Curve Fitting

The simplest method for fitting a curve to data is to plot the points and then sketch a line

- (a) Characterize the general upward trend of the data with a straight line
- (b) Use straight-line segment or linear interpolation
- (c) Use curves to try to capture the meanderings

Simple statistics

- Arithmetic mean

\[
\bar{y} = \frac{\sum y_i}{n}
\]

- Standard deviation: the measure of spread of a sample
\[ s_y = \sqrt{\frac{S_t}{n-1}} \]

where \( S_t \) is the total sum of the squares of the residual between the data points and the mean, or

\[ S_t = \sum (y_i - \bar{y})^2 \]

- **Variance**: The square of the standard deviation

\[ s^2 = \frac{S_t}{n-1} \]

- **Coefficient of variation (c.v.)**: The spread of data

\[ c.v. = \frac{s_y}{\bar{y}} \times 100\% \]

**Least-Squares Regression**

Least-squares regression is derived from a curve that minimized the discrepancy between the data points and the curve.

**Linear Regression**

A least-squares approximation is fitting a straight line to a set of paired observation. The mathematical expression for the straight line is

\[ y = a_0 + a_1 x + e \quad (5.1) \]

The error, or residual, is the discrepancy between the true value of \( y \) and the approximate value, \( a_0 + a_1 x \) and that is

\[ e = y - a_0 + a_1 x \quad (5.2) \]

The criterion for least-squares regression is

\[ \min S_r = \sum_i^m e_i^2 = \sum_i^m (y_{i,\text{measured}} - y_{i,\text{model}})^2 = \sum_i^m (y_i - a_0 - a_1 x_i)^2 \quad (5.3) \]

To determine values of \( a_0 \) and \( a_1 \), differentiate (5.3)

\[ \frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i) \quad (5.4) \]

\[ \frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i)x_i] \quad (5.5) \]

And setting these derivatives equal to zero, we get the so-called normal equations
The coefficients of a straight line are

\begin{align}
0 &= y_i - \sum \alpha_0 - \sum \alpha_1 x_i \\
0 &= \sum x_i y_i - \sum \alpha_0 x_i - \sum \alpha_1 x_i^2
\end{align}

(5.6)
(5.7)

Quantification of error of linear regression

- The sum of the square of the residual
  - A sampled data system
    \[ S_t = \sum (y_i - \bar{y})^2 \]
  - A linear regressioned system
    \[ S_r = \sum (y_i - \alpha_0 - \alpha_1 x_i)^2 \]

- Standard deviation
  - A sampled data system
    \[ s_y = \sqrt{\frac{S_t}{n - 1}} \]
  - A linear regressioned system
    \[ s_{y/x} = \sqrt{\frac{S_r}{n - 2}} \]

\( s_y \) quantifies the spread around mean.
\( s_{y/x} \) quantifies the spread around the regression line.

- The goodness of a fit

\[ r^2 = \frac{S_t - S_r}{S_t} \]

where \( r^2 \) is called the coefficient of determination and \( r^* \) is the correlation coefficient.

See the figure 17.4 in the textbook.
General Linear Least-Squares

The general linear least-square model:

\[ y = a_0 z_0 + a_1 z_1 + \cdots + a_m z_m + e \]  \hfill (5.15)

In matrix notation

\[ Y = ZA + E \]  \hfill (5.16)

Note that \( Z \) is not a square matrix but we want to know about \( A \).

\[ Z^T Z A = Z^T Y \]  \hfill (5.17)

Now \( A \) is

\[ A = (Z^T Z)^{-1} Z^T Y \]  \hfill (5.18)

Nonlinear Regression

Gauss-Newton method

1. Use a Taylor series to linearize a nonlinear function
2. Apply least-square theory to obtain new estimate of the parameters that move in the direction of minimizing the residual.

