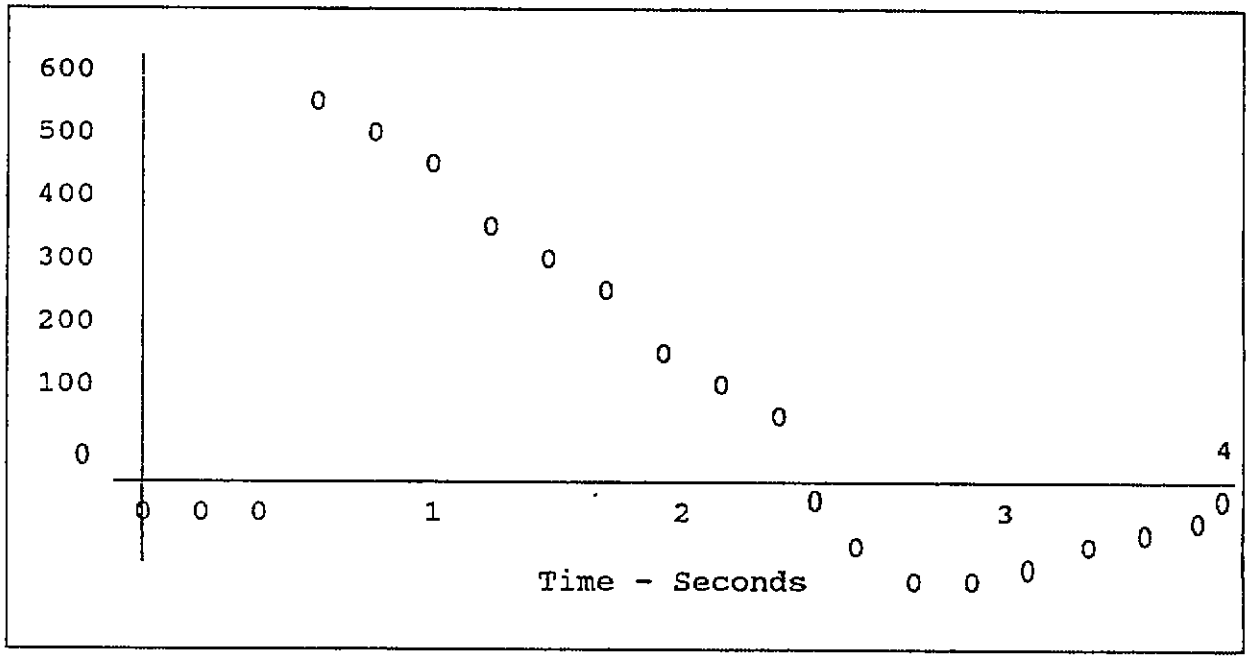


Figure 2-2. Real-time response. Point span - 20; sampling period - 0.05 seconds.

2.5 Simulation

Ordinarily, it is assumed that a signal process s(t) and a noise process n(t) defined for a discrete or continuous variable t are additive resulting in a process denoted by s(t) + n(t). A filter is used to suppress n(t) and extract s(t) with the least amount of distortion under a given set of operational criteria, for example, variance reduction factor, impulse response, and computation time. These factors usually vary from one process to another depending largely on the values of s(t) and n(t) even for a particular filter. It is therefore often desired to examine various response characteristics of a given filter applied to a specific signal plus noise process s(t) + n(t). The objective is to compare the filtered data, say $\hat{s}(t)$, obtained from s(t) + n(t) against the originally known signal process s(t). The act of perturbing s(t) with an additive noise process n(t) or filtering s(t) + n(t) or both is called the simulation of the signal plus noise process or of the filtering.

The simulation techniques are classified into two major categories: hardware simulations and numerical simulations. Simulated equipment tests and scale model experiments are typical of the hardware simulations. In the case of the numerical simulations, the entire system to be analyzed must first be represented by an appropriate mathematical model in the form of s(t) + n(t), in which the contribution of various perturbations is treated as the noise process n(t). The purpose here is to concentrate on numerical simulations.

The numerical simulations are divided into digital, analog, and hybrid. The digital simulations use digital computers as the primary simulation tools and are effective for discrete processes of the form s(i) + n(i), i=1,..., m. They are particularly suited for highly accurate simulation. Similarly, the analog simulations are made using analog computers or similar analog devices and are suited for the simulation of a process continuous in time. The hybrid simulations involve the mixture of both the digital and analog simulations incorporating digital-to-analog or analog-to-digital conversion processes or both. By the nature of their construction, neither the analog nor the hybrid computers can achieve a high degree of accuracy. Digital simulations have become more frequently used as the primary tools for analyzing systems performance and operational characteristics, even for those processes with continuous time variables.

Various digital filters may be applied to the output process of hardware simulations, digital simulations, or digitized values of analog/hybrid simulations. The performance of the filters is best evaluated under the conditions of the closed-loop simulation setup in figure 2-3. In a closed-loop simulation, the signal process s(t) is given a priori. The noise process n(t) is

generated using random noise generators or random number generators and is added to s(t), thus forming s(t) + n(t). The filter is applied to s(t) + n(t), yielding the filtered process $\hat{s}(t)$. The original signal process s(t) is fed into a comparator with appropriate delay to form the difference process $\hat{s}(t) - s(t)$, which can be evaluated for the performance of the filter.

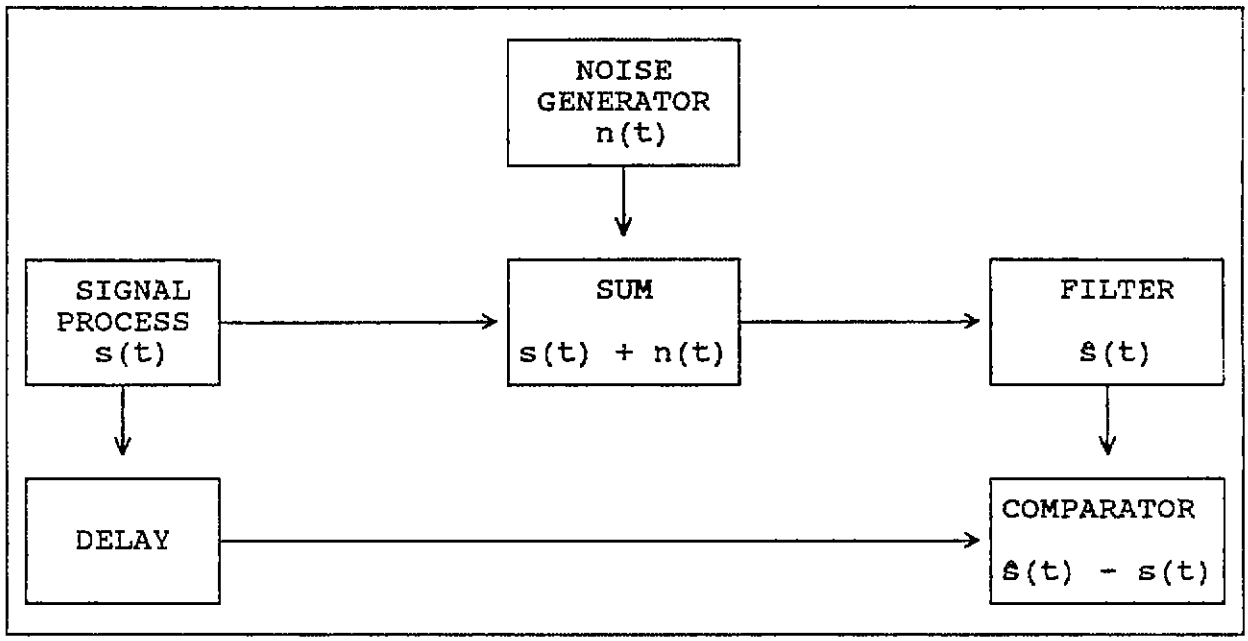


Figure 2-3. A closed-loop simulation logical diagram.

2.6 Monte Carlo Methods

An analogy of the random numbers used in digital simulations to the random outcomes of gambling devices such as roulette, dice, and cards has led to the use of the celebrated term "Monte Carlo" since the inception of digital simulations. The simulation of a random process by random numbers, followed by the calculation of statistical parameters of the end result of this process, is called a Monte Carlo technique. The random numbers in question can be generated by a computer. In the digital simulations of s(i) + n(i), i=1,..., m, the signal process s(i) is assumed to be known a priori and is available in the computer storage. The noise process n(i) is frequently assumed, unless facts to the contrary are known, to be independently and identically distributed random variables for all i=1,..., m with a given probability distribution function. The problem then reduces to the generation of random numbers with these qualifications.

2.6.1 Generation of Random Numbers From a General Distribution Function

Suppose {x(k): k = 1,..., }, are random numbers generated with a probability density function f(x), where f(x) can be either discrete or continuous in x. Basic concepts for the descrete and continuous cases are described in the following subparagraphs.

2.6.1.1 Definition of Discrete Case

If f(x) is discrete, it is then described as P{X = x}, where x is a discrete random variable (r.v.). There exists a probability value f(x(j)) for each random number x(j), j = 1, 2,..., ∞ such that

$$f(x(j)) \geq 0, \quad j = 1, 2, \ldots, \infty \tag{2-9}$$

and

$$\sum_{j=1}^{\infty} f(x(j)) = 1 \tag{2-10}$$

The cumulative probability F(x(n)) is defined to be

$$F(x(n)) = \sum_{i=1}^{n} f(x(i)) \tag{2-11}$$

It is the probability that the random variable X equals any one of the random numbers x(1), x(2),..., x(n). Notice that

$$\lim_{n \to \infty} F(x(n)) = 1.$$

2.6.1.2 Method of Random Number Generation

In the case of f(x) being continuous in x over the range of interest [a,b], f(x) is then described as having these three properties

$$f(x) \geq 0 \text{ for all } x \text{ in } [a,b] \tag{2-12}$$

$$\int_a^b f(x)dx = 1 \tag{2-13}$$

$$\int_c^d f(x)\,dx = P\{c \leq X \leq d\},$$

where X is a continuous r.v. and [c,d] is contained in $\subseteq$ [a,b]. The cumulative distribution function F(t) is defined to be

$$F(t) = \int_a^t f(x)\,dx, \quad a \le t \le b \tag{2-14}$$

Obviously, F(a) = 0, and F(b) = 1. (2-15)

2.6.1.3 Method of Random Number Generation

Let y(k), k=1,..., n, be random numbers uniformly distributed between 0 and 1. Random numbers x(k) can be generated from y(k) by finding a cumulative distribution function F mapping x to y, so that

$$y(k) = F(x(k)). \tag{2-16}$$

Pick F so that the new random numbers x(k) will be distributed in the manner desired. Now generate the new random numbers by using

$$x(k) = F^{-1}(y(k)). \tag{2-17}$$

In this manner, it is assured that the source y(k) of the new random numbers x(k) is random and uniformly distributed. The uniqueness of x(k) and the inverse mapping $F^{-1}y(k)$ can be seen immediately from figure 2-4. Equation (2-17) is applied often, since random number generators of many compilers produce numbers uniformly distributed between 0 and 1. The problem then narrows down to the generation of the random numbers y(k) from a uniform distribution over [0,1].

2.6.2 Random Numbers From a Uniform Distribution

Computers have mathematical routines which generate uniformly distributed random numbers. Various different earlier pseudo-random number generators have been examined in H. A. Meyer, Symposium on Monte Carlo Methods, John Wiley, New York, 1956.

One technique called the "congruence method" uses the following scheme. Given an initial random number x(k), $0 \le x(k) \le 2^n-1$, x(k+1) is computed recursively by

$$x(k+1) = (x(k)[1+2^a] \pm 1) \pmod{2^n} \tag{2-18}$$

where a = [n/2] ([·] is the greatest integer function). This technique is frequently used to generate random numbers in an n-bit binary machine.

2.6.3 Random Normal Deviates

Another kind of random number often used in digital simulation is that of the random normal deviates. A random normal deviate is a number in a set of random numbers that is normally distributed. Random normal deviates are used to simulate a random sample from a standard normal distribution. These random numbers can be generated using the mapping technique described in subparagraph 2.6.1 with the probability density function

$$f(x) = (2\pi)^{-1/2} \exp(-x^2/2). \qquad (2\text{-}19)$$

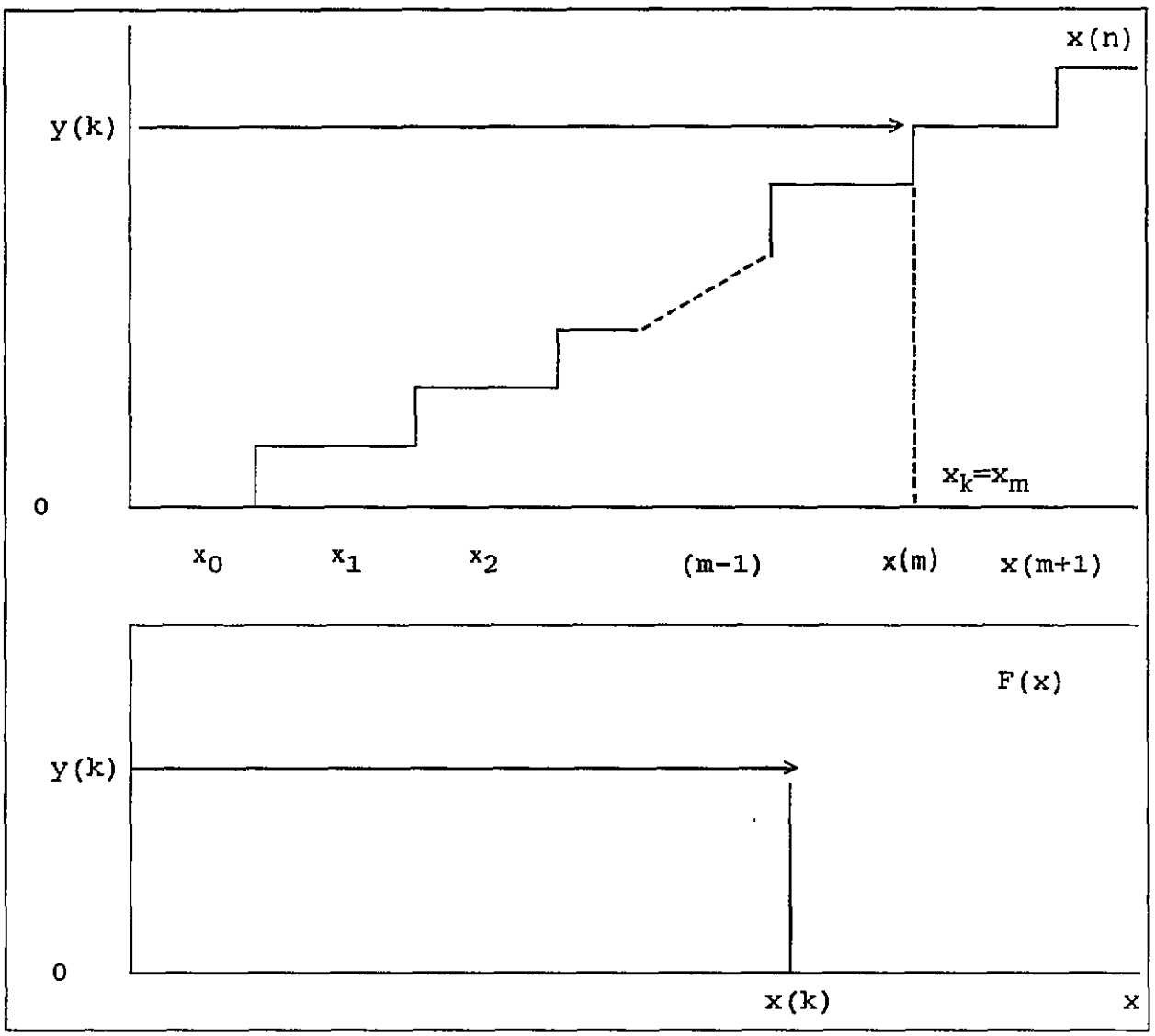


Figure 2-4. Graphic description of how the random number x(k) is determined. It is obtained with a cumulative distribution function F(x(n)) (in top graph) or F(x) (in bottom graph) and a random variable y(k) uniformly distributed over [0,1].

The resultant random numbers possess zero mean and unit variance. For the generation of random numbers $z(k)$, $k = 1,\ldots, N$, with a (nonstandard) normal distribution with mean m and variance v, use the transformation

$$z(k) = m + v(x(k)), \quad k = 1,\ldots, N. \tag{2-20}$$

2.7 Residuals

Some filter characteristics may be determined from using real test data by analyzing the residuals about the fit, that is, the differences between the filtered data values and the nonfiltered data values. A statistical analysis may be made for the autocorrelation and for the amount of variance or standard deviation. In addition, if the residual distribution includes a randomness about zero, this randomness is a good indication that the systematic error because of improper filtering is small. If these residuals do not display this randomness about zero, then the filter may be introducing bias or systematic error. The feature of randomness, however, in the residual analysis is only for the case when the noise is uncorrelated. Often the noise is correlated. Then the residual analysis becomes more complicated and must be undertaken with caution. Spectral analysis may be performed on the residuals as defined in RCC Document 153-71, Error Analysis and Methods For Estimating Errors in Position, Velocity, and Acceleration Data.

CHAPTER 3

FILTER ANALYSIS AND DESIGN IN THE FREQUENCY DOMAIN

3.1 Analysis of Filter Characteristics in the Time Domain

In the previous chapter, the analysis of filter characteristics in the time domain was considered. Also of interest are the characteristics of the filter in the frequency domain. To cite an instance, it may be necessary to know how the filter treats different frequencies which might be present in the data being filtered. Probably the most useful tool in any kind of analysis in the frequency domain is the Fourier transform as defined by

$$X(f) = \int_{-\infty}^{\infty} \exp(-2\pi ift)x(t)dt \qquad (3\text{-}1)$$

Here x(t) is some function of time, f is frequency, and $i = -1$. The function X(f), the Fourier transform of x(t), gives an indication of how different frequencies are distributed in the function x(t). To transfer from the frequency domain to the time domain, use the inverse Fourier transform, given by

$$x(t) = \int_{-\infty}^{\infty} \exp(i2\pi tf)X(f)df \qquad (3\text{-}2)$$

X(f) and x(t) are known as Fourier transform pairs.

3.2 Discrete Fourier Transform

Because the digital computer requires that input data be in sampled form, it is often more appropriate to use the discrete version of the Fourier transform given by

$$X(f) = t \sum_{j=-\infty}^{\infty} \exp(-2\pi ifj)x(j) \qquad (3\text{-}3)$$

where t is the time increment at which samples are taken. The variable t is sometimes referred to as the sampling period. Its reciprocal is sometimes referred to as the sampling frequency or sampling rate. The function X(f) is known as the discrete Fourier transform (DFT) of x(t). For a finite sequence

$$x(0), x(1),..., x(N-1) \qquad (3\text{-}4)$$

the discrete Fourier transform is given by the sequence

$$X(k) = \sum_{j=0}^{N-1} \exp(-2\pi ikj/N) \cdot t(x(j)) \quad k=0, 1, \ldots, N-1 \qquad (3\text{-}5)$$

The inverse discrete Fourier transform is given by

$$x(j) = (1/N) \sum_{j=0}^{N-1} \exp(-2\pi ikj/N) \cdot t(x(j)) \quad k=0, 1, \ldots, N-1 \qquad (3\text{-}6)$$

An efficient algorithm, developed by J. W. Tukey and J. W. Cooley published in 1965 and commonly used to calculate the discrete Fourier transform, is the Fast Fourier Transform (FFT). (A paper by I. J. Good a decade earlier describes a very similar algorithm.) Essentially, the FFT splits the sequence into two subsequences with each containing every other element of the original sequence, computes the discrete Fourier transforms (DFT) of those sequences, and rearranges the values of their DFTs back in their proper order to form the DFT of the original sequence. In this manner, the DFT can then be computed by using fewer complex operations (multiplications and additions).

Normally, the subsequences are divided again in the same manner as the original sequence to further reduce the number of operations required to compute the DFT of the original sequence. The subsequences are divided still further until there are only one-element sequences. When this subdividing has been completed, the ratio of the number of operations used for the FFT compared to the number of operations used to calculate the DFT directly is

$$\frac{\log_2 N}{N} \qquad (3\text{-}7)$$

where N is the order (length) of the original sequence.

It is often desirable that the original sequence have an order equal to a power of two, so it can be divided in the manner described earlier. In the event that the original sequence is not of such order, the sequence is often padded with zeroes, so that it can be completely reduced. The extra space used for padding the sequence with zeroes is more than compensated for by the savings afforded by the algorithm. Table 3-1 and figure 3-1 (obtained from references 47 and 53) demonstrate the difference between the number of multiplications used when calculating the DFT directly and that used when applying the FFT. See reference 52 for a brief and clear explanation of the Fast Fourier Transform.

TABLE 3-1. COMPARISON OF REQUIRED MULTIPLICATION OPERATIONS USING THE FFT AND THE DIRECT DFT.

N	N^2 (direct DFT)	$2N \log_2$ (FFT)
64	4,096	768
128	16,384	1,792
256	65,536	4,096
512	262,144	9,216
1,024	1,048,576	20,480

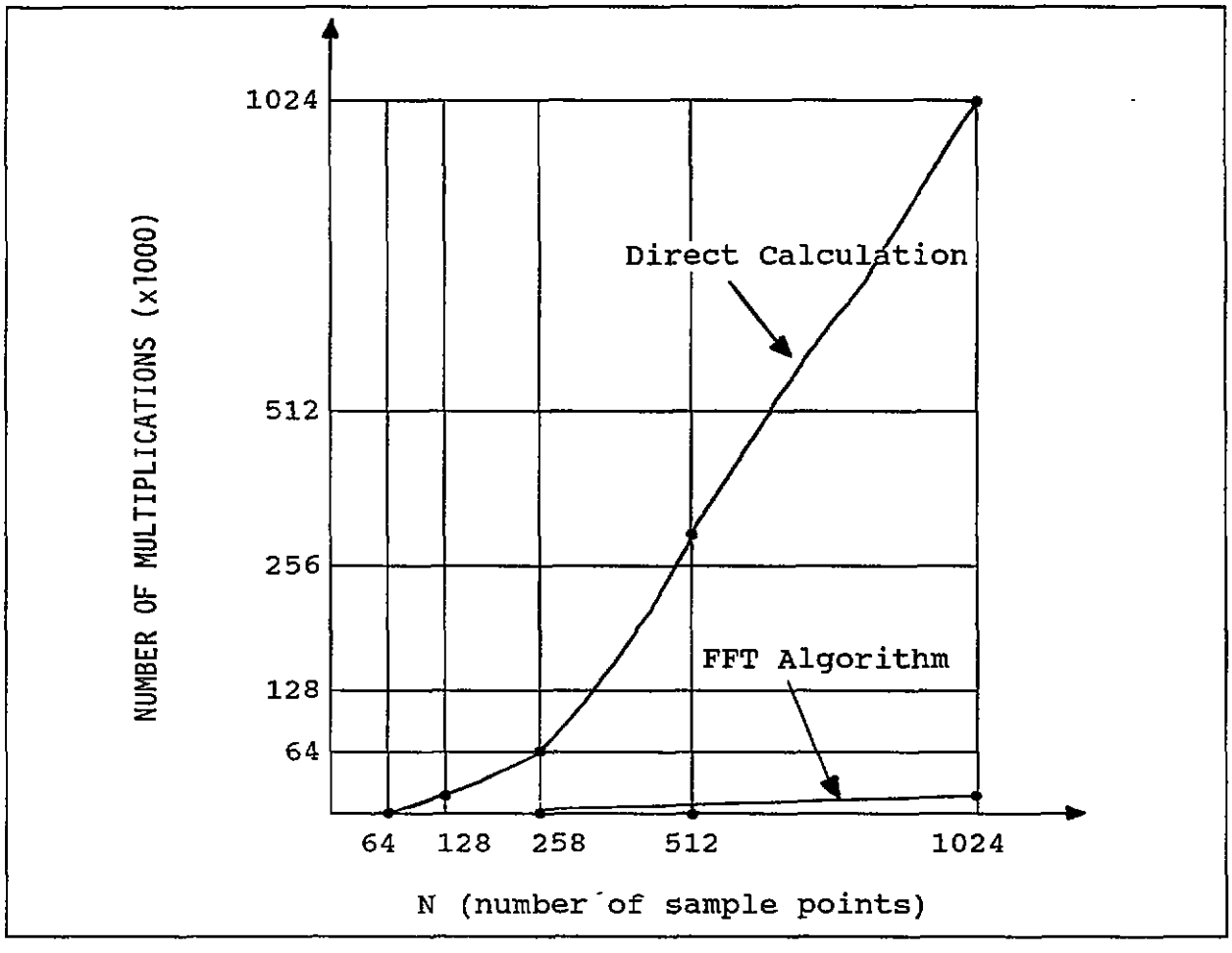


Figure 3-1. Comparison of multiplications required by direct calculation and FFT algorithm.

Another useful tool in the analysis of a filter frequency characteristics is the power spectral density function (psd) or power spectrum. This function's name came from a widely used term in the field of electrical engineering where it gives distribution of the signal power for each frequency. In this case, the power spectrum is measured in terms of watts per hertz. (In electrical engineering, the Fourier transform as a function of time is measured in hertz. Hertz is cycles per second.) The power spectral density function is the Fourier transform of the autocorrelation function, which is defined in chapter 2. In the statistical sense, it indicates how the variance is distributed as a function of frequency. For the discrete case, the power spectral density function of {x(j)} is given by the following sum

$$G_x(f) = \sum_{j=-\infty}^{\infty} \exp(-2\pi i f(j)) \cdot t(R(j)) \qquad (3\text{-}8)$$

where R(j), the autocorrelation function, is given by

$$R(j) = \sum_{i=-\infty}^{\infty} x(i)x(i-j) \qquad (3\text{-}9)$$

One of the best ways of analyzing a filter's frequency properties is to look at the frequency content of the output when the input is white noise. As mentioned in chapter 1, the power spectral density of white noise is a constant for all frequencies. Thus, by taking the power spectral density of the output of a filter whose input is white noise, it can be determined how the filter treats all the frequencies in the data. For example, it might be desired that the filter pass frequencies in a band $f_1 \leq f \leq f_2$ while rejecting all frequencies outside that band. Then, if the input is white noise, the ideal power spectral density of the output would be a translated rectangle as shown in the figure 3-2. In other words, if the psd of the output is that shown in the figure 3-2, then the appropriate frequencies (those less than f_1 or greater than f_2) were filtered out.

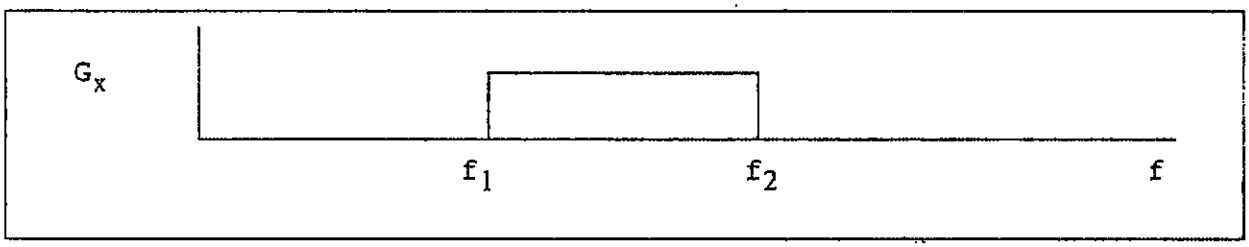


Figure 3-2. Power spectrum of output, where the input is white noise and where the filter passes only the frequencies between f_1 and f_2.

An indication of how good the filter is can be seen by comparing the power spectral density of the filter output, when the input is white noise, to this idealized power spectral density.

3.3 Transfer Function of a Filter

In chapter 1, it was noted that the output of a filter equals the input convolved with the filter weight function. A convolution for a time-invariant system takes the form

$$y(t) = \int_{-\infty}^{\infty} h(T)x(t-T)\ dT \qquad (3\text{-}10)$$

where h(T) is the weight function.

By taking the Fourier transform of the above equation, the equation becomes

$$Y(f) = H(f)X(f) \qquad (3\text{-}11)$$

The Fourier transform of the weight function, H(f), is known as the transfer function of the filter and represents the ratio of the filter output to the filter input in the frequency domain.

Repeating equation (1-18), the equation for a nonrecursive, time-invariant digital filter is

$$y(j) = \sum_{k} h(k)x(j-k) \qquad (3\text{-}12)$$

The transfer function H(f) for this filter is given by

$$H(f) = \sum_{k} h(k)\exp(-2\pi ifk)\cdot\Delta t \qquad (3\text{-}13)$$

Likewise, we repeat equation (1-19), the equation for a recursive time-invariant digital filter:

$$y(j) = \sum_{k=0}^{N} a(k)x(j-k) + \sum_{k=1}^{M} b(k)y(j-k) \qquad (3\text{-}14)$$

The transfer function H(f) for this filter is given by

$$H(f) = \frac{\sum_{k=0}^{N} a(k)\exp(-2\pi ifk)}{1 + \sum_{k=1}^{M} b(k)\exp(-2\pi ifk)} \qquad (3\text{-}15)$$

The modulus of the transfer function is called the gain of the filter and represents the amplification of the input.

Filters can be categorized according to the characteristics of their transfer functions. A filter is called an ideal low-pass filter if

$$|H(f)| = \begin{cases} 1 \text{ for } 0 \leq |f| < f_c \\ 0 \text{ for } f_c < |f| \end{cases} \qquad (3\text{-}16)$$

The plot of $|H(f)|$ is as shown in figure 3-3.

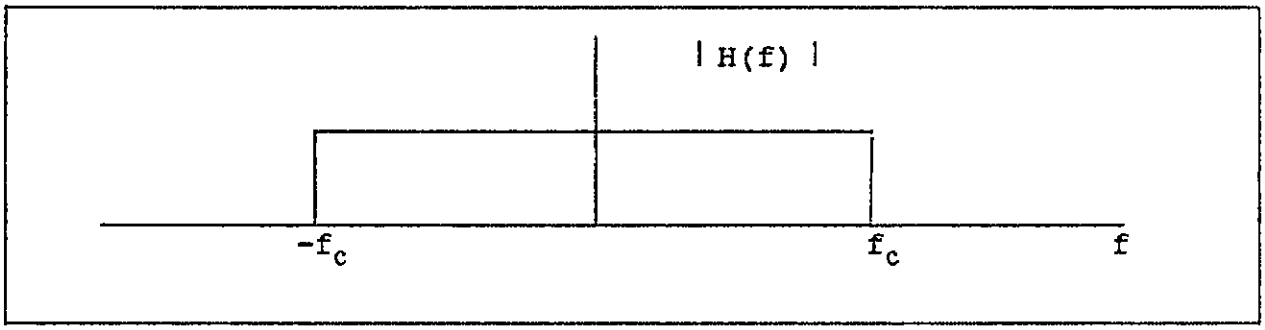


Figure 3-3. Ideal low-pass filter.

A filter is called an ideal high-pass filter if the transfer function is given by

$$|H(f)| = \begin{cases} 0 \text{ for } 0 \leq |f| \leq f_H \\ 1 \text{ for } f_H < |f| \end{cases} \qquad (3\text{-}17)$$

and the plot of $|H(f)|$ is as shown in figure 3-4.

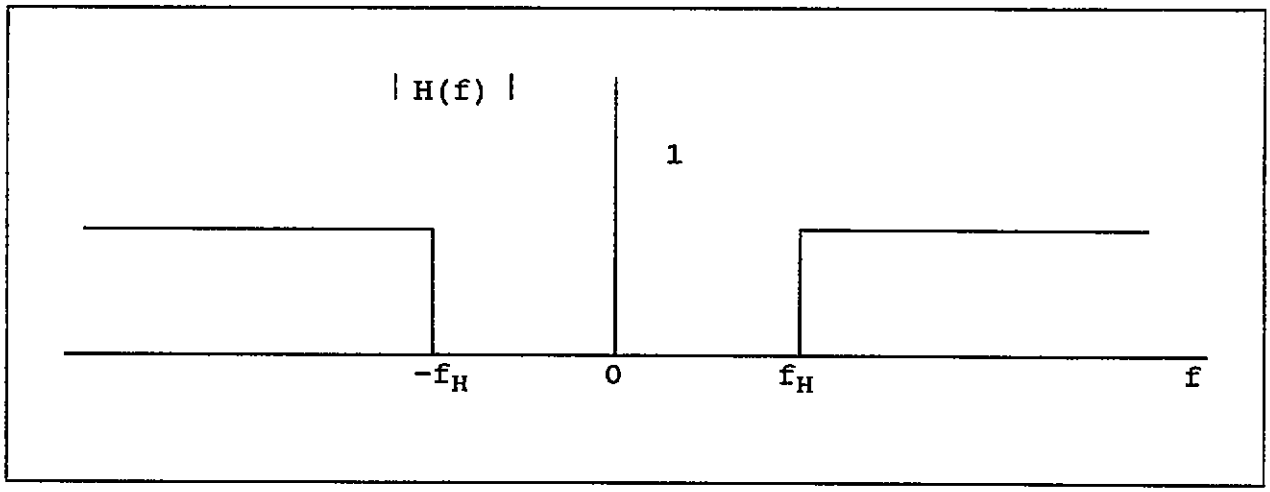


Figure 3-4. Ideal high-pass filter.

A filter is called an ideal band-pass filter if

$$|H(f)| = \begin{cases} 0 \text{ for } 0 < |f| < f_H \\ 1 \text{ for } f_H \leq |f| \leq f_L \\ 0 \text{ for } f_L < f \end{cases} \qquad (3\text{-}18)$$

and the plot is as shown in figure 3-5.

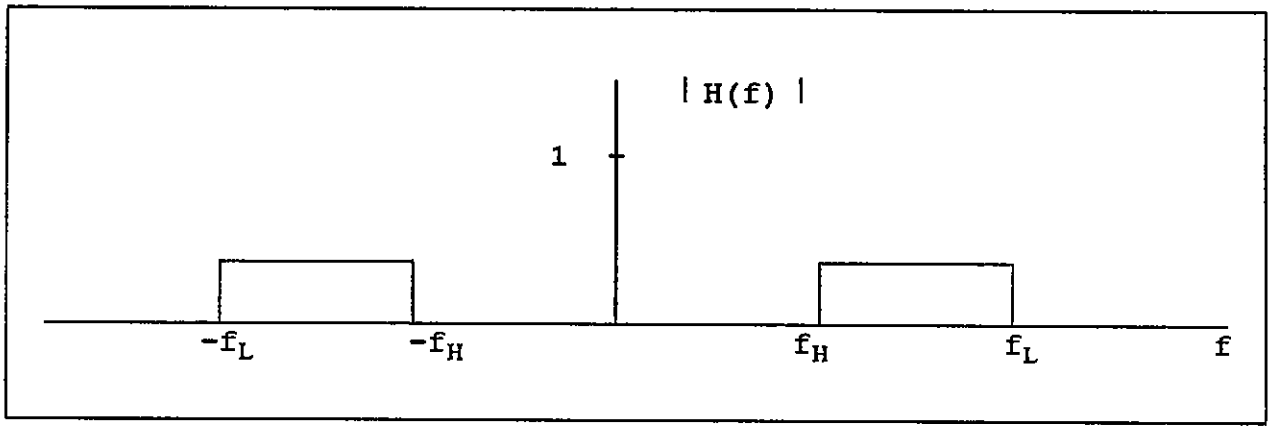


Figure 3-5. Ideal band-pass filter.

These ideal filters cannot be attained because there is, in practice, a finite number of samples. The impulse response in equation (3-10) is multiplied by a "boxcar" or window function which covers the time interval being sampled. Alternatively, the limits of integration can be changed to the appropriate, finite time limits. As stated in paragraph 1.5, the transform of the product of two functions is the convolution of the transforms of the two functions. Now the transform of the boxcar function is a sinc function. (By definition, sinc $x = \sin(\pi x)/\pi x$.) The resulting transfer function is the sinc function convolved with the transfer function of an ideal filter. The graph of the resulting gain is a boxcar function with ripples and not the ideal boxcar function illustrated in this section.

Methods for simulating the ideal low-pass filter are given in chapter 5. High-pass, band-pass, and band-rejection filters (discussed in chapter 4) are normally derived by first designing a low-pass filter. Methods for obtaining all of these filters are given in chapter 4.

3.4 Cutoff Frequency, Roll-Off, and Quality of Filter

In the previous section, a low-pass filter was defined in terms of a frequency denoted as f_c. This frequency f_c is called the cutoff frequency (see equation 3-16). The cutoff frequency is used as a criterion for designing digital filters in the frequency domain. It also gives a standard for analyzing the performance of filters in the frequency domain.

As explained in paragraph 3.3, ideal filters are simulated and not used directly. The simulating filters do not have gains that look like the boxcar functions drawn in paragraph 3.3. The gain of a simulating filter does not suddenly drop off from 1 to 0 the way the gain of an ideal filter does. Instead, it normally starts off (at f=0) with a value that is within a predetermined tolerance δ_1 of 1 and tapers off to within a predetermined tolerance δ_2 of 0.

The band of frequencies in which the gain is between $1-\delta_1$ and $1+\delta_1$ is called the passband. The greatest frequency value in the passband is the cutoff frequency for the passband. The band of frequencies in which the gain is $\leq\delta_2$ is called the stopband. The lowest frequency value for the stopband is the cutoff frequency for the stopband. The band in which the gain is $< 1-\delta_1$ and $> \delta_2$ is called the transition band. These items are illustrated in figure 3-6, which is obtained from reference 55.

One commonly used simulation to the ideal low-pass filter that is not designed on the basis of a transition band and two cutoff frequencies is the Butterworth filter. This filter and other simulations are given and explained in chapter 5.

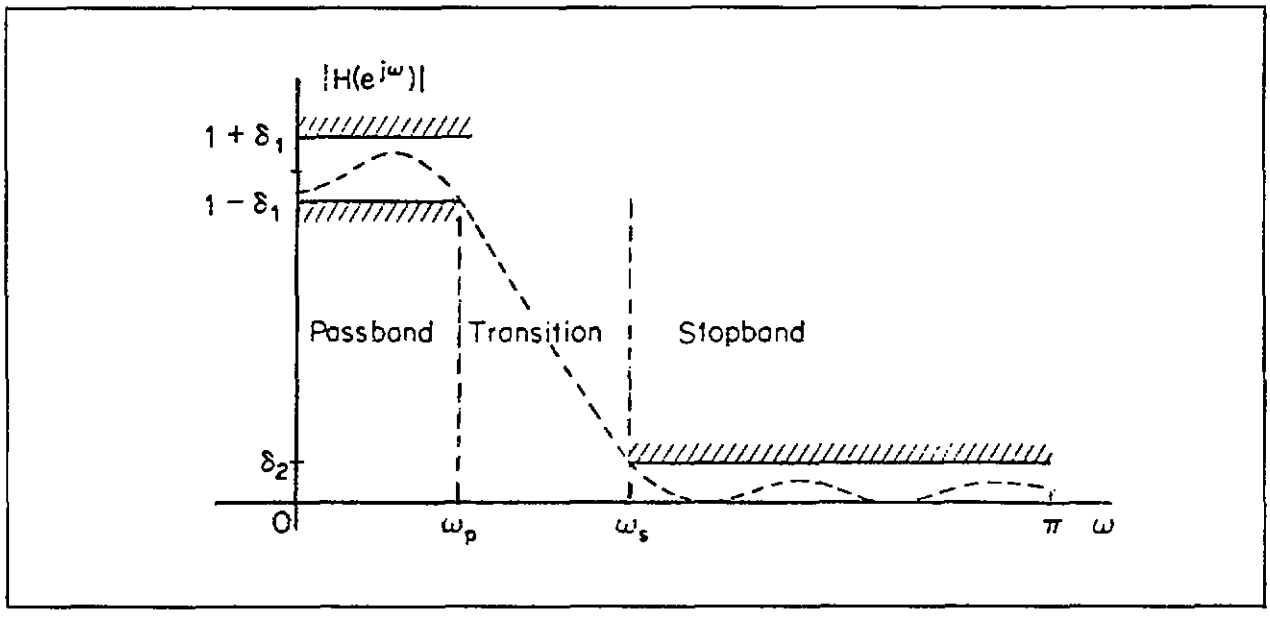


Figure 3-6. Tolerance limits for approximation of ideal low-pass filter.

Another important concept in the analysis is the roll-off. This quantity indicates to what extent the slope of the transfer function of a simulating (nonideal) filter (see paragraph 3.3) approaches the 90° slope of the transfer function of the ideal filter. Roll-off is related to the intuitive notion of "sharpness" of the cutoff. This quantity is defined as

$$p = \frac{-f}{|H|} \frac{d|H|}{df} \qquad (3\text{-}19)$$

A value for p can be obtained at any frequency desired. For purposes of comparison and analysis of filters, this quantity is often calculated at $|H| = 1/\sqrt{2}$. In figure 3-7, the plots of $|H(f)|$ for two filters, A and B, which approximate the ideal low-pass filter are given. Both filters have the same cutoff frequency, but A has a higher rolloff than B.

