Signal to Noise Instrumental Excel Assignment

Instrumental methods, as all techniques involved in physical measurements, are limited by both the precision and accuracy. The precision and accuracy of a measurement are ultimately limited by two factors imposed by nature- matter has thermal fluctuations and charge, and light and energy are quantized (1). Measurement devices can be made smaller and smaller, but ultimately the size of the transducer will be limited by the physical parameter that it is intended to measure. For example, the electrochemical measurement of an equilibrium constant utilizing a large surface area electrode (>1 mm diameter) is routinely accomplished for solutions at relatively high concentrations (> 100 μM), and the precision of such a device is usually very good. However, only a finite number of molecules will interact with the probe when this same measurement is made using an ultramicroelectrode (< 25 μm diameter) on a dilute solution (< 1 μM). This latter experiment poses significant problems with noise, and precision is likely to be poor.

You are already familiar with processing data with random noise. For example, spreadsheets are used routinely to determine the best fit line through data that follows a linear trend. The line represents the smallest deviation of the data from the best fit line, as calculated via a least squares method. But what if the data is noisy and not linear (e.g. an infrared spectrum of a dilute solution)? What are the options for reducing the noise?
The figure below illustrates noise that is superimposed on a hypothetical signal. The noise is a measure of precision, and a lot of effort is taken to reduce its influence on a signal. For example, an average of many measurements over time in many instances can reduce thev influence of noise. Note that for a static measurement, the noise is the standard deviation ($s = (\Sigma(x_i-x)^2)/(n-1))^{1/2}$). Some instruments are designed to take advantage of a particular frequency region, and then the signal-to-noise (S/N) is manipulated. There are a number of methods that can be used to improve the S/N ratio.

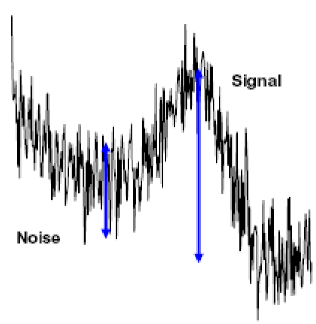

Figure 1. Illustration of a signal containing noise.

Signal processing entails manipulating the data, and often times enhancing the signal-to noise ratio. There are both physical (electronic) and digital methods for enhancing the signal-to noise ratio. Physical methods include, but are not limited to grounding, analog filtering, and modulation. More detailed information about these electronic methods is found in the literature

(1, 2). Digital methods include, but are not limited to moving averages, signal averaging, and Fourier transforms. Two common moving average methods are unweighted moving and weighted (Savitzky-Golay) averaging.

Moving Average. Smoothing algorithms involving unweighted averages can be performed in various ways. One method, known as boxcar averaging, involves collecting a specified number of points and averaging them to produce a single point. Another method involves a "moving" average, where a specified number of successive points (n) are collected and averaged, then the next measurement is averaged with the previous n-1 measurements, and this process continues through the data set.

Many instruments automatically take a number of readings, and digital displays or data file output represent the average of these multiple measurements. Usually signal averaging can be controlled by either the software or hardware, or both. For example, for a continuous reading from an atomic absorption spectrometer, an experimenter may manually set the instrument to display the average of 100 measurements made over 1-2 seconds.

Moving averages are appropriate for continuous output or static measurements (e.g. absorbance or voltage reading) that are not changing with time. Most modern digital acquisition (DAQ) boards can easily collect 100,000 data points per second (every 10 μs), average a specified number of these points, and report a single data point representing the average. Since DAQ boards are capable of obtaining data so rapidly, it is often appropriate to use moving averages of dynamic data. Moving averages are only appropriate if the data is collected rapidly relative to the history (change in data). Figure 2 illustrates the effect of averaging 100 points, 10 points and 1 point on voltage data collected versus time.

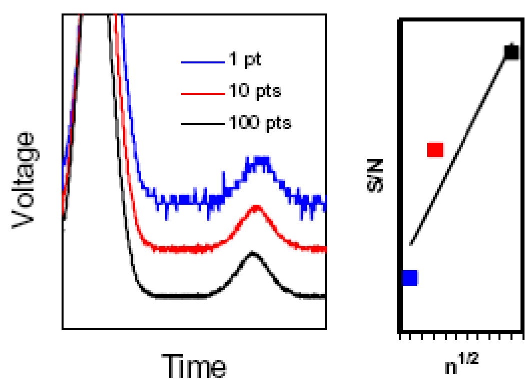

Figure 2. Signals composed of sequential points averaged and the signal-to-noise ratio for this data as a function of the square root of the number of points averaged (1-top, 10-middle, 100-lower trace). The solid line is a linear regression of the data.