Interpolation

Newton's Divided-Difference Interpolating Polynomials

Linear interpolation: connect two data points with a straight line

\[ f_1(x) = f(x_0) + \frac{f(x_1) - f(x_0)}{x_1 - x_0}(x - x_0) \]  \hfill (5.19)
Quadratic interpolation: connect three data points with a second-order polynomial

\[ f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1) \]  
\text{(5.20)}

where

\[
\begin{align*}
  b_0 &= f(x_0) \\
  b_1 &= \frac{f(x_1) - f(x_0)}{x_1 - x_0} \\
  b_2 &= \frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}
\end{align*}
\]

Newton's interpolating polynomial: connect \( n + 1 \) data with \( n \)-th-order polynomial

\[ f_n(x) = b_0 + b_1(x - x_0) + \cdots + b_n(x - x_0) \cdots (x - x_{n-1}) \]  
\text{(5.21)}

where the coefficients are

\[
\begin{align*}
  b_0 &= f(x_0) \\
  b_1 &= f[x_1, x_0] \\
  &\vdots \\
  b_n &= f[x_n, x_{n-1}, \ldots, x_0]
\end{align*}
\]

where the bracket function evaluations are finite divided differences.

\( n \)-th finite divided difference is

\[
f[x_n, x_{n-1}, \ldots, x_0] = \frac{f[x_n, \ldots, x_1] - f[x_{n-1}, \ldots, x_0]}{x_n - x_0} \]  
\text{(5.22)}

Newton's divided-difference interpolating polynomial is

\[
f_n(x) = f(x_0) + (x - x_0)f[x_1, x_0] + \cdots (x - x_0)(x - x_1) \cdots (x - x_{n-1})f[x_n, \ldots, x_0] \]  
\text{(5.23)}

### Lagrange Interpolating Polynomial

The Lagrange interpolating polynomial is simply a reformulation of the Newton polynomial that avoids the computation of divided differences.

\[
f_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i) \]  
\text{(5.24)}

where
where \( \prod \) designates the "product of."

### Spline Interpolation

Spline interpolation is an alternative approach that lower-order polynomial is applied to subsets of data point. Especially, when third-order curves are employed to connect each pair of data points, it is called cubic spline.

**Linear splines**: the simplest connection between two points is a straight line.

\[
f(x) = f(x_0) + m_0(x - x_0) \quad x_0 \leq x \leq x_1
\]

\[
f(x) = f(x_1) + m_1(x - x_1) \quad x_1 \leq x \leq x_2
\]

\[\vdots\]

\[
f(x) = f(x_{n-1}) + m_1(x - x_{n-1}) \quad x_{n-1} \leq x \leq x_n
\]

where \( m_i \) is the slope of the straight line

\[
m_i = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} \quad (5.26)
\]

**Quadratic splines**: connect three points with second-order polynomials.

- The function values of adjacent polynomials must be equal at the interior knots.
- The first and last functions must pass through the end points.
- The first derivatives at the interior knots must be equal.
- Assume that the second derivative is zero at the first point.

**Cubic splines**: derive a third-order polynomial for each interval between knots

\[
f_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \quad (5.27)
\]

### Fourier Approximation

In early 1800s, the French mathematician Fourier proposed that "any function can be represented by an infinite sum of sine and cosine terms." There are functions that do not have a representation as a Fourier series, however, most functions can be so represented. Fourier approximation is another representation of a function with trigonometric series.

**Trigonometric identities**

- \( \sin A \sin B = \frac{1}{2} [\cos(A - B) - \cos(A + B)] \)
- \( \sin A \cos B = \frac{1}{2} [\sin(A - B) + \sin(A + B)] \)
- \( \cos A \cos B = \frac{1}{2} [\cos(A - B) + \cos(A + B)] \)

Fourier series
Assume that \( f(x) \) is a periodic function of period \( 2\pi \) and is integrable over a period.

\[
f(x) \simeq A_0 + \sum_{n=1}^{\infty} [A_n \cos(nx) + B_n \sin(nx)]
\] (5.28)

- \( A_0 \): integrating on both sides of (5.28) from \(-\pi\) to \(\pi\)

\[
\int_{-\pi}^{\pi} f(x) dx = \int_{-\pi}^{\pi} A_0 dx + \sum_{n=1}^{\infty} A_n \int_{-\pi}^{\pi} \cos(nx) dx + \sum_{n=1}^{\infty} B_n \int_{-\pi}^{\pi} \sin(nx) dx
\]

The last two integrations of trigonometric terms are equal to zero. Hence

\[
A_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) dx
\]

- \( A_n \): multiply both sides of (5.28) by \( \cos(mx) \) and integrate

\[
\int_{-\pi}^{\pi} \cos(mx)f(x) dx = \int_{-\pi}^{\pi} A_0 \cos(mx) dx + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} A_n \cos(nx) \cos(mx) dx + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} B_n \sin(nx) \cos(mx) dx
\] (5.1)

