Lecture 8: Least Squares Overview

Today:

- Function approximation using ***least squares*** (instead of interpolation)

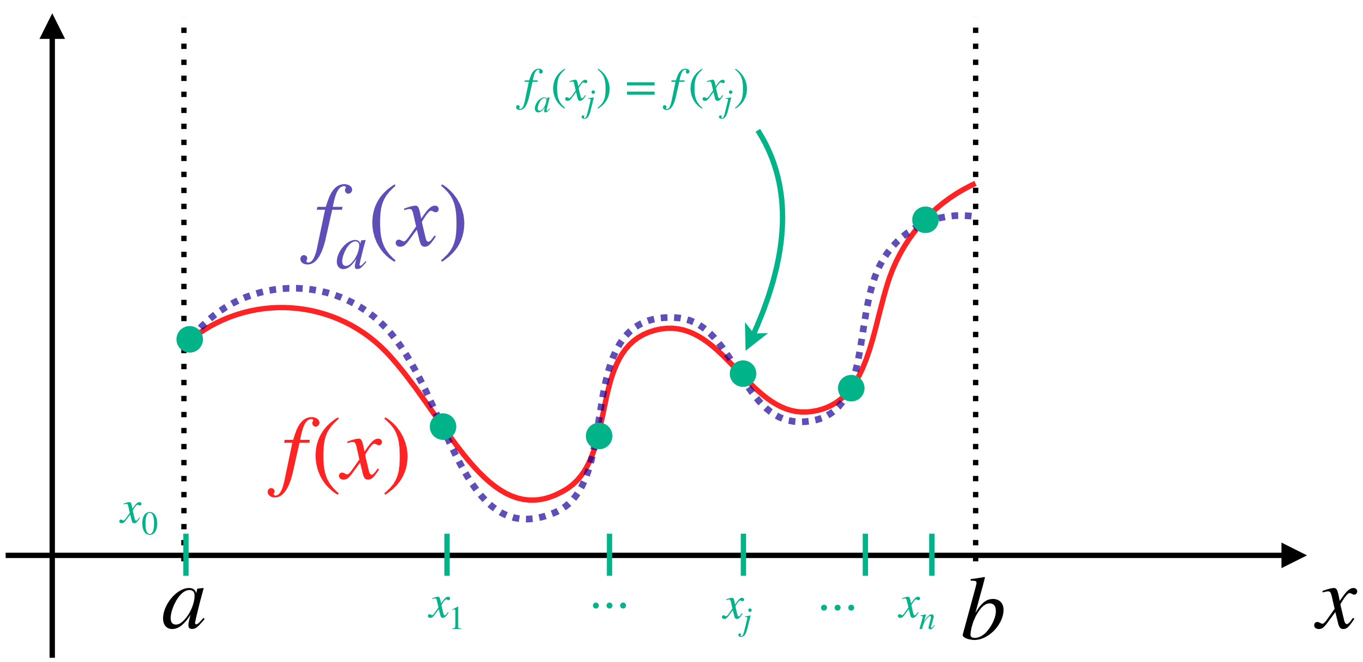
Reminder from last week

Goal: find a function, $f_a(x)$, that approximates a given function, $f(x)$, *accurately* on $x \in [a, b]$

Last time:

(A) Write $f_a(x)$ as a linear combo of ***basis*** functions

(B) Solve for the c_j using ***interpolation***



This time:

(A) We will ***still*** express our $f_a(x)$ as a linear combination of ***basis*** functions in terms of unknown coefficients

(B) We will solve for the coefficients using a ***different*** approach than interpolation: ***least squares***

The premise for least squares function approximation

(A) Pick a ***subspace*** that we want to approximate onto (e.g., $\mathscr{P}^n[a,b]$ or $\mathscr{T}^n[a,b]$)

(B) Pick a ***basis*** $\mathscr{B} = \{b_0(x), b_1(x), \ldots, b_n(x)\}$for that subspace and express the approximate function, $f_a(x)$, as a linear combination of the basis functions in terms of unknown coefficients