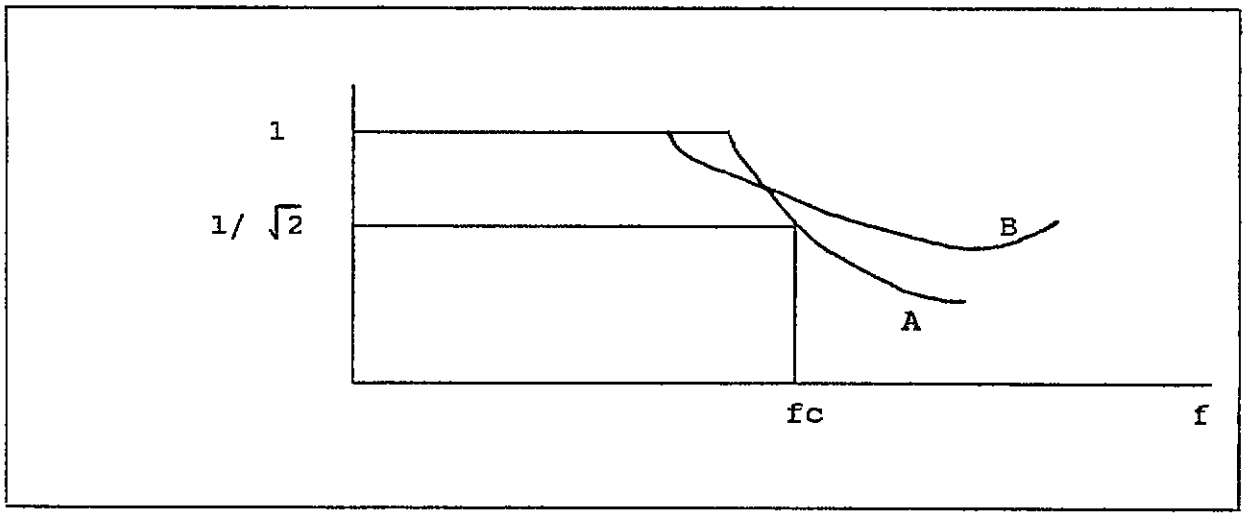


Figure 3-7. Comparison of rolloffs for filters A and B.

A third concept used in analyzing low-pass filters in the frequency domain is quality, which is defined as the extent to which the filter attenuates frequencies well above the cutoff frequency. (An attenuating function in the frequency domain is one that approaches zero as the frequency increases.) There is no standard mathematical definition for quality, so some subjective judgment must be used to determine the quality of a filter. One tool that can be used in making this judgment would be the power spectral density function as discussed in paragraph 3.2. For example, if two filters, A and B, yield the power spectral density functions shown in figure 3-8, then it might be concluded that filter A has the better quality.

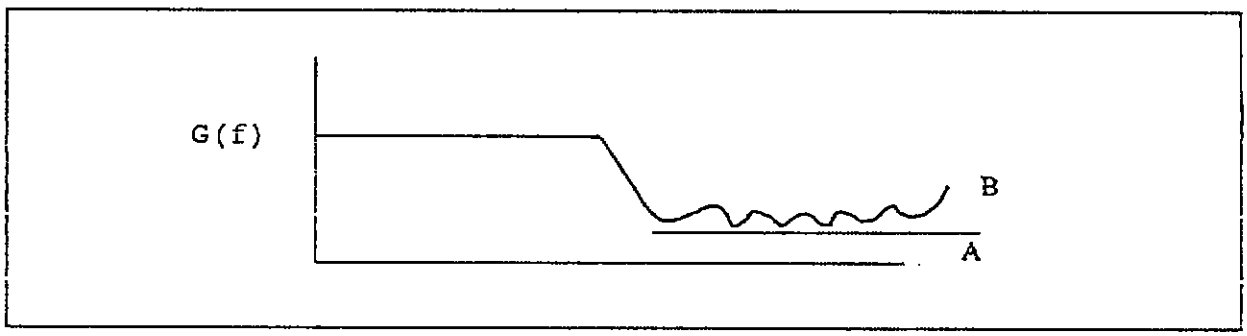


Figure 3-8. Comparison of the quality of filter A with the quality of filter B.

3.5 Phase Shift of a Filter

The transfer function H(f) of a filter is generally a complex number and can be expressed as

$$H(f) = |H(f)| \exp(i\theta(f)) \tag{3-20}$$

$\theta(f)$ is called the phase shift of the filter. Alternate names are phase response and phase angle. A time shift in the time domain corresponds to a phase shift in the frequency domain.

Equation (3-20) can be expressed as

$$H(f) = |H(f)| \cos(\theta(f)) + i\,H(f)\,\sin(\theta(f)) \tag{3-21}$$

or

$$H(f) = \mathrm{Re}(H(f)) + i \cdot \mathrm{Im}(H(f)) \tag{3-22}$$

where

$\mathrm{Re}(H(f)) = |H(f)| \cos(\theta(f)$ and $\mathrm{Im}(H(f)) = |H(f)| \sin(\theta(f))$. Re and Im denote real and imaginary parts of a complex number.

$$\text{Now} \quad \frac{\mathrm{Im}(H(f))}{\mathrm{Re}(H(f))} = \tan(\theta(f)) \tag{3-23}$$

The phase shift of the filter is then

$$\theta(f) = \arctan \frac{\mathrm{Im}(H(f))}{\mathrm{Re}(H(f))} \tag{3-24}$$

The output of the filter is in phase with the input only if $\mathrm{Im}(H(f)) = 0$. In most applications, this phase shift is important, and the analyst should be aware of its magnitude.

3.6 Construction of Filter Weights

Again consider a linear, nonrecursive digital filter whose defining equation is

$$y(k) = \sum_{j=o}^{M} h(j)x(k-j) \tag{3-25}$$

In this section, a general discussion of the way the $\{h(j)\}$ are determined will be given so that the filter will have a desired frequency response function. The frequency response is the same as the transfer function except that it is applied to a specific class of inputs in the form

$$x(n) = \exp(i2\pi fn). \tag{3-26}$$

$$x(n) = \exp(i2\pi fn). \tag{3-26}$$

One way of approaching this problem is to assume a desired frequency response function, $H(\omega)$, then take the inverse Fourier transform to get $h(t)$ as shown in equation (3-27). The variable ω is the angular frequency and is equal to $2\pi f$.

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(i\omega t)H(\omega)\, d\omega \tag{3-27}$$

The weights for the discrete filter are then determined by evaluating this continuous weight function at specific instants in time,

$$h_n = h(t_n)\Delta t. \tag{3-28}$$

There are problems with constructing these weights if the ideal frequency response is used. For example, the ideal frequency response for the low-pass filter is the unit rectangle as given in figure 3-3. The inverse Fourier transform must be truncated in the applied case, so only a finite number of weights are used in equation (3-25). The truncated Fourier transform of the unit rectangle is very "bumpy" (see paragraph 3.3, page 3-7) because of the discontinuity at f_c and is not useful in most applications.

Approaches to avoiding this dilemma are given in references 7, 11, and 13. In reference 7, Gennery eliminates the dilema by using various modifications of the Gaussian distribution function for the ideal response function. In reference 11, Ormsby replaces the ideal frequency response function with a function having a finite slope. In reference 13, Stirton multiplies the $h(t)$ for the ideal filter by certain "apodizing" functions which eliminate the bumps in the frequency response function.

In certain applications, the z-transform rather than the Fourier transform is used for designing filters in the frequency domain. The z in the z-transform is often used as a replacement for the exponential function in the discrete Fourier transform (DFT). It is used as such for computational ease and to facilitate transfer function analysis. For example, it is easier to determine stability of a filter by looking at the z-transform rather than by looking at the DFT. Determining filter stability in this manner will be discussed again in paragraph 3.8 and in chapter 5.

The z-transform approach is used when the transfer function of an analog filter is known, and when construction of a digital filter having that same transfer function is desired. The z-transform of an analog function $f(t)$ is defined as

$$\mathcal{Z}(f(t)) = \sum_{n=0}^{\infty} f(nT)z^{-i} \qquad (3\text{-}29)$$

It can be shown that the z-transform of the weight function is the ratio of two polynomials in the form

$$H(z) = \frac{\sum_{n=0}^{N} a(n)z^{-i}}{1 + \sum_{n=1}^{M} b(n)z^{-i}} \qquad (3\text{-}30)$$

if the filtered output {y(k)} is given by the recursive relation

$$y(k) = \sum_{n=0}^{N} a(n)x(k-n) + \sum_{n=1}^{M} b(n)y(k-n) \qquad (3\text{-}31)$$

It can also be shown that the z-transform of the weight function is in the form

$$H(z) = \sum_{n=0}^{N} a(n)z^{-i} \qquad (3\text{-}32)$$

if the filtered output {y(k)} is given by the nonrecursive relation

$$y(k) = \sum_{n=0}^{N} a(n)x(k-n) \qquad (3\text{-}33)$$

Other ways in which H(z) can be expressed are discussed in chapter 5.

3.7 Aliasing

In performing with digital filters, samples are taken from an analog sinusoidal waveform, and the waveform is reconstructed from the samples. The samples are typically taken at equal-spaced time intervals. The problem arises about how short these sampling intervals should be so that the original analog waveform can be exactly reconstructed. If the sampling intervals are too large, that is, if the samples are too infrequent, the original analog waveform may be reconstructed to have a lower frequency than it actually has. This misrepresentation of a frequency by a lower frequency is known as aliasing.

The appearance of wheel spokes turning slower than they actually are is an example of aliasing. The alternate term "foldover" can be expressed by the next example. Suppose an observer is taking visual samples at a constant time rate of the fast-turning wheel spokes. As the wheel's rotational speed increases, the wheel spokes seem to be rotating faster up to a certain point. Then the wheel spokes appear to slow down until they seemingly rotate faster and faster in the backward direction up to a certain point. The spokes then seemingly slow down until it looks as though they are going in a forward direction. If the observed rotational motion is placed on a plot of observed frequency versus actual frequency, the graph "folds over" at a frequency value known as the Nyquist or folding frequency (see figure 3-9). The F_o axis represents the observed frequency while the f_a axis represents the actual frequency, and the value f_N is the Nyquist frequency.

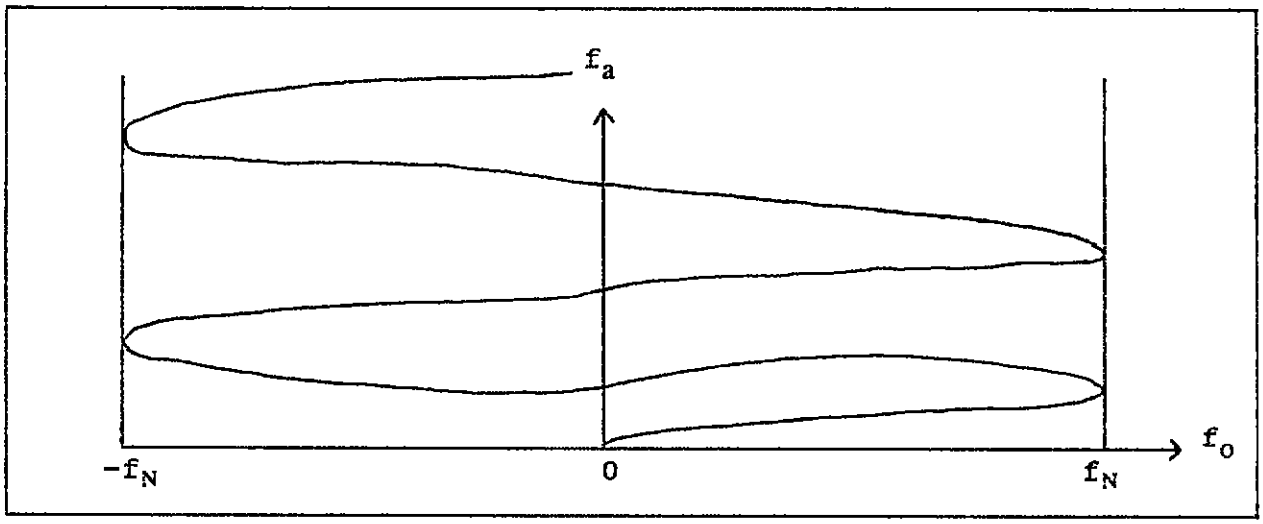


Figure 3-9. Foldover or aliasing resulting from observing a wheel's increasing rotational speed.

3.7.1 Aliasing in the Time Domain

An illustration of aliasing in the time domain is given in figure 3-10. The original analog waveform is the thinner sinusoid and is sampled at the points shown. With these sampled points, the original waveform is reconstructed so that the result is the thicker line which, of course, has a lower frequency than the original waveform. In producing new waveform, the original one was misreconstructed because the samples were taken too infrequently (see figure 3-10).

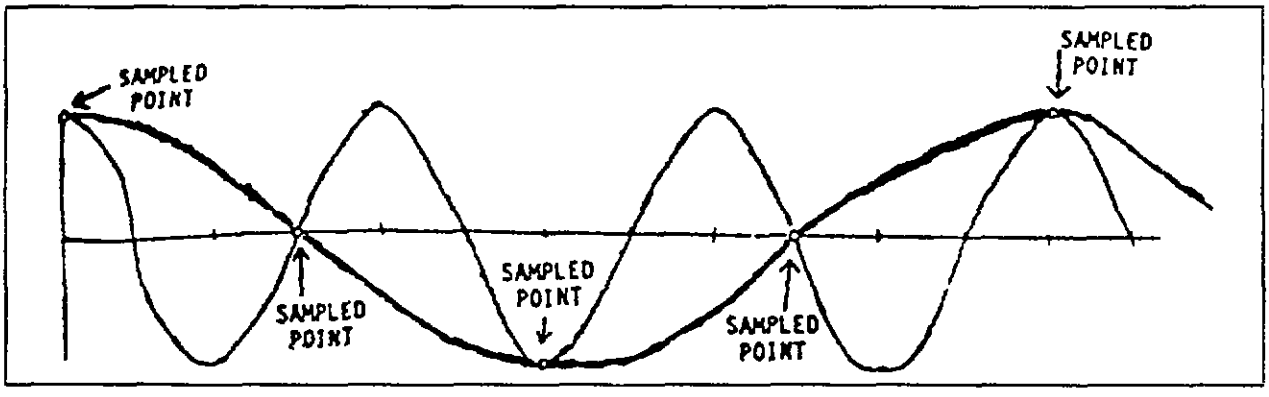


Figure 3-10. Aliasing in the time domain.

More often than not, the original waveform represents the sum of many sine and cosine functions. Figure 3-11, obtained from reference 52, illustrates this notion. The bottom sinusoid is the sum of the top three sinusoids. The frequency of each of the top three sinusoids is called a "frequency component" of the bottom sinusoid.

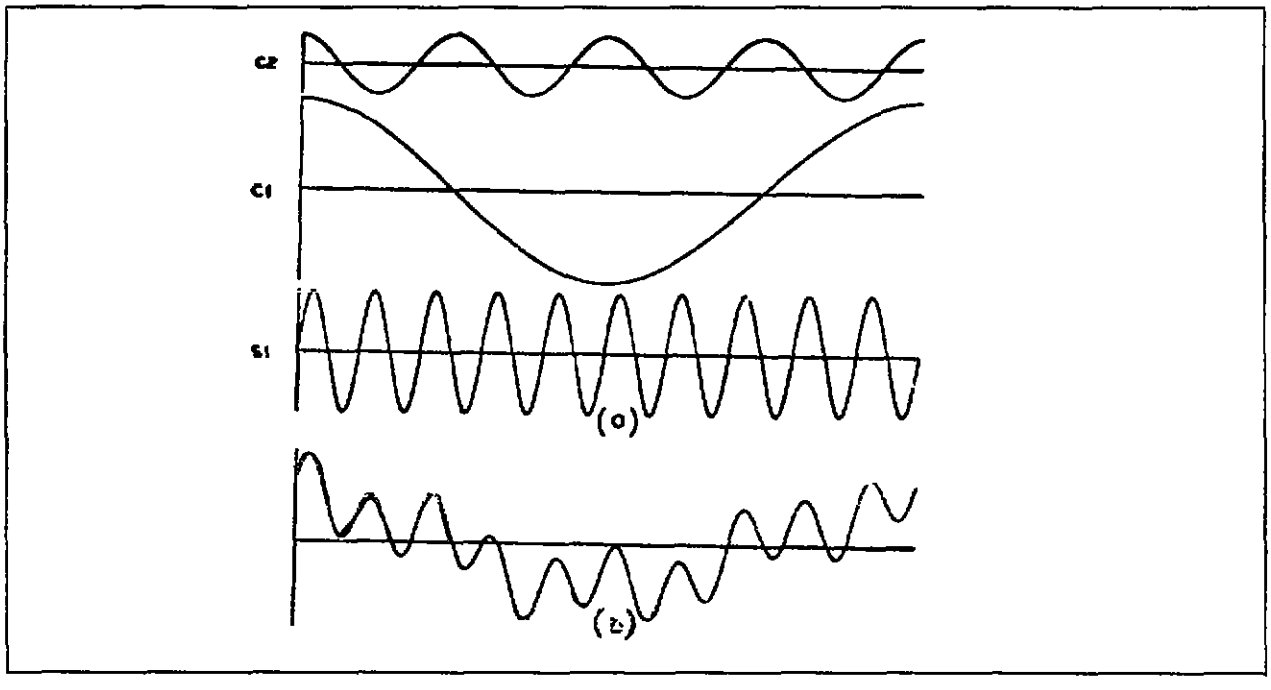


Figure 3-11. Bottom sinusoid shown as the sum of three frequency components, C2, C1, and S1.

3.7.2 Aliasing in the Frequency Domain

The discrete Fourier transform of a signal can be expressed in terms of the analog Fourier transform by

$$X(e^{i\omega T}) = (1/T) \sum_{m=-\infty}^{\infty} X_A(\omega + (2\pi/T)m) \qquad (3\text{-}34)$$

where $X_A()$ (in the right-hand side) is the analog Fourier transform, and $X()$ (in the left-hand side) is the discrete Fourier transform. This equation is derived in reference 55.

The frequency interval of each term in the summation sign of equation (3-34) will not intersect with its adjacent interval if it is at most $2\pi/T$ wide. Figure 3-12, obtained from Reference 54, point to this fact. Figure 3-12 illustrates the analog Fourier transform of one of the terms, which covers a frequency bandwidth of $2\pi/T$ and illustrates the infinite sum of such terms. As can be seen, $2\pi/T$ is the widest that these terms can be without intersecting each other.

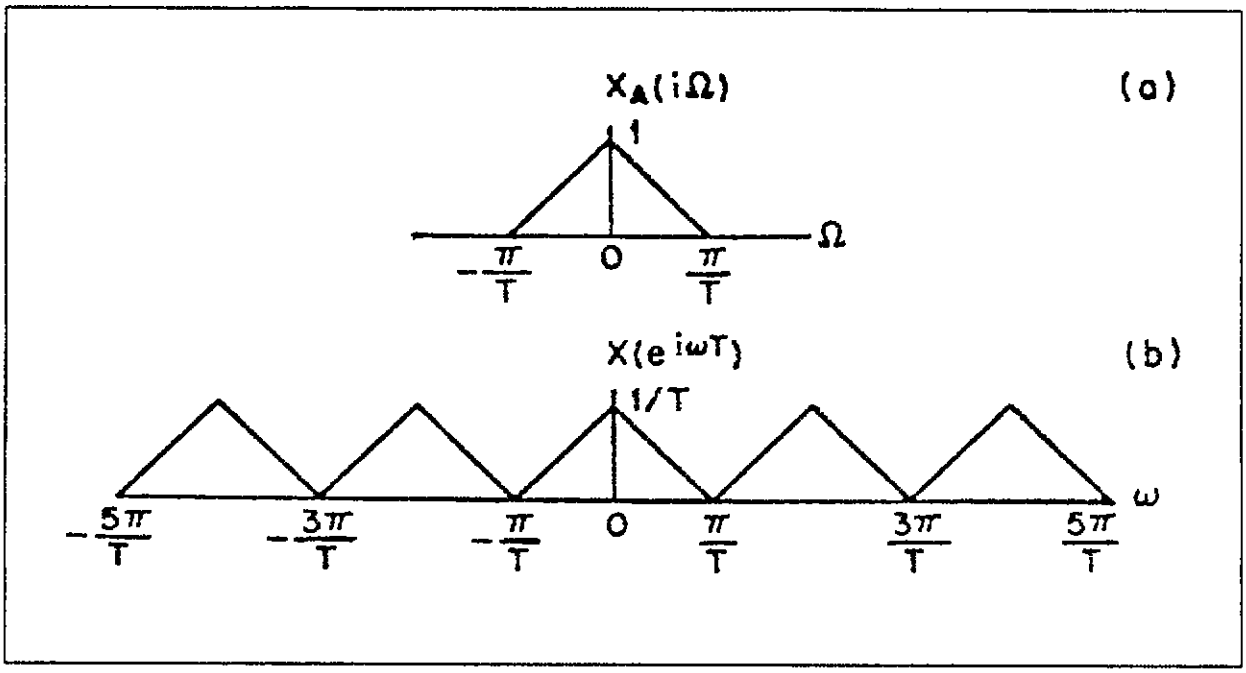


Figure 3-12. Sampling relations for analog and digital systems for properly sampled inputs.

Figure 3-13, obtained from reference 54, gives an example where samples are taken too infrequently, in particular, where they are taken at a rate of $w/3\pi$, which is less than $w/2\pi$. In other words, the frequency bandwidth for each term, which is $3\pi/T$, is greater than $2\pi/T$.) Figure 3-13a shows the graph of one of the terms in the summation of equation (3-34), when substituting $3\pi/T$ for w. Figure 3-13b explains how the adjacent terms overlap each other, and figure 3-12c displays the resulting graph of the Fourier transform. It is easily seen how the frequency $3\pi/2T$ can be mistaken for the frequency $\pi/2T$. This phenomenon, where in effect a frequency component takes on the identity of a lower frequency, is aliasing or foldover.

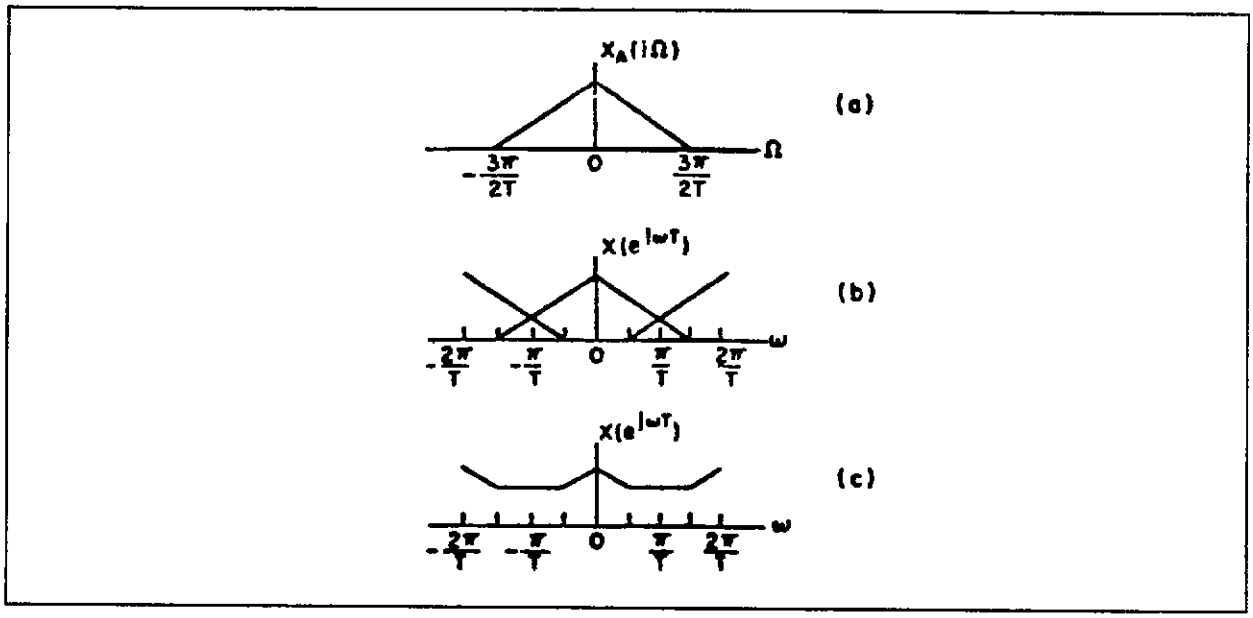


Figure 3-13. The effects of undersampling on the digital frequency response.

3.7.3 Sampling Theorem and the Nyquist Frequency

The Sampling Theorem offers a way to avoid aliasing. It states that if the Fourier transform of a signal is bandlimited, then the original signal can be exactly reconstructed if samples are taken at a frequency of at least twice that of the highest frequency component in the signal. In mathematical terms, the Sampling Theorem states if $F(\omega)$, the Fourier transform of a periodic time function f(t), is bandlimited so that

$$F(\omega) = 0 \text{ for } |\omega| \geq \omega_c \quad (\omega_c = 2\pi f_c) \tag{3-35}$$

then f(t) can be uniquely determined if it is sampled at intervals no greater than $1/2f_c = \pi/\omega_c$, that is, at a sampling rate of no less than $2f_c = \omega_c/\pi$.

Figures 3-11 and 3-12 imply the validity of this theorem. From these figures, it can be seen that the frequency ω_c should be at most π/T to avoid foldover. In other words, the sampling rate $1/T$ should be no less than ω_c/π.

For samples f_n which equals $f(t_n)$, where $t_n \leq n/2f_c = n\pi/\omega_c$, then the original analog signal f(t) can be reproduced by

$$f(t) = \sum_{n=-\infty}^{\infty} f_n \frac{\sin(\omega_c t - n\pi)}{\omega_c t - n\pi} \qquad (3\text{-}36)$$

$2f_c$ is known as the Nyquist or folding frequency. (This theorem is stated and elegantly proven in subparagraph 9.1.1 of reference 47.)

3.8 Stability of Filter

A digital filter is stable if every bounded input sequence yields a bounded output sequence through this filter. A sequence is bounded if all its terms are less than a specific positive integer. It can be shown that a necessary and sufficient condition for stability is that the impulse response $|h(k)|$ be such that

$$\sum_{k=-\infty}^{\infty} |h(k)| < \infty \qquad (3\text{-}37)$$

From this condition, it can be shown that another indication of filter stability is that all the poles of the transfer function be inside the unit circle in the (complex) z plane. If poles are on the unit circle, then the filter may or may not be stable. If a pole is outside the unit circle, the filter is not stable. See chapter 5 for the relation between pole location and filter stability.

CHAPTER 4

CATEGORIES OF DIGITAL FILTERS

In Chapter 1, some of the fundamental principles of digital filters were discussed. As stated earlier, the basic reason for using a digital filter is to separate or suppress errors and to pass signals without significant distortion. As discussed in Chapter 3, the frequency response of the filter should be such that those frequencies consisting mainly of the desired signal should be passed through the filter, and those frequencies consisting mainly of noise (error) should be rejected. If the complete statistical characteristics of the signal and noise are known, then using the principles discussed in chapter 3, a filter can be constructed to separate signal from noise in an optimum manner.

4.1 Low-Pass Digital Filters

The data from missile trajectory work usually contains a signal which consists mainly of large low-frequency components. The noise error is usually assumed distributed throughout the frequency spectrum with the desired signal being much greater in amplitude than the noise at low frequencies but smaller than the amplitude of the noise at the high frequencies. In this case, a low-pass filter is designed to pass the desired signal.

A low-pass filter has a frequency response of exact unity at zero frequency, approximate unity at low frequencies, and approximately zero at higher frequencies (see figure 4-1). The low-pass filter is simulated by the ideal low-pass filter defined in paragraph 3.3). The frequency at which the transition occurs between the high and low frequencies is called the cutoff frequency (f_c) of the filter. Because a low-pass filter removes the high frequency fluctuations from the data, it is often referred to as smoothing the data.

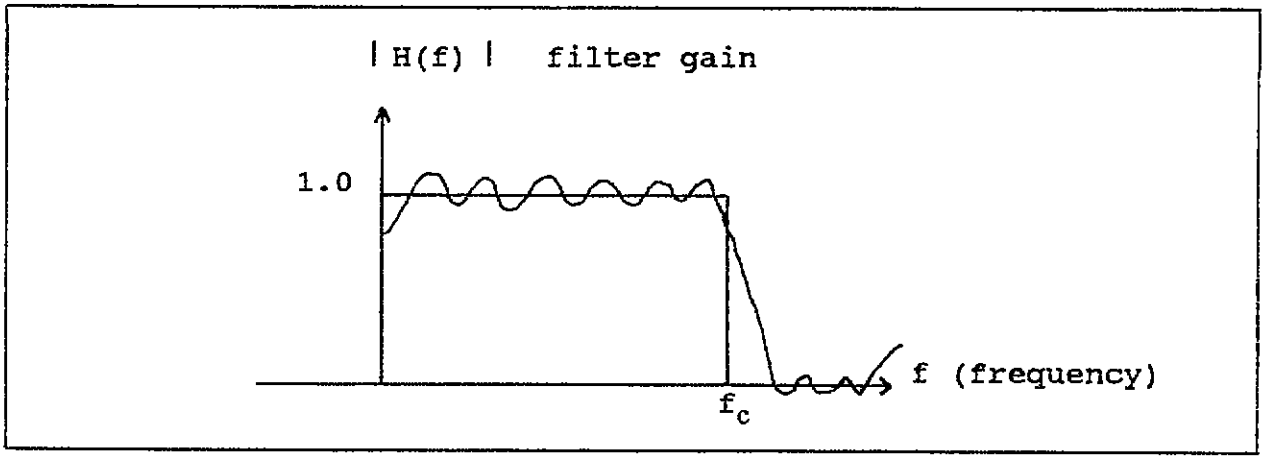


Figure 4-1. Typical frequency response of low-pass filter (represented by the curved line).

4.2 High-Pass Digital Filters

The category of low-pass filters discussed in the previous section can be transformed into filters which perform other tasks. In analysis work, one of the most commonly used is the high-pass filter. As the name implies, a high-pass filter is designed to pass high frequencies and reject the low frequencies. A low-pass filter can be transformed into a high-pass filter in several conceptually equivalent ways. One of the most common way is to subtract its frequency response from unity at all frequencies. That is, if $H_L(f)$ is the frequency response of a low-pass filter, then

$$H_H(f) = 1 - H_L(f) \qquad (4\text{-}1)$$

defines the frequency response of a high-pass filter.

Since

$$H_L(0) = 1 \qquad (4\text{-}2)$$

for a low-pass filter, then it follows that

$$H_H(0) = 0 \qquad (4\text{-}3)$$

Equation (4-1) shows that output from a high-pass filter is the same as the difference between the unfiltered data, that is, output from an "all-pass" filter and the output from a low-pass filter. Clearly, since a filter whose response is unity over all frequency bands, an all-pass filter passes the data without

change; that is, it does not filter the data at all. Therefore, the effects of a high-pass filter can be duplicated by filtering the raw data with a low-pass filter and then subtracting the results from the raw data as in

$$Y = H_H X = (1-H_L)X = X - H_L X \qquad (4-4)$$

X is the raw data (in the frequency domain). $H_L X$ is the low-pass filtered data. This technique is often used in analysis of high-frequency noise on missile trajectory data; however, the procedure of subtracting the low-pass or smoothed data from the raw data can be cumbersome and time consuming. A generally faster approach is to construct the weights of the high-pass filter and apply them directly to the raw data.

The weights for this high-pass filter can be obtained by taking the difference between the weights for the two filters given by the right hand side of equation (4-1). Remember from paragraph 2.3 that these weights amount to the unit impulse responses of their respective filters. For an all-pass filter, the response to a unit impulse is a unit impulse, because the input data is left unchanged by this filter. Consequently, all the weights of an all-pass filter are unit impulse responses. The weights for the low-pass filter are then subtracted from the weights of the all-pass filter to obtain the weights for the high-pass filter. Letting $W_{L,k}$ represent the weights for a low-pass filter and $W_{H,k}$ represent the weights for the corresponding high-pass filter, then

$$W_{H,k} = \delta_k - W_{L,k} \qquad (4-5)$$

where

$$\delta_k = \begin{cases} 1 \text{ if } k = 0 \\ 0 \text{ if } k \neq 0. \end{cases} \qquad (4-6)$$

4.3 Band Pass Filters

The band-pass filter passes a set of adjacent frequencies while rejecting the frequencies above and below the set of frequencies. A band-pass filter is said to be narrow if the frequency band contains very few frequencies, and wide if it contains many frequencies. If the desired frequency bandwidth is large compared to the center frequency of the band, a wide band-pass filter would be desirable to use for filtering the band. If the bandwidth is small compared to the center frequency, it would be better to filter it with a narrow band-pass filter.

The wide band-pass filter is constructed by combining a high-pass filter with a low-pass filter. The high-pass filter, used for this construction, has a cutoff frequency f_c in the area of the spectrum where the low-pass frequency response is approximately unity. The low-pass filter that is used has a cutoff frequency f_{c_H} where the frequency response of the high-pass filter is approximately unity. There is a resulting band of frequencies where the low-pass and high-pass filters are both unity, so that $f_{c_L} > f_{c_H}$. The frequency response of the resulting filter would have a low end of f_{c_H}, a high end of f_{c_L} and a center frequency of

$$f_o = \frac{f_{c_L} + f_{c_H}}{2} \quad \text{(see figure 4-2).} \qquad (4\text{-}7)$$

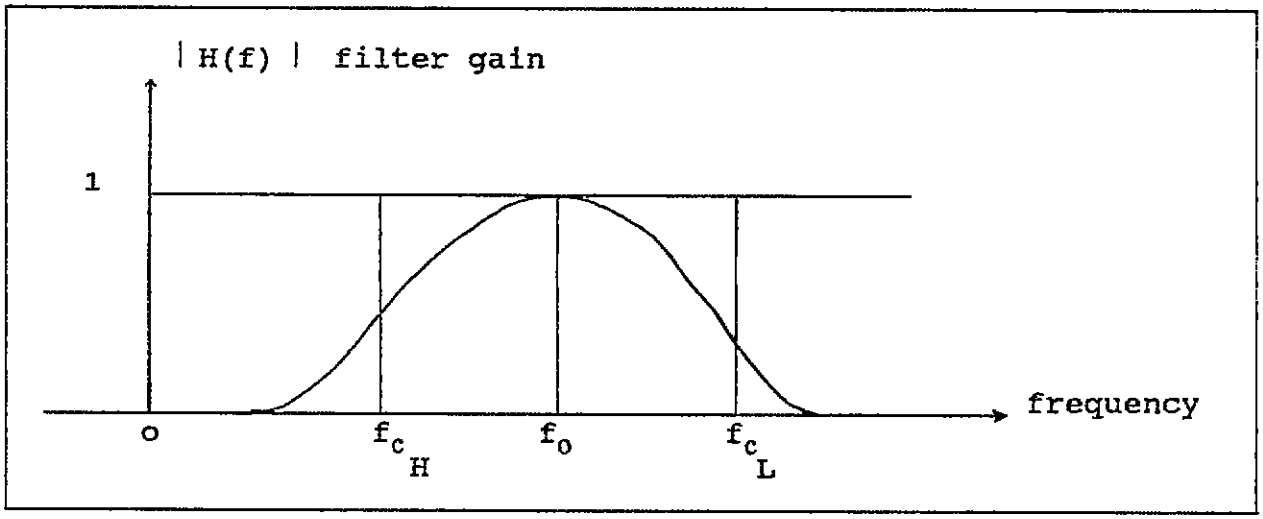


Figure 4-2. Frequency response of wide band pass filter.

To perform the band-pass filtering, the high-pass and low-pass filters can be applied separately to the data in series, which is usually time consuming. It is more desirable to combine the two filters into a single band-pass filter by convolving the weights of the two filters with the input data in the time domain. The same results can be obtained by multiplying the frequency response of each filter with the Fourier transform of the input data in the frequency domain and then taking the inverse Fourier Transform to get the band-pass filtered data in

the time domain. The methods mentioned here are essentially the same as the methods described in paragraph 4.2 for converting a low-pass filter into a high-pass filter.

If the frequency response curves of the high-pass and low-pass filters are sharp and closely approximate the ideal frequency response of a step function, the techniques discussed previously will work for a narrow band-pass filter. If this is not the case and the roll-off is not steep, a narrow band-pass filter can be constructed using the following technique. The frequency response of a low-pass filter is translated so that the center frequency of the low-pass filter, f=0, is moved to the desired center frequency of the passband with its own mirror image reflected about the zero frequency, thus its mirror image is at the corresponding negative frequencies (see figure 4-3).

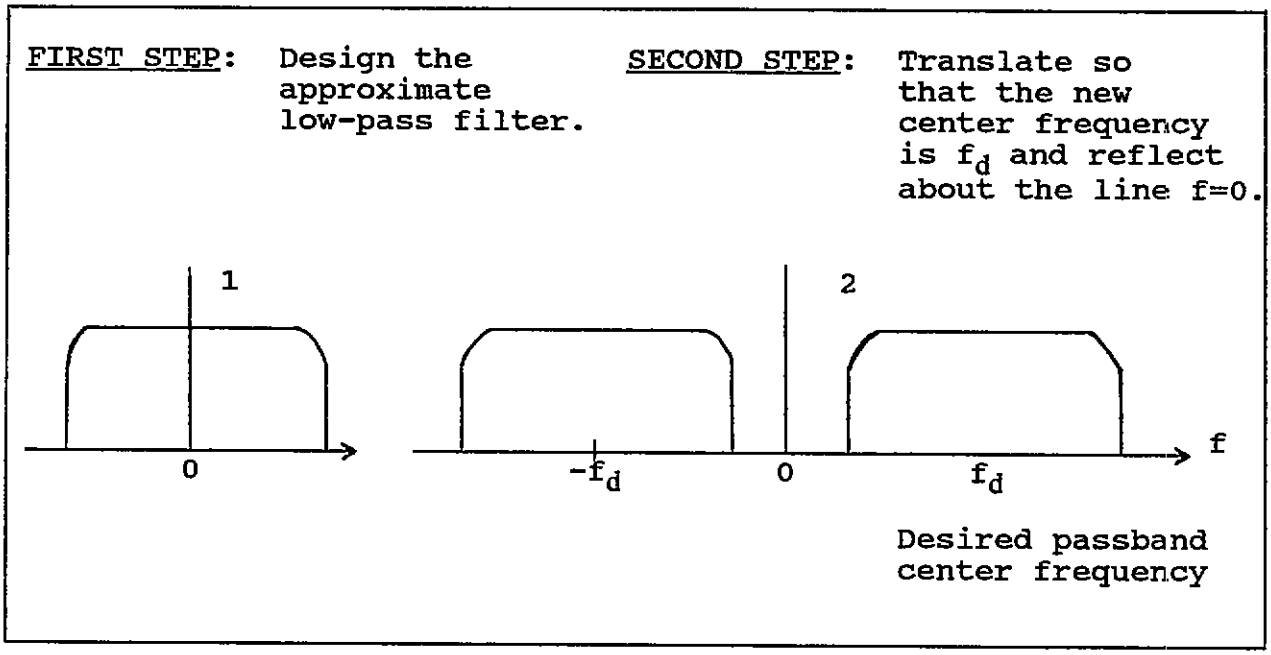


Figure 4-3. Method of converting a low-pass filter with shallow roll-off into a band-pass filter.

4.4 Band-Rejection Filters

To remove a particular band of frequencies from the data and, at the same time, to preserve the lower and higher frequencies, a band-rejection filter can be constructed. The band-rejection filter is the opposite of the band-pass filter. There

are various methods of constructing a band-rejection filter from a band-pass filter. Each method is analogous to an approach discussed in paragraph 4.2 in obtaining a high-pass filter from a low-pass filter.

The frequency response of the band-rejection filter can be obtained by subtracting the band-pass response from unity as shown in figure 4-4. In the time domain, the band-rejection results can be obtained by band-pass filtering and then subtracting the band pass-filter results from the raw data.

Another approach is to construct band-rejection weights and then filter the data directly in a one-step operation. The weights are constructed by

$$W_{R,k} = \delta_k - W_{B,k} \tag{4-8}$$

where

$W_{B,k}$ are the band pass weights (4-9)

and

$$\delta_k = \begin{cases} 1 \text{ if } k = 0 \\ 0 \text{ if } k \neq 0 \end{cases} \tag{4-10}$$

Filtering can be accomplished by convolving the weights of the band-rejection filter with the input data to obtain the filter.

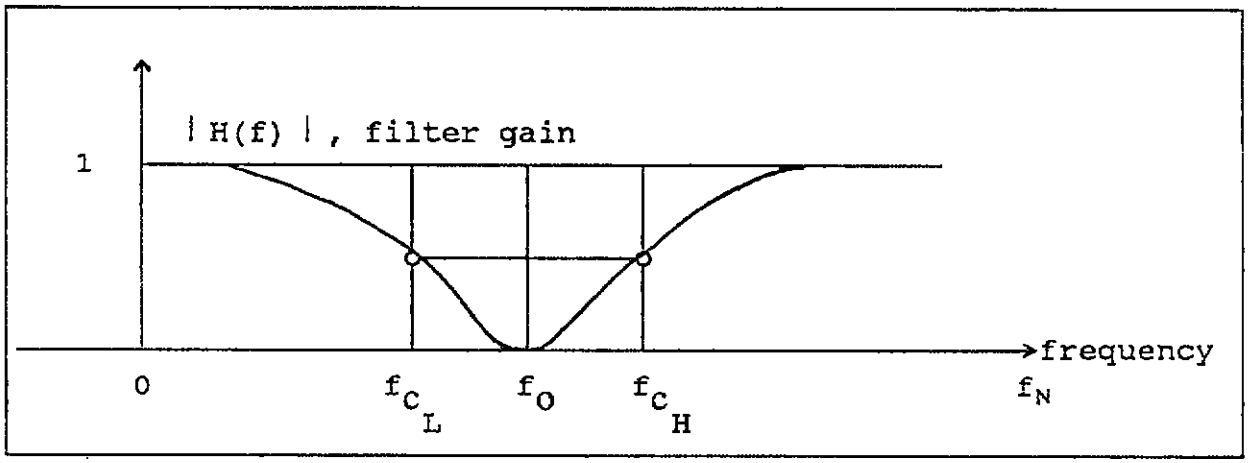


Figure 4-4. Frequency response of band-rejection filter.