The only nonzero term on the right is when \( m = n \) in the first summation

\[
A_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx
\]

- \( B_n \): multiply both sides of (5.28) by \( \sin(mx) \) and integrate

\[
\int_{-\pi}^{\pi} \sin(mx)f(x) dx = \int_{-\pi}^{\pi} A_0 \sin(mx) dx + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} A_n \cos(nx) \sin(mx) dx + \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} B_n \sin(nx) \sin(mx) dx
\] (5.2)

The only nonzero term on the right is when \( m = n \) in the second summation

\[
B_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx
\]

Fourier series for any period \( p = 2L \)

Consider the function whose period is \( p = 2L \)
where the Fourier coefficients of $f(x)$ are given by the Euler formulas

$$
A_0 = \frac{1}{2L} \int_{-L}^{L} f(x) \, dx \\
A_n = \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi}{L} x \, dx \\
B_n = \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi}{L} x \, dx
$$

(5.34)

Fourier series for even and odd functions

- Even function:

  $$
g(-x) = g(x)
  $$

  And integral value of a even function is

  $$
\int_{-L}^{L} g(x) \, dx = 2 \int_{0}^{L} g(x) \, dx
  $$

- Odd function:

  $$
h(-x) = -h(x)
  $$

  And integral value of a even function is

  $$
\int_{-L}^{L} h(x) \, dx = 0
  $$

- Fourier cosine series: the Fourier series of an even function of period $2L$.

  $$
f(x) = A_0 + \sum_{n=1}^{\infty} A_n \cos \frac{n\pi}{L} x
  $$

- Fourier sine series: the Fourier series of an odd function of period $2L$.

  $$
f(x) = A_0 + \sum_{n=1}^{\infty} B_n \sin \frac{n\pi}{L} x
  $$

Complex form of Fourier series: Real sines and cosines can be expressed in terms of complex exponentials by the formulas

(5.41)
\[
\sin nx = \frac{e^{inx} - e^{-inx}}{2i}, \\
\cos nx = \frac{e^{inx} + e^{-inx}}{2}
\]

From this

\[
A_n \cos nx + B_n \sin nx = \frac{1}{2} A_n (e^{inx} + e^{-inx}) + \frac{1}{2i} B_n (e^{inx} - e^{-inx}) \\
= \frac{1}{2} (A_n - iB_n) e^{inx} + \frac{1}{2} (A_n + iB_n) e^{-inx} \\
= c_n e^{inx} + c_{-n} e^{-inx}
\]

(5.42)

(5.43)

(5.44)

With the above equation

\[
f(x) = \sum_{-\infty}^{\infty} c_n e^{inx}
\]

(5.45)

where

\[
c_n = A_n - iB_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) (\cos(nx) - i\sin(nx)) dx = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-inx} dx
\]

This is the so-called complex form of the Fourier series, or complex Fourier series of \( f(x) \).

Sinusoidal function: represent any waveform with a sine or cosine

\[
f(t) = A_0 + C_1 \cos(\omega_0 t + \theta)
\]

(5.46)

where \( A_0 \) is the mean value, \( C_1 \) is the amplitude, \( \omega_0 \) is the angular frequency, and \( \theta \) is the phase angle or phase shift.

The angular frequency is related to frequency \( f \) (in cycles/time)

\[
\omega_0 = 2\pi f
\]

(5.47)

and frequency is

\[
f = \frac{1}{T}
\]

(5.48)

The trigonometric identity gives

\[
f(t) = A_0 + A_1 \cos(\omega_0 t) + B_1 \sin(\omega_0 t)
\]

(5.49)

where \( A_1 = C_1 \cos(\theta) \), \( B_1 = -C_1 \sin(\theta) \).
Curve Fitting with Sinusoidal Functions

Least-squares fit of a sinusoidal function is to determine coefficient values that minimize

\[ S_r = \sum_{i=1}^{N} \left\{ y_i - \left[ A_0 + A_1 \cos(\omega_0 t_i) + B_1 \sin(\omega_0 t_i) \right]\right\}^2 \]  

