Computational Science:
Computational Methods in Engineering

Introduction to Curve Fitting

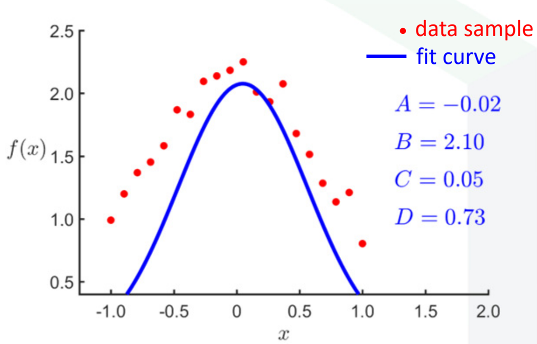
Outline

- What is curve fitting and why do we care?
- Statistics of Data Sets

What is Curve Fitting & Why Do We Care?

What is Curve Fitting?

Curve fitting is simply fitting an analytical equation to a set of measured data.



$$f(x) = A + Be^{-\left(\frac{x-C}{D}\right)^2}$$

"Curve fitting" determines the values of A, B, C, and D so that $f(x)$ best represents the given data.

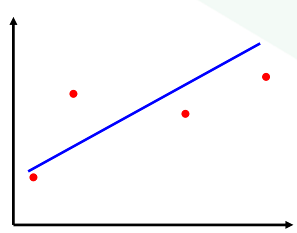
EMPossible

Why Fit Data to a Curve?

- Observe and quantify general trends.
- Remove noise from a function.
- Fit measured data to an analytical equation to extract meaningful parameters.
- Estimate data between discrete values (interpolation)
- Find a maximum or minimum.
- Deriving finite-difference approximations.

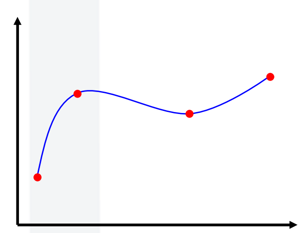
Two Categories of Curve Fitting

Best Fit – Measured data has noise so the curve does not attempt to intercept every point.



- Linear regression (ugly math)
- Linear least-squares (clean math)
- Nonlinear regression (moderate math)

Exact Fit – Data samples are assumed to be exact and the curve is forced to pass through each one.



- Fitting to polynomials

Statistics of Data Sets

Arithmetic Mean

If there was a single number that best represents an entire set of data, the arithmetic mean would be it.

$$f_{\text{avg}} = \frac{f_1 + f_2 + \cdots + f_M}{M} = \frac{1}{M}\sum_{m=1}^{M} f_m$$

Geometric Mean

The geometric mean is defined as

$$f_{\text{gm}} = \sqrt[M]{f_1 f_2 \cdots f_M}$$

The arithmetic mean tends to suppress the significance of outlying data samples. With the geometric mean, even a single small value among many large values can dominate the mean.

This is useful in optimizations where multiple parameters must be maximized at the same time and it is not acceptable to have any one of them low.

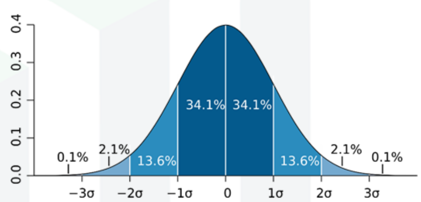
EMPossible 9

Variance & Standard Deviation

Standard Deviation σ_f

The standard deviation is a measure of the "spread" of the data about the mean. It is convenient because it shares the same units as the data.

$$\sigma_f = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(f_m - f_{\text{avg}}\right)^2}$$



Variance v_f

Variance is used more commonly in calculations, but carries the same information as the standard deviation.

$$v_f = \sigma_f^2 = \frac{1}{M}\sum_{m=1}^{M}\left(f_m - f_{\text{avg}}\right)^2$$

EMPossible 10

Coefficient of Variation

The *coefficient of variation* (CV) is the standard deviation normalized to the mean.

Think of it as "relative standard deviation."

$$\mathrm{CV} = \frac{\sigma_f}{f_{\mathrm{avg}}}$$