4.5 Phase-Shift Filters

All categories of digital filters have a phase-shift curve which characterizes the amount of phase shift the filter induces for each frequency. It is sometimes desirable to construct a filter which will shift the phase of a signal by a constant amount over a narrow band of frequencies while maintaining constant amplitude. For example, a sinusoidal signal with amplitude $\propto$ and phase ϕ can be resolved into two components with amplitude $\propto \cos \phi$ and phase 0° and amplitude $\propto \sin \phi$ and phase 90°. In this case, a filter is constructed which produces a phase shift of 90° and combines the output of the filter with the unshifted data in the appropriate manner to produce any desired resultant phase shift. A phase lead of 90° and no amplitude change corresponds to frequency response equal to $\sqrt{-1}$. An ideal digital-phase shifter maintains this value for all frequencies between zero and the Nyquist frequency.

4.6 Filter Combinations

Any or all of the filters discussed in this chapter can be combined to form filters which perform more complex tasks. In any case, whatever effects are wanted can be accomplished in the time domain by applying the filters independently in series or by getting a one-step complex filter by convolving the filter weights together in the time domain. On the other hand in the frequency domain, the frequency response curves can be multiplied, added, or subtracted to give whatever response is desired, then transformed to the time domain via the appropriate use of the inverse Fourier transformation.

One special combination filter is referred to as a "comb filter." This filter consists of a series of n narrow band-pass filters having a frequency response curve with n narrow frequency bands at equal spaces between zero and the Nyquist frequency, which was discussed in subparagraph 3.7.3. This type of filter is constructed by convolving the raw data with the weights of n narrow band pass filters that can be accomplished in a one-step operation. (A similar one-step operation is discussed in paragraph 4.4, in converting a band-pass filter into a band-rejection filter.) The filter is then usually used to aid in determining significant frequencies and cut-off levels for other low-pass or high-pass filters. The Fast Fourier Transform is usually considered a better tool for this analysis than the comb filter.

CHAPTER 5

FREQUENCY DOMAIN FILTER DESIGN

There have been many filters developed for digital signal processing in the frequency domain and various ways to express these filters. In this chapter, a synopsis of the filters is given along with their various modes of expression. For recursive filters, methods of conversion from the analog form to the digital form are given as well. Also discussed are the advantages and disadvantages of using each type of filter, mode of expression, and method of conversion. Since the approach to designing infinite impulse response (IIR) (recursive) and finite impulse response (FIR) (nonrecursive) filters, described earlier in this document, are completely different, the chapter is divided into two parts: IIR and FIR filter design.

5.1 Infinite Impulse Response (IIR) Filter Design

The IIR filters have been extensively analyzed in the analog form. Consequently, digital IIR filters are usually described first in analog form and then converted to digital form.

The problem of approximating an IIR filter involves a complex transfer function of the frequency ω. The transfer function can be written as

$$H(z) = |H(z)| e^{iB(z)}, \tag{5-1}$$

where $z=e^{i\omega}$, and $B(z)=B(e^{i\omega})$ is the phase angle, which can be defined as

$$B(e^{i\omega}) = \arctan \frac{Im(H(e^{i\omega}))}{Re(H(e^{i\omega}))}. \tag{5-2}$$

Thus both the magnitude and the phase angle (also called the phase response) are examined to determine the nature of the filter approximations.

Advantages to using an IIR filter are that it (1) can be expressed in closed form and thus can generally be computed more efficiently, (2) does not require powerful computational facilities to be calculated, and (3) achieves a superior amplitude response. An undesirable consequence of using an IIR filter is that it yields a nonlinear phase response. More on the linear phase characteristic is covered in paragraph 5.2.

5.1.1 Modes of IIR Filter Expression

Recall from chapter 3 that an IIR filter can be expressed as

$$y(k) = \sum_{i=0}^{N} a(i)\, x(k-i) + \sum_{i=1}^{M} b(i)\, y(k-i) \qquad (5\text{-}3)$$

Three popular modes of expression for the transfer function of an IIR filter are

(1) the direct form,

(2) the cascade form, and

(3) the parallel form.

The forms assume those names because when they are illustrated in network diagrams, they are depicted in the forms mentioned. In the following subparagraphs, the mathematical forms and illustrations are given as well as the advantages and disadvantages.

5.1.1.1 Direct Form

As mentioned in paragraph 3.6, the z-transform of a weight function, that is, the transfer function of an IIR filter can be written as

$$H(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{1 - \sum_{k=0}^{N} a_k z^{-k}}. \qquad (5\text{-}4)$$

The network design illustration for this equation is known as the direct form realization of the IIR filter, which is expressed by equation (5-3). This illustration is given in figure 5-1 and is obtained from reference 56.

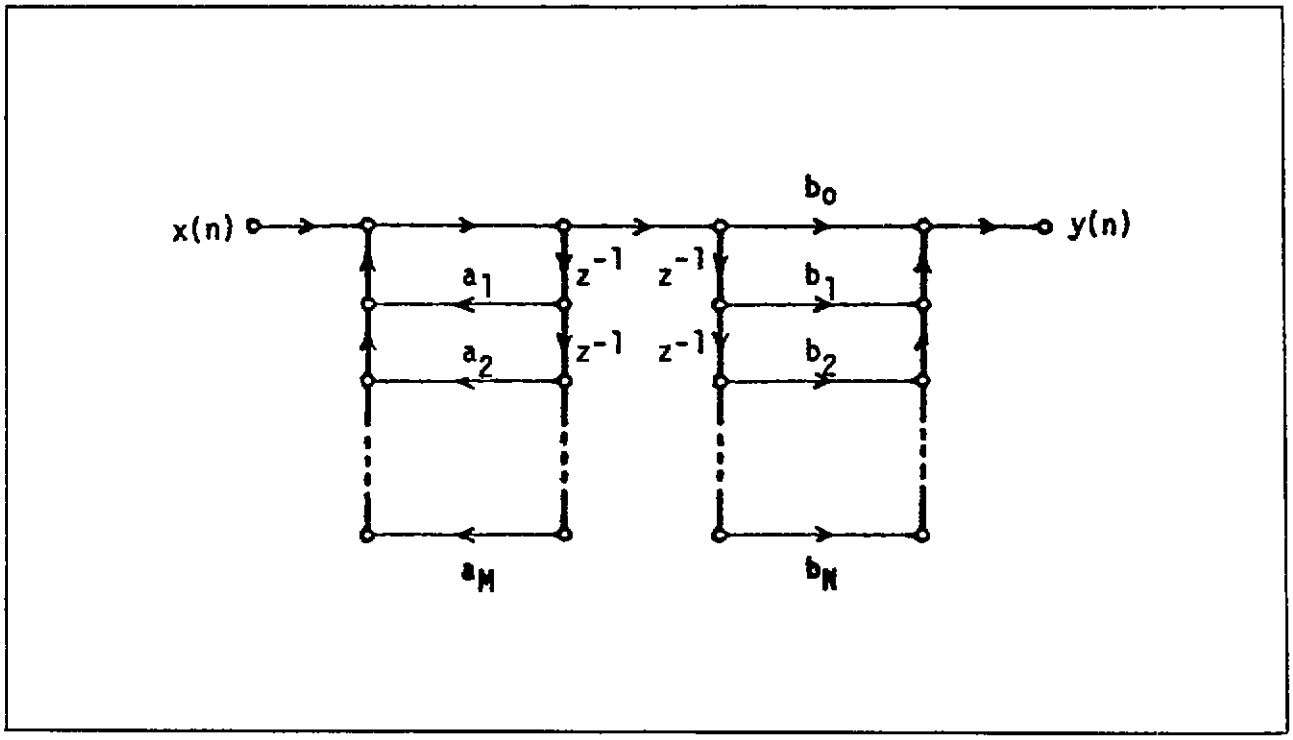


Figure 5-1. Direct form realization of IIR filter.

The advantages of this form are that the direct relation between the IIR filter equation (5-3) and the transfer function equation (5-4) is easily seen, and it is easy to formulate equation (5-4) given equation (5-3). The disadvantage of this form is that coefficients a_k and b_k are very sensitive to discretization errors. These errors will affect the accuracy and possibly the stability of the calculations. This form is generally not recommended for filters of order three or higher.

5.1.1.2 Cascade Form

Equation (5-4) can be rewritten as

$$H(z) = A \frac{\prod_{k=1}^{M_1} (1 - g_k z^{-1}) \prod_{k=1}^{M_2} (1 - h_k z^{-1})(1 - h_k^* z^{-1})}{\prod_{k=1}^{N_1} (1 - c_k z^{-1}) \prod_{k=1}^{N_2} (1 - d_k z^{-1})(1 - d_k^* z^{-1})} \qquad (5\text{-}5)$$

where $M_1 + M_2 = M$ and $N_1 + N_2 = N$. The numbers g_k and c_k are the real zeroes and poles of the transfer function. The numbers h_k, h_k^*, d_k, and d_k^* are the complex conjugate zeroes and poles of the transfer function. (A is a real number.)

This equation can be written as

$$H(z) = A \sum_{k=1}^{[(N+1)/2]} \frac{1 + {}_{1k}z^{-1} + {}_{2k}z^{-2}}{1 + {}_{\alpha 1k}z^{-1} + {}_{\alpha 2k}z^{-2}} \qquad (5\text{-}6)$$

(where [(N+1)/2] is the greatest integer less than or equal to (N+1)/2). Assume that $M \leq N$. If $N \leq M$, replace N in equation (5-6) with M.

The network design illustration for this equation is known as the "cascade form realization" of the IIR filter, which is expressed by equation (5-3). This illustration is given in figure 5-2 and is obtained from reference 56.

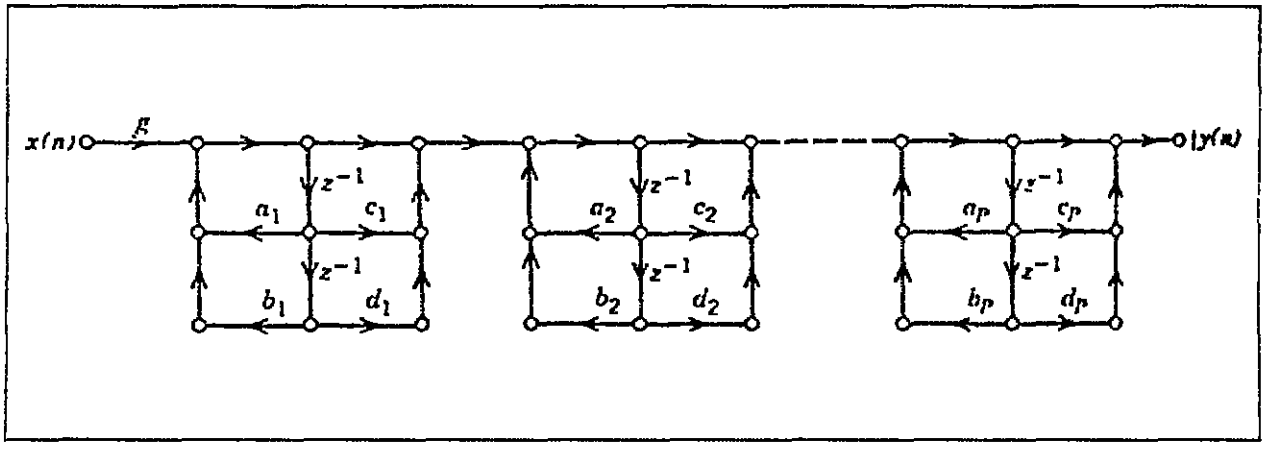


Figure 5-2. Cascade form realization of the IIR filter.

The advantages of this form are

(1) the zeroes (roots) of H(z) are easy to find,

(2) it is convenient to use for filters whose parameters are to be computed and changed in real time, and

(3) the poles (roots of the denominator) are easy to find.

The poles are important because they determine the stability of a filter. The filter is stable if and only if the poles are inside the unit circle. If a pole is on the unit circle, other criteria have to be used to determine stability.

A disadvantage of this form is that it is heavily subject to underflow and overflow problems.

5.1.1.3 Parallel Form

Equation (5-4) can be rewritten in terms of a partial fraction expansion as

$$H(z) = \sum_{k=1}^{N_1} \frac{A_k}{1 - c_k z^{-1}} + \sum_{k=1}^{N_2} \frac{B_k(1 - e_k z^{-1})}{(1 - d_k z^{-1})(1 - d_k^* z^{-1})} + \sum_{k=1}^{M-N} C_k z^{-1} \qquad (5\text{-}7)$$

The first two terms in equation (5-7) can be combined so that the equation becomes

$$H(z) = \sum_{k=1}^{M-N} C_k z^{-1} + \sum_{k=1}^{[(N+1)/2]} \frac{T_{0k} + T_{1k} z^{-1}}{1 - \alpha_{1k} z^{-1} - \alpha_{2k} z^{-2}} \qquad (5\text{-}8)$$

The network design illustration for this equation is known as the "parallel form realization" of the IIR filter, which is expressed by equation (5-3). This illustration is given in figure 5-3 and is obtained from reference 55.

An advantage of this form is that the problems mentioned for the direct and cascade forms are normally not encountered. A disadvantage is that while pole locations can be easily located (as in the cascade form), the zeroes cannot.

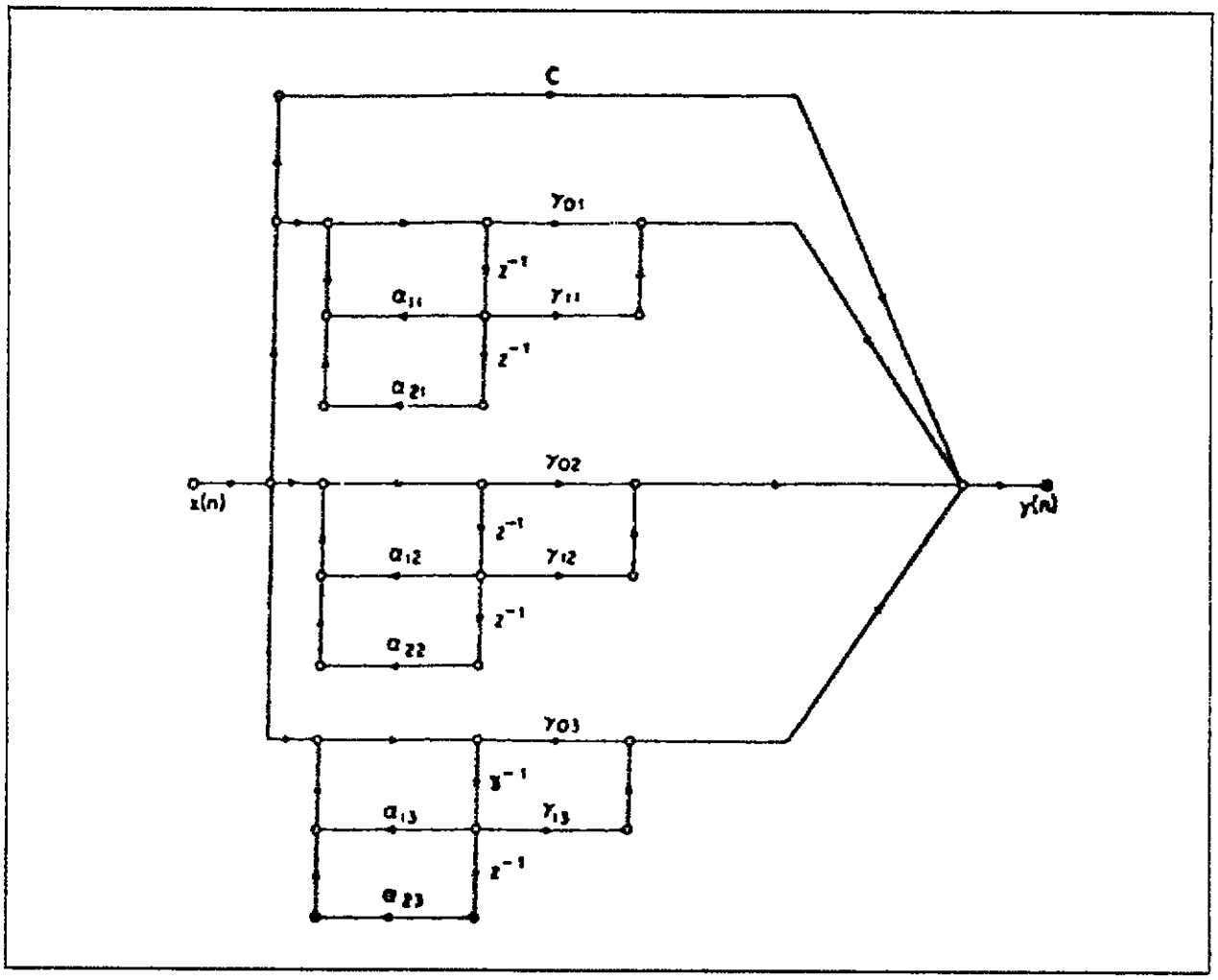


Figure 5-3. Parallel-form realization with the real and complex poles grouped in pairs.

5.1.2 Converting From Analog to Digital

The filters to be discussed in subparagraph 5.1.3 were originally derived in the analog form. Since these filters were discovered, a need developed to use frequency domain filters in the digital form. Because the art of analog filter design is highly advanced, it is considered advantageous to adopt the design procedures developed for the analog form to convert to the digital form. In this section four methods of analog-to-digital conversion are given along with their advantages and disadvantages.

The two desirable properties of any conversion method are

(1) The frequency axis in the "s plane" (the analog plane) be mapped to the unit circle in the "z plane" (the digital plane), ensuring that there is a one-to-one mapping between frequencies in the s plane and frequencies in the z plane. If this property is satisfied, the "frequency selective properties" of the analog filter are said to be preserved. With a conversion method that has this characteristic, the digital filter should filter through all and only those frequencies that the analog filter filters through.

(2) The left half of the s plane {Re[s]<0} be inside the unit circle in the z plane (|z|<1), which ensures the preservation of filter stability.

All the recommended conversion methods described in these subparagraphs have at least the second property.

5.1.2.1 Impulse Invariance

The main idea of the impulse invariance method is to preserve the impulse response when converting from analog to digital. That is, the purpose is to ensure

$$h(n) = h_a(nT) \tag{5-9}$$

where T is the sampling period, h is the digital impulse response, and h_a is the analog impulse response. In other words, the characteristic property of this transformation is that the impulse response of the resulting digital filter is a sampled version of the impulse response of the analog filter. Use the following method:

(1) set up analog transfer function H(s) in direct, cascade, or parallel form. (Generally, the parallel form is the most preferable, because it is easy to perform step (2) when using a table for transform pairs);

(2) obtain the inverse Laplace (or Fourier) transform of H(s), giving the analog impulse response function h(t);

(3) get your digital impulse response function h(n) using equation (5-9); and

(4) take the z-transform of h(n) to obtain your digital transfer function H(z).

Steps (2) and (4) are jointly justified, because it can be shown that the Laplace transform of the analog function h_a is related to the z-transform of the digital function h(t) by

$$H(z)\Big|_{z=e^{sT}} = (1/T)\sum_{k=-\infty}^{\infty} H_a(s + i(2\pi/T)k). \qquad (5\text{-}10)$$

With the z-transform, the relationship $z=e^{sT}$ is used in going from the s plane (for the analog function) to the z plane (for the digital function). Each horizontal strip of the left half of the s plane of width $2\pi/T$ is mapped into the unit circle in the z plane (see figure 5-4, obtained from reference 55). This multiple mapping may lead to aliasing (explained in chapter 3) because distinct frequencies from different strips in the s plane can be translated into one frequency in the z plane.

The occurrence of aliasing can be seen by noting that if $s = \sigma + i\pi$, then

$$z = e^{sT} = e^{\sigma T}(\cos\pi T + i*\sin\pi T)$$

$$= e^{\sigma T}(\cos(\omega + (2\pi n/T))T + i*\sin(\omega + (2\pi n/T))T) \qquad (5\text{-}11)$$

for any integer n.

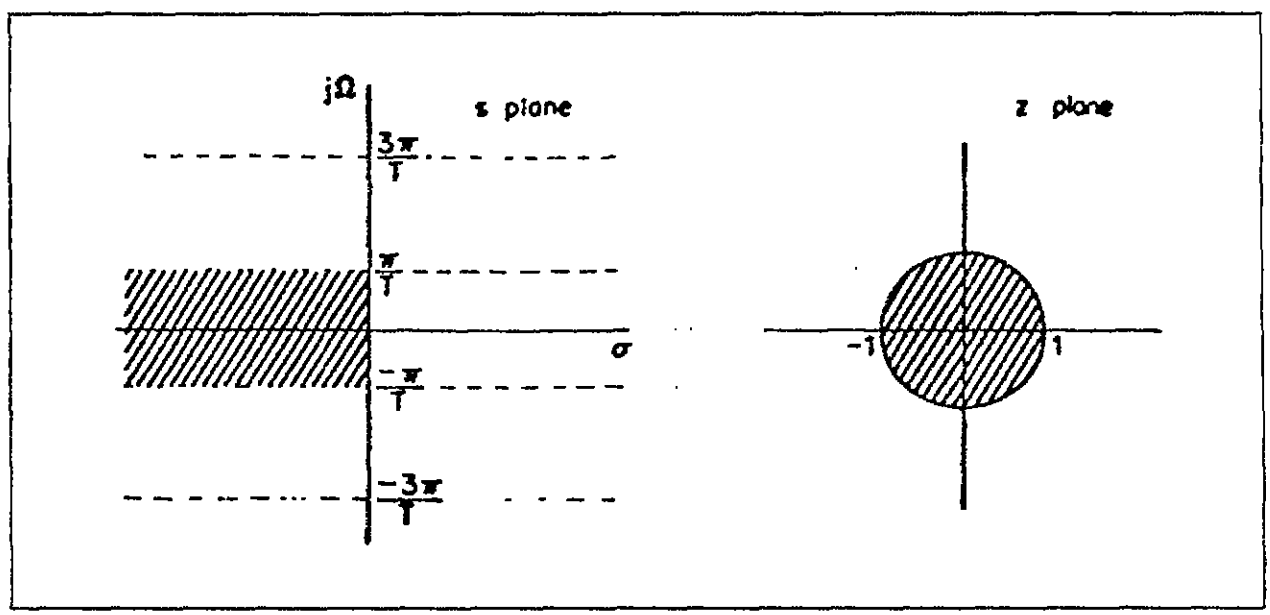


Figure 5-4. The mapping of a horizontal strip of width $2\pi/T$ in the s plane to the unit circle of the z plane, through the impulse invariance method.

A disadvantage in using this method is that it is subject to aliasing. This method works best when used with bandlimited filters such as low pass and band pass.

5.1.2.2 Bilinear Transformation

With the bilinear transformation, to go from the s plane (which represents the analog filter) to the z plane (which represents the digital filter), the following equation is used:

$$z = \frac{1 + (T/2)s}{1 - (T/2)s} \tag{5-12}$$

Likewise, s can be expressed in terms of z by

$$s = \frac{2}{T}\,\frac{1 - z^{-1}}{1 + z^{-1}} \tag{5-13}$$

That is, the transfer function of the digital filter, H(z), is set equal to the transfer function of the analog filter, H(s), where s is expressed in terms of z as in equation (5-13), or where z is expressed in terms of s as in equation (5-12). These equations were obtained by the procedure outlined next.

A digital filter is essentially a difference equation. For an analog filter, a differential equation is used. The analog transfer function is obtained by using Fourier or Laplace transforms. To obtain the bilinear transformation, integrate both sides of the differential equation and use a numerical approximation to the integral so that the integrated equation is expressed in discrete terms. Then take the z-transform of this equation to get the digital transfer function. When comparing the analog and digital transfer functions, it can be seen that s and z are related as stated in equations (5-12) and (5-13). Reference 55 gives the derivation of these equations in more detail.

Illustrated in figure 5-5, which was obtained from reference 55, is the relationship between s and z.

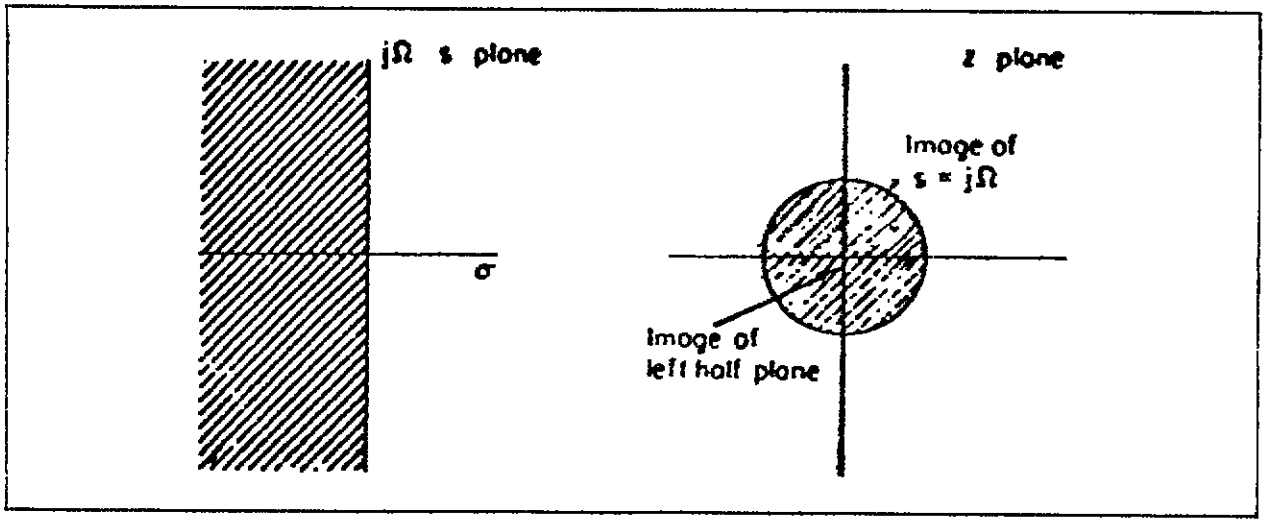


Figure 5-5. The mapping of the left half of the s plane to the unit circle of the z plane, using the bilinear transformation.

Reference 55 shows that the analog frequency Ω can be expressed in terms of the digital frequency ω as follows:

$$\Omega = (2/T)\tan(\omega/2) \qquad (5\text{-}14)$$

The graphic relationship between Ω and ω is given in figure 5-6, which was obtained from reference 54.

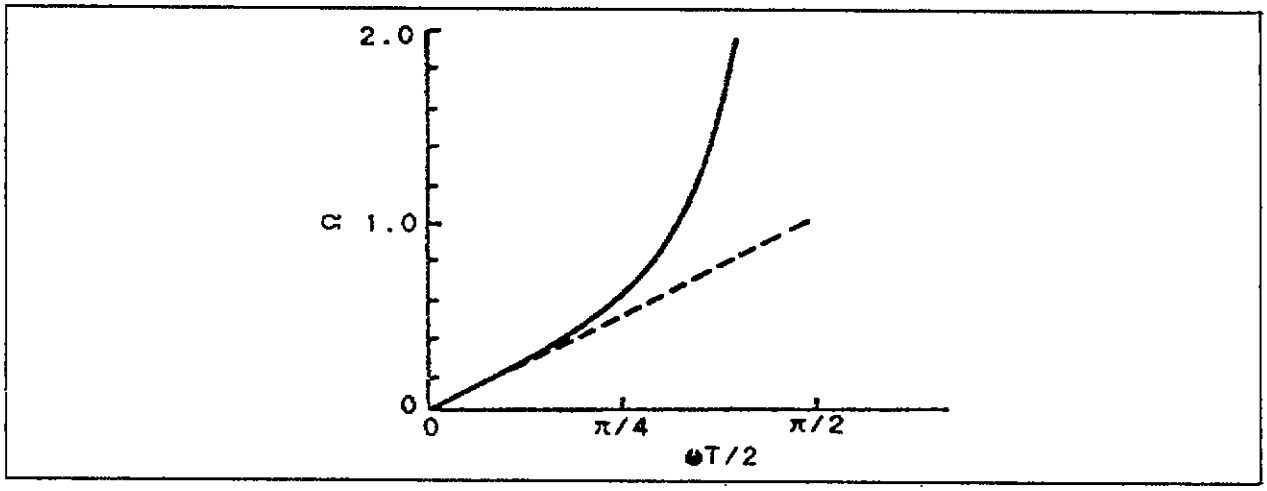


Figure 5-6. The relation between analog and digital frequency scales for the bilinear transformation.

As seen from the last two figures, the bilinear transformation maps an infinite range of frequencies onto the unit circle, representing a finite range of frequencies. This mapping, consequently, distorts the frequency scale.

The frequencies of interest in the s plane must be predistorted to make sure they will come out in the right places in the z plane. The predistortion formula is

$$\Omega = \frac{2}{T*\tan(\omega T/2)}. \qquad (5\text{-}15)$$

To sum up, this method will probably produce warping in the frequency axes. Frequencies in the s plane are mapped into frequencies in the z plane in a distorted manner. To make up for this distortion, the frequencies can be prewarped. As given in equation (5-15), this method satisfies property (1) of subparagraph 5.1.2. This method works best when used with nonband-limited filters such as band stop or high pass.

5.1.2.3 Direct Mapping of Differentials

There are three ways to employ the direct mapping of differentials: backward difference, forward difference, and generalized difference. Consider the following differential equation representing the analog filter:

$$\sum_{i=0}^{N} a_i \frac{d^i y(t)}{dt^i} = \sum_{i=0}^{M} b_i \frac{d^i x(t)}{dt^i} \qquad (5\text{-}16)$$

Equation (5-16) is then discretized by

$$\sum_{i=0}^{N} a_i \Delta_i[y(n)] = \sum_{i=0}^{M} b_i \Delta_i[x(n)] \qquad (5\text{-}17)$$

where $\Delta_i[y(n)]$ is the ith difference defined by the recursion

$$\Delta_{i+1}[y(n)] = \Delta_1\{\Delta_i[y(n)]\} \qquad (5\text{-}18)$$

and by the initialization

$$\Delta_i[y(n)] = \begin{cases} (1/T)[y(n) - y(n-1)] & \text{backward difference} \\ (1/T)[y(n+1) - y(n)] & \text{forward difference} \\ \\ \frac{1}{T}\sum_{i=1}^{L} \alpha_i[y(n+i) - y(n-i)] & \text{generalized difference} \end{cases} \tag{5-19}$$

and $\Delta_i[x(n)]$ is defined in the same way. The factor α_i for the generalized difference is a constant of the user's choice.

Backward Differences: When using backward differences, make the replacement

$$\frac{dy}{dt} \longleftrightarrow \frac{y(n) - y(n-1)}{T} \tag{5-20}$$

which, in terms of the relationship between s and z, corresponds to

$$s = \frac{1 - z^{-1}}{T} \tag{5-21}$$

and

$$z = \frac{1}{1 - sT} \tag{5-22}$$

The relationship between s and z is illustrated in figure 5-7, (obtained from reference 54). As can be seen property (1) mentioned in subparagraph 5.1.2 (one-to-one frequency correspondence) is not satisfied and that property (2) (stability preservation) is satisfied.

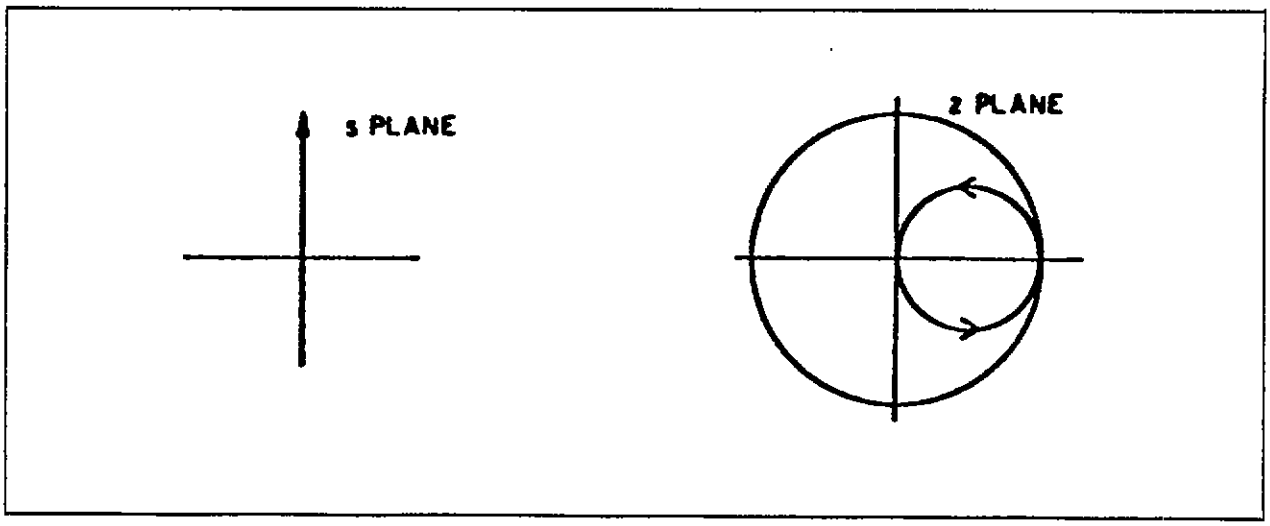


Figure 5-7. s plane to z plane mapping of jΩ axis for method of backward differences.

Reference 55 describes how property (1) becomes closer to being satisfied the higher the sampling rate. The high sampling rate required to make this technique adequate for analog-to-digital conversion is said to result in a very inefficient representation of the filter and the input signal. This technique is considered usable for low-pass filters only. An example for which this technique can be used is air flight control, where frequencies are normally at 100 Hz or less. An advantage of this method is the simplicity of the design.

Forward Differences: When using backward differences, we make the replacement

$$\frac{dy}{dt} \longleftrightarrow \frac{y(n+1) - y(n)}{T} \tag{5-23}$$

which, in terms of the relationship between s and z, corresponds to

$$s = \frac{z - 1}{T} \tag{5-24}$$

and

$$z = 1 + sT \tag{5-25}$$

The mapping of the frequency axis from the s plane to the z plane is illustrated in figure 5-8, as obtained from reference 54. As can be seen property (1) mentioned in subparagraph 5.1.2 (one-to-one frequency correspondence) is not satisfied. Also shown from either equation (5-16) or equation (5-17) property (2) (stability preservation) is not satisfied either. This method is not recommended.

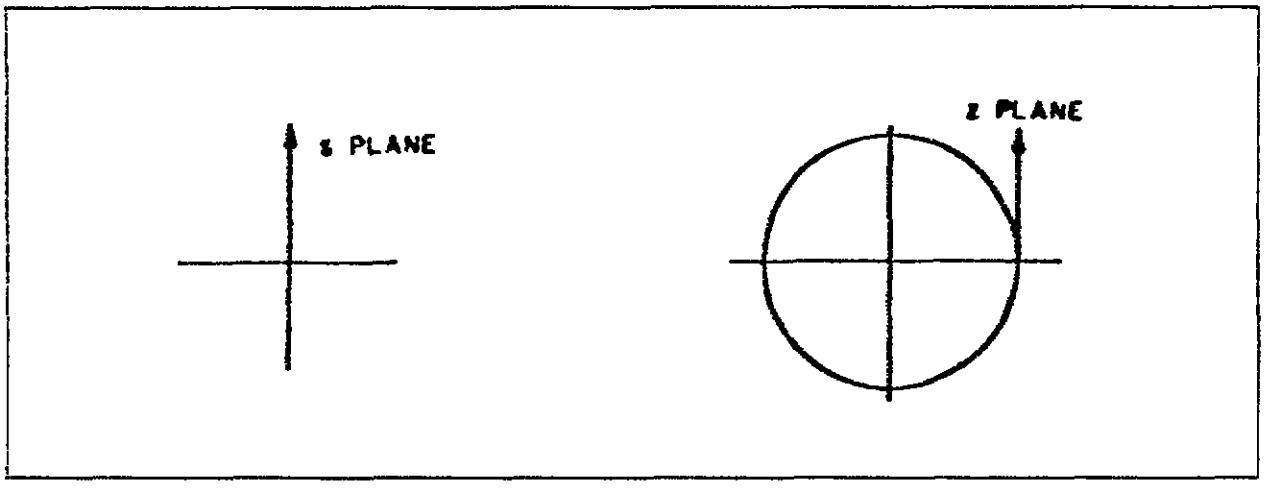


Figure 5-8. s plane to z plane mapping of jΩ axis for method of forward differences.

Generalized Differences: As can be seen from equation (5-19), this method uses higher-order differences to replace lower-order differentials. The mapping between the s plane and the z plane for this method is

$$s = \frac{1}{T} \sum_{i=1}^{L} \alpha i (z^{i} - z^{-i}) \qquad (5\text{-}26)$$

where L is the order of difference to be used. Reference 54 shows that with the proper choice of coefficients α_i, the frequency axis in the s plane is mapped monotonically to the unit circle in the z plane, thereby satisfying property (1) in subparagraph 5.1.2. The mapping of equation (5-26) can be shown to be conformal. A conformal mapping preserves angles and relative locations of points from the domain to the range. (From this conformality, the left half of the s plane is mapped to the inside of the unit circle of the z plane, thereby, showing that property (2) is satisfied.)

This method is considered efficient and accurate, if the appropriate coefficients α_i can be found. It is normally difficult to determine these coefficients. As a result, other techniques for digitizing filters are sought.

5.1.2.4 Matched Z-Transformation

The matched z-transformation is motivated by the following facts, stated without proof:

(1) to maintain stability, all the poles of H(z), the transfer function in the z plane, must lie within the unit circle, and

(2) all poles and zeroes must either be real or occur in complex conjugate pairs.

As such, this method matches poles and zeroes in the s plane to poles and zeroes in the z plane. For a real pole or zero, say -a, the transformation is

$$s + a \longrightarrow 1 - z^{-1}e^{-aT} \qquad (5\text{-}27)$$

where T is the sampling period.

For complex conjugate poles or zeroes, say $a \pm bi$, the transformation is

$$(s+a-bi)(s+a+bi) =$$

$$(s+a)^2 + b^2 \longrightarrow 1 - 2z^{-1}e^{-aT}\cos(bT) + z^{-2}e^{-2aT} \qquad (5\text{-}28)$$

(Since b=0 in the real case, it is easily seen that equation (5-28) can be simplified to equation (5-27).) The continuous transfer function H(s) must be in factored form to apply the transformation. The advantages of this method are that it ensures stability (it was designed to) and that it is very easy to implement. One disadvantage is that it may lead to aliasing. For instance,

$(s+a)^2 + b^2$ yields the same transformation as $(s+a)^2 + (b+2\pi)^2$. (5-29)

This transformation is considered unsuitable where H(s) is an all-pole system; that is, H(s) has only poles and no zeroes. Quite often for this case, H(z) is an all-pole system that does not adequately represent the desired continuous system. Generally, the bilinear transformation and impulse invariance methods are said to be preferred over the matched z-transformation.

5.1.3 Low-Pass IIR Filters

Discussed next are three classes of low-pass IIR filters: the Butterworth, the Chebyshev, and the elliptic (or "Cauer") filters. Each of these filters takes the form

$$H(\omega)^2 = \frac{1}{1 + \epsilon^2 f_N(\omega)} \tag{5-30}$$

where H is the frequency response, ω is the frequency variable, N is the order of the filter, ϵ is a factor in the interval [0,1] that determines the height of the passband ripples, and f is an nth-order polynomial containing only odd or only even powers of ω. Bandpass, high-pass, and band-rejection filters can then be designed from any of the aforementioned low-pass filters using the methods described in chapter 4.

5.1.3.1 Butterworth Filters

Butterworth filters take the form

$$|H(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2N}} \tag{5-31}$$

where ω_c is the cutoff frequency of the filter.

The Butterworth filter contains the following characteristics:

(1) It is defined by the property that the magnitude response $|H(\omega)|$ is maximally flat in the passband, meaning that the maximum number of derivatives, 2N - 1, of the squared magnitude function $|H(\omega)|^2$ are equal to zero at ω=0.

(2) $|H(\omega)| = 1/\sqrt{2}$ at the cutoff frequency ω_c since, when attempting to simulate the ideal low-pass filter, it is desired that $\lim_{n\to\infty} |H(\omega)| = 1$ for all ω in ω

$[-\omega_c, \omega_c]$. ($\lim_{N\to\infty} |H(\omega)| = 1$ for all $|\omega| < \omega_c$.)

(3) It is computationally and conceptually simpler than the Chebyshev and elliptic filters.

(4) The higher N is, the better this filter simulates the ideal low-pass filter.

(5) The magnitude of the frequency response is monotonic in both the passband and the stop band.

Illustrations of the Butterworth filter, obtained from reference 55, can be seen in figure 5-9.