(5.50)

\[
\begin{bmatrix}
N & \sum \cos(\omega_0 t) & \sum \sin(\omega_0 t) \\
\sum \cos(\omega_0 t) & \sum \cos^2(\omega_0 t) & \sum \cos(\omega_0 t) \sin(\omega_0 t) \\
\sum \sin(\omega_0 t) & \sum \cos(\omega_0 t) \sin(\omega_0 t) & \sum \sin^2(\omega_0 t)
\end{bmatrix}
\begin{bmatrix}
A_0 \\
A_1 \\
B_1
\end{bmatrix}
=
\begin{bmatrix}
\sum y \\
\sum y \cos(\omega_0 t) \\
\sum y \sin(\omega_0 t)
\end{bmatrix}
\]

(5.51)

For equispaced system

\[
\int_0^T \cos(\omega_0 t) dt = \left. \frac{1}{\omega_0} \sin(\omega_0 t) \right|_0^T = 0
\]

(5.52)

where \( \omega_0 T = \frac{2\pi}{T} = 2\pi \). These relationships give

\[
\begin{bmatrix}
N & 0 & 0 \\
0 & N/2 & 0 \\
0 & 0 & N/2
\end{bmatrix}
\begin{bmatrix}
A_0 \\
A_1 \\
B_1
\end{bmatrix}
=
\begin{bmatrix}
\sum y \\
\sum y \cos(\omega_0 t) \\
\sum y \sin(\omega_0 t)
\end{bmatrix}
\]

(5.53)

or

\[ A_0 = \frac{\sum y}{N} \]  

(5.54)

\[ A_1 = \frac{2}{N} \sum y \cos(\omega_0 t) \]  

(5.55)

\[ A_2 = \frac{2}{N} \sum y \sin(\omega_0 t) \]  

(5.56)

The above equations are similar with the determination of Fourier series.
Fourier Integral and Transform

Some of phenomenon does not occured repeatedly or it will be a long time until it occurs again. In this case we use Fourier integral that can be used to represent nonperiodic functions, for example a single voltage pulse not repeated, or a flash of light, or a sound which is not repeated. The transition from a periodic to a nonperiodic function can be effected by allowing the period to approach infinity. In other words, as $T$ becomes infinite, the function never repeats itself and thus becomes aperiodic.

From Fourier series to the Fourier integral

Consider any periodic function $f_L(x)$ of period $2L$

$$f_L(x) = A_0 + \sum_{n=1}^{\infty} \left( A_n \cos \omega_n x + B_n \sin \omega_n x \right)$$

(5.57)

where $\omega_n = n\pi / L$. Insert $A_n$ and $B_n$ which are given by the Euler formulas.

$$f_L(x) = \frac{1}{2L} \int_{-L}^{L} f_L(v) dv + \frac{1}{L} \sum_{n=1}^{\infty} \left[ \cos \omega_n x \int_{-L}^{L} f_L(v) \cos \omega_n v dv + \sin \omega_n x \int_{-L}^{L} f_L(v) \sin \omega_n v dv \right]$$

(5.3)

Now set

$$\Delta \omega = \omega_{n+1} - \omega_n = \frac{(n + 1)\pi}{L} - \frac{n\pi}{L} = \frac{\pi}{L}$$

(5.58)

Then $1/L = \Delta \omega / \pi$, and

$$f_L(x) = \frac{1}{2L} \int_{-L}^{L} f_L(v) dv + \frac{1}{\pi} \sum_{n=1}^{\infty} \left[ \cos \omega_n x \Delta \omega \int_{-L}^{L} f_L(v) \cos \omega_n v dv + \sin \omega_n x \Delta \omega \int_{-L}^{L} f_L(v) \sin \omega_n v dv \right]$$

(5.4)
Let $L \to \infty$ and assume a periodic function $f_L(x)$ to be a aperiodic function.

$$f(x) = \lim_{L \to \infty} f_L(x) \quad (5.59)$$

Then $1/L \to 0$ and the first term of function approaches zero.

$$f_L(x) = \frac{1}{\pi} \sum_{n=1}^{\infty} \left[ \cos \omega_n x \Delta \omega \int_{-L}^{L} f_L(v) \cos \omega_n v dv + \sin \omega_n x \Delta \omega \int_{-L}^{L} f_L(v) \sin \omega_n v dv \right] \quad (5.60)$$