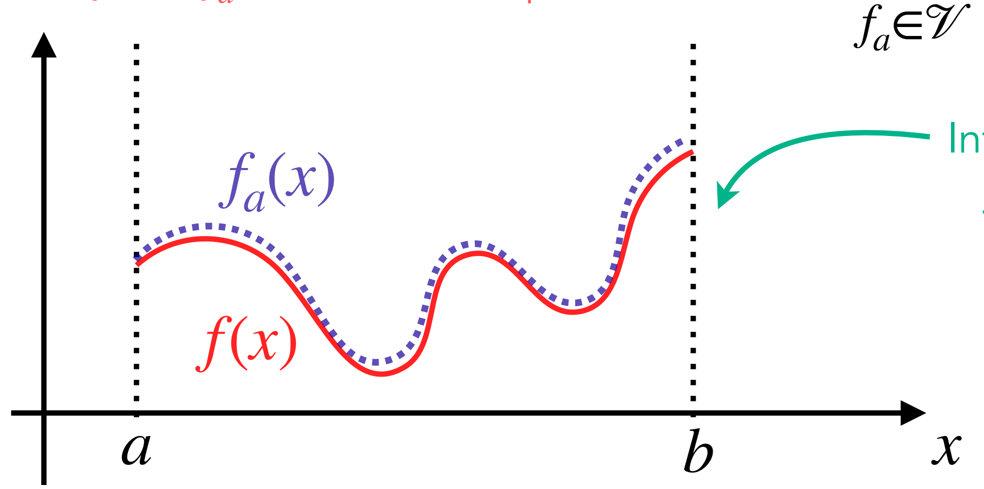
$$f_a(x) = \sum_{j=0}^{n} c_j b_j(x) \qquad (1)$$

(C) Need $n+1$ equations to solve for the $n+1$ unknown coefficients. Unlike in interpolation, we will develop these equations by trying to ***minimize the overall error***. That is, we will look for an $f_a(x)$ that satisfies

Remember: the $||\cdot||$ symbol means "norm", which is defined in terms of the inner product (review the error lecture if needed!)

In words: seek an f_a belonging to the desired subspace $\mathscr{W}$ such that the norm of the difference between f and f_a is the smallest possible

$$\min_{f_a \in \mathscr{V}} ||f - f_a||^2 \qquad (2)$$



Intuition: don't explicitly require $f = f_a$ at any specific points. Just try to get an f_a that is close to f in a meaningful way.

OK, that may be a nice sentiment. But how do we construct $n+1$ equations from that?

Constructing the $n+1$ equations in least squares approximation

The key to constructing the $n+1$ equations revolves around the following (non-obvious) fact:

The special f_a that obeys the minimization criteria (2) satisfies

$$(f - f_a, r) = 0 \text{ for any } r \in \mathscr{W} \qquad (3)$$

i.e., the inner product between the error $e = f - f_a$ and any r in the subspace $\mathscr{W}$ is zero

i.e., the error $e = f - f_a$ is ***orthogonal*** to any r in the subspace $\mathscr{W}$

We will interpret the meaning of (3) later. But first, let's identify why (3) is useful:

If (3) is true, then since each $b_i \in \mathscr{W}$ for $i = 0, \ldots, n$:

$$(f - f_a, b_i) = 0, \quad i = 0, \ldots, n \qquad (4)$$

Subbing equation (1) into (4) gives

$$\left(f - \sum_{j=0}^{n} c_j b_j, b_i \right) = 0, \quad i = 0, \ldots, n$$

And with some rearranging:

This gives $n+1$ equations to solve for the $n+1$ unknowns $c_0, \ldots, c_n$

$$\sum_{j=0}^{n} c_j \left(b_j, b_i \right) = (f, b_i), \quad i = 0, \ldots, n \qquad (5)$$

Breakout room activity

Turn equation (5) into a matrix system of equations $\mathbf{A}\mathbf{c} = \mathbf{b}$, where $\mathbf{c}$ is the vector of unknown coefficients, $\mathbf{b}$ is a righthand side that involves known quantities, and $\mathbf{A}$ is the appropriate matrix associated with the coefficient vector $\mathbf{c}$

Answer.

$$\begin{bmatrix} (b_0, b_0) & (b_1, b_0) & \cdots & (b_{n-1}, b_0) & (b_n, b_0) \\ (b_0, b_1) & (b_1, b_1) & \cdots & (b_{n-1}, b_1) & (b_n, b_1) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ (b_0, b_{n-1}) & (b_1, b_{n-1}) & \cdots & (b_{n-1}, b_{n-1}) & (b_n, b_{n-1}) \\ (b_0, b_n) & (b_1, b_n) & \cdots & (b_{n-1}, b_n) & (b_n, b_n) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \\ c_n \end{bmatrix} = \begin{bmatrix} (f, b_0) \\ (f, b_1) \\ \vdots \\ (f, b_{n-1}) \\ (f, b_n) \end{bmatrix} \tag{6}$$

Which inner product to use?