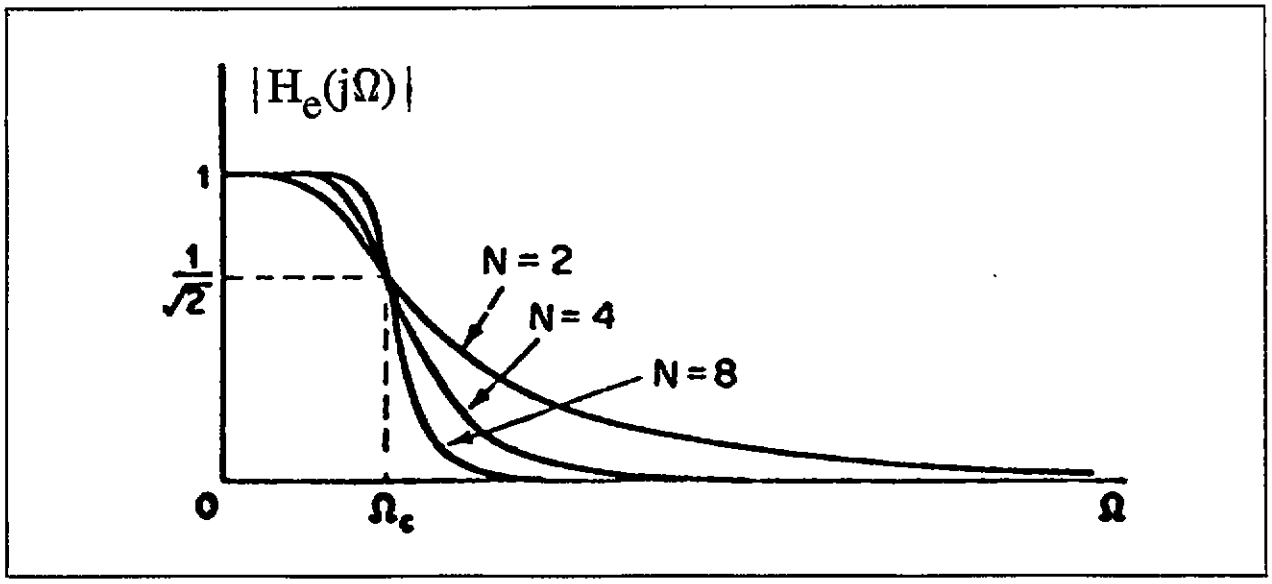


Figure 5-9. Dependence of Butterworth magnitude characteristic on the order N.

5.1.3.2 Chebyshev Filters

Chebyshev filters take the form

$$|H(\omega)|^2 = \frac{1}{1 + \epsilon^2 V_N^2(\omega/\omega_c)} \qquad (5\text{-}32)$$

where $V_N(x)$ is the nth order Chebyshev polynomial defined by

$$V_N(x) = \cos(N\cos^{-1}x). \qquad (5\text{-}33)$$

(For example, for N=2, $V_N(x)=V_2(x)=2x^2-1$.) (5-34)

The Chebyshev filter contains the following characteristics:

(1) It distributes uniformly the inaccuracy in simulating the ideal LP filter in either the passband or the stop band (but not both). That is, it produces an equiripple curve either in the passband (about the line $|H(\omega)|^2=1$) or in the stop band (about the line $|H(\omega)|^2=0$).

(2) It usually leads to a lower-order polynomial than does the Butterworth filter to accomplish the same result.

There are two types of Chebyshev filters. Chebyshev Type I filters are equiripple in the passband and monotonic in the stop band. Chebyshev Type II filters (sometimes called inverse Chebyshev filters) are monotonic in the passband and equiripple in the stop band. Figure 5-10, obtained from reference 54, illustrates Type I and Type II Chebyshev filters of odd and even orders. (The A near the vertical axes is what reference 54 uses at the parameter related to stop-band loss.)

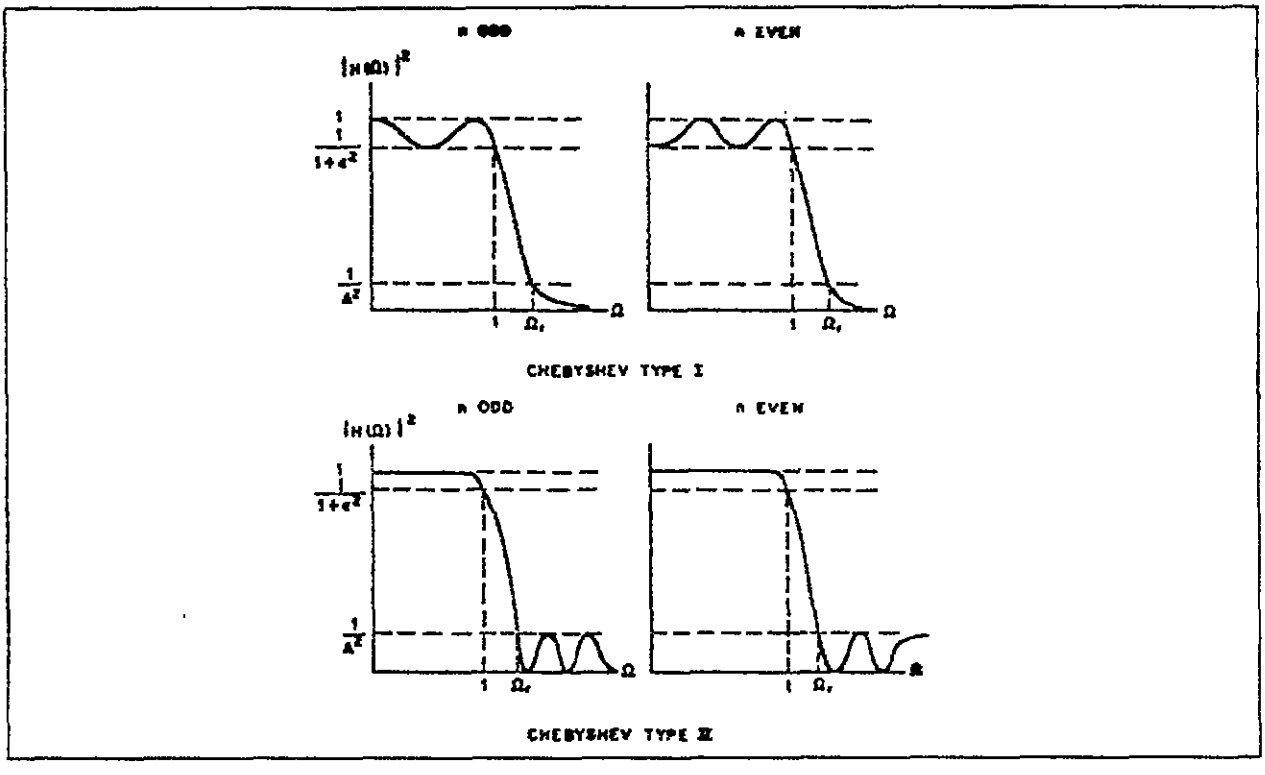


Figure 5-10. Type I and II Chebyshev filters of odd and even orders.

5.1.3.3 Elliptic Filters

Elliptic (or Cauer) filters take the form

$$|H(\omega)|^2 = \frac{1}{1 + \epsilon^2 U_N^2(\omega, L)} \qquad (5\text{-}35)$$

where $U_N(\omega,L)$ is a Jacobian elliptic function. Discussion of this function is highly intricate and is beyond the realm of this paper. Those interested in studying this function are referred to reference 58. Those interested in further studying the design of elliptic filters are referred to references 59, 60, and 61.

The elliptic filter contains the following characteristics:

(1) It distributes uniformly the inaccuracy in simulating the ideal LP filter in both the passband and the stop band. In other words, it produces an equiripple curve in the passband (about the line $|H(\omega)|^2=1$) and in the stop band (about the line $|H(\omega)|^2=0$).

(2) It yields a smaller transition band than does the Chebyshev. (Note that the Butterworth filter does not yield a transition band.)

Figure 5-11, obtained from reference 55, illustrates the elliptic filter.

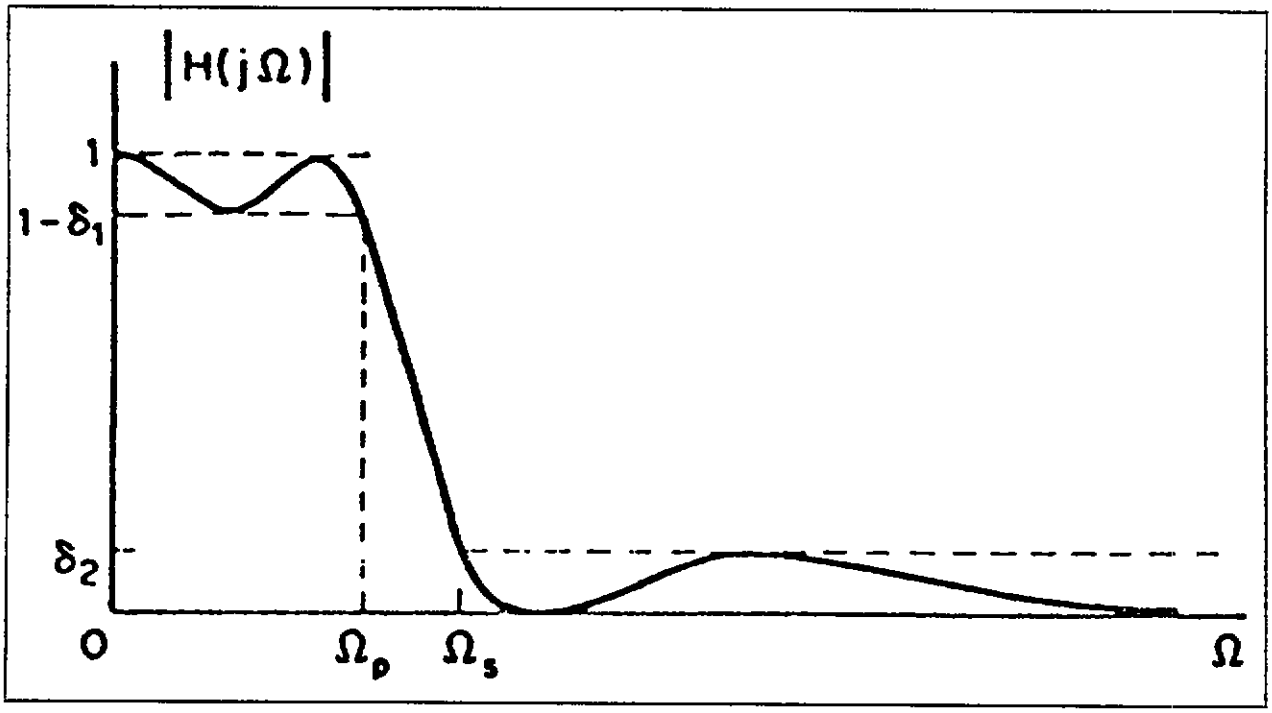


Figure 5-11. Equiripple approximation in both passband and stop band.

5.2 Finite Impulse Response (FIR) Filter Design

The FIR filters were briefly described in paragraphs 1.7 and 1.8. In this section, modes of expression for FIR filters and FIR filter design techniques are given. A disadvantage in implementing an FIR filter is that it cannot be expressed in closed form. Instead, it must be calculated through iterative procedures, and thus generally requires a great deal of computational time and the use of powerful computational facilities.

In many signal-processing applications, phase relations are important and must not be disturbed by filtering. For such purposes, a zero phase shift would be ideal; in practice, however, a filter whose phase shift is proportional to frequency is generally used. Such a filter is called linear phase. An advantage in using FIR filters is that they can have exact linear phase. Referring to figure 3-6, it has been found that the most favorable conditions for an FIR design are large values of δ_1, small values of δ_2, and large transition widths.

5.2.1 Modes of FIR Filter Expression

An FIR filter can be expressed as

$$y(k) = \sum_{i=0}^{N} a(i)\, x(k-i) \qquad (5\text{-}36)$$

There are four popular modes of expression for the transfer function of an FIR filter. They are

(1) the direct form,

(2) the cascade form,

(3) the frequency sampling form, and

(4) the linear-phase form.

The first two forms assume those names because when they are illustrated in network diagrams, they are depicted in the forms mentioned. The direct and cascade FIR forms may be derived from the corresponding IIR forms by simply omitting the pole-producing portions of the IIR forms. (The second term on the right-hand side of equation (5-3) generates the pole-producing terms in the IIR modes of expression.) The advantages and disadvantages of the direct and cascade forms are essentially the same for FIR filters as they are for IIR filters. As a result, it will be sufficient just to give their forms in the following subparagraphs.

The frequency sampling form for FIR filters is a type of parallel form of IIR filters but is derived in an entirely different way from the parallel form. The fact that FIR filters can have exact linear phase is used for the linear-phase form. When given its form, it will be easy to see how it can produce considerable savings in computations.

5.2.1.1 Direct Form

The direct form of the transfer function for FIR filters can be found to be

$$H(z) = \sum_{k=0}^{N} a_k z^{-k} \tag{5-37}$$

5.2.1.2 Cascade Form

The cascade form for the transfer function of FIR filters is

$$H(z) = \frac{Y(z)}{X(z)} = a_0 \prod_{i=1}^{K} H_i(z) \tag{5-38}$$

where $H_i(z)$ is either a second-order cascade section, that is,

$$H_i(z) = 1 + a_{1i} z^{-1} + a_{2i} z^{-2} \tag{5-39}$$

or a first-order cascade section, that is,

$$H_i(z) = 1 + a_{1i} z^{-1} \tag{5-40}$$

and K is the integer part of (N+1)/2, where N is the order of the filter.

It should mention here that if linear-phase filters are realized in this form, sensitivity to discretization errors in the coefficients will be less, but the errors may destroy the phase linearity.

5.2.1.3 Frequency Sampling Form

As mentioned in subparagraph 5.2.1, this form is a type of parallel form but is derived differently from the parallel IIR form. The parallel IIR form was derived from a partial-fraction expansion of the transfer function. Since partial-fraction expansions are based on poles, this approach cannot be considered for FIR filters. This form is derived by a design technique called the frequency sampling design technique and is explained in subparagraph 5.2.2.2.

5.2.1.4 Linear-Phase Form

Many applications require filters whose phase response is linear with frequency. It can be shown that linear-phase FIR filters have a symmetrical impulse response, that is, for an N-order filter

$$h(n) = h(N-1-n) \tag{5-41}$$

This symmetry requirement leads to certain economies in implementation. In an FIR filter, the h's correspond one-to-one with the a's of equation (5-37). Because of the symmetry in h(n), equation (5-37) becomes

$$y(n) = a_0x(n) + a_1x(n-1) + \ldots + a_1x(n-N-1) + a_0x(n-N) \tag{5-42}$$

which can be written for N even, as

$$y(n) = \sum_{i=0}^{N/2 - 1} b_n[x(n-1)+x(n-N+i)] \tag{5-43}$$

or for N odd, as

$$y(n) = b_{N/2}x(n - N/2) + \sum_{1=0}^{(N-1)/2 - 1} b_n[x(n-1)+x(n-N+i)] \tag{5-44}$$

Since the number of terms in the sum is reduced by approximately one half, considerable economies in computation are achieved. Because of the symmetry in this form, errors in quantizing coefficients, that is, discretization errors of the coefficients, will not disturb the linear-phase characteristic although performance but may still degrade performance.

5.2.2 FIR Filter Design Techniques

The techniques covered in this section are

(1) windowing,

(2) frequency sampling design, and

(3) equiripple design.

There are many types of windowing and some will be briefly described here. The second two techniques are computer-aided design techniques and require a great deal of iterative calculation.

5.2.2.1 Windowing

It is natural to design a filter by first simulating the ideal low-pass filter, defined in equation (3-16). In doing so, the impulse response h(n) should meet these two conditions: (a) it is finite, and (b) it is causal. In particular, that $h(n) = 0$ for $n < 0$.

The impulse response is the inverse Fourier transform of the transfer function. The impulse response of the ideal low-pass filter transfer function can be found to be

$$h(n) = \frac{\sin(\omega_c n)}{n} \qquad (5\text{-}45)$$

Neither condition (a) or (b) above is met. The ideal low-pass filter and h(n) are shown in figures 5-12a and 5-12b, obtained from reference 56.

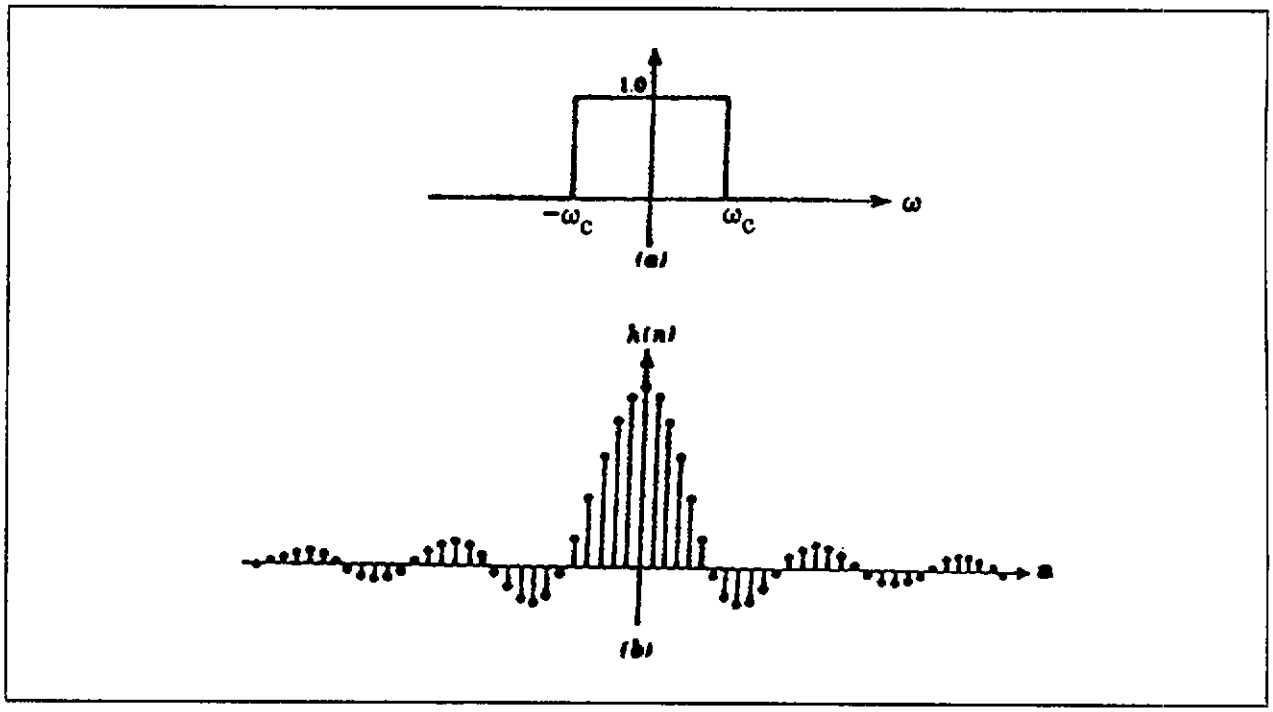


Figure 5-12. (a) Ideal low-pass filter characteristic with cutoff at ω_c and (b) impulse response corresponding to ideal low-pass filter.

A solution to meeting conditions (a) and (b) is to find a finite, causal approximation to h(n) which can be done by windowing the impulse response, (by truncating it for | n | greater than some cutoff time) and by shifting the response in time until the system is causal. The two steps are shown in figures 5-13a and 5-13b, obtained from reference 56.

As explained towards the end of paragraph 3.3, truncating the impulse response would result in a boxcar function with ripples (see figure 5-13c). Three deviations from the ideal low-pass filter emerge:

(1) the passband response is no longer flat but shows ripples that steadily increase in amplitude until the cutoff frequency,

(2) the stop band response is no longer zero, and

(3) the transition between passband and stop band is no longer abrupt.

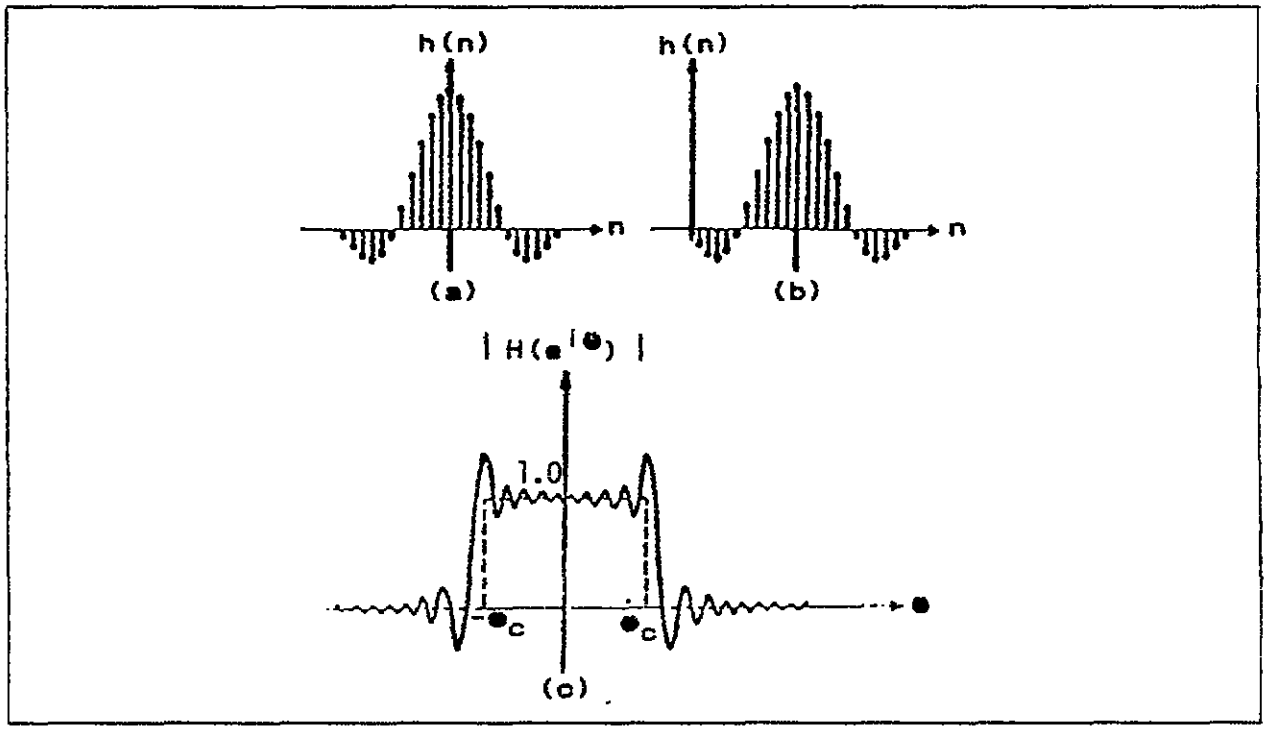


Figure 5-13. (a) Truncated version of impulse response in figure 5-12, (b) Truncated response shifted so as to make system causal, and (c) Filter frequency response resulting from truncation of impulse response.

To minimize these ripples, multiply the original infinite-impulse response by a windowing function (other than the rectangular function that was initially introduced). There is no finite window function whose transform has no side lobes, but functions can be found whose transforms have very small side lobes. If one of these functions is used, the ripples in the frequency response will be correspondingly reduced.

Some of the best-known windowing functions are listed in table 5-1 and shown in figure 5-14, along with their transforms. (The table and figure are obtained from reference 56.)

TABLE 5-1. COMMON WINDOWING FUNCTIONS

Name	Description[a]
Rectangular	$w(k) = 1$
Fejer-Bartlett	$w(k) = 1 - \lvert 2k/N \rvert$
Hanning	$w(k) = (1 + \cos \pi k/N)/2$
Hamming	$w(k) = 0.54 + 0.46 \cos \pi k/N$
Kaiser	$w(k) = \dfrac{I_0\left[N\theta_a\sqrt{1-(k/N)^2}\right]}{I_0[N\theta_a]}$

[a] For all windows, $w(k) = 0$ for $\lvert k \rvert > N$.

For the Kaiser window, I_0 is the zeroth-order Bessel function and θa is a constant that specifies a frequency response tradeoff between peak height of the side lobe ripples and the width or "energy" of the main lobe. (The Fejer-Bartlett window is also called the triangular window.) Reference 55 also mentions the Blackman window, which is $\omega(n) = 0.42 - 0.5\cos(2\pi n/(N-1)) + 0.08\cos(4\pi n/(N-1))$, $0 \le n \le N-1$.

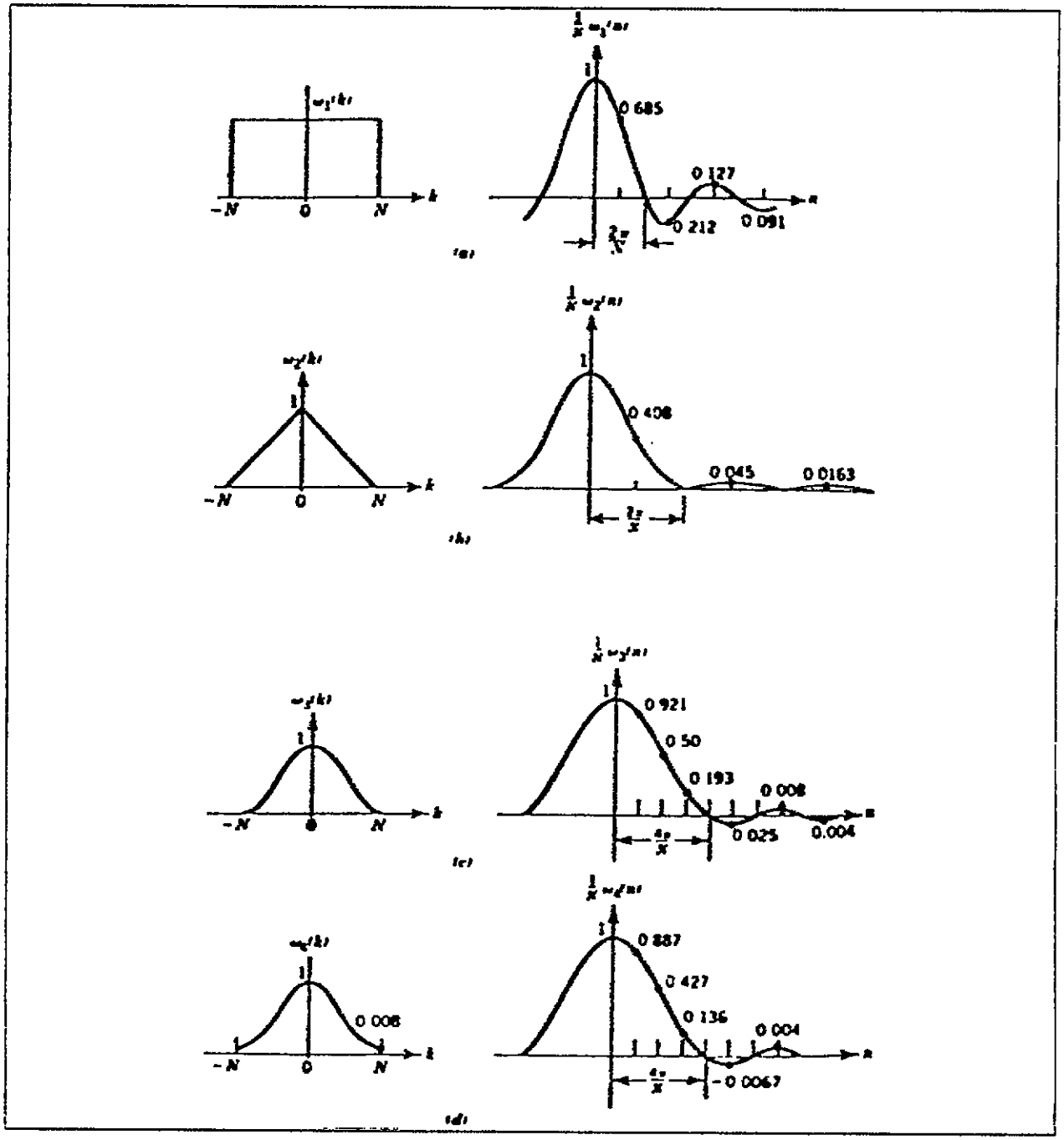


Figure 5-14. Common window functions and their transforms: (a) rectangular, (b) triangular (Fejer-Bartlett), (c) Hanning, (d) Hamming, and (e) Kaiser.

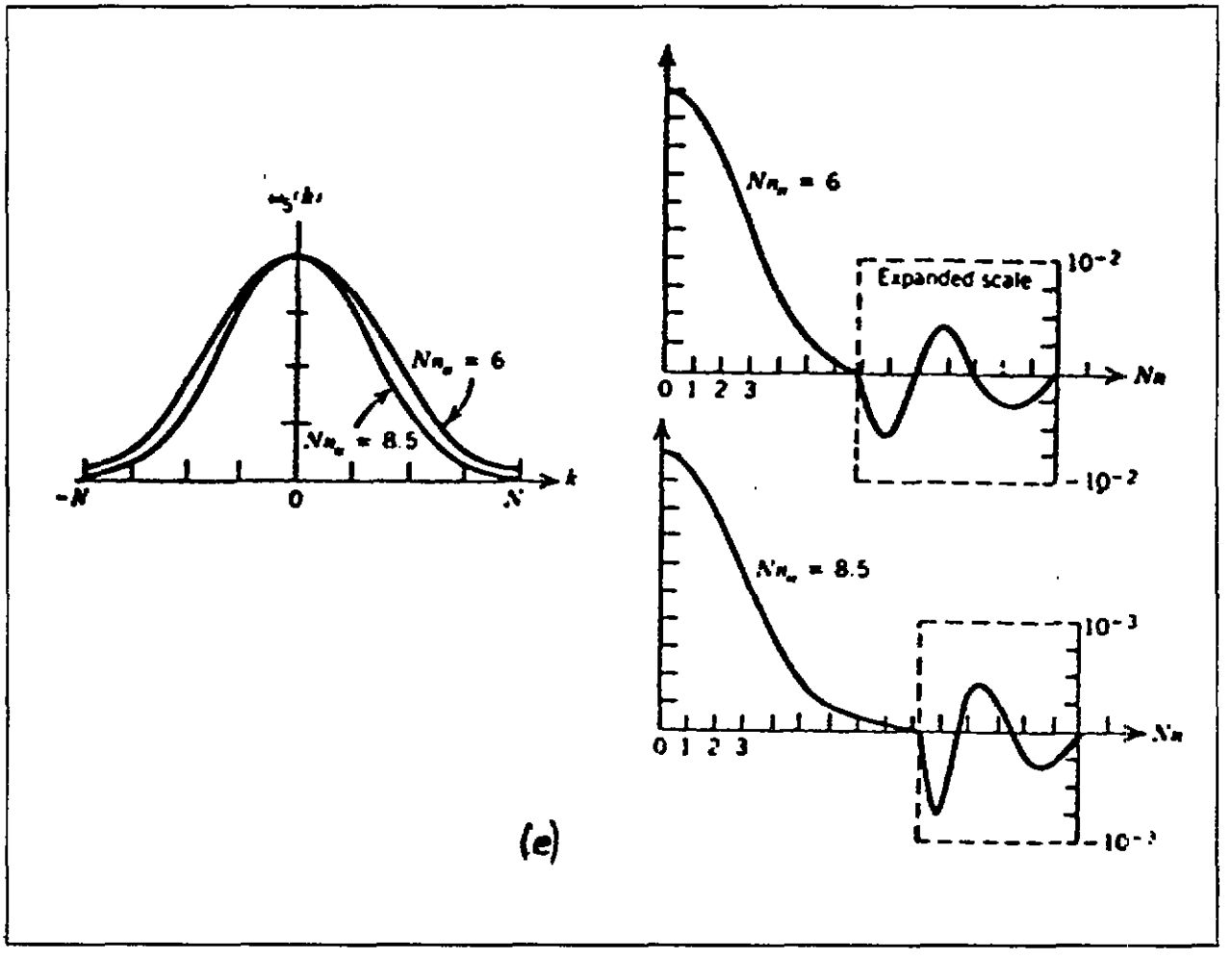


Figure 5-14 (Con.). Common window functions and their transforms: (a) rectangular, (b) triangular (Fejer-Bartlett), (c) Hanning, (d) Hamming, and (e) Kaiser.

Reference 55 compares various windowing functions in one graph, which is inserted here as figure 5-15.

In all of these windowing functions, the side lobes are much smaller than those resulting from the rectangular window, and the main lobes are all wider than those resulting from the rectangular window, producing a much closer approximation to the ideal low-pass filter. The search for the ideal windowing function is a search for the best tradeoff between side lobe amplitude and main lobe width. No FIR filter designed by Fourier transformation and windowing is optimal. The appeal of the technique lies in its simplicity and economy.

Finding the desired transfer function using windows can be summarized as follows:

(1) write the equation for the desired frequency response,

(2) find the impulse response (taking the inverse Fourier transform of the frequency response),

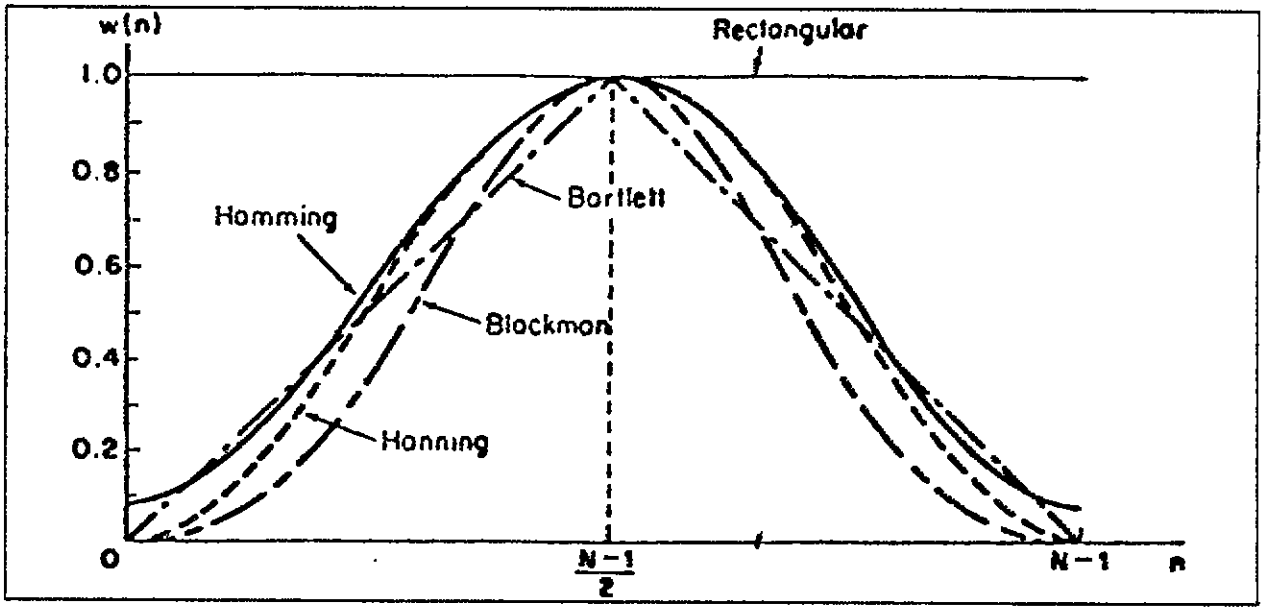


Figure 5-15. Commonly used windows for FIR filter design.

(3) select a windowing function and a window width to meet the required ripple and transition-width specifications, (Window the impulse response accordingly.)

(4) shift the impulse response to make it causal, and

(5) take the Fourier transform of the product of the windowing function and the new impulse response. The result is the desired transfer function.

5.2.2.2 Frequency Sampling Design

The approach of this method is to take samples of the frequency response and to design an FIR transfer function based on these samples. The approach starts by considering the transfer function H(z) of a digital filter, which can be found by taking the z-transform of the impulse response h(n)

$$H(z) = \sum_{n=0}^{N-1} h(n) z^{-n} \qquad (5\text{-}46)$$

The impulse response can be found from the frequency response by applying the inverse DFT as

$$h(n) = (1/N) \sum_{k=0}^{N-1} H(k) W^{nk} \qquad (5\text{-}47)$$

where $W = \exp(i2\pi/N)$.

Combining these two relations gives

$$H(z) = (1/N) \sum_{n=0}^{N-1} z^{-n} \sum_{n=0}^{N-1} H(k) W^{nk} \qquad (5\text{-}48)$$

which can be rewritten as

$$H(z) = \frac{1-z^{-N}}{N} \sum_{k=0}^{N-1} \frac{H(k)}{1-z^{-1}W^{k}} \qquad (5\text{-}49)$$

As can be seen, this design procedure consists simply of substituting samples of the desired frequency response into equation (5-49).

As seen from the last equation, there is a pole on the unit circle which leads to marginal stability (see paragraph 3.8.) To ensure stability, multiply $z^{-1}W^{k}$ by a number that is almost one (say $1 - 2^{-12}$). Good accuracy for this method requires many closely spaced samples. This design works particularly well for narrow-band filters in which only a few samples are nonzero. (Reference 55 states that even if more than a few samples are nonzero, the frequency-sampling design method yields excellent results.) A disadvantage of this method is that it lacks flexibility in specifying the passband and stopband cutoff frequencies. In addition, the ripple response for this design procedure is poor. Reference 65 suggests a method in which the ripple response can be greatly improved.

5.2.2.3 Equiripple Design

The approximation error of the frequency sampling design tends to be highest around the transition region and smaller in areas remote from the transition region. The equiripple design affords a way in which the approximation error can be spread out more uniformly.

Of concern are the zero-phase FIR filters with frequency responses of the form

$$H(e^{i\omega}) = \sum_{n=-M}^{M} h(n)e^{-i\omega n} \qquad (5\text{-}50)$$

For zero-phase filters, symmetricalness, h(n) = h(-n), and causality are required. Shift the summands in equation (5-50) are shifted to obtain causality, so that

$$H(e^{i\omega}) = h(0) + \sum_{n=1}^{M} 2h(n)\cos(\omega n) \qquad (5\text{-}51)$$

To be specified for the equiripple curve are the parameters M, δ_1, δ_2, ω_p, and ω_s, where δ_1 and δ_2 are the upper and lower ripple tolerances, and ω_p and ω_s are the passband and stop band cutoff frequencies, (see figure 3-6 or figure 5-16).

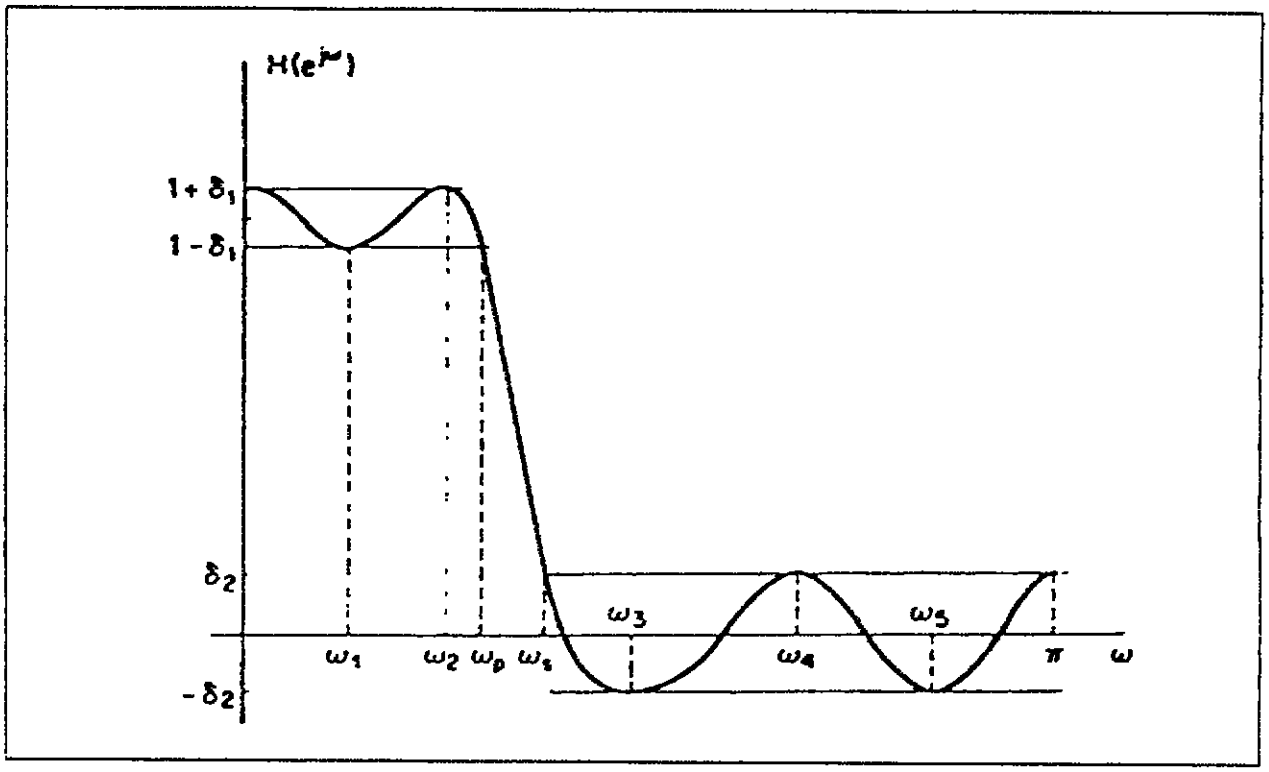


Figure 5-16. Equiripple approximation of a low-pass filter.

Messrs. O. Herrmann and H. W. Schuessler, (references 66 and 67); E. Hofstetter, A. V. Oppenheim, and J. Seigel (references 68 and 69); and J. Seigel (reference 70) developed procedures in which M, δ_1, and δ_2 are held fixed and ω_p and ω_s are solved. Messrs. T. W. Parks and J. H. McClellan, (references 71 and 72) and L. R. Rabiner (references 73 and 74) developed procedures for which M, ω_s, and ω_p are held fixed and for which δ_1 and δ_2 are solved.

In an example using a method by Herrmann and Schuessler, there are five frequencies greater than 0 and less than π where there are maxima and minima in the ripples (see figure 5-16). It can be easily shown that for a symmetric filter of order $N = 2M+1$, there will be at most M+1 local extrema in the interval $0 \leq \omega \leq \pi$.

