ENGR 695 Advanced Topics in Engineering Mathematics Fall 2024

Lecture Outline for Monday, Sept. 9

1. Application of overdetermined systems: Curve-fitting and the method of least squares

   a. Example from previous lecture: The following small data set has been collected. Use it to estimate $y(3)$, that is, the value of y at $x = 3$.

i	x_i	y_i
1	1.0	1.1
2	2.0	3.2
3	4.0	5.2

   b. One possible approach: Compute the coefficients of the quadratic expression for a curve that passes through the data points.

$$y = c_0 + c_1 x + c_2 x^2$$

   The data set leads to a 3 × 3 system for finding c_0, c_1, and c_2 (unique solution).

   c. Another possible approach: Compute the coefficients of the linear expression for a line that passes through the data points.

$$y = c_0 + c_1 x$$

   The data set leads to a 3 × 2 system for finding c_0 and c_1 (inconsistent – no solution).

2. Curve-fitting: How to find the "best" solution to an inconsistent system

   a. Given a data set: (x_i, y_i), $i = 1$ to M → data vectors $\mathbf{x}$ and $\mathbf{y}$
   b. Define model: a set of functions $\{f_j(x_i)\}_{j = 1 \text{ to } N}$ and coefficients $\{c_j\}_{j = 1 \text{ to } N}$ that yield the best approximations to $\{y_i\}_{i = 1 \text{ to } M}$:

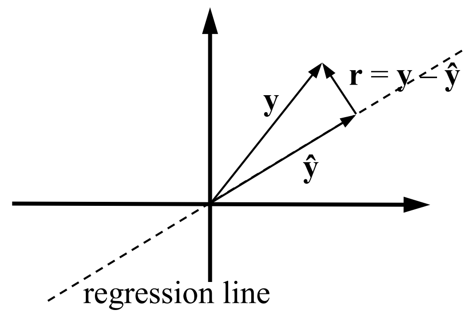
      $$y(x) \approx \hat{y}(x) = \sum_{j=1}^{N} c_j f_j(x) \qquad \hat{y}(x) \text{ is the best fit to the actual curve } y(x)$$

   c. In matrix form, $\hat{\mathbf{y}} = F\mathbf{c}$, where $F_{ij} = f_j(x_i)$ and $\hat{\mathbf{y}}$ contains best fit
   d. Functions $\{f_j(x)\}$ (often called basis functions) can be almost anything; popular choices are 1 and x (linear fit), polynomials (including quadratic and cubic), sin/cos, exponentials, and logarithms
   e. Least squares approach:
      i. Residual vector: $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ (r_i = distance from actual y_i to approximation $\hat{y}_i$ for each data point i; $\mathbf{r}$ has M rows)
      ii. Minimize $|\mathbf{r}|^2 = \mathbf{r}^T\mathbf{r}$ or make residual orthogonal to approximation ($\mathbf{r}^T\hat{\mathbf{y}} = 0$)
      iii. Either way, the *normal equation* results (LS solution)

(*continued on next page*)

3. Derivation of normal equation from $\mathbf{r}^T\hat{\mathbf{y}} = 0$:

$$\mathbf{r}^T\hat{\mathbf{y}} = (\mathbf{y}-\hat{\mathbf{y}})^T\,\hat{\mathbf{y}} = 0$$
$$(\mathbf{y}-F\mathbf{c})^T\,F\mathbf{c} = 0$$
$$\left[\mathbf{y}^T - (F\mathbf{c})^T\right]F\mathbf{c} = 0$$
$$\left(\mathbf{y}^T - \mathbf{c}^T F^T\right)F\mathbf{c} = 0$$
$$\left(\mathbf{y}^T F - \mathbf{c}^T F^T F\right)\mathbf{c} = 0$$



The elements of **c** should not be zero, so

$$\mathbf{y}^T F - \mathbf{c}^T F^T F = 0$$
$$F^T\mathbf{y} - F^T F\mathbf{c} = 0$$
$$F^T F\mathbf{c} = F^T\mathbf{y} \quad\rightarrow\quad \mathbf{c} = \left(F^T F\right)^{-1} F^T\mathbf{y}$$

This result is called the *normal equation* (sometimes the plural *normal equations*) because it is derived from the fact that **r** is normal (orthogonal) to $\hat{\mathbf{y}}$. Since it is equivalent to the result obtained by minimizing $|\mathbf{r}|^2 = \mathbf{r}^T\mathbf{r}$, it is also sometimes called a *least squares* solution.

4. Practical considerations:

   a. In *Matlab*, use: `c = F\y`; automatically forms solution using normal equation (or its functional equivalent)
   b. Could also use: `c = (F'F)\(F'y)` (academic interest only)
   c. F^TF is symmetric and nonsingular if there are no repeated data points
   d. F is $M \times N$, so F^TF is $N \times N$

5. Back to the simple data set example: Apply the normal equation

   a. Find quadratic fit $y = c_0 + c_1x + c_2x^2$ to the following small data set. Note that $f_0(x) = 1, f_1(x) = x, f_2(x) = x^2$.

i	x_i	y_i
1	1.0	1.1
2	2.0	3.2
3	4.0	5.2

Form matrix F and data vector **y**, then solve normal equation:

$$F = \begin{bmatrix} 1 & 1.0 & 1.0 \\ 1 & 2.0 & 4.0 \\ 1 & 4.0 & 16.0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1.1 \\ 3.2 \\ 5.2 \end{bmatrix} \quad\rightarrow\quad \mathbf{c} = \begin{bmatrix} -1.7333 \\ 3.2000 \\ -0.3667 \end{bmatrix}$$

(*continued on next page*)

b. Find linear fit $y = d_0 + d_1 x$ to the data set. Note that $f_0(x) = 1, f_1(x) = x$.

Form matrix F and data vector $\mathbf{y}$, then solve normal equation:

$$F = \begin{bmatrix} 1 & 1.0 \\ 1 & 2.0 \\ 1 & 4.0 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1.1 \\ 3.2 \\ 5.2 \end{bmatrix} \quad \rightarrow \quad \mathbf{d} = \begin{bmatrix} 0.1000 \\ 1.3143 \end{bmatrix}$$