A New Method for Data Pre-Processing

S.C. Raisinghani* and K.D. Bilimoria†
Indian Institute of Technology, Kanpur, India

Introduction

IN various engineering measurements, the signals obtained from various transducers may be contaminated by interference or noise of unspecified origin. For example, in any launch vehicle mission or aircraft flight testing program, amounts of data are recorded through the telemetry link or by an onboard data acquisition system; we shall refer to such data as "raw data." Many times it is desirable and even essential to process this data prior to subsequent interpretation or analysis.

In the literature various methods are available for smoothing discrete raw data traces. One of the popular and effective ways is the Fast Fourier Transform (FFT) method[1,2] which allows transformation from the time-domain to the frequency-domain. One requirement of the FFT method is that the basic frequency content of the signal be known a priori. In some situations, the frequency content may be known approximately, e.g., in case of flight test data. In any case, the frequency contents of the signal and noise components can be arrived at on the basis of spectral analysis through the FFT.[3] Thus the method requires that the raw data be transformed into the frequency domain using the FFT. The power spectral density of the signal is plotted against frequency, and the proper cut-off frequencies are chosen so as to set the Fourier coefficients of the unwanted (noise component) frequencies equal to zero. Finally, inverse transformation through the FFT will yield the time-domain filtered (smoothed) data.

During the investigation of parameter identification through the Gauss-Newton scheme[4] in the time-domain, it was observed that data smoothing through the FFT method worked well as long as the basic frequency of the signal was comparable to or larger than the inverse of the duration of the given signal. For low frequency signals, the filtered data showed fluctuations around the original signal. One possible remedy seemed to be to use the Packing Theorem[2,5] to modify the given set of data points so as to artificially increase the total length of the signal and thereby obtain a better representation of it in the frequency domain, where the filtering is done. This approach did improve the results, but a discrepancy persisted between the original and filtered signals.

The present method was then used on these low-frequency signals, and it was found that the resulting smoothed set of data is in good agreement with the original signal. We propose to call this method the "Method of Reflection." The method is general and can be utilized in any application requiring that a smoothed set of data be obtained from noise-corrupted data.

Method of Reflection

The given discrete raw data are to be rearranged so as to double the number of points by adding mirror-image points of the given set at the end of the given sequence. Thus, if $\{G(i)\}$ is a sequence of N discrete data points, then the required sequence $\{H(i)\}$ is defined by:

$$\begin{aligned} H(i) &= G(i), & 0 \le i \le N-1 \\ &= G(2N-1-i), & N \le i \le 2N-1 \end{aligned} \qquad (1)$$

Submitted Feb. 24, 1983; revision received April 25, 1983.

*Assistant Professor, Aeronautical Engineering Department.
†Undergraduate Student, presently Graduate Student at Virginia Polytechnic Institute and State University, Blacksburg, Va.

The sequence of data points $\{H(i)\}$ is used for obtaining the filtered data by applying the *FFT* methods as mentioned earlier. After the final step of inverse transformation is carried out to recover the filtered signal in the time-domain, only the first N points are relevant and belong to the original signal; the last N points are to be ignored in any further analysis of the data.

Application of Method and Results

To illustrate the advantage of the proposed method, flight data representing the response of a business jet airplane to control inputs were simulated on a digital computer. The data were contaminated with simulated noise by generating successively uncorrelated pseudo-random numbers. These data are "raw data," to be processed to obtain smoothed data. We consider only three perturbed motion variables: yaw rate (r), bank angle (ϕ), and forward speed (u). As will be shown, the last two motion variables represent typical cases of low-frequency signals, wherein the present method is used advantageously to improve the quanlity of smoothed data.

Figure 1 shows the original and noise-contaminated signals for the three motion variables r, ϕ, and u. All signals are sampled at the rate of 50/s over a time interval of 5 s. Noise contamination was produced by adding pseudo-random numbers having a normal distribution with zero mean and assigned standard deviation, the standard deviation being chosen to correspond approximately to 10% of the maximum range of the assumed measuring instrument for each motion variable.

Using the discrete form of the FFT method, the power spectral density of each of the motion variables was calculated using method C of Ref. 3 and plotted against frequency. To facilitate the choice of the upper cut-off frequency (UCF), it was found expedient to enlarge the scale for power spectral density by a factor of 100 at higher frequencies; one such plot for ϕ is shown in Fig. 2a. This figure indicates that a low-pass filter is required. The nature of the plot for ϕ (and for u) does not sharply define the UCF to be used; the selected UCF for ϕ is indicated in Fig. 2. Once the UCF is defined, the Fourier coefficients for frequencies larger than the UCF were set to zero, and inverse transformation yielded the filtered time-domain signals. Figure 3 shows these filtered signals along with the original signals. The filtered signal for r is in good agreement with the original signal; the other two cases of u and ϕ show fluctuations of the filtered response about the original response.

Next, the Packing Theorem[2,5] was utilized for the u and ϕ signals—the signal length was doubled by adding N trailing zeroes to the given N data points, and the above procedure was repeated. The final results are shown in Fig. 3. For comparison, the power spectrum is also shown in Fig. 2b, and the earlier remarks about an ambiguity in the selection of the

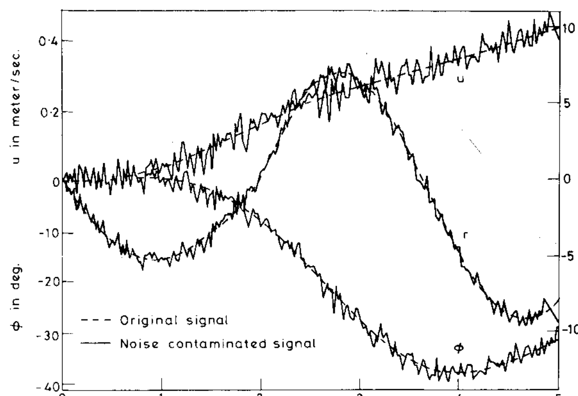

Fig. 1 Pure and noise-contaminated signals.