Consider the fact that at the passband cutoff frequency, the frequency response curve is at the lower tolerance limit about 1 and that at the stop band cutoff frequency, the curve is at the higher tolerance limit about 0. This fact can be used to obtain the following two equations:

$$H(e^{i\omega_p}) = 1-\delta. \tag{5-52}$$

and

$$H(e^{i\omega_s}) = \delta_2. \tag{5-53}$$

By observing either figure 5-16 or equations (5-52) and (5-53), the following set of equations can be written:

$$H(e^{i0}) = 1 + \delta_1, \qquad H(e^{i\pi}) = \delta_2 \tag{5-54}$$

$$H(e^{i\omega_1}) = 1 - \delta_1, \qquad H'(e^{i\omega_1}) = 0 \tag{5-55}$$

$$H(e^{i\omega_2}) = 1 + \delta_1, \qquad H'(e^{i\omega_2}) = 0 \tag{5-56}$$

$$H(e^{i\omega_3}) = -\delta_2, \qquad H'(e^{i\omega_3}) = 0 \tag{5-57}$$

$$H(e^{i\omega_4}) = \delta_2, \qquad H'(e^{i\omega_4}) = 0 \tag{5-58}$$

$$H(e^{i\omega_5}) = -\delta_2, \qquad H'(e^{i\omega_5}) = 0 \tag{5-59}$$

For this method, there is the flexibility to decide which of the M-1 frequencies properly between 0 and π should be in the pass-band and which ones should be in the stop band. In this case, M+1=7, there are seven unknown coefficients (h(n)) in equation (5-51). There are 5 unknown frequencies $\omega_1,\ldots,\omega_5$ at which extrema occur, so there are 12 equations in 12 unknowns. These equations are nonlinear and must be solved by an iterative procedure.

Generally, there are 2M equations in 2M unknowns. This approach has been found to be satisfactory for orders of M = 30 or lower, and it provides the narrowest transition between passband and stop band.

CHAPTER 6

KALMAN FILTERS

A Kalman filter is a linear, recursive algorithm for computing an optimal estimate from measurements, some of which may contain noise. The noise on the measurements is assumed to be white; namely, the noise values are not correlated over time. Also assumed is knowledge of the statistics of the noise on the measurements. The algorithm is recursive and thus requires an initial estimate to start the filter as well as a guess as to the correctness of that estimate. Finally, the algorithm is linear, consisting of matrix equations. (Reference 3 has an excellent intuitive introduction to Kalman filtering in its first chapter.)

The Kalman filter is most often used as a data processing algorithm (a computer program) and can be extremely efficient, requiring a minimum of computer storage and using all available data by weighing the data measurements. The filter has the ability to take several different types of data and generate an estimate of a totally different quantity. Because it is a predictor-corrector, it can generate its next estimate based on less current data than would be needed for a directly calculated solution. The data need not be entered at equally spaced time intervals, at the same time, or in certain sequences. The filter allows the user to apply the knowledge of the behavior and statistics of both the measurements and the quantities to be estimated to obtain the solution; in fact, these models and statistics may vary with time. In addition, a self-contained error analysis is included in its equations. Finally, as a predictor, it is useful for real-time control.

A few caveats apply, however. To begin with, there is no "general" Kalman filter. Each algorithm is dependent on the quantities to be estimated and their dynamics, the measurements available and their statistics, and the initial values needed to start the algorithm. Secondly, the algorithm is most efficient if matrix inversions can be avoided. Matrix inversions can usually be done, although it depends on the application, that is, the particular Kalman filter written. Thirdly, a Kalman filter will update with less current data than is required for a calculated solution, but if updating with this data is done for too long, the estimate may become grossly invalid. The filter is said to have diverged. Fourthly, the order of the input of data is unimportant only for the basic Kalman filter: a filter whose dynamic and measurement models are both linear. Though all derivations and claims of optimality are valid only for such filters, linearity is rare in the real world. Hence, suboptimality is often settled for using an extended Kalman filter. Sometimes it is found that permuting the order in which data is

entered may significantly change the estimate. Finally, the validity of the self-contained error analysis is dependent on the degree to which the noise statistics fit the theoretical assumptions. For example, biases in the data will not be indicated by the output error covariance matrix.

Before describing a Kalman filter, the description of some terms are in order.

The "expected value" of g(x) of a random variable X whose frequency function is f(x) is

$$E[g(x)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Loosely speaking, the expected value of g(x) is the average of g(x).

An "estimate" is a computed value of a quantity. For example, the sample mean

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{\eta}$$

is an estimate of true population mean μ.

A statistic t is called an "unbiased estimate" or "unbiased estimator" of the parameter ν if $E[t] = \nu$. For example, the sample mean x can be shown to be an unbiased estimate of the true population mean μ.

An "optimal estimate" is one that minimizes the variance of an estimate and is unbiased. When using an optimal filter such as a filter computing an optimal estimate, it is assumed that the exact descriptions of the system dynamics and the measurement process are known. In addition, an optimal filter must model <u>all</u> error sources in the system including unmodeled parameters, linearization errors, leaky attitude controls, and solar winds.

Error sources mentioned in the previous paragraph constitute "process noise," noise that stems from mismodeling. Other error sources include those of "measurement noise," which stems from faulty measuring devices or the misuse of those devices. (Because the Kalman filter is based on least squares, an error source inherent in the Kalman filter is the difference between the estimate sample measurement values and the true population measurement values.) An example dealing with unmodeled parameters is filtering a position parameter without including the

velocity or acceleration parameter into the model, which where including these parameters could facilitate filtering the position parameter.

Leaky attitude controls and solar winds are factors that can lead a spacecraft off course. They are usually considered negligible and as such are normally not included in the models. The errors that may result from their not being included constitute process noise.

A suboptimal filter is a filter that does not take into account all the factors that have been mentioned in describing an optimal filter. A "predictor-corrector" is a recursive algorithm that has two steps to each recursion: a predictor and a corrective step. A predictor step is the state of the system predicted by using the output data of the preceding recursions in which the nth state of the system is predicted using the first n-1 states. A corrective step is the given input nth state of the system used to correct the predicted nth state of the system, thus producing the final output nth state of the system. (The nth state of the system is the state of the system at the nth recursion.)

6.1 Linear Discrete Kalman Filter

In the case of a linear discrete Kalman filter, the dynamics of the quantities to be estimated may be described by linear difference equations. Additionally, the relationship between the measurements, taken at discrete times, and the estimated quantities is linear.

6.1.1 Definitions

To write such a Kalman filter, begin with these definitions.

(1) A model describing the dynamics of the quantities to be estimated. This model will be of the form

$$\underline{x}(k) = \Phi(k,k-1)\underline{x}(k-1) + G(k,k-1)\underline{w}(k-1) \qquad (6\text{-}1)$$

where

k refers to the kth time point t_k;

$\underline{x}(k)$ is the set of quantities to be estimated, arranged in a vector, and called the state vector;

$\Phi(k,k-1)$ is called the transition matrix and describes the change in the state vector from time t_{k-1} to t_k;

$G(k,k-1)$ is called the input matrix; for practical purposes, it is generally taken to be equal to the identity matrix;

and

$\underline{w}(k-1)$ is called the plant or process noise. The vector $\underline{w}(k-1)$ is assumed to contain zero-mean white noise with Gaussian distribution. The vector resulting from the product $G(k,k-1)\underline{w}(k-1)$ represents the unknown portion of the dynamic model. The covariance matrix of $\underline{w}(k)$ is designated by $Q(k)$.

(2) An equation relating the state vector to the measurements. The form of this equation is

$$\underline{y}(k) = H(k)\underline{x}(k) + \underline{v}(k) \qquad (6\text{-}2)$$

where

$\underline{y}(k)$ is the measurement at time t_k;

$H(k)$ is called the measurement matrix and linearly relates the measurements to the state variables;

and

$\underline{v}(k)$ is the measurement noise. The vector $\underline{v}(k)$ is assumed to contain zero-mean white noise with Gaussian distribution. The covariance matrix of $\underline{v}(k)$ is designated by $R(k)$.

(3) Finally, the following quantities are needed:

$\hat{\underline{x}}(0)$ the initial state estimate; and

$P(0)$ the corresponding error covariance matrix. (In the Kalman filter algorithm, P_k is the covariance matrix of the estimate error of the kth state vector. That is, $P_k = E[(\underline{x}(k)-\hat{\underline{x}}_k)(\underline{x}(k)-\hat{\underline{x}}_k)^T]$.)

To write a Kalman filter, three matrices, I, G, and H are needed to define the dynamic and measurement structures as well as the matrices, Q and R, to define the dynamic and measurement statistics. Also needed are the initial values, $\hat{\underline{x}}(0)$ and $P(0)$.

Given these items, the filter is written using the equations described next. In these equations, the notation $\hat{\underline{x}}$ denotes the estimate of the state vector $\underline{x}$. The notation $\hat{\underline{x}}(a/b)$ signifies the estimate of the state vector at time t_a given measurements taken up to and including time t_b.

6.1.2 Kalman Filter Algorithm

The first step in this Kalman filter algorithm is the extrapolation or time update for the estimate and its covariance matrix. These equations are

$$\hat{\underline{x}}(k/k-1) = \Phi(k,k-1)\hat{\underline{x}}(k-1/k-1) \tag{6-3}$$

and

$$P(k/k-1) = \Phi(k,k-1)P(k-1/k-1)\Phi^T(k,k-1) + G(k,k-1)Q(k-1)G^T(k,k-1). \tag{6-4}$$

The next step is to compute the weighing matrix K, called the Kalman gain, by

$$K(k) = P(k/k-1)H^T(k)[H(k)P(k/k-1)H^T(k)+R(k)]^{-1} \tag{6-5}$$

Finally, the output is obtained by

$$\hat{\underline{x}}(k/k) = \hat{\underline{x}}(k/k-1) + K(k)[\underline{y}(k) - H(k)\hat{\underline{x}}(k/k-1)] \tag{6-6}$$

and

$$P(k/k) = [I - K(k)H(k)]\ P(k/k-1), \tag{6-7}$$

where I is the proper size identity matrix.

$\hat{\underline{x}}(k/k)$ in equation (6-6) is, ultimately, the result for the kth step in the algorithm. P(k/k) of equation (6-7) is used as P(k-1/k-1) in equation (6-4) in a recursive step. Equations (6-3) through (6-7) represent one recursive step in the Kalman filter algorithm.

Table 6-1 is obtained from reference 2 and gives a summary of the entire algorithm. The top two boxes give the original model equations, initial conditions, and assumptions. Equations (6-1) and (6-2) are equivalent to the system and measurement model equations. (The $\underline{z}$ in the table is the same as the $\underline{y}$ vector of equation (6-2).) The "other assumptions" equation in the second box means that the system model errors and the measurement noises are uncorrelated. The equations in the bottom two boxes correspond to equations (6-3) through (6-7), which comprise one recursive step in the algorithm.

TABLE 6-1. SUMMARY OF DISCRETE KALMAN FILTER EQUATIONS

System Model	$\underline{x}_k = \Phi_{k-1}\underline{x}_{k-1} + \underline{w}_{k-1}, \quad \underline{w}_k \sim N(\underline{0}, Q_k)$
Measurement Model	$\underline{z}_k = H_k\underline{x}_k + \underline{v}_k, \quad \underline{v}_k \sim N(\underline{0}, R_k)$
Initial Conditions	$E[\underline{x}(0)] = \hat{\underline{x}}_0, E[(\underline{x}(0) - \hat{\underline{x}}_0)(\underline{x}(0) - \hat{\underline{x}}_0)^T] = P_0$
Other Assumptions	$E[\underline{w}_k\underline{v}_j^T] = 0$ for all j, k
State Estimate Extrapolation	$\hat{\underline{x}}_k(-) = \Phi_{k-1}\hat{\underline{x}}_{k-1}(+)$
Error Covariance Extrapolation	$P_k(-) = \Phi_{k-1}P_{k-1}(+)\Phi_{k-1}^T + Q_{k-1}$
State Estimate Update	$\hat{\underline{x}}_k(+) = \hat{\underline{x}}_k(-) + K_k[\underline{z}_k - H_k\hat{\underline{x}}_k(-)]$
Error Covariance Update	$P_k(+) = [I - K_kH_k]P_k(-)$
Kalman Gain Matrix	$K_k = P_k(-)H_k^T[H_kP_k(-)H_k^T + R_k]^{-1}$

(-) and (+) are used to denote the times immediately before and immediately after a discrete measurement. For example, $P_k(+)$ and $P_k(-)$ in the table are equivalent to $P(k/k-1)$ in equation (6-7).

Several quantities in equations (6-3) through (6-7) are of particular interest. The vector $H(k)\hat{\underline{x}}(k/k-1)$ in equation (6-6) is called the predicted measurement and its difference with the actual measurement, given by

$$\underline{y}(k) - H(k)\underline{x}(k/k-1) \tag{6-8}$$

is called the innovation. This vector is often used for testing for filter divergence or for data editing.

The matrix

$$H(k)P(k/k-1)H^T(k) + R(k) \tag{6-9}$$

in equation (6-5) is considered the covariance matrix of the innovations and is usually chosen to be one-dimensional, namely, a scalar, in the following manner. If only one measurement is input at time t_k, then in equation (6-2) $\underline{y}(k)$ is a scalar, $H(k)$ has only one row, and $\underline{v}(k)$ is a scalar. The matrix (6-9) is then a scalar, whatever the size of $P(k/k-1)$. If more than one

measurement is available at time t_k, it is usually more economical in computer time to process each measurement individually, since matrix inversions are avoided. Hence, the first measurement used would be preceded by a time update (equations (6-3) and (6-4). (The estimate $\hat{\underline{x}}$ and the estimate error covariance matrix P being is updated from time t_{k-1} to time t_k. (equations (6-3) and (6-4)) The other measurements at time t_k would not need a time update, so only equations (6-5), (6-6), and (6-7) would be used with the old $\hat{\underline{x}}(k/k-1)$ and $P(k/k-1)$ of the previous step being replaced in each equation by the new ones.

As mentioned before, for a Kalman filter with linear models and the assumed noise statistics, the order in which all the measurements at time t_k are processed is not significant. Most importantly, processing the measurements individually will give the same output as processing them together in a measurement vector if there is no cross correlation among the measurements at time t_k. Usually this is the case or at least can be assumed true with a minimum of error.

The preceding equations will easily yield a computer program which is a working Kalman filter. The problem remains, however, to make sure the filter does the job for which it was intended. Once written, the filter must be subjected to extensive simulation testing. Testing is necessary to detect modeling errors, statistical assumption errors, inappropriate initial conditions, biases, correlated noise sequences, and finally to determine tuning value.

Filter tuning is the process of achieving the best possible estimation performance from a filter once its structural form has been specified. In a Kalman filter, the structural form is specified by the matrices Φ, G, and H. The initial estimate error covariance matrix P(0), the model error covariance matrix Q, and the measurement noise covariance matrix R are the variables modified during tuning. The use of Q and R is normally based on the knowledge of $\underline{w}$ (the system model error) and $\underline{v}$ (the measurement noise). These vectors, $\underline{w}$ and $\underline{v}$, account for actual noises and disturbances in the physical system as well as inadequacies in the dynamic and measurement models.

6.2 The Linear Continuous Kalman Filter

For a linear continuous Kalman filter, start with the dynamic model

$$\dot{\underline{x}}(t) = F(t)\underline{x}(t) + G(t)\underline{w}(t) \qquad (6\text{-}10)$$

and the measurement model

$$\underline{y}(t) = H(t)\underline{x}(t) + \underline{v}(t) \qquad (6\text{-}11)$$

with initial values $\underline{x}(0)$ and $P(0)$, where $\underline{w}(t)$ and $\underline{v}(t)$ are zero-mean white noise processes uncorrelated with $\underline{x}(0)$ and have covariance matrices $Q(t)$ and $R(t)$ (see table 6-2).

TABLE 6-2. SUMMARY OF CONTINUOUS KALMAN FILTER EQUATIONS (WHITE MEASUREMENT NOISE)

System Model	$\dot{\underline{x}}(t) = F(t)\underline{x}(t) + G(t)\underline{w}(t), \quad \underline{w}(t) \sim N(\underline{0}, Q(t))$
Measurement Model	$\underline{y}(t) = H(t)\underline{x}(t) + \underline{v}(t), \quad \underline{v}(t) \sim N(\underline{0}, R(t))$
Initial Conditions	$E[\underline{x}(0)] = \hat{\underline{x}}_0, \; E[(\underline{x}(0) - \hat{\underline{x}}_0)(\underline{x}(0) - \hat{\underline{x}}_0)^T] = P_0$
Other Assumptions	$R^{-1}(t)$ exists
State Estimate	$\dot{\hat{\underline{x}}}(t) = F(t)\hat{\underline{x}}(t) + K(t)[\underline{y}(t) - H(t)\hat{\underline{x}}(t)], \; \hat{\underline{x}}(0) = \hat{\underline{x}}_0$
Error Covariance Propagation	$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + G(t)Q(t)G^T(t) - K(t)R(t)K^T(t), \; P(0) = P_0$
Kalman Gain Matrix	$K(t) = P(t)H^T(t)R^{-1}(t)$ when $E[\underline{w}(t)\underline{v}^T(\tau)] = 0$ $= [P(t)H^T(t) + G(t)C(t)]R^{-1}(t)$ when $E[\underline{w}(t)\underline{v}^T(\tau)] = C(t)\delta(t-\tau)$

Then the filter equations are

$$\dot{\hat{\underline{x}}}(t) = F(t)\hat{\underline{x}}(t) + K(t)[\underline{y}(t) - H(t)\hat{\underline{x}}(t)], \qquad (6\text{-}12)$$

and

$$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + G(t)Q(t)G^T(t) - K(t)R(t)K^T(t) \qquad (6\text{-}13)$$

with

$$K(t) = [P(t)H^T(t) + G(t)C(t)]\,R^{-1}(t) \qquad (6\text{-}14)$$

where $E\{\underline{w}(t)\underline{v}^T(s)\} = C(t)\delta(t-s)$, with E representing the expected value and δ, the Dirac delta function. The three equations comprise one recursive step in the continuous Kalman filter algorithm.

The problem is, in part, to solve P(t) in equation (6-13), which is known as the "matrix Riccati equation." (Reference 2 gives methods for solving this equation.) Then plug P(t) into equation (6-14) to solve for K(t), which is used in equation (6-12) to obtain the final desired result $\hat{\underline{x}}(t)$ for a particular recursive step. (See reference 2 for further discussion.)

6.3 Extended Kalman Filters

An extended Kalman filter is a data processing algorithm which is based on the Kalman filter algorithm and which efficiently provides estimates for nonlinear problems. Such filters are suboptimal because no theoretical optimality of the estimate can be proven. Either or both of the given models (continuous or discrete, otherwise known as "dynamic" or "measurement") of a Kalman filter may be nonlinear for an extended Kalman filter to be required. The extended Kalman filter is still a linear algorithm, however.

For an extended Kalman filter, assume the dynamic model is

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t),t) + G(\underline{x}(t),t)\underline{w}(t) \qquad (6\text{-}15)$$

with initial conditions $\underline{x}(t_0)$ and $P(0) = \text{cov}\{\underline{x}(t_0),\underline{x}(t_0)\}$, where $\underline{w}(t)$ is a zero-mean, Gaussian noise vector, such that $\text{cov}\{\underline{w}(t),\underline{w}(s)\} = Q(t)\delta(t-s)$, and $\text{cov}\{\underline{x}(t_0),\underline{w}(t)\} = 0$ for $t \geq t_0$. The notation cov{.,.} represents the covariance matrix of the two vectors. The δ is the Dirac delta function. The vector $\underline{f}$ is a nonlinear function of the state vector $\underline{x}(t)$ and of the time t. Notice that the model (6-15) is not completely general since the dynamic noise is assumed additive.

The measurement model is given by

$$\underline{y}(t_k) = \underline{h}(\underline{x}(t_k),t_k) + \underline{v}(t_k),\ k = 0,1,\ldots, \qquad (6\text{-}16)$$

where $\underline{v}(t_k)$ is the zero-mean measurement noise vector, with $\text{cov}\{\underline{v}(t_k),\underline{v}(t_k)\} = R(k)$, $\text{cov}\{\underline{v}(t_k),\underline{v}(t_j)\} = 0$, $\text{cov}\{\underline{x}(t_0),\underline{v}(t_k)\} = 0$, and $\text{cov}\{\underline{w}(t),\underline{v}(t_k)\} = 0$.

Note that the measurement model, though nonlinear, is still discrete since the measurements are taken at discrete times. Also assume the initial values $\hat{\underline{x}}(0)$ and P(0) are given.

Then the extended Kalman filter update from the measurement at time t_k to the measurement at time t_{k+1} is accomplished in the following manner:

(1) The state vector update is achieved by using the vector function $\underline{f}$ of equation (6-15). Thus, the equation

$$\dot{\hat{\underline{x}}}(t/t_k) = \underline{f}(\hat{\underline{x}}(t/t_k),t) \tag{6-17}$$

is integrated for $t_k \le t \le t_{k+1}$, resulting in $\hat{\underline{x}}(t/t_k)$.

(2) The matrix

$$F(\hat{\underline{x}}(t),t) = \left.\frac{\delta \underline{f}(\underline{x}(t),t)}{\delta \underline{x}}\right|_{\underline{x}(t) = \hat{\underline{x}}(t/t_k)} \tag{6-18}$$

is calculated.

For the rest of the chapter, $F(\hat{\underline{x}}(t),t)$ will be denoted $F(t)$, $\underline{f}(\underline{x}(t),t)$ will be denoted $\underline{f}$, and $\underline{x}(t)$ will be denoted $\underline{x}$ for the sake of convenience.

(3) The transition matrix $\Phi(t_{k+1},t_k)$ is calculated using

$$\frac{\partial \Phi(t,t_k)}{\partial t} = F(t)\Phi(t,t_k) \tag{6-19}$$

where $t_k \le t \le t_{k+1}$ and $\Phi(t_k,t_k)$ is the identity matrix.

(4) The state noise covariance matrix Q* is calculated using

$$Q^*(t_{k+1}) = \int_{t_k}^{t_{k+1}} \Phi(t_{k+1},t)G(t)Q(t)G^T(t)\Phi^T(t_{k+1},t)dt. \tag{6-20}$$

(5) The measurement matrix $H(t_{k+1})$ is calculated by

$$H(t_{k+1}) = \left.\frac{\partial \underline{h}}{\partial \underline{x}}\right|_{\underline{x} = \hat{x}(t_{k+1}/t_k),} \tag{6-21}$$

where $\hat{\underline{x}}(t_{k+1}/t_k)$ was obtained in step (1).

The update is not ready to be performed. To obtain the gain matrix $K(t_{k+1})$, first extrapolate the state covariance matrix using equation

$$P(t_{k+1}/t_k) = \Phi(t_{k+1},t_k)P(t_k/t_k)\Phi^T(t_{k+1},t_k) + Q^*(t_{k+1}). \tag{6-22}$$

Then

$$K(t_{k+1}) = P(t_{k+1}/t_k)H^T(t_{k+1})[H(t_{k+1})P(t_{k+1}/t_k)H^T(t_{k+1}) + R(t_{k+1})] \quad (6\text{-}23)$$

The state update is then

$$\hat{x}(t_{k+1}/t_{k+1}) = \underline{\hat{x}}(t_{k+1}/t_k) + K(t_{k+1})[\underline{y}(t_{k+1}) - \underline{h}(\underline{\hat{x}}(t_{k+1}/t_k), t_k)]. \quad (6\text{-}24)$$

Equation (6-24) gives the desired update at time t_{k+1}. To complete the filter update, compute the state covariance matrix at time t_{k+1} by

$$P(t_{k+1}/t_{k+1}) = [I - K(t_{k+1})H(t_{k+1})]P(t_{k+1}/t_k), \quad (6\text{-}25)$$

where I is the appropriately sized identity matrix.

Further modifications of the basic Kalman filter equations are possible and often necessary for the proper functioning of a Kalman filter. If the noise vectors are not zero mean, the state vector may be expanded to solve for bias estimates. If the noise sequences are not white, shaping filters may be used. If the noise statistics are unknown, adaptive estimation schemes may be added to determine these statistics. If computer word length affects the numerical stability of the filter, square root type filters can be written. The literature is filled with such Kalman filter applications and more are being added every day as Kalman filters are used to solve harder and harder problems.

A disadvantage of the extended Kalman filter is that the Kalman gain $K(t_{k+1})$ and the error covariance update matrix $P(t_{k+1}/t_{k+1})$ must be computed in real time. They cannot be precomputed before the measurements are collected and stored in computer memory as can be done when using the basic Kalman filter because $K(t_{k+1})$ and $P(t_{k+1}/t_{k+1})$ are both dependent on $\underline{\hat{x}}(t_{k+1}/t_k)$. In equations (6-23) and (6-25), it is seen that $H(t_{k+1})$ is needed to calculate $K(t_{k+1})$ and $P(t_{k+1}/t_{k+1})$. $H(t_{k+1})$, given in equation (6-21), is dependent on $\underline{\hat{x}}(t_{k+1}/t_k)$. $H(t_{k+1})$ is actually shorthand for $H(\underline{\hat{x}}(t_{k+1}/t_k))$. Needed is the estimate update, and hence the measurements themselves, to calculate the Kalman gain and the error covariance update.

After the extended Kalman filter is designed, a "sensitivity analysis" is sometimes performed. A sensitivity analysis comprises a set of analyses to determine the sensitivity of the filter design to any possible differences between this suboptimal filter and a filter that fits the optimal mold exactly.

References 1 and 2 discuss solutions to several of the problems mentioned in the preceding paragraphs. Since the discussion here was limited to conventional real time Kalman filters, the reader may also wish to consult references 1 and 2 for a state-of-the-art exposition on post-flight Kalman smoothers.

6.4 Example

A simple example will illustrate an extended Kalman filter as described in paragraph 6.3. A moving target is tracked by N stations, each measuring at times t_k, k=1,2,..., m, the distance of the target from the station. Estimate the target's position and velocity at time t_j, $1 \le j \le m$, using all the range measurements up to and including those at time t_j.

In this example, the state vector $\underline{x}(t)$ is chosen to be the six-vector consisting of the position and velocity of the target in a geocentric coordinate system, so

$$\underline{x}(t) = \begin{bmatrix} x(t) \\ y(t) \\ z(t) \\ vx(t) \\ vy(t) \\ vz(t) \end{bmatrix} \tag{6-26}$$

The continuous dynamic model corresponding to equation (6-15) is

$$\dot{\underline{x}}(t) = \underline{f}(\underline{x}(t)) + G(\underline{x}(t))\underline{w}(t) \tag{6-27}$$

where

$$\underline{f}(\underline{x}(t)) = \begin{bmatrix} vx(t) \\ vy(t) \\ vz(t) \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad G(\underline{x}(t)) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6-28}$$

and

$$\underline{w}(t) = \begin{bmatrix} ax(t) \\ ay(t) \\ az(t) \end{bmatrix} \tag{6-29}$$

where ax, ay, and az are Gaussian white noise errors. They can also be considered geocentric components of target acceleration. Of course, $\underline{f}(\underline{x}(t))$, G, and $\underline{w}(t)$ could have been assigned other values to satisfy equation (6-27). In real-life situations, however, the values assigned to $\underline{f}$, G, and $\underline{w}$ are as given above.

The initial conditions $\hat{\underline{x}}(t_0)$ and P(0) are assumed to be provided before time t_1. A simple (and usually inadequate) method is to take three ranges prior to time t_1 and triangulate to get $\hat{x}(t_0)$, $\hat{y}(t_0)$, and $\hat{z}(t_0)$ and set $\hat{vx}(t_0) = \hat{vy}(t_0) = \hat{vz}(t_0) = 0$. The 6x6 matrix P(0) can be chosen to be diagonal with values on the main diagonal to reflect the filter designer's confidence or lack thereof in the method of determining $\hat{\underline{x}}(t_0)$. (Generally, the greater the confidence, the larger the values.)

The matrix Q(t) where cov $\{\underline{w}(t),\underline{w}(s)\} = Q(t)\delta(t-s)$ is assumed to be $\sigma^2 I_3$, where I_3 is the 3x3 identity matrix and σ^2 is chosen to compensate for unmodeled accelerations. The value of σ may be determined by subsequent tuning studies.

The measurement model is given in equation (6-16) as

$$\underline{y}(t_k) = \underline{h}(\underline{x}(t_k),t_k) + \underline{v}(t_k).$$

For this example, the vector $\underline{y}(t_k)$ is N-dimensional consisting of the N ranges available at time t_k. For any i, $1 \le i \le N$, the ith component of $\underline{h}(\underline{x}(t_k),t_k)$ is

$$\sqrt{(x(t_k) - X_i)^2 + (y(t_k) - Y_i)^2 + (z(t_k) - Z_i)^2} \qquad (6\text{-}30)$$

where $x(t_k)$, $y(t_k)$, and $z(t_k)$ are components of the state vector at time t_k and (X_i, Y_i, Z_i) is the location of the ith tracking station in geocentric coordinates. The matrix R(k) = cov $\{\underline{v}(t_k),\underline{v}(t_k)\}$ is assumed to be diagonal with main diagonal element i, $1 \le i \le N$, equal to the expected variance of the measurements provided by the ith ranging station.

The measurements in this example are going to be processed individually as described in paragraph 6.1. Hence, for each t_k, there are N measurements to process which satisfy the scalar equation

$$y_i(t_k) = h_i(\underline{x}(t_k)) + v_i(t_k), \quad i = 1,2,\ldots, N, \qquad (6\text{-}31)$$

where

$$h_i(\underline{x}(t_k)) = \sqrt{(x(t_k) - X_i)^2 + (y(t_k) - Y_i)^2 + (z(t_k) - Z_i)^2}. \qquad (6\text{-}32)$$

The matrix $R_i(k) = \text{cov}\{\epsilon_i(t_k), v_i(t_k)$ is then a scalar.

Now proceed with the steps described in paragraph 6.3. These steps describe the quantities needed to estimate the position and velocity of the target at time t_{k+1}, given the estimate at time t_k, and the measurements at time t_{k+1}.

(1) The equation

$$\dot{\underline{\hat{x}}}(t/t_k) = \underline{f}(\underline{\hat{x}}(t/t_k)) \tag{6-33}$$

must be integrated for $t_k \leq t \leq t_{k+1}$ to get $\underline{\hat{x}}(t/t_k)$. According to equation (6-27), equation (6-33) is written out as

$$\begin{bmatrix} \hat{x}(t/t_k) \\ \hat{y}(t/t_k) \\ \hat{z}(t/t_k) \\ \hat{vx}(t/t_k) \\ \hat{vy}(t/t_k) \\ \hat{vz}(t/t_k) \end{bmatrix} \qquad \begin{bmatrix} \hat{vx}(t/t_k) \\ \hat{vy}(t/t_k) \\ \hat{vz}(t/t_k) \\ 0 \\ 0 \\ 0 \end{bmatrix} . \tag{6-34}$$

To integrate the vector equation (6-34), this example begins with the lower three components of the vector. The fourth component of the vector equation (6-34) is

$$\dot{\hat{vx}}(t/t_k) = 0. \tag{6-35}$$

Integrate equation (6-35) for $t_k \leq t \leq t_{k+1}$ to obtain

$$\hat{vx}(t_{k+1}/t_k) = \hat{vx}(t_k/t_k). \tag{6-36}$$

Similar computations hold for $\hat{vy}$ and $\hat{vz}$. Since the velocity estimate is constant over the interval $[t_k, t_{k+1}]$, the first component of equation (6-34) can be rewritten as

$$\dot{\hat{x}}(t/t_k) = vx\ (t_k/t_k). \tag{6-37}$$

Integrate both sides to obtain

$$\int_{t_k}^{t_{k+1}} \dot{\hat{x}}(t/t_k)\ dt = \hat{vx}(t_k/t_k) \int_{t_k}^{t_{k+1}} dt \tag{6-38}$$

so that

$$\hat{x}(t_{k+1}/t_k) - \hat{x}(t_k/t_k) = \hat{vx}(t_k/t_k)(t_{k+1} - t_k) \tag{6-39}$$

or

$$\hat{x}(t_{k+1}/t_k) = \hat{x}(t_k/t_k) + \hat{vx}(t_k/t_k)(t_{k+1} - t_k). \quad (6\text{-}40)$$

Similar computations yield $\hat{y}(t_{k+1}/t_k)$ and $\hat{z}(t_{k+1}/t_k)$. (6-41)

Thus, the state vector time update equation is

$$\begin{bmatrix} \hat{x}(t_{k+1}/t_k) \\ \hat{y}(t_{k+1}/t_k) \\ \hat{z}(t_{k+1}/t_k) \\ \hat{vx}(t_{k+1}/t_k) \\ \hat{vy}(t_{k+1}/t_k) \\ \hat{vz}(t_{k+1}/t_k) \end{bmatrix} = \begin{bmatrix} \hat{x}(t_k/t_k) + \hat{vx}(t_k/t_k)(t_{k+1} - t_k) \\ \hat{y}(t_k/t_k) + \hat{vy}(t_k/t_k)(t_{k+1} - t_k) \\ \hat{z}(t_k/t_k) + \hat{vz}(t_k/t_k)(t_{k+1} - t_k) \\ \hat{vx}(t_k/t_k) \\ \hat{vy}(t_k/t_k) \\ vz(t_k/t_k) \end{bmatrix} \quad (6\text{-}42)$$

(2) The next step is to compute the matrix F(t) of equation (6-18). Using equation (6-27), the i,jth element of the matrix F(t) is

$$\left[\frac{\partial \underline{f}}{\partial \underline{x}}\right]_{ij} = \frac{\partial (\underline{f})_i}{\partial (\underline{x})_j}, \qquad 1 \le i,j \le 6, \quad (6\text{-}43)$$

evaluated at $\underline{x} = \hat{\underline{x}}(t/t_k)$. For $1 \le i \le 3$ and $1 \le j \le 3$,

$$\frac{\partial (\underline{f})_i}{\partial (\underline{x})_j} \text{ is of the form } \frac{\partial vx(t)}{\partial x(t)}, \quad (6\text{-}44)$$

which is zero, since velocity is not dependent on position.

For $1 \le i \le 3$ and $4 \le j \le 6$, if $i \neq j$.

$$\frac{\partial (\underline{f})_i}{\partial (\underline{x})_j} \text{ is of the form } \frac{\partial vx(t)}{\partial vy(t)}, \quad (6\text{-}45)$$

which is zero. If i=j,

$$\frac{\partial (\underline{f})_i}{\partial (\underline{x})_j} \text{ is of the form } \frac{\partial vx(t)}{\partial vx(t)}, \quad (6\text{-}46)$$

which is 1. For $4 \leq i \leq 6$ and all j,

$$\frac{\partial (\underline{f})_i}{\partial (\underline{x})_j} = 0 \qquad (6\text{-}47)$$

Hence,

$$F(t) = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} . \qquad (6\text{-}48)$$

(3) The next step is to compute the transition matrix Φ according to the matrix differential equation (6-19) which is

$$\frac{\partial\ \Phi(t,t_k)}{\partial t} = F(t)\Phi(t,t_k)$$

where $t_k \leq t \leq t_{k+1}$ and $\Phi(t_k,t_k)$ is the 6x6 identity matrix I_6. The solution to equation (6-19) is

$$\int_{t_k}^{t_{k+1}} F(t)dt$$

$$\Phi(t_{k+1},t_k) = e$$

$$= I_6 + \int_{t_k}^{t_{k+1}} F(t)dt + 1/2\left[\int_{t_k}^{t_{k+1}} F(t)dt\right]^2 + \ldots . \qquad (6\text{-}49)$$

Then, since $\int_{t_k}^{t_{k+1}} F(t)dt$

$$= \begin{bmatrix} 0 & 0 & 0 & (t_{k+1}-t_k) & 0 & 0 \\ 0 & 0 & 0 & 0 & (t_{k+1}-t_k) & 0 \\ 0 & 0 & 0 & 0 & 0 & (t_{k+1}-t_k) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and $\left(\int_{t_k}^{t_{k+1}} F(t)dt \right)^n$ is the zero matrix for n>1,

$$\Phi(t_{k+1},t_k) = \begin{bmatrix} 1 & 0 & 0 & (t_{k+1}-t_k) & 0 & 0 \\ 0 & 1 & 0 & 0 & (t_{k+1}-t_k) & 0 \\ 0 & 0 & 1 & 0 & 0 & (t_{k+1}-t_k) \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (6\text{-}50)$$

(4) The next calculation is equation (6-20) to obtain the state noise covariance matrix Q*. To evaluate equation (6-20), compute the product $\Phi(t_{k+1},t)G(t)Q(t)G^T(t)\Phi^T(t_{k+1},t)$. Recall that in defining the dynamic model, it was assumed $Q(t) = \sigma^2 I_3$. Furthermore, the matrix G may be partitioned as $\left[\begin{array}{c} 0_3 \\ \hline I_3 \end{array}\right]$,

where 0_3 is the 3x3 zero matrix. Then

$$G(t)Q(t)G^T(t) = \left[\begin{array}{c} 0_3 \\ \hline I_3 \end{array}\right] \left[\sigma^2 I_3 \right] \left[0_3 \mid I_3 \right]$$

$$= \left[\begin{array}{c|c} 0_3 & 0_3 \\ \hline 0_3 & \sigma^2 I_3 \end{array}\right] . \qquad (6\text{-}51)$$

Then, if $D = t_{k+1} - t$, then

$$\Phi(t_{k+1},t)G(t)Q(t)G^T(t) = \left[\begin{array}{c|c} I_3 & DI_3 \\ \hline 0_3 & I_3 \end{array}\right] \left[\begin{array}{c|c} 0_3 & 0_3 \\ \hline 0_3 & \sigma^2 I_3 \end{array}\right]$$

$$= \left[\begin{array}{c|c} 0_3 & \sigma^2 DI_3 \\ \hline 0_3 & \sigma^2 I_3 \end{array}\right] ,$$

and

$$\Phi(t_{k+1},t)G(t)Q(t)G^T(t)\Phi^T(t_{k+1},t) = \left[\begin{array}{c|c} 0_3 & \sigma^2 DI_3 \\ \hline 0_3 & \sigma^2 I_3 \end{array}\right] \left[\begin{array}{c|c} I_3 & 0_3 \\ \hline DI_3 & I_3 \end{array}\right]$$

$$= \left[\begin{array}{c|c} \sigma^2 D^2 I_3 & \sigma^2 DI_3 \\ \hline \sigma^2 DI_3 & \sigma^2 I_3 \end{array}\right] .$$

(6-52)

To evaluate equation (6-20), the integrals of the four blocks of the partitioned matrix (6-52) are calculated. For every nonzero element in the block $\sigma^2 D^2 I_3$,

$$\int_{t_k}^{t_{k+1}} \sigma^2 D^2 dt = \int_{t_k}^{t_{k+1}} \sigma^2 (t_{k+1} - t)^2 dt$$

$$= \sigma^2 \int_{t_k}^{t_{k+1}} (t_{k+1}^2 - 2t_{k+1}t + t^2) dt$$

$$= \sigma^2 (t_{k+1}^2 t - t_{k+1} t^2 + t^3/3) \Big|_{t_k}^{t_{k+1}}$$

$$= \sigma^2 (t_{k+1}^3 - t_{k+1}^2 t_k - t_{k+1}^3 + t_{k+1} t_k^2 + t_{k+1}^3/3 - t_k^3/3)$$

$$= \sigma^2(t_{k+1}^3/3 - t_{k+1}^2 t_k + t_{k+1}t_k^2 - t_k^3/3)$$

$$= \frac{\sigma^2}{3}(t_{k+1} - t_k)^3. \qquad (6\text{-}53)$$

Hence,

$$\int_{t_k}^{t_{k+1}} \sigma^2 D^2 I_3 \, dt = \frac{\sigma^2}{3}(t_{k+1} - t_k)^3 I_3. \qquad (6\text{-}54)$$

Similarly,

$$\int_{t_k}^{t_{k+1}} \sigma^2 D I_3 \, dt = \frac{\sigma^2}{2}(t_{k+1} - t_k)^2 I_3,$$

and

$$\int_{t_k}^{t_{k+1}} \sigma^2 I_3 \, dt = \sigma^2(t_{k+1} - t_k) I_3. \qquad (6\text{-}55)$$

Thus,

$$Q^*(t_{k+1}) = \sigma^2 \begin{bmatrix} \frac{(1/3)(t_{k+1}-t_k)^3 I_3}{(1/2)(t_{k+1}-t_k)^2 I_3} & \Big| & \frac{(1/2)(t_{k+1}-t_k)^2 I_3}{(t_{k+1}-t_k) I_3} \end{bmatrix}. \qquad (6\text{-}56)$$

(5) This step defines the measurement matrix for the measurements at time t_{k+1}. In processing the measurements individually, for each t_{k+1} there are N matrices $H_i(t_{k+1})$, i=1,..., N. Each H_i is defined by equation (6-21) as

$$H_i(t_{k+1}) = \left.\frac{\partial h_i}{\partial \underline{x}}\right|_{\underline{x} = \hat{\underline{x}}\,(t_{k+1}/t_k)}, \qquad (6\text{-}57)$$

where h_i is defined in equation (6-32) and $\hat{\underline{x}}\,(t_{k+1}/t_k)$ was obtained in step (1). The matrix H_i is defined by equation (6-57) to be 1x6. The first component of H_i is

$$\frac{\partial h_i}{\partial x(t)} = \frac{\partial\left(\sqrt{(x(t) - X_i)^2 + (y(t) - Y_i)^2 + (z(t) - Z_i)^2}\right)}{\partial x(t)}$$

$$= \frac{x(t) - X_i}{\sqrt{(x(t) - X_i)^2 + (y(t) - Y_i)^2 + (z(t) - Z_i)^2}}$$

which, evaluated at $\hat{\underline{x}}(t_{k+1}/t_k)$ is

$$\frac{(\hat{x}(t_{k+1}/t_k) - X_i)}{\sqrt{(\hat{x}(t_{k+1}/t_k) - X_i)^2 + (\hat{y}(t_{k+1}/t_k) - Y_i)^2 + (\hat{z}(t_{k+1}/t_k) - Z_i)^2}}. \quad (6\text{-}58)$$

Similar calculations can be made for y and z. Since h_i is not a function of vx, vy, or vz, then

$$\frac{\partial h_i}{\partial vx} = \frac{\partial h_i}{\partial vy} = \frac{\partial h_i}{\partial vz} = 0 \quad \text{for } i = 1,\ldots,N. \quad (6\text{-}59)$$

Hence, if r denotes the denominator of equation (6-58), the measurement matrix H_i, i = 1,..., N, is

$$H_i(t_{k+1}) = \left[\frac{(\hat{x}(t_{k+1}/t_k) - X_i)}{r} \quad \frac{(\hat{y}(t_{k+1}/t_k) - Y_i)}{r} \quad \frac{(\hat{z}(t_{k+1}/t_k) - Z_i)}{r} \quad 0 \; 0 \; 0\right] \quad (6\text{-}60)$$

Now all the values have been obtained that are needed to update the estimate of the position and velocity of the target at time t_{k+1}, given the estimate at time t_k, and the measurements at time t_{k+1}. Assuming that the initial values $\hat{x}(t_0)$ and P(0) are given, then the following sequence will give the Kalman filter for this example. This sequence can be used to write the filter program, in FORTRAN for example, with steps (1) through (5) as subroutines which are called at the appropriate times in the sequence.