There are many good choices for an inner product. One option is the IP we considered in week 1, the function $(\cdot\,,\cdot)$ defined by

$$(f, g) = \int_a^b f(x)g(x)dx \text{ for any } f, g \in \mathscr{V}$$

We will consider another IP in the next lecture

So for this specific IP, the entry of the matrix in equation (6) involving b_2, b_3 is

$$(b_2, b_3) = \int_a^b b_2(x)b_3(x)dx$$

Pro tip: it might be useful to fill in what some of the other entries of the matrix would be as an exercise!

OK great, so we now have what we need to approximate a function with least squares!

(A) Solve (6) for the unknown coefficients after filling in $\mathbf{A}$ and $\mathbf{b}$ using our chosen IP

(B) Now we have $f_a(x)$ via equation (1)

But remember that this all assumed that (3) was true. Is it?! Yes, and let's build some intuition for what it means!

Showing that equation (3) is true

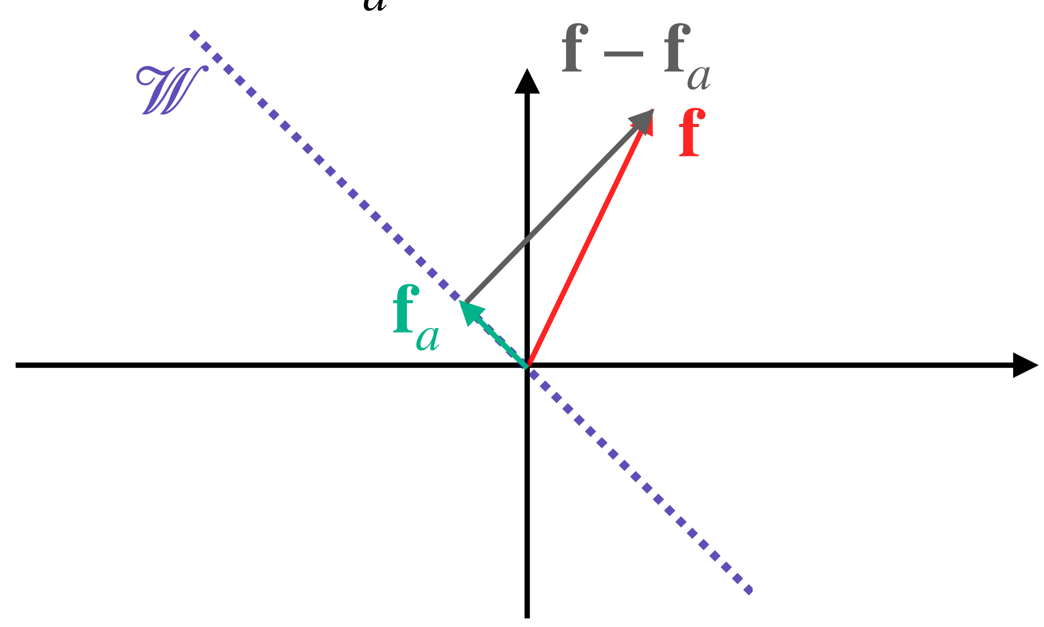
We want to understand the statement that $(f - f_a, r) = 0$ for any $r \in \mathscr{W}$

The statement is most easily understood with finite-dimensional vectors. Let's say $\mathbf{f} \in \mathbb{R}^2$ is some vector that we want to approximate

And let's say that we want to approximate this onto the subspace

$$\mathscr{W} = \left\{ \alpha \begin{bmatrix} -1 \\ 1 \end{bmatrix} : \alpha \in \mathbb{R} \right\}$$

From linear algebra and/or geometry, you might recall that the $\mathbf{f}_a$ that best approximates $\mathbf{f}$ on $\mathscr{W}$ satisfies $\mathbf{f} - \mathbf{f}_a \perp \mathscr{W}$



Math appreciation break: by creating a framework for defining inner products and norms, we have a way for intuiting best approximation of functions using visualizable vectors!