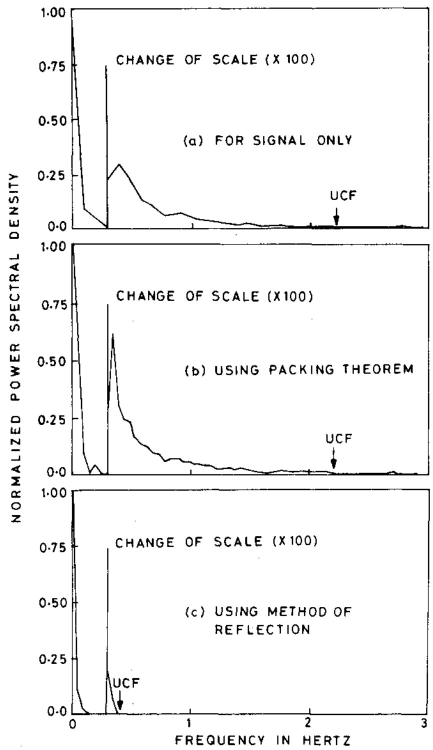

Fig. 2 Spectral estimation of roll angle (ϕ) response.

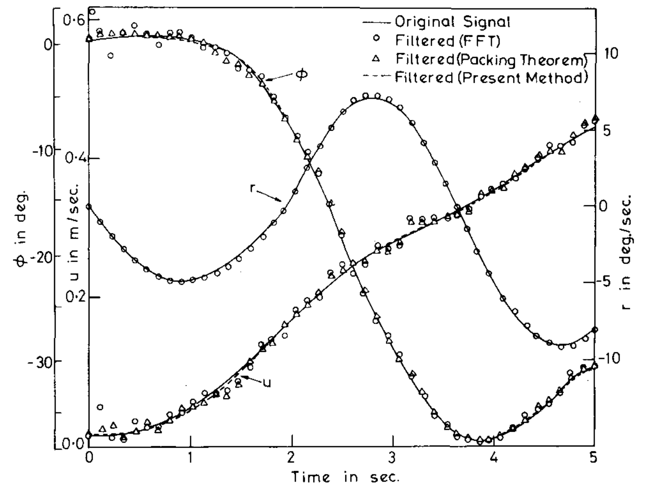

Fig. 3 Comparison of pure signals with filtered signals obtained by application of the three filtering procedures.

UCF are still valid. Although there is an improvement for both u and ϕ, the fluctuations about the original signal still persist, particularly for u.

The proposed method of reflection was now utilized for the u and ϕ signals. The resulting power spectrum is shown in Fig. 2c, while the filtered signals are shown in Fig. 3. The agreement between the filtered and original signals for both u and ϕ is indeed good, the improvement for u being more marked than for ϕ. In addition, Fig. 2c shows that the selection of the UCF for ϕ is no longer ambiguous. A similar result was obtained for u.

It appears that the lower the basic frequency of the original signal, the more beneficial the improvements obtained through application of the present method. In our illustrations, u was the signal of lowest frequency among the three signals considered, and its filtered response was improved the most by the Method of Reflection.

This observation prompted a quick-look study that consisted of generating sinusoidal signals of varying frequency, corrupting these with noise, and then applying the three filtering procedures discussed above. From this study, it was concluded that the Method of Reflection is warranted when the basic frequency of the signal is below $0.5/NT$ Hz, where N is the total number of data points and T is the sampling interval.

The above study showed that the Method of Reflection resulted in a power spectrum which packed the signal content in a well-defined frequency range (Fig. 2c), in direct contrast to the other two methods where the power spectra showed "spreads" over a large frequency range (Figs. 2a-b). Various suitable upper cut-off frequencies (e.g., from 0.8-2.5 Hz for ϕ) were tried from this "spread" but the results were always unsatisfactory. It is our conjecture that a high UCF selected from the "spread" leads to the leakage of the noise-components into the filtered signal whereas a low UCF chops off a part of the signal itself. The unambiguous selection of the UCF from the power spectrum generated by the present method seems to result in an efficient filtering-out of the noise components while preserving the signal components intact.

In conclusion, the proposed method uses the given data points in a modified format and recovers the original signal, to a good approximation, by filtering out noise. The method is quite general and can be applied to any noise-contaminated set of recorded data to obtain smoothed data, the smoothing process being more profitable as compared to other similar methods for signals of low frequency.

References

[1] Brigham, E.O., *The Fast Fourier Transform,* Prentice-Hall, Inc., N.J., 1974, pp. 148-196.

[2] Bracewell, R.N., *The Fourier Transform and its Applications,* 2nd ed., McGraw-Hill, New York, 1978, pp. 356-384.

[3] Yuen, C.K., "A Comparison of Five Methods for Computing the Power Spectrum of a Random Process Using Data Segmentation," *Proceedings of the IEEE,* Vol. 65, June 1977, pp. 984-986.

[4] Adak, A.K., "Aircraft Parameter Estimation from Flight Test Data by Gauss-Newton Optimization Technicque," M.Tech. Thesis, Indian Institute of Technology, Kanpur, India, 1982.

[5] Wells, W.R. and Banda, S.S., "Data Analysis for Aircraft Parameter Estimates," *AIAA Atmospheric Flight Mechanics Conference,* Albuquerque, New Mexico, Aug. 19-21, 1981.