First, extrapolate the state vector $\hat{\underline{x}}(t_k/t_k)$ using equation (6-42)in step (1). Also extrapolate the state covariance matrix $P(t_k/t_k)$ using equation (6-22). For this, $I(t_{k+1},t_k)$ is needed from step (3) and $Q^*(t_{k+1})$ from step (4).

The data at time t_{k+1} is input as an N-vector $\underline{y}(t_{k+1})$. However, process the measurement from each station individually. To do so, process them in sequence by considering each measurement y_i, i=1,...,N as a one-dimensional measurement vector. Then go from y_i to y_{i+1} in a similar manner as going from $\underline{y}(t_k)$ to $\underline{y}(t_{k+1})$. Once the calculations of all the y_i's have been completed, the measurement vector for time t_{k+1} will be known, which is designated as

$$\underline{y}(t_{k+1}) = (y_1, y_2, \ldots. y_N). \qquad (6\text{-}61)$$

Now, begin the sequence of measurement updates for time t_{k+1} with y_1.

The gain matrix is computed using equation (6-23). The matrix H_1 from step (5) is used here. Note then that the resulting $K(t_{k+1})$ will be the first of N gain matrices, denoted $K_i(t_{k+1})$, i = 1,...,N and that they are all 6x1.

The measurement update of equation (6-24) is then completed for the one-dimensional vector y_1, using y_1 as the $y(t_{k+1})$ of equation (6-24), and plugging $h_1(\hat{\underline{x}}(t_{k+1}/t_k))$ from equation (6-32) into the $\underline{h}(\hat{\underline{x}}(t_{k+1}/t_k),t_k)$ of equation (6-24). The output $\hat{\underline{x}}(t_{k+1}/t_{k+1})$ will be designated as $\hat{\underline{x}}_1(t_{k+1}/t_{k+1})$ to indicate its calculation from measurement y_1. The covariance matrix is also updated, using equation (6-25) and resulting in $P_1(t_{k+1}/t_{k+1})$.

Only the last three steps consisting of gain calculation, measurement update, and covariance update are repeated for measurements $y_2, y_3, \ldots, y_N$. No time update is needed because these measurements are all at the same time as y_1. Thus, for measurement y_2, $\hat{\underline{x}}(t_{k+1}/t_k)$ is replaced in equation (6-24) by $\hat{\underline{x}}(t_{k+1}/t_{k+1})$, and $P(t_{k+1}/t_k)$ is replaced in equations (6-23) and (6-25) by $P_1(t_{k+1}/t_{k+1})$, since $\hat{\underline{x}}_1$ and P_1 are the best estimates currently available. The output of equation (6-24) will then be $\hat{\underline{x}}_2(t_{k+1}/t_{k+1})$ and that of equation (6-25) will be $P_2(t_{k+1}/t_{k+1})$. These three steps are repeated until $\hat{\underline{x}}_N(t_{k+1}/t_{k+1})$ is obtained which is the estimate of the target's position and velocity at time t_{k+1} given all the measurements up to and including those at time t_{k+1}, and $P_N(t_{k+1}/t_{k+1})$ which is its corresponding covariance matrix. The updates have now been completed for the $(k+1)^{st}$ time point.

CHAPTER 7

FILTERS IN CURRENT RANGE USE

This chapter contains a description of some of the most frequently used filters. The filters sampled from the survey provide a good cross section of range data applications, and no attempt has been made to select a filter which is best for any application. There are seven least squares filter applications in use identified by the survey, and probably several more that were not identified. Discussion of several variations of the least squares filters is given in this chapter. Also discussed is the Quadratic Digital (QD) filter which has been in use for many years, and at the time of the survey, there were five QD users. The originator of the QD filter is Mr. W. A. McCool, White Sands Missile Range (WSMR). A section on the Digital Filter X (DFX) is included to provide support for the QD filter.

Appendix A contains a summary of filter information based on a survey of the Data Reduction and Computer Group of the Range Commanders Council. The appendix essentially contains all of the information received. In addition, the appendix provides areas of application used at the contributing ranges and observations about the suitability of the filters.

7.1 Least Squares Filters

There are numerous filters based on least squares. In this section, one basic and most widely used least squares filters is the least squares polynomial moving arc filter. Although the simple average filter is not usually recognized as a type of least squares filter, it is addressed here. The least squares methods described throughout the rest of the chapter are variations of the least squares polynomial moving arc filter.

7.1.1 Simple Moving Average Filters

Simple moving average filters give the average of the N most recent observations and are equivalent to fitting a zeroth degree polynomial evaluated at the midpoint. The formulation is quite elementary; it is

$$M_t = \frac{x_t + x_{t-1} + \cdots + x_{t-N+1}}{N} \qquad (7\text{-}1)$$

$$= M_{t-1} + \frac{x_t - x_{t-N}}{N}.$$

This filter is very simple and straightforward and is useful for observing a constant process. However, if the process is changing, a small value of N is needed for rapid response.

The following is an explanation of a simple moving average filter as a least squares filter. It is assumed that the average is a constant value (call it a) expressed in terms of the parameter in consideration (call it x). The problem starts out with the simple equation

$$x = a \tag{7-2}$$

The problem now is to find the value for a so that the N data points $x_{t-N+1},\ldots,x_t$ collectively deviate from equation (7-2) as little as possible (in the least squares sense). Then, to find the best least squares estimate for a over the last N values of x, the value

$$S = \sum_{i=t-N+1}^{t} (x_i-a)^2 \tag{7-3}$$

must be minimized.

Taking the partial derivative of S with respect to a,

$$\frac{\partial S}{\partial a} = 2 \sum_{i=t-N+1}^{t} (x_i - a) = 0 \tag{7-4}$$

It can be seen that S is minimized when

$$\sum_{i=1}^{t} x_i = na \tag{7-5}$$

from which it is found that the best least squares estimate for a is

$$\frac{\sum_{i=t-N+1}^{t} x_i}{N} \tag{7-6}$$

the average value of the parameter at the most recent N points.

7.1.2 Least Squares Polynomial Moving Arc Filters

Least squares polynomial moving arc filters are based on the assumption that the true function can be expressed over a finite

span of time by a polynomial of fairly low degree and that the errors in the measurements are random with zero mean and finite rms value and are serially uncorrelated. Noted here, the smaller the time intervals are, the less the degree the polynomial needs to be to provide a good estimating curve for the data. A sampling interval of .05 seconds is said to be sufficiently small for good data fitting. It is desired to solve for the coefficients of the polynomial which best fits the data in a least squares sense, that is, to minimize the sum

$$S = \sum_{i=1}^{n} (a_0 + a_1 t_i + a_2 t_i^2 + \ldots + a_d t_i^d - y_i)^2 \qquad (7\text{-}7)$$

where

a_k, k = 0, 1,...,d are the coefficients of the polynomial of degree d

t_i = time of each sample referenced to the midpoint of the span under consideration.

Y_i = sampled measurements

n = number of points in the span (constrained to be odd and greater than d).

To minimize S, take the partial derivatives of S with respect to each of the unknowns, a_k, and solve the resulting set of d+1 linear equations. In matrix notation, write this as

$$CA = B, \text{ where} \qquad (7\text{-}8)$$

$$C \text{ is composed of } C_{ij} = \sum_{k=1}^{n} t_k^{i+j-2}, \quad \begin{matrix} i=1, 2,\ldots,d+1, \\ j=1, 2,\ldots,d+1 \end{matrix}$$

$$A = (a_0, a_1,\ldots,a_d)^T, \text{ and}$$

$$B \text{ is a vector with elements } b_i = \sum_{k=1}^{n} y_k t_k^{i-1}, \; i=1, 2,\ldots,d+1.$$

Then $A = C^{-1}B$. If the polynomial $a_0 + a_1 t + a_2 t^2 + \ldots + a_d t^d$ represents position, then velocity and acceleration are found by taking the derivatives of the polynomial with respect to time.

The vector A of coefficients has been found for the first n time points. Normally, the polynomial with these coefficients is applied just to the midpoint of the time span. The first (n-1)/2 points are generally used just for finding the new data value for

the midpoint, but new values may be generated (using that polynomial) for the first (n-1)/2 points also. Once the new midpoint data value has been found, the next step is to redetermine the coefficients of the polynomial using points 2 to n+1, so a new data value for the new midpoint can be found. Repeat this process until finished with the polynomial moving from points m,...,m+n-1 to m+1,...,m+n with each step (where m ranges from 1 to the total number of time points minus n), thus describing a moving arc.

If the data are sampled at equally spaced intervals of time, the t can be considered to be unity and C^{-1} is uniquely determined when d and n are chosen. By use of C^{-1} and t_i, a set of weights can be computed for determining the a_k such that

$$a_k = \sum_{i=1}^{n} w_{k\,i}\, y_{N-n+1,} \text{ where} \qquad (7\text{-}9)$$

N is the subscript of the last point in the span.

These weights are useful for determining the expected error in position, velocity, and acceleration for data with known input error since

$$\frac{\sigma_{a_k}}{\sigma_y} = \left(\sum_{i=1}^{n} w_{ki}2 \right)^{1/2}. \qquad (7\text{-}10)$$

It is assumed that the standard deviations of all the measurements y_i are the same, and it is called σ_y. Thus the sampling interval can be chosen, as well as the number of points smoothing and the degree of the polynomial which will accomplish the desired results, with a minimum expense of computational effort (see reference 3). There are methods of calculating C^{-1} that do not entail finding the inverse directly, which would be very time consuming. One of these methods will be discussed later in this chapter.

An advantage of using a polynomial is that much need not be known about the true process, because if the polynomial is of high enough degree, it will tend to seek out the signal in the presence of noise and give a reasonable estimate of it. For this reason, the higher degree of the polynomial, the more flexibility is allowed in accurately constructing the true process. Another reason is by using least squares, the polynomial is not forced to pass exactly through any of the observations, which would result in a certain amount of smoothing.

7.2 Position and Velocity Constrained Least Squares Filters

Position and velocity constrained least squares filters are based on the philosophy that the best estimate of the true function and its derivatives can be obtained by fitting a polynomial to fixed time spans of the measurements by the method of least squares while imposing constraints on the polynomial which force continuity between successive spans. The assumptions concerning the data are the same as for least squares polynomial filters, but the additional requirement is made that the polynomial being fit to the filter span must pass through a point on the previous curve with the same slope. For example, consider fitting a quadratic to data sampled at evenly spaced intervals. Suppose there are five points in the span, that the midpoint of the span is at t=0, and that each point in the span is one unit apart from the one next to it. For the first span, the time values are t_i, i=1,...,5 where t_i= i-3. It would be desirable to have

$$\hat{y}_i = a_0 + a_1 t_i + a_2 t_i^2 \qquad (7\text{-}11)$$

evaluated at i=3 to be the filtered data value for the midpoint. Thus, the filtered data value is equated with a_0, since $t_3 = 0$. Once a_0, a_1, and a_2 for this span have been solved for using the regular least squares method, the first step is completed.

The next step is to solve for $\hat{y}_i$ at $t_i = 1$. Move the span one point over so that the new span is for points i=2,...,6. The new midpoint is at i=4. Remember there are two constraints: position and velocity. Note now that t_4 is the middle time value for the new span. The idea of the filter is that the polynomials of both spans have the same value (position) and the same slope (velocity) at time t_4.

For both the old and new spans, consider the middle time value to be 0 and the following time value to be 1. So for the old time span, $t_3 = 0$ and $t_4 = 1$. For the new time span, $t_3 = -1$, $t_4 = 0$, and $t_5 = 1$. Let

$$p(t_4) = a_0' + a_1' t_4 + a_2' t_4^2 \qquad (7\text{-}12)$$

be the polynomial for the midpoint of the new span, and let

$$q(t_4) = a_0 + a_1 t_4 + a_2 t_4^2 \qquad (7\text{-}13)$$

be the polynomial for the same point of the old span. Use the position constraint to obtain a_0' and the velocity constraint to obtain a_1'.

7.2.1 Position Constraint

Because $p(t_4)$ is for the midpoint of the new span, the method equates t_4 to 0, so that $p(t_4)$ takes on the value a_0' Because $q(t_4)$ is for the point to the right of the midpoint of the old span, the method equates t_4 to 1, so that $q(t_4)$ assumes the value $a_0 + a_1 + a_2$. To ensure that these two polynomials have the same value at t_4, equate a_0' with $a_0 + a_1 + a_2$. Now the value has been found for the first coefficient for the new span.

7.2.2 Velocity Constraint

Again, consider the polynomial $p(t_4)$ for the midpoint of the new span. The slope for this polynomial is $a_1' + 2a_2't_4$. Because this slope is for the midpoint of the new span, the method evaluates this slope at $t_4 = 0$, whence the slope takes on the value a_1'. Also consider the polynomial $q(t_4)$ for the same point of the old span. The slope for this polynomial is $a_1 + 2a_2t_4$. Because this slope is for the midpoint of the old span, the method evaluates the slope at t=1, whence the slope assumes the value $a_1 + 2a_2$. To ensure that the slopes of these two polynomials have the same value at t_4, we equate a_1 with $a_1 + 2a_2$. Now the value has been found for the second coefficient of the polynomial for the new span.

So from the position constraint,

$$a_0' = a_0 + a_1 + a_2 \tag{7-14}$$

and from the velocity constraint,

$$a_1' = a_1 + 2a_2. \tag{7-15}$$

Now proceed to find the value for a_2'.

Start by minimizing the sum

$$S = \sum_{i=N-n+1}^{N} (a_0'+a_1't+a_2't_i^2 - y_i)^2 \tag{7-16}$$

where

a_0', a_1', and a_2' are coefficients for the second span

n = number of points in the span (should be odd and greater than 2; in this case, n=5)

N = subscript of last point in the span (in this case, N=6)

t_i = time of each sample referenced to the midpoint of the span such that $t_i - t_{i-1} = 1$

y_i = sampled measurements.

Taking the partial derivative of S with respect to the unknown, a_2, equating it to zero, and solving for a_2' gives

$$a_2' = \frac{1}{\Sigma t_i^4} \left(\Sigma\, y_i t_i^2 - a_0' \Sigma\, t_i^2 - a_1' \Sigma\, t_i^3\right). \qquad (7\text{-}17)$$

Similarly to the first span, the filtered data value for the midpoint of the second span is

$$\hat{y}_i = a_0' + a_1' t_i + a_2' t_i^2 \qquad (7\text{-}18)$$

evaluated at i = 4. That is, equate the filtered data value with a_0, since $t_4 = 0$. The subsequent steps are done in the same manner as the second step. Sometimes it is desirable to constrain position only to be obtained from the previous fit. In this case both a_1' and a_2' in equation (7-16) are unknown, and the solution would involve two equations with two unknowns.

This method has a distinct computational advantage over unconstrained least squares, because there are fewer summations and simpler equations. It also produces smoother output because of its recursive nature. However, it is slower to react to a change in the input and under certain conditions results in oscillation in the output.

7.3 Orthogonal Polynomials

In smoothing data by the usual method of least squares, it is necessary to choose in advance the degree of polynomial which will be used to approximate the data. This choice is necessary because the coefficients found are dependent upon the degree of curve being fitted. Often, however, it is not known in advance what degree curve will best fit the data. In such a case, it is desirable to fit several polynomials, each time increasing the degree used, until it is seen that any further increases would not produce a significantly better fit. The computation of successive polynomials is greatly simplified by the use of the orthogonal polynomial procedure. This method determines the approximating polynomial in terms of another variable, so chosen that each coefficient found is independent of the others, making it possible to increase the degree of curve used without making it necessary to recompute the previously found coefficients.

This program is generally used to smooth position data. The degree of curve fitted is increased until an F-test indicates that additional coefficients of the polynomial would not be significantly different from zero. The smoothed positions are then differentiated to obtain velocities, and the velocities differentiated to obtain accelerations. The error estimates of the smoothed data and derivatives are computed in the form of standard deviations for each point. Coefficients of the original polynomial are derived in terms of the new polynomial. Orthogonal polynomials are discussed in reference 46. The method is given here. First, the term "orthogonal polynomials" is defined.

Two polynomials $P_i(x)$ and $P_j(x)$ (of degrees i and j) are orthogonal on a set of points $x_1,\ldots,x_n$ provided that

$$\sum_{k=1}^{n} P_i(x_k) P_j(x_k) = 0 \qquad \text{for } i \neq j. \tag{7-19}$$

Given a polynomial of order k, k orthogonal polynomials can be found whose linear combination expresses the original polynomial. The method of orthogonal polynomials follows. Assume that the polynomial

$$p(t_i) = a_0 + a_1 t_i + a_2 t_i^2 + \ldots + a_k t_i^k \tag{7-20}$$

expresses the true function. Reference 46 shows directly and by example that this polynomial may be re-expressed as

$$p(t_i) = b_0 P_{0,t_i} + b_1 P_{1,t_i} + \ldots + b_2 P_{2,t_i}$$

where $P_{0,t_i} = 1$

$$P_{1,t_i} = t_i$$

$$P_{j,t_i} = t_i P_{j-1,i} - \frac{Q_{j-1}}{Q_{j-2}} P_{j-2,i}$$

$$b_j = \frac{y_i P_{j,t_i}}{Q_j}$$

$$\text{where } Q_j = \sum_i P_{j\,t_i}^2 \tag{7-21}$$

The time variable t_i ranges from $-(N-1)/2$ to $(N-1)/2$ in steps of one. (The index i could have been used instead of t_i, but t is the more conventional variable for time.) Reference 46 shows that the P_{j,t_i}'s are orthogonal to each other.

Once the original polynomial of equation (7-20) has been re-expressed in terms of orthogonal polynomials (in equation (7-21)), p(t) can be expanded to be of one more degree through the Gram-Schmidt orthogonalization process,

where

$$\sum_{k=1}^{n} P_i(x_k)P_j(x_k) \text{ is the inner product and} \qquad (7\text{-}22)$$

where the polynomials $P_i(x)$ and $P_j(x)$ are vectors in the vector space of polynomials spanned by $1,x,x^2,\ldots,x^n$, where n is the degree of the higher degree polynomial. In this manner, the values of any of the previously obtained b_j's or $P_{j,t}$'s are not changed. Instead, another such term is merely added on.

At what degree should expanding stop? Reference 46 suggests a clever criterion for deciding when to stop. The variance V of the data can be expressed in the following two ways:

$$V = \sigma_a^2 = \frac{\sum_{i=-(N-1)/2}^{(N-1)/2} \left(\sum_{j=1}^{k} a_j t_i^j - y_i\right)^2}{N - 1 - k}$$

$$V = \sigma_b^2 = Q_j(b_j - B_j)^2 \qquad (7\text{-}23)$$

where b_j is the computed coefficient and B_j is the true coefficient. Reference 46 gives the derivations for the last two equations. Of course, $\sigma_a^2 = \sigma_b^2$, but two different notations are given for the sake of future reference. The object now is to test the hypothesis that the true value of the jth coefficient B_j is zero. Test the hypothesis using an F-test. (The F-test tests the proportion of two variances.) Test the proportion,

$$\frac{\sigma_b^2 \Big|_{B_j 0}}{\sigma_a^2} \qquad (7\text{-}24)$$

Of course, the numerator $\sigma_b^2\Big|_{B_j=0} = Q_j b_j^2$. If B_j is ideally 0, then this proportion should be close to 1.

In the test, if

$$\frac{\sigma_b^2\Big|_{B_j=0}}{\sigma_a^2} \leq F, \qquad (7\text{-}25)$$

then the hypothesis is accepted as true, and b_j is set equal to zero. The next coefficient is tested in the same manner. When two consecutive coefficients are set equal to zero, then the degree k of the polynomial is determined from the last coefficient which was not set equal to zero; that is, k=j-2.

Stopping after two consecutive zero coefficients assumes that succeeding terms in the polynomial will carry negligible weight. For practical purposes, most parameters can be modeled using a lower-order polynomial. If the polynomial is of higher degree, the higher coefficients are close to zero. Normally, data generally does not shoot up or down suddenly enough to warrant higher degree, high- magnitude coefficients.

7.4 Least Squares Polynomial Moving Arc Filters Using Recursive Sums

A recursive scheme for computing the sums required to perform least squares polynomial moving arc smoothing has been developed at the Air Force Development Test Center which permits this type of smoothing to be accomplished considerably faster than by the usual linear combinations of observations. This is not a recursive filter in the usual sense in that previous filter output is not used in determining current output.

In paragraph 7.1 it was stated that the equation for this type of smoothing may be written in matrix notation as

$$A = C^{-1}B, \qquad (7\text{-}26)$$

where the elements of A, B, and C have been previously defined. Each element of B and C consists of a summation of n terms (n = number of points smoothing) which must be performed each time a point is dropped and another is picked up in the moving arc

process. However, these summations can be avoided by computing the sums for a given span in terms of those for the previous span.

For instance, suppose it was found that through the method discussed in paragraph 7.1

$$b_i^{(1)} = \sum_{k=1}^{n} y_k \, t_k^{i-1}, \qquad t_k = T_k - T_{\frac{n+1}{2}}, \tag{7-27}$$

where T_i is the ith time value. So t_k is the length of the interval between the kth time point and the midpoint of the span. To find

$$b_i^{(2)} = \sum_{k=2}^{n+1} y_k \, (t_k - \Delta)^{i-1}, \text{ where } \Delta = T_{\frac{n+3}{2}} - T_{\frac{n+1}{2}}, \tag{7-28}$$

do so in terms of $b_i^{(1)}$ by means of the recursion

$$b_i^{(2)} = \sum_{j=0}^{i-1} (-\Delta)^j \binom{i-1}{j} b_{i-j}^{(1)} + y_{n+1} \sum_{j=0}^{i-1} (-\Delta)^j \binom{i-1}{j} t_{n+1}^{i-j-1} \tag{7-29}$$

$$- y_1 \sum_{j=0}^{i-1} (-\Delta)^j \binom{i-1}{j} t_1^{i-j-1} .$$

The elements of C are computed similarly by considering the $y_k=1$. However, as previously stated, if the data are evenly spaced, C is uniquely determined when the degree of the polynomial and number of points smoothing have been selected. In this case, C^{-1} can be precomputed and stored for use in the smoothing. In addition, Δ can be considered unity, and the equations for the recursion are simplified.

With this formulation no lengthy summations are required and the number of operations, once initiated, is independent of the number of points used in the smoothing. Furthermore, by computing recursively $\sum . y_k^2$, reference 63 shows that with very little extra effort the residual sum of squares of each span can be obtained as an estimate of the error in the data by

$$s = y_k^2 - A^T B. \tag{7-30}$$

7.5 Derivative Information Recovery by a Selective Integration Technique (DIRSIT)

The data smoothing technique known as Derivative Information Recovery by a Selective Integration Technique (DIRSIT) was developed in 1962 at White Sands Missile Range. The technique has been modified somewhat to permit more control of interval program parameters to satisfy the requirements of a greater number of users. The basic steps in the DIRSIT smoothing philosophy are

(1) the sign of the difference between the raw and smoothed position data is examined at each point within the filter span;

(2) the smoothed position data is satisfactory if the number of positive signs differs from the number of negative signs by one; (The total number of signs is odd.) and,

(3) if (2) above is not satisfied, the acceleration history within the filter span is modified as much as necessary to force the difference to one.

It is assumed that sampled position data is being entered with associated time. The output is smoothed position, velocity, and acceleration with associated time. For the filter span 2m+1, which is odd, table 7-1 shows a starting condition.

TABLE 7-1. A STARTING CONDITION FOR A DIRSIT FILTER WHOSE SPAN IS 2m+1 POINTS

Raw Position	Smoothed Position	Velocity	Acceleration
X_1	$\bar{X}_1$	$\dot{X}_1$	$\ddot{X}_1$
X_2			$\ddot{X}_2$
.			.
.			.
.			.
$Xm+1$			$\ddot{X}m+1$
.			.
.			.
.			.
X_{2m+1}			$\ddot{X}_{2m+1}$

The parameters $\bar{x}_1$, $\dot{x}_1$, x_i, i=1,...,2m+1 are generated during initialization; consequently, the initial estimates are obtained independently. Columns $\bar{x}_i$ and x_i are computed as

$$\bar{x}_{i+1} = \bar{x}_i + \Delta t_{i+1}\dot{x}_i + (1/2)\,\Delta\, t_i^2\ddot{x}_1 \qquad (7\text{-}31)$$

$$\dot{x}_{i+1} = \dot{x}_i + \Delta t_{i+1}\ddot{x}_1 \qquad (7\text{-}32)$$

where i=1,...,2m+1; $\Delta t_{i+1} = t_{i+1} - t_i$

All accelerations $\ddot{x}_i$,i=m+1,...,2m+1 are equal.

Having completed the columns of table 7-1, examine how closely the smoothed and raw positions agree. Generate m values of xi by computing $\Delta x_i = x_i - x_i$, i=m+1,...,2m+1. Hence, the number of negative Δx_i and the number of nonnegative Δx_i cannot both be odd, neither can they both be even. Let NL be the number of Δx_i which are negative; that is, the $\bar{x}_i$ is larger than the x_i. Let NS be the number of Δx_i which are nonnegative. NL + NS = m and the Δx_i (i=m+2,...,2m+1) are acceptable if $|\,NL-NS\,| \geq 3$. These conditions are corrected by changes in the $\ddot{x}_i$, i=m+1,...,2m+1, which will also produce changes in the $\dot{x}_i$ and $\bar{x}_i$, i=m+2,...,2m+1.

Examine the procedure for changing the $\ddot{x}_i$. To be specific, assume that NS $\geq$ NL + 3; then the x_i must be increased. Starting at some time t, the acceleration will be increased with slope 1, to time t_{m+1} which will be called t_p, the pivot point. Then for times t_p, $t_{m+2},\ldots,t_{2m+1}$ the acceleration increase will be constant. This type of increase in acceleration will just result in a change in only one $\bar{x}_k$ so that Δx_k would just barely become negative where formerly it was nonnegative. The first step is to equate Δx_k as shown in the following equation:

$$\Delta x_k = (1/6)(t_p-\bar{t})^3 + (1/2)(t_p-\bar{t})^2(t_k-t_p) + (1/2)(t_p-\bar{t})(t_k-t_p)^2 \qquad (7\text{-}33)$$

Note that the change in distance is obtained by a double integration of acceleration. First integration of figure 7-1 would yield a velocity curve of the shape shown in figure 7-2.

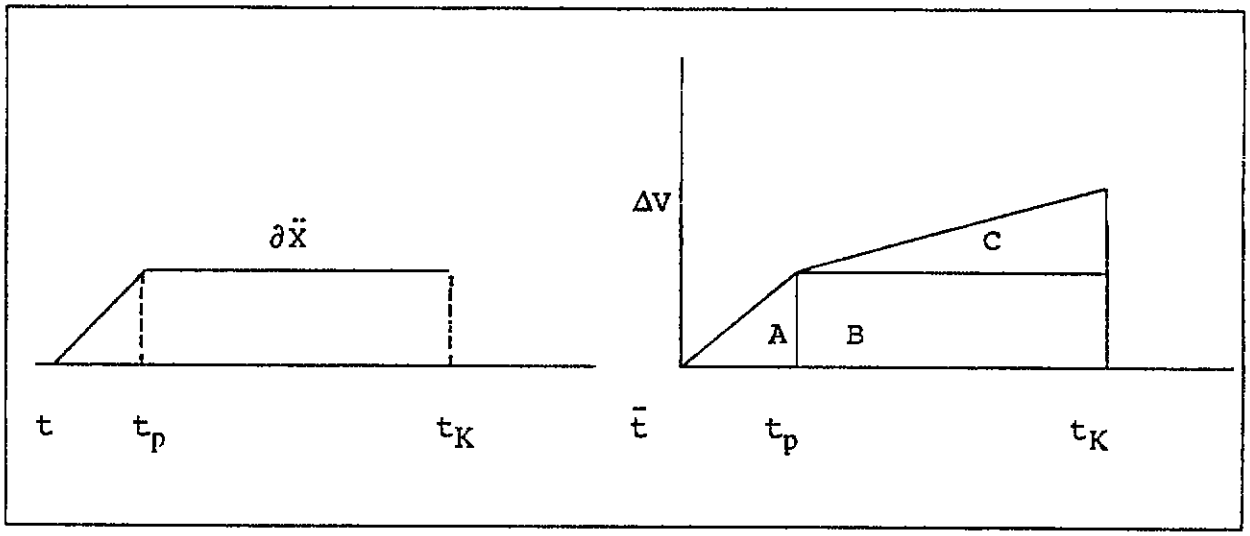


Figure 7-1. Example acceleration curve.

Figure 7-2. Corresponding velocity curve.

Where Area A = $(1/6)(t_p-\bar{t})^3$, Area B = $(1/2)(t_p-\bar{t})^2(t_k-t_p)$,

and Area C = $(1/2)(t_p-\bar{t})(t_k-t_p)^2$ (7-34)

$$\Delta x_k = (1/6)\left[(t_p-\bar{t})^3 + 3(t_p-\bar{t})^2(t_k-t_p) + 3(t_p-\bar{t})(t_k-t_p)^2\right] \quad (7\text{-}35)$$

$$= (1/6)\left[(t_k-\bar{t})^3 - (t_k-t_p)^3\right]$$

from which $(t_p-\bar{t}) = [6\Delta x_k \div (t_k-t_p)^3]^{1/3} - (t_k-t_p)$ (7-36)

Thus, equation (7-36) gives the distance between t- and t_p. This change made in x_i and the changed x_k would decrease NS by 1 and increase NL by 1. For each positive Δx_j, compute a t_i but only the largest $\bar{t}_1$ is used. Thus only one Δx_i would change sign. For this selected value of $\bar{t}$, the changed $\ddot{x}_i$, $\dot{x}_i$, and x_i are calculated for each time beyond $\bar{t}$. A check is made again: if NL-NS$\leq$1, then the output at t, of $\bar{x}$, $\dot{x}$, $\ddot{x}$, is the output of DIRSIT; if NL-NS>1, the above computation of $\bar{t}$ is repeated.

Suppose that NL $\geq$ NS+3 or that the number of negative Δx_j is three or more greater than the number of nonnegative Δx_j. Under these conditions equation (7-36) becomes

$$(t_p-\bar{t}) = [-6\Delta x_k \div (t_k-t_p)]^{1/3} - (t_k-t_p) \quad (7\text{-}37)$$

To this point it has been assumed that $\bar{t} \leq t_1$, when $\bar{t} > t_1$ another procedure is employed. Increase the slope of the acceleration as shown in figure 7-3.

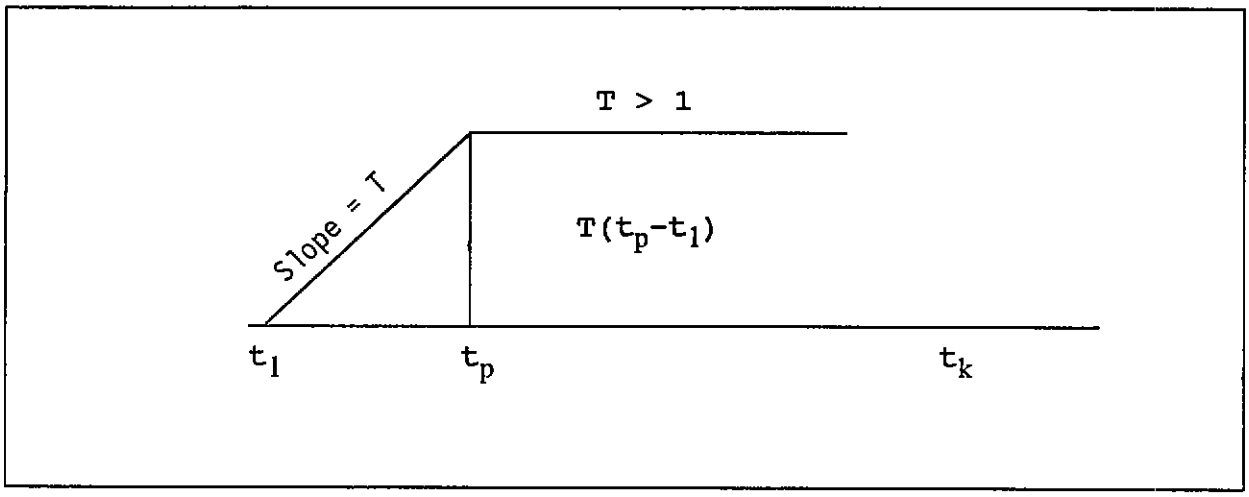


Figure 7-3. Example acceleration curve for $t > t_1$

$$T = \frac{6\Delta^x k}{(t_k - t_1)^3 - (t_k - t_p)^3} \qquad (7\text{-}38)$$

With this new value of T the new $\ddot{\bar{x}}_i$, $\dot{\bar{x}}_i$, $\bar{x}_i$, i=2,...,2m+1 are calculated. If | NL-NS | =1, the data output is finished; if not, repeat calculation of t. Now redefine terms by shifting their previous values down one, that is, shift

$$\begin{aligned} t_{i+1} &\longrightarrow t_i \\ x_{i+1} &\longrightarrow x_i \\ \bar{x}_{i+1} &\longrightarrow \bar{x}_i \\ \dot{\bar{x}}_{i+1} &\longrightarrow \dot{\bar{x}}_i \\ \ddot{\bar{x}}_{i+1} &\longrightarrow \ddot{\bar{x}}_i \end{aligned} \qquad i=2,\ldots,m \qquad (7\text{-}39)$$

The next data point is now input to the filter and the above process is repeated.

The DIRSIT was the original real-time flight filter at WSMR and has been successfully used with data from such missile systems as the Nike Zeus, ARPAT, Pershing, and Athena. The filter has been used for Nike Zeus and ARPAT elevation data, to

generate acquisition data, and to initialize a re-entry simulation in the Athena's real-time operational program. It is the major filter being used at Holloman Air Force Base and WSMR to evaulate the Athena re-entry data and is the prime source of post-flight data for the flight safety impact prediction programs at WSMR. The post-flight data is then used for vectoring recovery aircraft.

Because DIRSIT, a self-editing filter, uses a median-smoothing criterion, random spikes will have no effect on the filter output. Operation is not dependent upon equal time increments between data points, and a missing or repeated data point will not cause problems. With existing subroutines, numerous internal parameters can be controlled to obtain the desired filter responses. However, DIRSIT was not without its problems. Position, velocity, acceleration, and time arrays must be stored, thus requiring a significant amount of storage space and computing time. In addition, the filter has a lag which makes it undesirable for real-time use. (For additional information on DIRST, see references 36 through 39).

7.6 Digital Filter X (DFX)

In 1963 at WSMR, real-time flight safety support was first provided from the central computer facility. Originally, the DIRSIT filter was used, but it became evident that DIRSIT was not a satisfactory real-time filter because of the excessive storage requirements, computing time, and inherent lag. Consequently, the original version of the DFX filter was developed. The DFX is a variation of a constrained second degree filter which defines "best fit" criterion to be that the algebraic sum of the differences between the raw and smoothed data is zero.

The filter is more efficient in both speed and storage requirements because of a development which avoids computing the constraints explicitly and which stores only the array of differences rather than arrays of both raw and smoothed data. Additionally, a mean rather than a median smoothing criterion is used. The filter response is controlled by a set of input parameters which, in turn, determine the shape of the acceleration correction curve. The smoothing criterion for the DFX filter is

$$S = \sum_{i=1}^{N} (\bar{X}_i - X_i) = 0 \qquad (7\text{-}40)$$

which leads to

$$S = \sum_{i=1}^{N} (\bar{X}_0 + \dot{X}_0 (i\,\Delta t) + \ddot{X}_0 \frac{(i^2\,\Delta t^2)}{2} - X_i) = 0 \qquad (7\text{-}41)$$

With the constraints $\bar{X}_0$ and $\dot{X}_0$ fixed, only the acceleration has to be adjusted with each step

$$\ddot{X}_0 = \frac{1}{\frac{\Delta t^2 \sum i^2}{2}} \sum_{i=1}^{N} X_i - N\bar{X}_0 - \dot{X}_0 \Delta t \sum_{i=1}^{N} i \qquad (7\text{-}42)$$

Equation (7-41) differs from the constrained least squares (CLS) equations only in the multiplying factor $\frac{(i^2 \Delta t^2)}{2}$, and this difference will affect only the computation of the correction coefficients used by the filter. The method of applying the correction to the acceleration terms is the same for both a CLS and DFX filter.

To see how this method is developed, first look at the expression for $\hat{X}_{i+1}$ in terms of $\hat{X}_1$ where $\hat{x}_k$ is the new data value of x_k after smoothing:

$$\hat{X}_{i+1} = \hat{X}_1 + \hat{\dot{X}}_1 (i \Delta t) + \hat{\ddot{X}}_1 \frac{(i^2 \Delta t^2)}{2} \qquad (7\text{-}43)$$

(If x_{i+1} were expressed in terms of x_i, Δt would be the step size. In this case, consider $i \Delta t$ to be the step size.) Redefine terms by shifting down the values for the next cycle, so that $\hat{x}_{j+1}$ of the previous step becomes what will be designated as $\hat{x}_j$ in the current step, where the bar indicates "prior to applying the correction factor."

This redefining of terms gives the following equation:

$$\bar{X}_i = \hat{X}_0 + \hat{\dot{X}}_0 (i \Delta t) + \hat{\ddot{X}}_0 \frac{(i^2 \Delta t^2)}{2} \qquad (7\text{-}44)$$

Compare the previous equation (7-44) with the following equation (7-45), which expresses the ith smoothed point based on the constraints.

$$\hat{X}_i = \hat{X}_0 + \hat{\dot{X}}_0 (i \Delta t) + \hat{\ddot{X}} \frac{(i^2 \Delta t^2)}{2} \qquad (7\text{-}45)$$

Subtracting equation (7-44) from equation (7-45)

$$(\hat{X}_i - \bar{X}_i) = \frac{(i^2 \Delta t^2)}{2} (\hat{\ddot{X}}_0 - \hat{\ddot{X}}_0) \qquad (7\text{-}46)$$

which will be written as

$$(\hat{X}_i - \bar{X}_i) = \frac{(i^2\ \Delta t^2)}{2}\ (\delta\ddot{X}_0) \qquad (7\text{-}47)$$

or

$$\hat{X}_i = \bar{X}_i + \frac{(i^2\ \Delta t^2)}{2}\ \delta\ddot{X}_0 \qquad (7\text{-}48)$$

That is, the smoothed position at the ith point equals the position before correction plus a constant factor depending only on the delta time increment and the position of the ith point in the N point span times the correction to the acceleration. The significance of this fact is that if there are predicted outputs and a means of computing the acceleration correction, then there is no need to compute the constraints explicitly or save arrays of both raw and smooth positions. Note that equation (7-48) can also be written

$$(\hat{X}_i - X_i) = (\bar{X}_i - X_i) + \frac{(i^2\ \Delta t^2)}{2}\ \delta\ddot{X}_0 \qquad (7\text{-}49)$$

or

$$\Delta X_i = \Delta\bar{X}_i + \frac{(i^2\ \Delta t^2)}{2}\ \delta\ddot{X}_0 \qquad (7\text{-}50)$$

The quantity $\sum_{i=1}^{N} \Delta X_i$ is set to be zero as the smoothing criterion.

The array ΔX_i can be used instead of the two arrays $\hat{X}_i$ and X_i for computing each cycle. The ΔX_i's, which are the differences between predicted and raw input at each time are saved, shifted down, and corrected after each succeeding computing cycle. As each ΔX_i is operated on in successive computing cycles, it becomes numerically smaller. Most of the correction is applied to the ΔX_i at the cycle when i=N, the real time span. This factor, it will be seen, is what makes the QD filter possible. Performance is almost identical to second order CLS smoothing; however DFX requires much less computation time and storage regardless of the span size. In addition, filter characteristics can be changed dynamically during flight to improve performance.

The basic DFX package includes editing, initialization, radar selection, and noise estimate routines. Versions of this filter were used in all WSMR real-time programs for over five years. References 40 and 41 are concerned with the DFX filter.