$L \to 0$ results in $\Delta \omega \to 0$ and the sum of infinite series become an integral from 0 to $\infty$.

$$f(x) = \frac{1}{\pi} \int_{0}^{\infty} \left[ \cos \omega x \int_{-\infty}^{\infty} f(v) \cos \omega v dv + \sin \omega x \int_{-\infty}^{\infty} f(v) \sin \omega v dv \right] d\omega \quad (5.61)$$

Introduce $A(\omega)$ and $B(\omega)$ as

$$A(\omega) = \int_{-\infty}^{\infty} f(v) \cos \omega v dv, \quad B(\omega) = \int_{-\infty}^{\infty} f(v) \sin \omega v dv \quad (5.62)$$

Finally Fourier series for an aperiodic equation become

$$f(x) = \int_{0}^{\infty} [A(\omega) \cos \omega x + B(\omega) \sin \omega x] d\omega \quad (5.63)$$

This is called a representation of $f(x)$ by a Fourier integral.

Alternatively, the Fourier integral can be written as complex Fourier series.

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n x} \quad (5.64)$$

$$c_n = \frac{1}{2L} \int_{-L}^{L} f(u) e^{-i\omega_n u} du$$

$$f(x) = \sum_{n=-\infty}^{\infty} \left[ \frac{1}{2L} \int_{-L}^{L} f(u) e^{-i\omega_n u} du \right] e^{i\omega_n x} \quad (5.65)$$

Use $1/L = \Delta \omega / \pi$

$$f(x) - \sum_{n=-\infty}^{\infty} \left[ \frac{\Delta \omega}{2\pi} \int_{-L}^{L} f(u) e^{-i\omega_n u} du \right] e^{in\omega} \quad (5.66)$$

(5.67)
where

\[ F(\omega_n) = \int_{-L}^{L} f(u) e^{i\omega_n(x-u)} du \]  \hspace{1cm} (5.68)

If \( \Delta \omega \) goes to zero, a limit of a sum becomes an integral

\[ f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\omega) d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(u) e^{i\omega(x-u)} du d\omega \]  \hspace{1cm} (5.69)

\[ = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{i\omega x} d\omega \int_{-\infty}^{\infty} f(u) e^{-i\omega u} du \]  \hspace{1cm} (5.70)

Define \( g(\omega) \) by

\[ g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(u) e^{-i\omega u} du \]  \hspace{1cm} (5.71)

Then

\[ f(x) = \int_{-\infty}^{\infty} g(\omega) e^{i\omega x} d\omega \]  \hspace{1cm} (5.72)

**Fourier Transform**

\[ f(x) = \int_{-\infty}^{\infty} g(\omega) e^{i\omega x} d\omega \]  \hspace{1cm} (5.73)

\[ g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(u) e^{-i\omega u} du \]

\( f(x) \) and \( g(\omega) \) are called a pair of Fourier transforms. Usually, \( g(\omega) \) is called the Fourier transform of \( f(x) \) and \( f(x) \) is called the inverse Fourier transform of \( g(\omega) \).

**Discrete Fourier Transform (DFT)**

In engineering, functions are often represented by finite sets of discrete values and data is often collected in or converted to such a discrete format. For the discrete time system, a discrete Fourier transform can be written as

\[ F_k = \sum_{n=0}^{N-1} f_n e^{-i\omega kn} \]  \hspace{1cm} (5.74)

and the inverse Fourier transform as
where $\omega_0 = \frac{2\pi}{N}$.

**Fast Fourier Transform (FFT)**

The fast Fourier transform (FFT) is an algorithm that has been developed to compute the DFT in an extremely economical fashion.

**The Power Spectrum**

A power spectrum is developed from the Fourier transform and it is derived from the analysis of the power output of electrical systems. The power of a periodic signal can be defined as

$$P = \frac{1}{T} \int_{-T/2}^{T/2} f^2(t) dt$$  \hspace{1cm} (5.76)

A power spectrum can be calculated by the power associated with each frequency component.

**Curve Fitting with Libraries and Packages**

- Matlab:
  - polyfit
  - polyval
  - poly2sym
  - interp1
  - spline
  - fft
- IMSL: various routines are exist to solve curve fitting and fft problems

**Engineering Applications: Curve Fitting**

See the textbook