Savitzky Golay. Moving averages enable the visualization of historical trends, and are appropriate to use when the sampling rate is much faster than the rate of change in the data. Fine structural features in the data may be lost if a moving average is inappropriately applied. The Savitzky-Golay algorithm was developed to smooth nonlinear data using a weighted moving average to minimize the loss of fine structural details. This algorithm was developed by chemists, not mathematicians, in order to improve the signal-to-noise ratio on the first infrared spectrometers. The Savitzky-Golay algorithm is a signal averaging/least squares method used to increase signal-to-noise ratio. This least squares computation uses a set of integers to minimize the influence of noise within the signal. Each data point is first multiplied by a convolution integer (C_i), and then summed with its neighbors, and the total is divided by a normalization integer. Thus, the resultant data point represents a weighted average. The convolution and normalization integers are found in the literature (3), and examples for five and seven point smooths are listed below:

	Convolution integer (C_i)							Normalization integer
5 pt		-3	12	17	12	-3		**35**
7 pt	**-2**	3	6	7	6	3	**-2**	**21**

Specifically, in a five point average, each number is multiplied by the appropriate convolution integer (e.g. C_i in the table above). The sum of these numbers is divided by the sum of these integers or the normalization integer:

$$(\Sigma_{i=1}^{\,i=5}\ C_iX_i)/35$$

For example, in a 5 point smoothing routine, the quantity

$$(-3x_1 + 12x_2 + 17\ x_3\ \ldots)/35,$$

is the smoothed value of the data at the point x_3 .This process is carried out on the next data point:

$$(\Sigma_{i=2}^{\,i=6}\ C_{(i-1)}X_i)/35 \text{ or } (-3x_2 + 12x_3 + \ldots)/35,$$

And continues to the end of the data:

$$(\Sigma_{i=n-4}^{\,i=n}\ C_{(i-n+5)}X_i)/35.$$

In this manner each data point represents a weighted average, with the central data point given the most significance. Contrast this with a moving average, in which all points are equally weighted.

Practical considerations of moving averages. The increase in S/N is directly related to the square root of the number of points used in the smooth for an unweighted smooth, whereas the Savitzky Golay is some fraction of that. The noise can be calculated by determining the standard deviation of a flat region of the signal. There is a tradeoff between increasing the S/N and signal distortion, and a detailed description of these tradeoffs are described in the literature (3, 4). Usually, multiple passes of a moving average with a small window (e.g. two passes of a 5 point Savitzky Golay) will retain low frequency signals, but the tradeoff is that the S/N improvement will not be as good as a single pass with a larger window (e.g. a single pass with a 9 point Savitzky Golay).

Ensemble averaging. Another type of averaging technique involves taking the average of entire ensembles of data. For example in nuclear magnetic resonance (NMR) spectroscopy, 10,000 scans can be routinely collected over the span of a few hours, and the average of the entire spectra is determined. This process can be represented as follows:

transforming back to the time domain (the "reverse" or "inverse" Fourier transform) we obtain a new set of data that appears to correspond to the original set, but with the periodic noise removed. In the problem set below, you will have the opportunity to carry out a Fourier transform on data containing periodic noise.

The following questions require data contained in Excel files (RSData***n***.xls), where n is specified in the problem. Utilize the help function if you have trouble executing any function.

1. ***Moving Averages***. Raw data is contained in RSData1.xls. To determine the S/N, find a flat region of the chromatogram and use the STDEV function to get the standard deviation. Divide a peak height (use the same one) by the standard deviation.

A. Smooth and plot the data using a 5 point moving average (hint: use the AVERAGE function in Excel). How much was the S/N improved?
B. Smooth and plot the data using and a 5 point Savitzky-Golay average. Use the convolution and normalization integers in the text. Formulate an equation in Excel from the equations described in the text (hint: you will need to use the "$" symbol as a place holder in your formula for your convolution and normalization integers).
How much was the S/N improved?

2. ***Ensemble Average***. Raw data of 5 replicate chromatograms are provided in a data file (RSData2.xls). Use an ensemble average to improve the signal-to-noise ratio. What is the advantage/disadvantages of using this method? How much was the signal/noise improved?

3. ***Fourier Transform***. Process the provided data using a Fourier transform by removing part of the periodic noise, and most of the periodic noise (RSData3.xls). Note that the data to be transformed must be some integral power of 2 (i.e. n = 2x, where x is some whole number). The

steps are as follows:

A. Select Tools → DataAnalysis →Fourier Analysis.
B. Select input and output range.
C. The fast Fourier Transform produces imaginary numbers ($x + iy$), and these numbers are converted to a real number via the IMABS function for plotting and manipulation.
D. To remove periodic noise, a column of 0's and 1's are created. Then multiply the low frequency data by 1 and the high frequency data by 0 using the IMPRODUCT function. Keep in mind that the Fourier transform produces a mirror image of your data. You need to keep this in mind when you filter your data (i.e. don't remove the signal at low frequencies and its mirror image).
E. The reverse transform is performed by repeating step A and B, and selecting the inverse box. The IMABS function is again need to convert imaginary numbers back to real numbers.
F. The frequency is calculated using the following steps:

1. Make a column of numbers from $n = 0 \ldots N-1$, where $N = 2x$ is the number of pts.
2. The points should be equally-spaced in time, with spacing Δt.
3. The maximum frequency is $f_{max} = 1/\Delta t$, which is not included because it corresponds to a frequency of zero.
4. Make a column of frequencies according to the following formula:
$f_n = (n/N) f_{max} = n/(N \Delta t)$.