7.7 Quadratic Digital (QD) Filter

The DFX equations can be derived using the CLS smoothing criterion. If an approximation is made concerning the effects of the terms of the sums of products, the necessity of even one stored array is eliminated and the second order DFX formulation can be reduced to a set of six equations. The performance of the resulting (QD) filter is practically identical to CLS performance, although computing time is independent of the filter span. In reality, the QD filter span is implied rather than actual because no arrays are stored. The QD filter is essentially a DFX filter which applies all the corrections needed to the real-time point only.

The QD formulas employ a predict-correct procedure using the estimates determined for the current given data value evaluated in the preceding step. They are, therefore, said to be recursive. The prediction formulas are

$$\bar{\ddot{X}}_{n+1} = \hat{\ddot{X}}_n \tag{7-51}$$

$$\bar{\dot{X}}_{n+1} = \hat{\dot{X}}_n + \Delta t\, \hat{\ddot{X}}_n, \tag{7-52}$$

$$\bar{X}_{n+1} = \hat{X}_n + t\, \hat{\dot{X}}_n + \frac{\Delta t^2}{2}\, \hat{\ddot{X}}_n, \tag{7-53}$$

or

$$\bar{X}_{n+1} = \hat{X}_n + \frac{\Delta t}{2}\, (\bar{\dot{X}}_{n+1} + \hat{\dot{X}}_n) \tag{7-54}$$

where

$\bar{\ddot{X}}_{n+1}$ - is predicted second derivative,

$\hat{\ddot{X}}_n$ - second derivative estimate from preceding step,

$\bar{\dot{X}}_{n+1}$ - predicted first derivative,

$\hat{\dot{X}}_n$ - first derivative estimate from preceding step,

$\bar{X}_{n+1}$ - predicted data value,

$\hat{X}_n$ - data value estimate from preceding step, and

Δt - time differential from preceding step.

The corrections of the predicted values are based upon the differences between the given data and the predicted data values. The correction formulas are

$$\hat{\ddot{X}}_{n+1} = \bar{\ddot{X}}_{n+1} + K_1 \Delta X, \tag{7-55}$$

$$\hat{\dot{X}}_{n+1} = \bar{\dot{X}}_{n+1} + K_2 \Delta X, \tag{7-56}$$

$$\hat{X}_{n+1} = \bar{X}_{n+1} + K_3 \Delta X, \tag{7-57}$$

where

$\hat{\ddot{X}}_{n+1}$ - correction of the predicted second derivative,

$\hat{\dot{X}}_{n+1}$ - correction of the predicted first derivative,

$\hat{X}_{n+1}$ - correction of the predicted data value, and

K_1, K_2, K_3 - correction coefficients which minimize the error of the estimates:

ΔX - difference between given and predicted data values.

The functional relationship among the correction coefficients K_1, K_2, and K_3 are revealed through study of predicted values obtained by use of the second order CLS filter. In the CLS filter, intercept and slope constraints are applied at the oldest value end of a given data span M, so that the polynominal fit to the M data values in the least squares sense must also contain the estimates of the true value and the corresponding first derivative at the oldest span point. Using these constraints in the second order CLS filter, the predicted data value and its derivative can be obtained with the truncated Taylor series:

$$\bar{\dot{X}}_{n+1} = \hat{\dot{X}}_{n-M+1} + M\Delta t \; \hat{\ddot{X}}_{n-M+1}, \tag{7-58}$$

$$\bar{X}_{n+1} = \hat{X}_{n-M+1} + M\Delta t \; \hat{\dot{X}}_{n-M+1} + \frac{(M \Delta t)^2}{2} \; \hat{\ddot{X}}_{n-M+1}. \tag{7-59}$$

where n-M+1 indicates the oldest value in the filter span, having been shifted from its position in the preceding step of n-M+2. Since $\hat{\ddot{X}}_{n-M+1}$ and $\hat{\dddot{X}}_{n-M+1}$ are fixed (because the position and velocity are constrained), only $\hat{\ddot{X}}_{n-M+1}$ can be adjusted to make equation (7-59) fit the M data values in the least squares sense. Let $\Delta\ddot{X}$ be the required correction. Then

$$\hat{\ddot{X}}_{n+1} = \hat{\ddot{X}}_{n-M+1} + \Delta\ddot{X} \tag{7-60}$$

Note that the second derivative is constant across the second order CLS filter span. Applying the correction to equations (7-58) and (7-59) holding x_{n-M+1} and $\dot{x}_{n-M+1}$ fixed, gives the estimates

$$\hat{\dot{X}}_{n+1} = \hat{\dot{X}}_{n-M+1} + M\,\Delta t(\hat{\ddot{X}}_{n-M+1} + \Delta\ddot{X}) \tag{7-61}$$

$$\hat{X}_{n+1} = \hat{X}_{n-M+1} + M\,\Delta t\hat{\dot{X}}_{n-M+1} + \frac{(M\,\Delta t)^2}{2}\,(\ddot{X}_{n-M+1} + \Delta\ddot{X}) \tag{7-62}$$

Subtracting equations (7-58) and (7-59) from equations (7-61) and (7-62) gives

$$\hat{\dot{X}}_{n+1} = \bar{\dot{X}}_{n+1} + M\,\Delta t\,\Delta\ddot{X}, \tag{7-63}$$

$$\hat{X}_{n+1} = \bar{X}_{n+1} + \frac{(M\,\Delta t)^2}{2}\,\Delta\ddot{X}. \tag{7-64}$$

Comparison of equations (7-60), (7-63), and (7-64) with equations (7-55), (7-56), and (7-57) reveals the relationship between K_1, K_2, and K_3 as desired. From equations (7-55) and (7-60)

$$\Delta\ddot{X} = K_1\Delta X \tag{7-65}$$

Then from equations (7-63), (7-56), and (7-65),

$$K_2 = M\Delta tK_1 \tag{7-66}$$

and from equations (7-64), (7-57), and (7-65),

$$K_3 = \frac{(M\Delta t)^2}{2}\,K_1 \tag{7-67}$$

In the QD theory, K_1 is developed as a function of the corresponding CLS filter span M, that is,

$$K_1 = K_1(M) = \frac{60M^2}{10M^3 + 33M^2 + 23M-6} \tag{7-68}$$

(See reference 45 for the derivation of this formula.) Thus, when M is arbitrarily specified, the QD correction coefficients are determined by equations (7-66), (7-67), and (7-68).

In its basic formulation, QD can be used as a real-time filter, because the argument (or subscript) of the estimates corresponds to the latest value accepted by the filter. On the other hand, estimates can be obtained for the data value at the oldest end of the span which are significantly better than the corresponding real-time estimates. Such estimates, which are often called "smoothed" data, correspond to the constraints computed in the CLS filter. Smoothed estimates are simply obtained with a Taylor series expanded about the real-time estimates, that is,

$$\hat{\hat{\hat{X}}}_{n+1} = \hat{\hat{X}}_{n+1} - (M-1)\ t\hat{\ddot{X}}_{n+1} \tag{7-69}$$

$$\hat{\hat{X}}_{n+1} = \hat{X}_{n+1} - \frac{(M-1)}{2}\ t\ (\hat{\hat{\hat{X}}}_{n+1} + \hat{\hat{X}}_{n+1}) \tag{7-70}$$

for the second order QD filter and

$$\hat{\hat{\ddot{X}}}_{n+1} = \hat{\ddot{X}}_{n+1} - (M-1)\ \Delta t\hat{\ddot{X}}_{n+1} \tag{7-71}$$

$$\hat{\hat{\dot{X}}}_{n+1} = \hat{\dot{X}}_{n+1} - \frac{(M-1)}{2}\ \Delta t\ (\hat{\hat{\ddot{X}}}_{n+1} + \hat{\ddot{X}}_{n+1}) \tag{7-72}$$

$$\hat{\hat{X}}_{n+1} = \hat{X}_{n+1} - \frac{(M-1)\Delta t}{2}\ (\hat{\hat{\dot{X}}}_{n+1} + \hat{\dot{X}}_{n+1}) - \frac{(M-1)^3}{12}\ (\Delta t)^3\ \hat{\ddot{X}}_{n+1} \tag{7-73}$$

for the third order QD filter.

The QD filter achieves smoothing effectiveness almost identical to that of the comparable constrained least squares filter without the usual array of saved input data samples. The QD is the real-time filter used in the majority of WSMR operational programs. References 42 through 45 discuss the QD filter in more detail.

APPENDIX A

1. Name or Acronym

Advanced Medium Range Air-to-Air Missile (AMRAAM) Parameter Estimation Program (APEP)

2. Contact

Mr. John E. Lindegren
3200 SPTW/KRTAR
Eglin AFB, FL 32542
(904) 882-4267
DSN: 872-4267

3. Documentation Sources

a. Bierman, G. J. Factorization Methods for Discrete Sequential Estimation. New York: Academic P, 1977.

b. Luenburger, D. G. Optimization by Vector Space Methods. New York: Wiley, 1969.

c. Sorenson, H. W. Kalman Filtering Techniques, Advances in Control Systems. Vol. 3, C. T. Leondes, ed. New York: Academic P, 1966.

4. Origin

Ball Systems Engineering (formerly VERAC, Incorporated)
10975 Torreyana Road, Suite 300
San Diego, CA 92121
14 February 1982

5. Comments

Strengths of the APEP filter are described in the following subparagraphs:

a. This filter provides optimally smoothed trajectory estimates in the same sense as an extended Kalman filter but with superior numerical accuracy because of echelonized square root information matrices with Householder orthogonal transformations of augmented state transition matrices. In addition, the matrices are partitioned for dynamic and measurement bias states and make explicit use of the block diagonal matrix structure to minimize computation steps for the sparse matrices.

b. The forward filter cycle is slightly slower than the standard Kalman filter, but the smoother cycle (backward filter cycle) is significantly faster than a standard fixed interval Kalman smoother. Smoothed estimates require only backward substitution, and smoothed covariances are generated with a UDU^T decomposition of the filter state covariances. Propagation of the smoothed square root information matrices does not require matrix inversions.

c. The procedure for "folding in" new measurements using Householder transformations to echelonize the measurement augmented state equations allows inclusion of asynchronous or irregular measurement data rates without reinitializing any of the filter solutions.

d. Trajectory propagation using the current filter estimates and state transition matrices avoids dependency on any particular measurement input.

e. All kinematic parameters are provided with 1-sigma error bounds for the estimates, which require propagation of the error covariances for indirect parameter derivations.

f. Data dropouts of short duration or changes in the number of available instruments are accommodated by APEP without catastrophic filter/smoother estimate errors or failures.

g. A review of filter cycle measurement residuals allows identification of wild points for edit and adjustment of a priori measurement covariances used in program input. (Automatic edit may be selected for a multiple of the expected residual standard deviation.) Deviations from the measurements error model used by the software are indicated by correlated measurement residuals.

Weaknesses of the APEP are described in the following subparagraphs:

a. If no inertial navigation system (INS) data are available for measurement input, the square root inverse filter (SRIF) cycle may have to be repeated to fine tune the a priori filter measurement and state noise covariances or to edit wild points which seriously perturb the estimates in the state propagation steps.

b. The smoothing cycle requires reversal of read order for the entire measurement interval, thus requiring all measurement data to be stored on disk files or other random access storage media. The algorithm uses fixed interval smoothing. The volume of data that can be processed at one continuous time interval is limited.

c. Data-processing time is significantly longer than the data-collection time interval if Time-Space-Position-Information (TSPI) from several instruments and from the inertial data are included as measurement inputs to the filter.

d. The present error model for TSPI instruments include only bias terms as the systematic error source. (The error model does provide for the Markov noise model.) For the short time intervals associated with air-to-air missile flights, bias terms may provide an adequate model. For longer time intervals, with large shifts in target range or angle relative to TSPI instruments, a more extensive error model would be required to ensure stationarity for the measurement residual.

FAST LEAST SQUARES

1. Name or Acronym

Least Squares Moving Arc Polynomial Using Recursive Sums

2. Contact

Mr. John E. Lindegren
3200 SPTW/KRTAR
Eglin AFB, FL 32542
(904) 882-4267
DSN: 872-4267

3. Documentation Sources

Sterrett, John K. "Manual for Moving Polynomial Arc Smoothing." Ballistic Research Laboratory Report 840. Aberdeen Proving Ground: Ballistic Research Laboratory, 1952.

4. Origin

Ms. Martha D. Everett
AFDTC/KRBA
Eglin AFB, FL 32542

5. Comments

The filter accomplishes its designed tasks very well. By changing the time span or the polynomial degree or both, the filter can be adapted to the trajectory, and by changing a single time constant, the polynomial can be evaluated at the end point for post-mission processing. No knowledge of measurement errors is required, and the wild point edit is adaptive because the distance off the curve considered defective is proportional to the noise in the data for the particular time span.

This filter has one weakness. It cannot handle a step function properly and thus filter gives poor results at missile booster ignition, burnout, or similar occurrences.

WEIGHTED LEAST SQUARES

1. **Name or Acronym**

   Weighted Least Squares Smoother

2. **Contact**

   Mr. Mike Dodgen
   Computer Sciences Corporation
   6545 Test Group/ENAC
   Hill AFB, UT 84056
   (801) 777-6497
   DSN: 458-6497

3. **Documentation Sources**

   Mr. Mike Dodgen
   Computer Sciences Corporation
   6545 Test Group/ENAC
   Hill AFB, UT 84056

4. **Origin**

   Kentron International, Incorporated

WLSRE in MITO26

1. Name or Acronym

Weighted Least Squares Recursive Estimation (WLSRE) contained in Multiple Instrument Trajectory Module (MITO26)

2. Contact

Mr. Robert Fierro
U.S. Army White Sands Missile Range
Attn: STEWS-NR-AM
White Sands Missile Range, NM 88002
(505) 678-2543
DSN: 258-2543

3. Documentation Sources

a. Fierro, Robert and Carolyn Nicholson. "Dynamic Optimized Smoothing Span (DYNOS)." Analysis and Computation Directorate Document. White Sands Missile Range: Analysis and Computation Directorate, 1970.

b. Comstock, D., M. Wright, and V. Tipton. "Handbook of Data Reduction Methods." Data Reduction Division Tech Rept. White Sands Missile Range: Data Reduction Division, 1964

c. Greene, Earl. "Edfil, A Routine for Editing and Filtering Data." Analysis and Computation Directorate Document. White Sands Missile Range: Analysis and Computation Directorate, 1977.

d. Comstock, D. "Introduction to Least Squares." Analysis and Computation Directorate Document. White Sands Missile Range: Analysis and Computation Directorate, 1968.

4. Origin

Software Branch
Data Sciences Division
Attn: NR-A
White Sands Missile Range, NM 88002
1970

5. Comments

The filter is embedded in several modules which make up the Modular Integrated Processing System (MIPS) of Data Sciences Division. The Multiple Instrument Trajectory Module (MITO26) exercises the filtering process most extensively. With this module,

a. observations are rotated to a common plane;

b. observations are filter-smoothed;

c. a set of predicted position, velocity, and acceleration components are obtained;

d. corrections for items such as refraction and velocity of propagation are made;

e. initial editing is performed to eliminate observations that have gross errors;

f. approximate position components are obtained using standard weights;

g. weights are determined for position-related observations;

h. best estimate of position components is obtained;

i. final editing of position-related observations is performed; (If any more position-related observations are rejected, steps f through h are repeated.)

j. approximate velocity components are obtained using standard weights;

k. weights are determined for velocity-related observations;

l. best estimate of velocity is obtained;

m. final editing of velocity-related observations are performed; (If any more velocity-related observations are rejected, steps j through l are repeated.)

n. approximate acceleration components are obtained using standard weights;

o. weights are determined for acceleration-related observations;

p. best estimate of acceleration components is obtained; and

q. final editing of acceleration-related observations is performed. (If any more acceleration-related observations are rejected, steps n through p are repeated.)

M-STATION

1. Name or Acronym

M-Station

2. Contact

Mr. Wen-Mi Liou
Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
(805) 989-7931
DSN: 351-7931

3. Documentation Sources

Liou, Wen-Mi. "Square Root Information Filter/Smoother for Multiple-Radar Tracking." Tech Note 3440-02-87. Point Mugu: Pacific Missile Test Center. March 1987.

4. Origin

Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
1987

5. Comments

The square root information filter/smoother had been designed to merge measurements from different radars to obtain the best estimate of trajectory. The measurements consist of range, azimuth, elevation, and range rate from radar tracking of up to 10 radars. This filter is a version of the conventional nine-state extended Kalman filter but has a fundamentally different approach to the optimal estimation problem. The whole updating process of the filter is founded on numerically stable, orthogonal transformation and preserves non-negativity of computed covariances. An ad hoc procedure for adaptive estimation had been implemented. In the implementation, the user has the options to select the time span for smoothing and to apply more weight on the measurements of some of the radars. Real radar tracking data were used to test the filter/smoother; the test results were satisfactory. In the testing, 100 data point smoothing was found adequate; consequently, more data point smoothing provided no significant improvement. This finding saves computer computation time and storage. As a result of this finding, a better divergence control method is being investigated.

PBS PROGRAM, M-STATION

1. **Name or Acronym**

Post Batch System (PBS) Program, M-Station

2. **Contact**

Mr. Wen-Mi Liou
Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
(805) 989-7931
DSN: 351-7931

3. **Documentation Sources**

Powers, W. "Multiple Station Radar Solution, Part I, Method of Calculation." Tech Note 3285-581. Point Mugu: Pacific Missile Test Center. September 1964.

4. **Origin**

Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
1964

5. **Comments**

Radar measurements are inversely proportional to estimates of their standard deviations. These weights are fixed and cannot change as the distance between target and site change; thus, separate passes are necessary during a tracking operation. The routine is fairly robust to individual site errors as each observation is edited if it exceeds a specified residual; however, to work well generally, individual radar data must be extensively pre-edited for on track (beacon) times, and bias errors should be corrected through calibration. The procedure produces a nine-state vector and its covariance matrix for each time point and, additionally can estimate individual radar biases.

PBS PROGRAM, DERIVE

1. Name or Acronym

Post Batch System (PBS) Program,DERIVE

2. Contact

Mr. Wen-Mi Liou
Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
(805) 989-7931
DSN: 351-7931

3. Documentation Sources

Morris, G. "Polynomial Smoothing and Differentiation by Least Squares." Working Note. Point Mugu: Pacific Missile Test Center, Code 3442, 1965.

4. Origin

Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
1965

5. Comments

This filter works well if data are well-approximated by a given quadratic expression. The technique may give poor estimation of end points where midpoint computations are not applied. The improved rate estimation is being examined through the PBS program, Optimum Finite Impulse Response Linear Phase Digital Filter (OPFILT).

PBS FILTER

1. **Name or Acronym**

Post Batch System (PBS) Filter

2. **Contact**

Mr. Wen-Mi Liou
Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
(805) 989-7931
DSN: 351-7931

3. **Documentation Sources**

Welch, M. Working Note 3442-22-79. Point Mugu: Pacific Missile Test Center, Code 3442, 1979.

4. **Origin**

Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
1965

5. **Comments**

This filter technique is superior to any time domain method for post-operations analysis; however, the results are poor when signal frequencies cannot be separated from noise frequencies or when the signal lies partly in transition zones. The usual end point problems occur because of the requirements for midpoint estimation. One solution to the end point problem is to collapse or expand the data at the ends, although the small filter size would result in a wide transition zone. Response function improvement is possible through a computer-aided design of filter weights. (See Post Batch System program Optimum Finite Impulse Response Linear Phase Digital Filter (OPFILT).)

PBS PROGRAM, OPFILT

1. Name or Acronym

Post Batch System (PBS) Program, Optimum Finite Impulse Response Linear Phase Digital Filter (OPFILT)

2. Contact

Mr. Wen-Mi Liou
Pacific Missile Test Center
Code 3400
Point Mugu, CA 93042
(805) 989-7931
DSN: 351-7931

3. Documentation Sources

a. Liou, Wen-Mi. "Optimum Finite Impulse Response (FIR) Linear Phase Digital Filters: Theory and Analysis of the McClellan-Parks Algorithm." Tech Note 3442-02-82. Point Mugu: Pacific Missile Test Center, May 1982.

b. McClellan, James A., Thomas W. Parks, and Lawrence R. Rabiner. "A Computer Program for Designing Optimum FIR Linear Filters." IEEE Trans. Audio Electroacoustics. AU-21.6 (1973): 506-26.

4. Comments

This general purpose filter design algorithm is capable of designing a large class of optimum (in the minimax sense) finite impulse linear phase digital filters such as low pass and high pass, as well as first and second differentiators. The algorithm can also be used to design filters which approximate arbitrary frequency specifications provided by the user. The user has control over the sizes of transition bands and ripples in both passband and stop band. The sharpening technique equipped with the filter design algorithm further reduces the sizes of small ripples in the designs of low-pass, high-pass, and band pass filters of any complexity. At the same time, this technique increases the sharpness of the frequency response in the transition bands.

ATAGAS KALMAN FILTER

1. **Name or Acronym**

Air-to-Air Gunnery Assessment System (ATAGAS) Kalman Filter

2. **Contact**

Mr. Lee Gardner
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-2628
DSN: 527-2628

3. **Documentation Sources**

Analytic Sciences Corporation. "Optimal Estimation for the Air-to-Air Gunnery Assessment System, Final Program Review." Reading: Analytic Sciences Corporation, n.d.

4. **Origin**

The Analytical Sciences Corporation
One Jacob Way
Reading, MA 01867
(617) 944-6850

5. **Comments**

The Analytical Sciences Corporation has a report which describes their evaluation of the filter's effectiveness. It was concluded that the filter's effectiveness was better but very costly computer time-wise.

MOTION TRACKING KALMAN FILTER

1. Name or Acronym

Motion Tracking Kalman Filter

2. Contact

Mr. Len Childers
6545 Test Group/ENAC
Hill AFB, UT 84056
(805) 777-8605
DSN: 458-8605

3. Documentation Sources

High Accuracy Multiple Object Tracking System (HAMOTS) Computer Software Maintenance Manual, December 1979.

4. Origin

Messrs. Doug Troxler and Mick Chaplin
General Dynamics
Electronics Division
San Diego, California

5. Comments

The filter does not handle wild points from input data as well as expected.

TEC TRACKER

1. Name or Acronym

U.S. Army TEXCOM Experimentation Center (TEC) Tracker (formerly Combat Developments Experimentation Center (CDEC) Kalman Filter)

2. Contact

Dr. Joseph M. Weinstein
Scientific Support Laboratory
P.O. Box 100
Fort Ord, CA 93491
(408) 384-2161

Ms. Julie Lemen
Scientific Support Laboratory
P.O. Box 898
Jolon, CA 93928
(408) 385-2880

3. Documentation Sources

a. Weinstein, Joseph M. "Position Location Logic." TEC Scientific Support Laboratory Document. Jolcn: TEC Scientific Support Laboratory, June 1988.

b. Weinstein, Joseph and others. "Kalman Filter User Manual." Revision I. CDEC Scientific Support Laboratory Document. Jolon: CDEC Scientific Support Laboratory, July 1981.

4. Origin

Kalman tracking software developed by General Dynamics staff in 1971 was furnished to TEC for use with Range Measuring System (RMS) hardware. The TEC has since upgraded tracking procedures, logic, and code, notably around 1976 and 1980 but especially since 1987.

5. Comments

Description. The TEC Tracker provides three-dimensional position and velocity of ground and air players during instrumented field combat simulations. This tracker is comprised of a Kalman filter and several other real-time routines which together process ranges between A and B units (RMS ranging and player units) as well as altimeter and other input. Post-operation routines smooth and analyze tracking output and apply alternative tracking algorithms to logged real-time input.

Performance. The tracker meets TEC's main real-time need for tracking as many as 100 players reasonably well with 10 meters horizontal error. As upgraded, the tracker reinitializes quickly as needed, does not require time-clumped ranges to any given B-unit, and prefilters (with outlier rejection) ranges between any given A,B pair before using them in the Kalman filter.

Prospects. The Kalman computations are intensive and must rely on predictive modeling of ground and air combat maneuver dynamics. Recent study suggests that throughput and accuracy may gain from replacing the Kalman filter by a simpler routine which relies just on good quality measured inputs, for example, a linearized least squares fit to the prefiltered ranges.

DRIFT TRACKING KALMAN FILTER

1. Name or Acronym

Drift Tracking Kalman Filter

2. Contact

Mr. Len Childers
6545 Test Group/ENAC
Hill AFB, UT 84056
(801) 777-8605
DSN: 458-8605

3. Documentation Sources

High Accuracy Multiple Object Tracking System (HAMOTS) Computer Software Maintenance Manual. December 1979.

4. Origin

General Dynamics
Electronics Division
San Diego, California
1979

5. Comments

This filter has not been used because score pod equipped vehicles are not available at Utah Test and Training Range.

EATS KALMAN FILTER

1. **Name or Acronym**

Extended Area Test System (EATS) Kalman Filter

2. **Contact**

Ms. L. Wilson
Pacific Missile Test Center
Code 3452
Point Mugu, CA 93042
(805) 984-8784
DSN: 351-8784

3. **Documentation Sources**

The General Dynamics, Electronics Division, System Controller Design document for EATS.

Any technical reference for Kalman filtering.

4. **Origin**

General Dynamics
Electronics Division
San Diego, California
1978 to 1980

5. **Comments**

The EATS design requirements called for real-time state vector estimation with accuracy constraints. This six-state Kalman design provided the best alternative while satisfying design goals. Real-time accuracies can be improved with finer tuning capability which is currently available with tri-level (low-medium-high) dynamic tuning parameter. This dynamic tuning parameter appears satisfactory for most purposes. Post-operations accuracy improvement techniques are currently under study.

TRIDENT KALMAN FILTER

1. Name or Acronym

Trident Missile Tracking Kalman Filter

2. Contact

Mr. Donald Olson
Pacific Missile Test Center, Code 1051
Point Mugu, CA 93042-5000
(805) 989-8804
DSN: 351-8804

Mr. Eric Senor
Pacific Missile Test Center, Code 3442
Point Mugu, CA 93042-5000
(805) 989-7931
DSN: 351-7931

3. Documentation Sources

a. Gelb, Arthur. Applied Optimal Estimation. Cambridge: MIT P, 1974.

b. Olson, Donald. "Filter and Smoother for Trident Missile Tracking." TP-64. Point Mugu: Pacific Missile Test Center, June 1988.

4. Origin

Designed by Pacific Missile Test Center personnel, the preliminary covariance simulations of the Trident Kalman filter began in 1977. These simulations led to the first operational version which supported four missile tests in the Pacific during 1983-1984. In 1984, a major redesign effort resulted in a more reliable and accurate second version which has supported 13 operational tests thus far.

5. Comments

For range safety redundancy, the nine-state extended Kalman filter operates on two CDC Cyber 175 mainframes. Square-root filtering was found to offer no advantage over standard covariance filtering in this application, presumably because of the large, 60-bit word size. Measurements consist of range sums and range rate sums from five transmitting stations operating 20 milliseconds apart. In addition, pedestal angles at the receiving telemetry antenna are used. While process noise is constant, the filter switches to a lower value at third stage burnout. Constants are also used for measurement noise standard deviations and edit limits.

Accurate initialization was found to be crucial to filter performance throughout missile flight. The initial state estimate is computed in a tangent plane, rectangular coordinate system. Ranging data from three stations are filtered with a recursive, first-degree polynomial filter, then used to solve for missile position and velocity. Because the z component of the ranging solution is inaccurate at low elevations, z position and velocity are obtained from missile telemetry. If telemetry is not available, the scheme defaults to a combination of nominal data and the ranging solution.

QD - WSMR

1. **Name or Acronym**

Quadratic Digital (QD) Filter

2. **Contact**

Mr. John Falke
U.S. Army White Sands Missile Range
Attn: STEWS-NR-AR
White Sands Missile Range, NM 88002
(505) 678-3458
DSN: 258-3458

3. **Documentation Sources**

McCool, W. A. "QD-A New Efficient Digital Filter." Analysis and Computation Directorate Internal Memorandum 60. White Sands Missile Range: Analysis and Computation Directorate, August 1967.

4. **Origin**

Mr. W. A. McCool
Acting Director
Analysis and Computation Directorate
White Sands Missile Range, NM 88002
August 1977

5. **Comments**

The quadratic digital filter's computing time is extremely small and invariant with point span. It achieves smoothing effectiveness almost identical to that of the comparable constrained least-squares filter.

QD - EGLIN

1. Name or Acronym

Quadratic Digital (QD) Filter

2. Contact

Mr. John E. Lindegren
3200 SPTW/KRTAR
Eglin AFB, FL 32542
(904) 882-4267
DSN: 872-4267

3. Origin

Eglin AFB, Florida

4. Comments

Although this filter is efficient, extremely fast, easy to use, and does not require extensive data arrays, its instabilities caused by unedited wild points limits its use.

QD - APO SAN FRANCISCO

1. **Name or Acronym**

Quadratic Digital (QD) Filter

2. **Contact**

Mr. Sonny Padayhag
Kentron International, Incorporated
Box 1207
APO San Francisco, CA 96555
(805) 238-7994, Ext. 8-2020
DSN:254-2020

3. **Documentation Sources**

Real-time program (RTP) or real-time impulse filter (RTIF) documentation.

4. **Comments**

This filter works well for speedy execution and extrapolation of small periods of time such as 200 milliseconds. It does not perform well with exoatmospheric, ballistic trajectories and long extrapolations or propagations.

QD - YUMA

1. Name or Acronym

Quadratic Digital (QD) Filter

2. Contact

Mr. Robert Mai
Yuma Proving Ground
Attn: STEYP-MT-TA
Yuma, AZ 85365
(602) 328-3295
DSN: 899-3295

3. Documentation Sources

a. McCool, W. A. "QD-A New Efficient Digital Filter." Analysis and Computation Directorate Internal Memorandum 60. White Sands Missile Range: Analysis and Computation Directorate, August 1967.

b. Mai, Robert W. "The QD Filter in YPG's Real-Time Laser Display System." STEYP-MAC Document. Yuma: Yuma Proving Ground.

4. Origin

Mr. Robert Mai
Yuma Proving Ground
Attn: STEYP-MAC
Yuma, AZ 85365-9102
1974

5. Comments

The quadratic digital filter does a much better job of estimating velocity than conventional least squares moving arc routines. Computation time is very small and does not depend on filter memory length. This easy-to-use filter is very simple to put into computer code. Filter response to impulse changes in acceleration is slow, that is, approximately two to three times the memory length selected, so filter estimates are useful primarily during segments where acceleration is minimal. The filter memory length, which determines the constant gain coefficients for position, velocity, and acceleration, can be adjusted to accommodate higher accelerations but then the noise on the filter estimates increases. Perhaps the major drawback of the QD filter is the difficulty in determining accuracy of velocity estimates which is dependent upon the complex interaction of sampling rate, target, dynamics, filter memory length, and noise on the measurements.

QUADRATIC POLYNOMIAL FILTER

1. **Name or Acronym**

Quadratic Polynomial Filter

2. **Contact**

Dr. Floyd Hall
Naval Weapons Center
Code 62303
China Lake, CA 93555
(619) 939-6346
DSN: 437-6346

3. **Documentation Sources**

Gossett, Eric. "On-Axis Tracking System."

4. **Comments**

This filter works quite well for real-time trajectory data from radar or laser data. The quadratic polynomial filter is an α-β-γ filter based on two parameters. This filter is a little different from what is expressed in Gossett's paper although it gives the same results. The filter is not considered adequate for close-in targets with tracking radar problems. Naval Weapons Center is currently developing a filter that rectifies this inadequacy. The new filter will have nonlinear constraints.

QD - NATC

1. Name or Acronym

Quadratic Digital (QD) Filter

2. Contact

Mr. John Shields
Naval Air Test Center
Computer Sciences Division (CS35)
Patuxent River, MD 20670-5304
(301) 863-3396
DSN: 326-3396

3. Documentation Sources

McCool, W. A. "QD-A New Efficient Digital Filter." Analysis and Computation Directorate Internal Memorandum 60. White Sands Missile Range: Analysis and Computation Directorate, August 1967.

4. Origin

Mr. Frederick K. H. Hoeck
Computer Services Directorate
Patuxent River, MD 20670-5304
1979

5. Comments

This filter is adequate for the application; however, as with all filters, there is some setting time. Comparisons made to other data sources were favorable. The filter requires little memory because samples need not be stored. Constants need to be calculated based on sample rate and number of points in the point span. It handles data errors and recovers with few problems.

AVRAGE

1. **Name or Acronym**

AVRAGE

2. **Contact**

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-0871
DSN: 527-0871

3. **Documentation Sources**

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Average Data Smoothing Subroutine." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Chapter 8. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. **Origin**

Systems Development Corporation
June 1972

5. **Comments**

The AVRAGE gives an average of up to 61 consecutive input data values.

1. Name or Acronym

Differentiation by Least Squares Subroutine (DUZ2)

2. Contact

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB, CA 93523
(805) 277-0871
DSN: 527-0871

Mr. Mike Tietz
Computer Sciences Corporation
P.O. Box 446
Edwards AFB, CA 93523
(805) 277-3800
DSN: 527-3800

3. Documentation Sources

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Differentiation by Least Squares Subroutine (DUZ2)." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Chapter 20. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. Comments

The DUZ2 calculates smooth values as well as first and second derivatives by fitting a least squares parabola. This subroutine works well with low-angular acceleration but not with high-angular accelerations. It is limited in how much data can be processed in one call to the subroutine by the array sizes, because an excessive amount of computer memory could be required.

SDIFFR

1. **Name or Acronym**

Differentiation Subroutine (SDIFFR)

2. **Contact**

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-0871
DSN: 527-0871

3. **Documentation Sources**

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Differentiation Subroutine (SDIFFR)." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Chapter 20. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. **Origin**

Messrs. R. C. Schramand R. T.Scott
White Sands Missile Range, New Mexico
July 1972

5. **Comments**

Basically, the SDIFFR functions as a single-parameter, cycling DIRSIT which creates parameter arrays containing up to 60 time points of smooth values, first derivatives, and second derivatives as desired. The SDIFFR is slightly different from DIRSIT in that it is called once for each desired parameter, initializing (if requested), and processing up to 60 points and then returns. Thus, SDIFFR cycles up to 60 points inside itself for 1 parameter instead of being called up to 60 times.

As a method of computation, the initialization is accomplished by fitting a least squares parabola and using a maximum likelihood procedure to the first user-designated interval input data points. The DIRSIT process then takes over for subsequent points. Because this process is not self-starting, at least the first interval of points should be allowed at the beginning for start-up. Once the DIRSIT process begins, a second-order Taylor series expansion is adjusted according to certain statistical criteria until the number of points above the curve differs from the number below the curve by a user-designated amount.

The first and second derivatives of the resulting second-order curve are taken as the derivatives at the first of the interval. The interval is then moved by one point, and the process continued until the end of data is reached. At the end of the data, the interval size is collapsed until only one point remains. The derivatives for the last point are set equal to the derivative of the next to the last point. Because of this interval collapsing, up to INPTS-1 points should be disregarded at the end of the data, where INPTS is the number of points in the interval.

A wild point has only minimal effect on the DIRSIT process since the only test made is whether it is above or below the curve. How far above or below is of little importance.

The conditions of validity are listed below.

a. Care should be taken to avoid processing large discontinuities; however, time increments need not necessarily be constant.

b. The appropriate slope parameter and tolerance criteria parameter must be provided to ensure valid results.

c. The number of input data points per parameter is limited to 60 for each call. More points can be accommodated through multiple calls to SDIFFR.

d. The SDIFFR is not a self-starting routine. The initialization process requires a number of points to get things moving, so the first several time points of derived values will be poor approximations.

e. The SDIFFR uses a collapsing technique to process the last interval of time points, but the results for these last several points (INPTS-1) become increasingly degraded as SDIFFR is forced to work on smaller and smaller intervals.

f. The calling program for SDIFFR uses an overlapping technique to avoid including the invalid results at the beginning and end of each array processed. A check should be made to ensure that the overlap is adequate so that no major discontinuities appear at each overlap point.

g. The calling program for SDIFFR uses a default interval size which the user may change.

DIRSIT

1. **Name or Acronym**

   Derivative Information Recovery by a Selective Integration Technique (DIRSIT)

2. **Contact**

   Mr. Len Childers
   6545 Test Group/ENAC
   Bldg 1284
   Hill AFB, UT 84056
   (801) 777-8605
   DSN: 458-8605

3. **Documentation Sources**

   Utah Test and Training Range postflight data reduction documents.

4. **Origin**

   Mr. Len Childers
   6545 Test Group/ENAC
   Hill AFB, UT 84056-5000
   (Adaptation of SDIFFR)

UFTAS FILTER OVERLAY

1. Name or Acronym

Uniform Flight Test Analysis System (UFTAS) Digital Filtering Primary Overlay (FILTER)

2. Contact

Dr. William G. Kitto
6521 Range Squadron/RCP
EdwardsAFB, CA 93523
(805) 277-3198
DSN: 527-3198

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-0871
DSN: 527-0871

3. Documentation Sources

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "UFTAS Digital Filtering Primary Overlay (FILTER)." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. Origin

Dr. William G. Kitto
1982

5. Comments

Currently, this filter is used heavily in the Frequency Response Analysis (FRA) program. (Butterworth or User Supplied Weight Type filter.)

BET

1. **Name or Acronym**

Best Estimate of Trajectory (BET)

2. **Contact**

Mr. Robert Mai
Yuma Proving Ground
Attn: STEYP-MT-TA
Yuma, AZ 85365
(602) 328-3295
DSN: 899-3296

3. **Documentation Sources**

Liebelt, Paul B. An Introduction to Optimal Estimation. Reading: Addison-Wesley, 1967.

4. **Origin**

Mr. Robert Mai
Yuma Proving Ground
Attn: STEYP-MAC
Yuma, AZ 85365
1976

5. **Comments**

The Kalman filter and Rauch-Tung-Striebel smoother take range, range rate, azimuth, elevation, and apparent angle measurements from up to 20 instruments and based upon a priori statistical information about the trajectory, the measurements, and the initial conditions, optimal estimates of the state vector are computed at each of the measurement times. The state vector size is expandable from the basic three components (x,y,z coordinate system) to include up to five derivatives: measurement source, pitch, yaw, pitch rate, and yaw rate. Estimates of the state vector are optimal in the sense that the uncertainty in each component of the state vector is minimized. This BET can provide excellent results; however, the results are only as good as the a priori statistical information. For applications where knowledge of measurement uncertainty and dynamics in the trajectory are well known or easily estimated, then the BET will undoubtedly provide very close to the absolute best estimate. If the a priori statistics distort the truth, then the BET trajectory will be distorted. In actual practice at Yuma Proving Ground, a BET trajectory is derived in an interactive fashion. Initial runs are made with a priori statistics implying more uncertainty than is actually present. Then analysis of the measurement residuals and estimated dynamics are used to refine the a priori estimates. A new BET trajectory is then generated and analysis of residuals and dynamics undertaken. This process continues until the a priori statistics match the derived estimates.

There are significant disadvantages to using the BET. Although the program setup is very complex and tedious, obtaining *a priori* statistical information is often perceived as an art rather than a science. Errors in setup or in judgment almost always have a detrimental effect on the final estimates. In addition, computation time is very long, and the required iterative process can cause data turnaround time to expand to longer than two weeks. So if you want the very best, you must be willing to pay for it!

1. **Name or Acronym**

Real-Time Update Filter (RTUF)

2. **Contact**

Mr. Andy Roy or Joe Warren
Computer Sciences Raytheon
Unit 2310
Eastern Space and Missile Center
Patrick AFB, FL 32925
(407) 494-7133
DSN: 854-7133

Mr. J. V. Copp
Computer Sciences Raytheon
Eastern Space and Missile Center
Patrick AFB, FL 32925
(407) 853-7783

3. **Origin**

Ms. Marie Colmer and Mr. Royal Pepple
RCA/MTP
Eastern Space and Missile Center
Patrick AFB, FL 32925
1971

4. **Comments**

This filter is used mainly for the Missile Precision Instruction Radar (MIPIR) class of radars. These radars are large scale. The RTUF is recursive, adaptive, exponentially weighted, and has fading memory. By adaptive, it is meant that the filter can adjust as the data gets seemingly noisier or less noisy. As the data becomes noisier, the bandwidth is decreased, and as the data becomes less noisy the bandwidth increases. In an adaptive filter, the coefficients of the terms involving the difference between the predicted value and the given raw value vary with time. In an exponentially weighted filter such as this one, the coefficients have the time variable as part of their exponents. In a fading-memory filter, the memory fades with time. In this case, the memory fades exponentially. In other words, newer data are weighted considerably more than older data when predicting a new value.

The RTUF is a simple filter because it does not work with partial derivatives or matrices. For this reason, Mr. Joe Warren says that this filter was good in the 1960s when computers were slower and had less memory. Mr. Warren predicts that no more RTUFs will be used within two to three years and suggests using a Kalman-type filter in lieu of this one.

RFILTR

1. Name or Acronym

Real-Time Software System (RFILTR)

2. Contact

Mr. Robert Crolene
Pacific Missile Test Center (formerly Pacific Missile Range (PMR))
Code 1074
Point Mugu, CA 93042
(805) 982-8073
DSN: 351-8073

3. Documentation Sources

Cragun, G. C. "Real-Time Data Filtering." PMR Tech Note 3285-576. Point Mugu: Pacific Missile Range.

4. Origin

NAMTC
Code 3400
1965

5. Comments

The RFILTR is a simple, exponential adaptive filter. The advantage of this filter is its simplicity. Upgrading can be achieved with some adaptive smoothing controls by basing it on residuals or by employing a Kalman algorithm. However, this step would complicate filter operations and may not be practical for the software system of the Univac 1230 for which this filter was originally designed. Upgrading the filter for use in a mainframe computer should present no problems.

Major problems with this filter are smoothing control and initialization variance control. Correcting one control usually makes the other one worse. Initialization errors are especially large for velocities and accelerations.

RCHECK

1. **Name or Acronym**

Rate Check Data Smoothing Subroutine (RCHECK)

2. **Contact**

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-0871
DSN: 527-0871

3. **Documentation Sources**

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Rate Check Data Smoothing Subroutine (RCHECK)." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Chapter 8. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. **Origin**

Mr. G. A. Lott
Lockheed-Georgia Company
June 1972

The RCHECK was successfully used by the Lockheed-Georgia Company in processing C-5A test data and was later modified to fit into UFTAS.

5. **Comments**

This subroutine searches for wild points by comparing DLY, the absolute difference between the current and previous point values, with TOL, a linear function of the average of successive differences within a given interval. The TOL represents the tolerance limit for the interval's data variations. The multiplicative and additive factors of this linear function are user-supplied. The TOL is increased or decreased from point to point according as DLY is greater than or less than TOL. In addition, if DLY is greater than or equal to TOL, the current data value is replaced by the last valid one.

EYBALL

1. Name or Acronym

Eyball Data Smoothing Subroutine (EYBALL)

2. Contact

Mr. Ly V. Tran
6521 Range Squadron/RCP
Edwards AFB CA 93523-5000
(805) 277-0871
DSN: 527-0871

3. Documentation Sources

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Eyball Data Smoothing Subroutine (EYBALL)." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. Origin

Mr. C. F. Carpenter
General Dynamics

Capt J. H. Pierson, USAF
Edwards AFB, CA
July 1972

5. Comments

The user specifies a value DELTA that represents the difference between the maximum and minimum values for a given parameter. The value of DELTA is based on the user's prior knowledge of what values the parameter should take.

If the actual range of values within the interval is less than or equal to DELTA, then the output parameter value is the average of the values in the interval. If the range is greater than DELTA, then the weighted average of the most common values in the interval, with the more central values given the greater weight, is the output parameter value.

The advantages of the EYBALL subroutine are

a. the EYBALL closely approximates intuitive treatment of data,

b. the user has control by using prior knowledge of the parameter,

c. the wild points are discarded before determination of the output parameter value, and

d. the excessive wild points are identified.

ON-AXIS

1. **Name or Acronym**

On-Axis Radar Target Tracking System (ON-AXIS)

2. **Contact**

Mr. John E. Lindegren
3200 SPTW/KRTAR
Eglin AFB, FL 32542
(904) 882-4267
DSN: 872-4267

3. **Documentation Sources**

"On-Axis: Philosophy/Technology/Development." Advanced Research Project Agency, Radar Microwave Link Tech Memo 211. Patrick AFB: Advanced Research Project Agency, Radar Microwave Link, 15 December 1970.

4. **Origin**

Advanced Research Project Agency
Radar Microwave Link
Patrick AFB, Florida
1969

Modified: Air Force Development Test Center
AFDTC/KR
Eglin AFB, Florida
1973

5. **Comments**

Strengths of this filter include self-calibration capability through stellar track; adaptive track gain with smooth, low-noise track; and predetermined missile trajectory slew at launch.

This filter's weaknesses consist of requiring a well-qualified crew and careful maintenance and software control of the computer system and data base. The On-Axis track accuracy (with pedestal position pick-offs) offers no significant improvement over normal

AN/FPS-16 TSPI for aerodynamic maneuvering targets if standard post-mission smoothing procedures (moving-arc-polynomial midpoint fits to raw data) are used.

DYNO43

1. Name or Acronym

DYNO43

2. Contact

Mr. Robert Fierro
U.S. Army White Sands Missile Range
Attn: STEWS-NR-AM
White Sands Missile Range, NM 88002
(505) 678-2543
DSN: 258-2543

3. Documentation Sources

a. Comstock, D., M. Wright, and V. Tipton. "Handbook of Data Reduction Methods." Data Reduction Division Tech Rept. White Sands Missile Range: Data Reduction Division, 13 August 1964.

b. Comstock D. Introduction to Least Squares. White Sands Missile Range: Analysis and Computation Directorate, 1968.

4. Origin

Classical Least Squares Smoothing
August 1970

5. Comments

The filter derives smooth data from observation using the least squares moving arc method. The smoothed positions are differentiated to obtain velocities which, in turn, are differentiated to obtain accelerations.

1. **Name or Acronym**

Gennery Filter (GNFL)

2. **Contact**

Dr. Kenneth Lane
CSR 3200
P.O. Box 4127
Patrick AFB, FL 32925

3. **Documentation Sources**

a. Computer Program 285, GNFL

b. Gennery, Donald B. "An Improved Digital Filter." Mathematical Services TM-63-8, Dec. 1963.

4. **Origin**

Mr. Donald B. Gennery
1963

5. **Comments**

This filter is designed to give a frequency response with sharper roll-off and with increased attenuation at higher frequencies when compared to standard least squares polynomial filters.

1. Name or Acronym

Gennery Smoother (GNSM)

2. Contact

Dr. Kenneth Lane
CSR 3200
P.O. Box 4127
Patrick AFB, FL 32925

3. Documentation Sources

a. Computer Program 586, GNSM

b. Gennery, Donald B. "Direct Digital Filters for General Purpose Use." RCA Document. Patrick AFB: RCA/MTP, January 1966.

4. Origin

Mr. Donald B. Gennery
1966

5. Comments

Designed as an improvement of the GNFL filter with sharper roll-off, this filter has the additional feature of being able to bridge discontinuities in the derivatives of the input data.

FIRFILT

1. **Name or Acronym**

Finite Impulse Response Filter (FIRFILT)

2. **Contact**

Mr. Jerry Biedscheid
Sandia National Laboratories
Division 7522
P.O. Box 5800
Albuquerque, NM 87185
(505) 844-4048
DSN: 244-4048

3. **Documentation Sources**

McClellan, James A., Thomas W. Parks, and Lawrence R. Rabiner. "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters." IEEE Trans. Audio Electroacoust. AU-21.6 (1973): 506-26.

4. **Origin**

Mr. D. J. Miller
Sandia National Laboratories
Division 1414
P.O. Box 5800
Albuquerque, NM 87185
1980

5. **Comments**

This filter gives good results in low dynamic situations, but the filter is difficult to control.

RECFLT

1. Name or Acronym

RECFLT

2. Contact

Mr. W. D.Swartz
Sandia National Laboratories
Division 7524
P.O. Box 5800
Albuquerque, NM 87185
(505) 844-2237

3. Documentation Sources

Stearns, S. D. Digital Signal Analysis. Rochelle Park: Hayden, 1975.

4. Origin

Mr. D. J.Miller
Sandia National Laboratories
Division 7524
P.O. Box 5800
Albuquerque, NM 87185
1980

5. Comments

This filter provides excellent results, particularly with large data sets using the phase shift removal option.

FILMAX

1. **Name or Acronym**

   FILMAX

2. **Contact**

   Mr. D. O. Smallwood
   Sandia National Laboratories
   Division 7544
   P.O. Box 5800
   Albuquerque, NM 87185
   (505) 844-1074
   DSN: 244-1074

3. **Documentation Sources**

   Smallwood, D. O. "An Improved Recursive Formula for Calculating Shock Response Spectra." Shock and Vibration Bulletin. 51.2 (1981): 211-17.

4. **Origin**

   Mr. D. O. Smallwood
   Sandia National Laboratories
   Division 7544
   1979

5. **Comments**

The older recursive models of this filter used for calculating the shock response spectra resulted in significant errors when the natural frequency exceeded one-sixth of the sample rate. This new filter avoids the problem.

1. Name or Acronym

   MDPT51

2. Contact

   Mr. J. A. Ward
   Eastern Space and Missile Center
   RCA/MTP
   Patrick AFB, FL 32935

3. Documentation Sources

   a. RAID Computer Program

   b. "Filtering Tracking Data for Range Safety Displays." RCA Document. Patrick AFB: RCA/MTP, January 1984.

1. **Name or Acronym**

BSMW

2. **Contact**

Mr. J. A. Ward
Computer Sciences Raytheon
Eastern Space and Missile Center
Patrick AFB, FL 32925

3. **Documentation Sources**

a. RAID Computer Program

b. "Filtering Tracking Data for Range Safety Displays." RCA Document. Patrick AFB: RCA/MTP, January 1984.

4. **Comments**

The BSMW is a quadratic filter that is recursive but not adaptive. (A nonadaptive filter is a filter with a fixed bandwidth. The coefficients of the terms involving the difference between the predicted value and the given raw value are constant; they do not vary with the perceived noise content of the data.)

1. Name or Acronym

SMW

2. Contact

Mr. J. A. Ward
Computer Sciences Raytheon
Eastern Space and Missile Center
Patrick AFB, FL 32925

3. Documentation Sources

a. RAID Computer Program

b. "Filtering Tracking Data for Range Safety Displays." RCA Document. Patrick AFB: RCA/MTP, January 1984.

4. Comments

The SMW is a linear filter that is recursive but not adaptive. (A nonadaptive filter is a filter with a fixed bandwidth. The coefficients of the terms involving the difference between the predicted value and the given raw value are constant; they do not vary with the perceived noise content of the data.)

FREQUENCY RESPONSE ANALYSIS

1. **Name or Acronym**

Frequency Response Analysis (FRA)

2. **Contact**

Mr. Barry Mishler
6521 Range Squadron/RCP
Edwards, CA 93523-5000
(805) 277-6040
DSN: 527-6040

3. **Documentation Sources**

Computer Sciences Branch, 6521 Range Squadron, 6510 Test Wing. "Frequency Response Analysis (FRA) Overlay." Uniform Flight Test Analysis System (UFTAS) Reference Manual. Version 3.1. Edwards Air Force Base: Air Force Flight Test Center, October 1990.

4. **Origin**

Mr. Tom Twisdale
6510 Test Wing/DOEF
Edwards AFB, CA 93523-5000
(805) 277-1248
DSN: 527-1248

Dr. William G. Kitto
6521 Range Squadron/RCP
Edwards AFB, CA 93523-5000
(805) 277-3198
DSN: 527-3198

5. **Comments**

The Frequency Response Analysis program transfers the dynamic time domain data into the frequency domain to do the frequency response analysis specified by the user. Power spectral densities, transfer functions, and coherence functions can be calculated in the frequency domain. The dynamic time domain data are contained on either a B-file or a C-file. After the desired analysis is complete, FRA can put the frequency domain results back into the time domain.

The FRA program is not a filter. This program is meant to aid the user in determining what filter to use and how to use it. For instance, the FRA can aid in determining what cutoff frequencies should be in the frequency spectrum or if the filter should be bandlimited or nonbandlimited. At the present time, the program cannot accept more than 1024 time points. It uses the first 1024 time points received and discards the rest of the points.

REFERENCES

1. Bierman, Gerald J. Factorization Methods for Discrete Sequential Estimation. New York: Academic P, 1977.

2. Gelb, Arthur. Applied Optimal Estimation. Cambridge: MIT P, 1974.

3. Maybeck, Peter S. Stochastic Models, Estimation and Control. 3 vols. New York: Academic P, 1979. Vol. 3.

4. Morrison, Norman. Introduction to Sequential Smoothing and Prediction. New York: McGraw-Hill, 1969.

5. Bucy, Richard S., and Peter D. Joseph. Filtering for Stochastic Processes with Applications to Guidance. New York: Interscience Publishers, 1968.

6. Data Reduction and Computing Group. Error Analysis and Methods for Estimating Errors in Position, Velocity, and Acceleration Data. Document 153-71. White Sands Missile Range: Range Commanders Council Secretariat, 1971.

7. Gennery, Donald B. "Direct Digital Filters for General Purpose Use." RCA Document. Patrick Air Force Base: RCA/MTP, 1966.

8. Berry, Floyd N. "QD Filters Compared with α-β-γ Filters." Minutes of 35th Data Reduction and Computing Working Group Meeting. 19-23 Oct 1970. White Sands Missile Range: Range Commanders Council Secretariat, 1970. 132-42.

9. Leondes, Cornelius T. Theory and Applications of Kalman Filtering. N.p.: Harford House, Feb. 1970.

10. Moore, Alan J. "A Study of the Optimum Techniques for Radar and Doppler Tracking Data Digital Filtering." Systems Development Corporation Document. Santa Monica: Systems Development Corporation, 1969.

11. Ormsby, Joseph F. A. "Design of Numerical Filters with Applications to Missile Data Processing." Space Technology Laboratories Document. Los Angeles: Space Technology Laboratories, 1960.

12. Smith, E. S. "A Collection of Linear Numerical Filters for Missile Trajectory Data with General Principles and Design Considerations." RCA Document. Patrick Air Force Base: RCA, 1966.

13. Stirton, R. J. "Discrete Numerical Filters." U.S. Naval Ordnance Test Station Document. China Lake: U.S. Naval Ordnance Test Station, 1966.

14. Taylor, James T. "Digital Filters for Non Real Time Data Processing." NASA Document. Huntsville: NASA, George C. Marshall Space Flight Center, 1967.

15. McCool, W. A. "Derivation of the Discrete Kalman Filter Formulation Based on the QD Filter and Classical Least Squares Theories." Analysis and Computation Directorate Tech Rept 9. White Sands Missile Range: Analysis and Computation Directorate, 1969.

16. Hansen, Allen J. "624A Velocity Filter." Federal Electric Corporation Document. Vandenberg Air Force Base: Federal Electric Corporation, 1965.

17. Siler, Henry K. "Procedure for Computation of Frequency Constrained Digital Filter Coefficients." RCA Document. Patrick Air Force Base: RCA, 1963.

18. Pauley, Richard F. "A Unified Approach to Data Smoothing." RCA Document MTC-TDR-10. Patrick Air Force Base: RCA, 1962.

19. Cox, Henry. "Recursive Nonlinear Filtering." David Taylor Model Basin. Washington, DC: n.p., n.d.

20. McCool, W. A. "Techniques for Comparing the Performance of General-Purpose Digital Filters." Analysis and Computation Directorate Internal Memorandum 63. White Sands Missile Range: Analysis and Computation Directorate, 1967.

21. Rader, Charles M. "Digital Filter Design Techniques." MIT Lincoln Laboratory Tech Note 1965-63. Lexington: Lincoln Laboratory, 1965. [AD 627146].

22. Ford, Wayne T., and James H. Hearne. "Least Squares Inverse Filtering." Geophysics. 31.5 (1966): 917-26.

23. Berry, Floyd N. "Evaluation of Several Recursive Random Error Estimation Techniques." Minutes of 35th Data Reduction and Computing Working Group Meeting. 19-23 Oct 1970. White Sands Missile Range: Range Commanders Council Secretariat, 1970. 143-62.

24. Holtz, Howard, and C. T. Leondes. "The Synthesis of Recursive Digital Filters." J. Assoc. Comput. Mach. 13.2 (1966): 262-80.

25. Smith, F. T. The Application of the Kalman Filter to Some Astrodynamical Problems. Rand Corporation Memorandum RM-4277-PR. Santa Monica: Rand Corporation, 1964.

26. Kaplan, Wilfred. Operational Methods for Linear Systems. Reading: Addison-Wesley, 1962. 391-95.

27. Quenouille, M. H. The Analysis of Multiple Time-Series. New York: Hafner, 1957. 65-81.

28. Lueg, Russell E. Basic Electronics for Engineers. Scranton: International Textbook, 1963. 140-45.

29. Bendat, Julius S., and Allan G. Piersal. Measurement and Analysis of Random Data. New York: Wiley, 1966.

30. D'Esopo, D. A. "A Note on Forecasting by the Exponential Smoothing Operator." Oper. Res. 9.5 (1961): 686-87.

31. Brown,Robert G., and Richard F. Meyer. "The Fundamental Theorem of Exponential Smoothing." Oper. Res. 9.5 (1961): 673-85.

32. Conte, Samuel D. Elementary Numerical Analysis: An Algorithmic Approach. New York: McGraw-Hill, 1965.

33. Hamming, Richard W. Numerical Methods for Scientists and Engineers. New York: McGraw-Hill, 1962.

34. Drylov, V. I. Approximate Calculation of Integrals. New York: MacMillan, 1962.

35. Ralston, Anthony. A First Course in Numerical Analysis. New York: McGraw-Hill, 1965.

36. Schramm, R., and M. Scott. "Derivative Information Recovery by a Selective Integration Technique." Army Missile Test and Evaluation Directorate Tech Memorandum 1040. White Sands Missile Range: Army Missile Test and Evaluation Directorate, Nov. 1962.

37. Agee, William. "A Multidimensional Fortran Subroutine for DIRSIT." Army Missile Test and Evaluation Directorate Internal Memorandum 3. White Sands Missile Range: Army Missile Test and Evaluation Directorate, Aug. 1965.

38. Foster, T. "Preliminary Documentation of Filter X." Analysis and Computation Directorate Internal Memorandum 40. White Sands Missile Range: Analysis and Computation Directorate, Jan. 1967.

39. Baker, C. "Program Description for ATHENA DCS, Post Flight Programs, DCSFIL and RTARC7." Analysis and Computation Directorate Internal Memorandum 95. White Sands Missile Range: Analysis and Computation Directorate, n.d.

40. McCool, W. A. "A Digital Computer Subroutine for the Automated Selection of Radar Data for Instantaneous Impact Prediction." Computer Directorate Internal Memorandum 8. White Sands Missile Range: Computer Directorate, Feb. 1966.

41. McCool, W. A. "DFX-A New Digital Filter." Analysis and Computation Directorate Internal Memorandum 14. White Sands Missile Range: Analysis and Computation Directorate, Mar. 1966.

42. McCool, W. A. "QD-A New Efficient Digital Filter." Analysis and Computation Directorate Internal Memorandum 60. White Sands Missile Range: Analysis and Computation Directorate, Aug. 1967.

43. McCool, W. A. "QDVSP, An Efficient Digital Data Processing Subroutine." Analysis and Computation Directorate Internal Memorandum 81. White Sands Missile Range: Analysis and Computation Directorate, Feb. 1969.

44. McCool, W. A. "A Matrix Vector Formulation of the QD Filter." Analysis and Computation Directorate Tech Rept 3. White Sands Missile Range: Analysis and Computation Directorate, Dec. 1968. [AD 746623]

45. McCool, W. A. "Zero and First Order QD Filters." Analysis and Computation Directorate Tech Rept 15. White Sands Missile Range: Analysis and Computation Directorate, Mar. 1970.

46. Comstock, D. W., M. Wright, and V. Tipton. "Handbook of Data Reduction Methods." Data Reduction Division Tech Rept. White Sands Missile Range: Data Reduction Division, 13 Aug 1964. [NTIS AD-754857]

47. Blinchikoff, Herman J., and Anatol I. Zverev. Filtering in the Time and Frequency Domains. New York: Wiley, 1976.

48. Papoulis, Athanasios. Probability, Random Variables, and Stochastic Processes. New York: McGraw-Hill, 1965.

49. Peled, Abraham, and Bede Liu. Digital Signal Processing. New York: Wiley, 1976.

50. Bracewell, Ronald N. The Fourier Transform and Its Applications. New York: McGraw-Hill, 1978.

51. Conover, W. J. Practical Nonparametric Statistics. New York: Wiley, 1980.

52. Weaver, H. Joseph. Applications of Discrete and Continuous Fourier Analysis. New York: Wiley, 1983.

53. Brigham, E. Oran. The Fast Fourier Transform and Its Applications. Englewood Cliffs: Prentice-Hall, 1988.

54. Rabiner, Lawrence R., and Bernard Gold. Theory and Application of Digital Signal Processing. Englewood Cliffs: Prentice-Hall, 1975.

55. Oppenheim, Alan V., and Ronald W. Schafer. Digital Signal Processing. Englewood Cliffs: Prentice-Hall, 1975.

56. Belove, Charles, ed. Handbook of Modern Electronics and Electrical Engineering. New York: Wiley, 1986.

57. Jackson, Leland B. Digital Filters and Signal Processing. Boston: Kluwer Academic Publishers, 1989.

58. Daniels, R. W. Approximation Methods for the Design of Passive, Active, and Digital Filters. New York: McGraw-Hill, 1974.

59. Guillemin, E. A. Synthesis of Passive Networks. New York: Wiley, 1957.

60. Storer, J. E. Passive Network Synthesis. New York: McGraw-Hill, 1957.

61. Gold, Bernard, and Charles M. Rader. Digital Processing of Signals. New York: McGraw-Hill, 1969.

62. Hoel, Paul G. Introduction to Mathematical Statistics. 4th ed. New York: Wiley, 1971.

63. "Comparison of Static and Dynamic Filters." White Sands Missile Range: Analysis and Computation Division, 8 Oct 1970.

64. Rabiner, Lawrence R., and Charles M. Rader. Digital Signal Processing. New York: IEEE P, 1972.

65. Rabiner, Lawrence R., Bernard Gold, and C. A. McGoneal. "An Approach to the Approximation Problem for Nonrecursive Digital Filters." IEEE Trans. Audio Electroacoust. AU-18.2 (1970): 83-106.

66. Herrmann, O. "On the Design of Nonrecursive Digital Filters with Minimum Phase." Electron. Lett. 6.11 (1970): 328-29.

67. Herrmann, O., and H. W. Schuessler. "Design of Nonrecursive Digital Filters with Minimum Phase." Electron. Lett. 6.11 (1970): 329-30.

68. Hofstetter, E., A. V. Oppenheim, and J. Siegel. "A New Technique for the Design of Nonrecursive Digital Filters." Proc. of 5th Annual Princeton Conf. Inform. Sci. Sys. 1971. 64-72.

69. Hofstetter, E., A. V. Oppenheim, and J. Siegel. "On Optimum Nonrecursive Digital Filters." Proc. of 9th Allerton Conf. Circuit Sys. Theory. Oct. 1971.

70. Siegel, J. "Design of Nonrecursive Approximations to Digital Filters with Discontinuous Frequency Responses." Ph.D. thesis, MIT, Jun. 1972.

71. Parks, Thomas W., and James H. McClellan. "Chebyshev Approximation for Nonrecursive Digital Filters with Linear Phase." IEEE Trans. Circuit Theory. CT-19.2 (1972): 189-94.

72. Parks, Thomas W., and James H. McClellan. "A Program for the Design of Linear Phase Finite Impulse Response Filters." IEEE Trans. Audio Electroacoust. AU-20.3 (1972): 195-99.

73. Rabiner, L. R. "The Design of Finite Impulse Response Digital Filters Using Linear Programming Techniques." Bell Sys. Tech. J. 51.6 (1972): 1177-98.

74. Rabiner, L. R. "Linear Program Design of Finite Impulse Response (FIR) Digital Filters." IEEE Trans. Audio Electroacoust. 20.4 (1972): 280-88.

75. Schwarz, H. R. Numerical Analysis: A Comprehensive Introduction. Chichester: Wiley, 1989.

76. Young, David M. and Robert Todd Gregory. A Survey of Numerical Mathematics. Vol. 1. Reading: Addison-Wesley, 1972.

RCCS DOCUMENT TARGET BOX 01

DOCUMENT #: 156-73

NAME: Physical Constants and Conversion Factors

TYPE: DOCUMENT

DATA REDUCTION AND COMPUTER GROUP

DOCUMENT 156-73

PHYSICAL CONSTANTS AND CONVERSION FACTORS

WHITE SANDS MISSILE RANGE
KWAJALEIN MISSILE RANGE
YUMA PROVING GROUND
DUGWAY PROVING GROUND
ELECTRONIC PROVING GROUND

ATLANTIC FLEET WEAPONS TRAINING FACILITY
NAVAL AIR WARFARE CENTER WEAPONS DIVISION
NAVAL AIR WARFARE CENTER AIRCRAFT DIVISION
NAVAL UNDERSEA WARFARE CENTER DIVISION NEWPORT

30TH SPACE WING
45TH SPACE WING
AIR FORCE FLIGHT TEST CENTER
AIR FORCE DEVELOPMENT TEST CENTER
AIR FORCE WEAPONS AND TACTICS CENTER
DETACHMENT 2, SPACE AND MISSILE SYSTEMS CENTER

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED

IRIG DOCUMENT 156-73

PHYSICAL CONSTANTS AND CONVERSION FACTORS

Prepared by

Data Reduction and Computing Working Group
Inter-Range Instrument Group
Range Commanders Council

Published by

Secretariat
Range Commanders Council
White Sands Missile Range, New Mexico 88002

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

TABLE OF CONTENTS

		Page
I.	Introduction	I-1
II.	Mathematical Constants	II-1
III.	Miscellaneous Constants	III-1
IV.	Conversion of Constants	IV-1
	Notes	IV-4
V.	Spheroid Constants	V-1
	Reference Datums	V-2
	Datum Shift Constants	V-3
	Notes	V-4
VI.	Solar System Constants	
	Earth and Moon	VI-1
	Earth and Sun	VI-2
	Planets	VI-3
	Notes	VI-4
	Glossary of Geodetic Terms	A-1
	References	R-1

I. Introduction

The objective of this document is to present a compilation of the more commonly used physical constants, mathematical constants, unit conversion factors and other fixed numerical values as may be required at the various member installations for data processing.

Where possible the name or symbol associated with the constant is given. The constant is given in a variety of units, where considered useful. If an exact or defining value is known, this is given and so indicated. Where possible the values are given to 16 digit precision which may be useful in some computer systems.

Reference numbers corresponding to the source material in the Bibliography are given for each constant listed. A brief description of usage or additional information pertaining to the constant may be found in the notes which follow each section.

II. Mathematical Constants

Constant	Value	Description	Ref.
2π	6.28318 53071 79586		1,8
$3\pi/2$	4.71238 89803 84690		1,8
π	3.14159 26535 89793		1,8
$\pi/2$	1.57079 63267 94897		1,8
$\pi/4$	.78539 81633 97448 3		1,8
$1/\pi$	.31830 98861 83790 7		1,8
e	2.71828 18284 59045		1,8
1/e	.36787 94411 71442 3		1,8
M	.43429 44819 03251 8	$\log_{10}$ e	1,8
1/M	2.30258 50929 94046	$\log_e$ 10	1,8
$\sqrt{2}$	1.41421 35623 73095		1,8

III. Miscellaneous Constants

Definition	Value	Units	Ref.
Speed of Light (Vac)	299792456.2 $\pm$ 1.1	m/sec	23
Speed of Sound*	340.292046	m/sec	22
Standard SL Pressure	1013.25	mb	22
Standard SL Density	1.2250	Kg/m^3	22
Standard SL Temp	15	^oC	22

*Speed of Sound is given for Standard Sea Level Pressure, Density and Temperature which follows.

IV. Conversion Constants

To Convert From	Multiply By	Units	Note	Ref.
Feet to Meters	.3048*	m/ft	1	2
Meters to Feet	3.28083 98950 13123	ft/m		2
Nautical Miles to Meters	1852.*	m/nmi	1	2
Meters to Nautical Miles	$.53995\ 68034\ 55723\ 5 \times 10^{-3}$	nmi/m		2
Nautical Miles to Feet	6076.11548 55643 04	ft/nmi		2
Feet to Nautical Miles	$.16457\ 88336\ 93304\ 5 \times 10^{-3}$	nmi/ft		2
Inches to Meters	.0254*	m/in	1	2
Statute Miles to Meters	1609.344*	m/mi	1	2
Astronomical Units to Meters	$.14959\ 80000 \times 10^{12}$	m/AU	2	3
Astronomical Units to Feet	$.49081\ 36483 \times 10^{12}$	ft/AU		3,2
Meters to Astronomical Units	$.66844\ 91979 \times 10^{-11}$	AU/m		3
Feet to Astronomical Units	$.20374\ 33155 \times 10^{-11}$	AU/ft		3,2
Knots to Feet per Second	1.68780 98571 01196	fps/kt		2

*EXACT VALUES

IV. Conversion Constants (cont'd)

To Convert From	Multiply By	Units	Note	Ref.
Pounds (mass) to Kilograms	.45359 237*	kg/lb	1	2
Kilograms to Pounds (mass)	2.20462 26218 48776	lb/kg		2
Pounds (force) to Newtons	4.44822 16152 605*	n/lbf	1	2
Newtons to Pounds (force)	.22480 89430 99710 5	lbf/n		2
Average Mean Gravity Field	9.80665* 32.174048 55643 045	m/sec^2 ft/sec^2	1	2 2
Slugs to Pounds (mass)	32.174048 55643 045	lbm/slug		2
Millibars to Newtons/Meter^2	100.*	$\frac{n}{m^2}/mb$	1	2
Millibars to Pounds (force)/Foot^2	2.08854 34233 15013	$\frac{lbf}{ft^2}/mb$		2
Grams/cm^3 to Slugs/ft^3	1.94032 03319 79716	$\frac{slugs}{ft^3}/\frac{gm}{cm^3}$		2
Radians to Degrees	57.29577 95130 8232	deg/rad		1

*EXACT VALUES

IV. Conversion Constants (cont'd)

To Convert From	Multiply By	Units	Note	Ref.
Degrees to Radians	.01754 32925 19943 29	rad/deg		1
Mils to Degrees	.05625*	deg/mil	1	1
Degrees to Mils	6400./ 360.* 17.77777 77777 7778	mil/dec mil/dec	1	1
Millibars to Inches of Mercury ($0^{o}C$)	.02952 998	in/mb		6
Inches of Mercury to Millibars ($0^{o}C$)	33.86389	mb/in		6

*EXACT VALUES

Notes

1. The exact (*) relations are exact by definition and come from an international agreement among English speaking nations and are applicable to systems oriented to WGS-60. All feet and nautical miles are international units. One meter = 3.28083333333333 American survey feet.

2. The astronomical unit is a measurement of the mean distance between the sun and a fictitious unperturbed planet such that the numerical value of the Gaussian constant k may be exactly .01720209895 when the unit of time is the ephemeris day.

Spheroid Constants

Spheroid	a(meters)	b(meters)	1/f	Notes	Ref.
Clarke 1866	6378206.4	6356583.8	294.978698	1	5
Fischer 1960	6378166.	6356784.284	298.3	1	5
Kaula 1961	6378165.	6356783.287	298.3	1	5
Bessel 1841	6377397.155	6356078.963	299.15281	1	5
Everest	6377276.345	6356075.413	300.8017	1	5
Hough	6378270.	6356794.343	297.	1	5
International	6378388.	6356911.946	297.	1	5
Australian National	6378160.	6356774.719	298.25	1	5
Clarke 1880	6378249.145	6356514.870	293.465	1	15
APL	6378144.	6356757.339	298.23	1	20
Geodetic Ref System	6378160.	6356774.719	298.25	1	14
SAO 69	6378155.	6356770.094	298.255	1	19
NWL 8E	6378145.	6356759.769	298.25	1	19
AIRY	6377563.396	6356256.909	299.324965	1	19

Reference Datums

Datum	Spheroid	Origin	Latitude(φ)	Longitude(λ)(E)	Ref.
Australian Geodetic	Australian National	Johnston Memoral Cairn	-25°56'54".55	133°12'30".08	21
Tokyo	Bessel 1841	Tokyo Observatory (old)	35°39'17".515	139°44'40".502	21
Arc-Cape (So. Africa)	Clarke 1880	Buffelofontein	-33°59'32".00	25°30'44".622	21
North American 1927	Clarke 1966	Meades Ranch	39°13'26".686	261°27'29".494	21
Old Hawaiian	Clarke 1866	Oahu West Base	21°18'13".89	202°09'04".20	21
Indian	Everest	Kalianpur	24°07'11".26	77°39'17".57	21
Mercury 1960	Fischer 1960	Earth Centered			7
Vanguard	Hough	Meades Ranch	39°13'26".686	261°27'29".494	7
Wake-Eniwetok 1960	Hough	Wake Island Astro 1952	19°17'19".991	166°38'46".294	21
Canton Astro 1966	International	1966 Canton Secor Astro	-02°46'28".99	188°16'43".47	21
European	International	Potsdam	52°22'51".447	13°03'58".741	21
Johnston Island 1961	International	Johnston Island 1961	16°44'49".729	190°29'04".781	21
Kaula	Kaula 1961	Earth Centered			9

Datum Shift Constants

(Local Datum to Earth Centered Datum)

Earth Centered Datum / a / 1/f / Local Datum	Mercury 1960 6,378,166 298.3 ΔX ΔY ΔZ (meters)	Kaula 1961 6,378,165 298.3 ΔX ΔY ΔZ (meters)	NWL 8E 6,378,145 298.25 ΔX ΔY ΔZ (meters)
NAD 1927	3 111 225	-23 142 196	-23 159 185
EUROPEAN	-71 -15 -122	-57 -37 -96	-81 -99 -118
TOKYO	-158 648 613	-89 551 710	-147 530 676
OLD HAWAIIAN	61-256 -153		52 -262 -183
SOUTH AFRICAN CAPE (ARC)	-125 -48 -315	-109 -70 -289	

Notes

1. The underlined values define the oblate spheroid approximating the shape of the earth.

2. In converting a ϕ, λ, h from one datum to another, the location is converted to the geocentric u, v, w system using the original datum (or Local Datum), the differences (ΔX, ΔY, ΔZ) are applied to the geocentric coordinates, and the new ϕ, λ, h is computed using the new datum (Earth Centered Datum).

VI. Solar System Constants

Earth and Moon

Definition	Symbol	Value	Units	Note	Ref.
Inertial Rotation Rate	ω_I	$360^o/23^h56^m04\ ^s09s95413$		1	10
		15.04106717859357	deg/hr		
		.004178074216275992	deg/sec		
		.2625161452801118	rad/hr		
		.00007292115146669772	rad/sec		
Sidereal Rotation Rate	ω_s	$360^o/23^h56^m04\ ^s09054$		2	10
		15.0410686387	deg/hr		
		.00417807462185	deg/sec		
		.262516170764	rad/hr		
		.0000729211585452	rad/sec		
Earth Gravity Constant	GM_E	$398603. \times 10^9$	m^3/sec^2	3	3
		$1.40765321 \times 10^{16}$	ft^3/sec^2		3,1
		398601.49 ± 0.4	km^3/sec^2	4	12
Moon Gravity Constant	GM_C	4902.778 ± 0.1	km^3/sec^2	4	12
Mass Ratio, Earth/Moon	M_E/M_C	81.30		3	3
		$81.30059 \pm .00083$		4	12

Earth and Sun

Definition	Symbol	Value	Units	Note	Ref.
Mass Ratio, Sun/Earth	M_S/M_E	332958.		3	3
		332945.6 ± 0.4		4	11
Ecliptic Angle		$23^{o}27'08.''26$		5	3
		.4093197551	rad		
Sun Gravity Constant	GM_S	132718×10^{15}	m^3/sec^2	3	4
		$(132712.499 \pm .015) \times 10^{6}$	km^3/sec^2	4	11

Planets

Planet	Mass Ratio (Note 3)	Ref	Mass Ratio (Note 4)	Ref	Gravity Constant $\frac{km^3}{sec^2}$	Ref
Mercury	6000000.	10	5983000. ± 25000.	11		
Venus	408000.	10	408522.66 ± 3.	12	324859.6 ± 3.0	12
Mars	3093500.	10	3098700. ± 100.	11	42830. ± 8.	13
Jupiter	1047.355	10	1047.340 ± 0.0156	17	$(1.26711 \pm 0.00003) \times 10^8$	18
Saturn	3501.6	10	3499.2 ± .4	11	$(3.7938 \pm 0.00152) \times 10^7$	18
Uranus	22869.	10	22930. ± 6.	11	$(5.821 \pm 0.0524) \times 10^6$	18
Neptune	19314.	10	19260. ± 100.	11	$(6.867 \pm 0.0343) \times 10^6$	18
Pluto	360000.	10	1812000. ± 40000.	11	$(7.32 \pm 1.83) \times 10^4$	18

Notes

1. The time shown is the Mean Solar Time (UTI) required for one rotation of the earth as determined by the fixed stars (inertial rotation). As of 1 January 1972 the duration of the second has been changed from one dependent upon the duration of the Mean Solar Day to Atomic Time. Because of this a more exact method of determining the earth's rotation period and rotation rate is shown here:

$$\text{period} = (\text{period as shown})/(1+\Delta UT)$$

$$\text{rate} = (\text{rate as shown})(1+\Delta UT)$$

where ΔUT is determined from the United States Naval Observatory time service publication, Series 7 or Series 11.

2. The time shown is the Mean Solar Time (UTI) required for one rotation of the earth as defined by the First Point of Aries (sidereal rotation). A more exact value of this time is:

$$\left[23^h56^m(04.09053821-.000005059T)^s\right]/(1+\Delta UT)$$

for rotation period, and

$$(.0000729211585468+.428\times10^{-14}T)(1+\Delta\dot{U}T)\ \text{rad/sec}$$

for rotation rate, where $\Delta\dot{U}T$ is defined in Note 1 above and T is the number of Julian Centuries of 36525 U.T. Days from mean noon U.T. on 0 January 1900.

Notes (cont'd)

3. The value agreed upon by the International Astronomical Union at Hamburg, 1964.

4. The value computed by Jet Propulsion Labs from the latest satellite and radar data.

5. A more exact value of this angle is:

   $23^{\circ}27'(08.26-46.845T-.0059T^{2}+.00181T^{3})"$

   where T is the number of Julian Centuries of 36525 Ephemeris Days from the epoch of 1900 January 0.5 E.T.

Glossary of Geodetic Terms

Datum Origin - The reference position from which angles and distances are reckoned.

Geocentric Coordinates - (terrestrial) Coordinates that define the position of a point with respect to the center of the earth. Geocentric Coordinates can be either Cartesian (X,Y,Z) or spherical (latitude, longitude and radial distance). Also called geocentric coordinate system; geocentric position.

Geodetic Datum - A reference surface consisting of five quantities: the latitude and longitude of an initial point, the azimuth of a line from this point, and two constants necessary to define the reference spheroid. It forms the basis for the computation of horizontal - control surveys in which the curvature of the earth is considered. Also called horizontal control datum; horizontal datum; horizontal geodetic datum.

Spheroid - A mathematical figure closely approaching the geoid in form and size and used as a surface of reference for geodetic surveys.

NOTE: Other geodetic terms can be found in reference 16.

REFERENCES

1. Handbook of Tables for Mathematics, Third Edition, Chemical Rubber Publishing Co., Cleveland, 1967.

2. The International System of Units (SI), Physical Constants and Conversion Factors, RCA Government Services, Field Support Material No. 65-FS10, March 1965.

3. The American Ephemeris and Nautical Almanac for the Year 1972, U.S. Government Printing Office, Washington, 1970.

4. Supplement to the American Ephemeris 1968, The Introduction of the IAU System of Astronomical Constants, U.S. Government Printing Office, Washington, 1966.

5. AFETR Geodetic Coordinates Manual, United States Air Force, Eastern Test Range, Pan American World Airways, Aerospace Services Division, and Radio Corporation of America, Missile Test Project, February 1972.

6. Annual Book of American Society for Testing and Materials Standards, Philadelphia, Pa., 1972.

7. Rosenfield, G.H., Geodetic Datum Conversion Techniques, RCA Mathematical Services Technical Note No. 7, 29 January 1964.

8. SP-7012, The International System of Units, Physical Constants and Conversion Factors, E.A. Mechtly, George C. Marshall Space Flight Center, October 1964.

9. A Review of Geodetic Parameters, William M, Kaula, Goddard Space Flight Center, Greenbelt, Md. X-640-63-55, May 1963.

10. Explanatory Supplement to the Ephemeris, Her Majesty's Stationary Office, London, 1961.

11. Melbourne, W.G., Mulholland, J.D., Sjogren, W.L., and Sturns, F.M., Jr., Constants and Related Information for Astrodynamic Calculations, 1968, National Aeronautics and Space Administration Technical Report 32-1306, Jet Propulsion Laboratory, Pasadena, 15 July 1968.

12. Pease, G., et al, The Mariner V Flight Path and its Determination from Tracking Data, National Aeronautics and Space Administration Technical Report 32-1363, Jet Propulsion Laboratory, Pasadena, 1 July 1969.

13. Null, G.W., Gordon, II.J., and Tito, D.A., The Mariner IV Flight Path and its Determination from Tracking Data, National Aeronautics and Space Administration Technical Report 32-1108, Jet Propulsion Laboratory, Pasadena, 1 August 1967.

14. Geodetic Ref System 1967 SP #3 of International Association of Geodesy, 1971, Paris, France.

15. Geodetic Distance and Azimuth Computations for Lines over 500 Miles, Department of the Air Force, Aeronautical Chart and Information Center Technical Report No. 80, St. Louis, December 1959.

16. Glossary of Mapping, Charting and Geodetic Terms, Department of the Army, Corps of Engineers, U.S. Army Topographic Command, Washington, D.C., December 1969.

17. Zielenbach, J.W., The Motion of (48) Doris and the Mass of Jupiter, National Aeronautics and Space Administration Technical Report 32-1382, Jet Propulsion Laboratory, Pasadena, 15 April 1970.

18. Newburn, R.L., Jr., and Gulkis, S., A Brief Survey of the Outer Planets Jupiter, Saturn, Uranus, Neptune, Pluto, and Their Satellites, National Aeronautics and Space Administration Technical Report 32-1529, Jet Propulsion Laboratory, Pasadena, 15 April 1971.

19. 1969 Smithsonian Standard Earth (II), Smithsonian Astrophysical Observatory Special Report No. 315, E.M. Gaposchkin and K. Lambeck, 18 May 1970.

20. Program Requirements for Two-Minute Integrated Doppler Satellite Navigation Solution, John B. Moffett, John Hopkins University Applied Physics Laboratory, September 1971.

21. NASA Directory of Observation Station Locations, Vol I, Goddard Space Flight Center, Greenbelt, Maryland, November 1971.

22. U.S. Standard Atmosphere, 1962, Supertendent of Documents, U.S. Government Printing Office, Washington, D.C.

23. Physical Review Letter, National Bureau of Standards, Volume 29, November 1972.