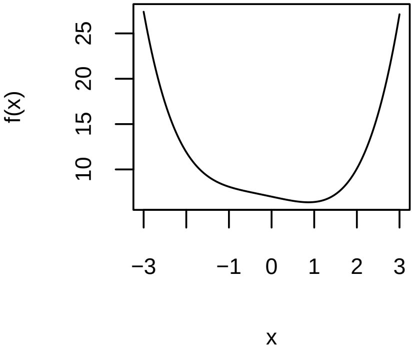
Lecture 17: Newton's method

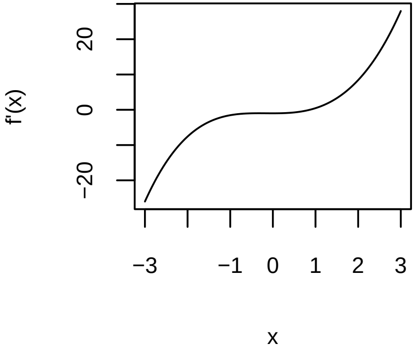
Ciaran Evans

Motivation

$$f(x) = \frac{x^4}{4} - \sin(x) + 7$$

$$f'(x) = x^3 - \cos(x)$$





Want to minimize $f(x)$, i.e. solve $f'(x) = 0$. Challenge: no closed form.

Idea: Approximate f' with something *easy* to solve

Taylor expansion

First-order Taylor expansion of function $g(x)$ around point a:

$$g(x) \approx g(a) + (x - a)g'(a)$$

Question: What is required of x for this approximation to be reasonable?

Taylor expansion

First-order Taylor expansion of function $g(x)$ around point a:

$$g(x) \approx g(a) + (x - a)g'(a)$$

- Let x^* be the location of the true minimum
- Let $x^{(k)}$ be our current guess for x^*
- Taylor expansion:

$$f'(x^*) \approx$$

Newton's method (in one dimension)

Iterative updates:

$$x^{(k+1)} = x^{(k)} - (f''(x^{(k)}))^{-1} f'(x^{(k)})$$

Question: How does this compare with gradient descent?

Newton's method (in one dimension)

Iterative updates:

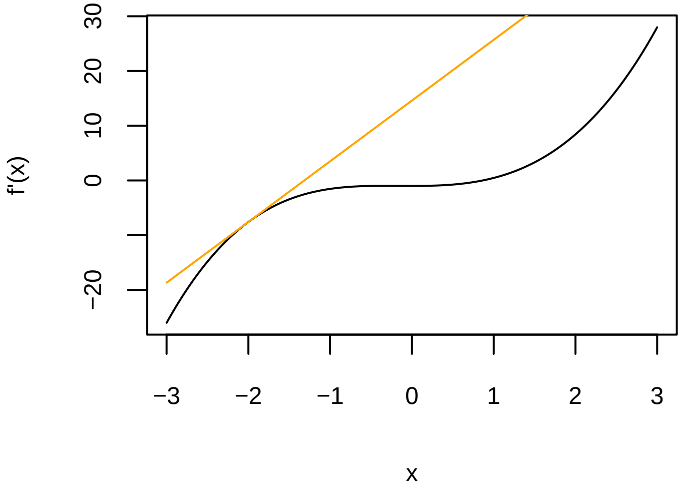
$$x^{(k+1)} = x^{(k)} - (f''(x^{(k)}))^{-1} f'(x^{(k)})$$

More generally:

$$x^{(k+1)} = x^{(k)} - \alpha_k (f''(x^{(k)}))^{-1} f'(x^{(k)})$$

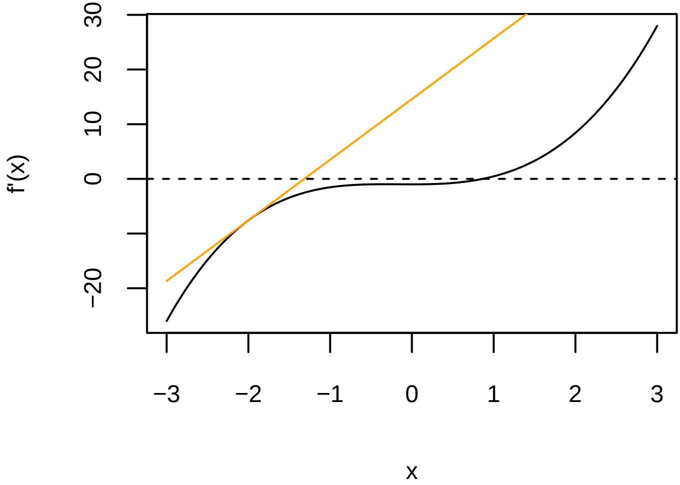
- $\alpha_k = 1$ is the "natural" value
- Only use a different value of α_k when $\alpha_k = 1$ doesn't work

Illustration



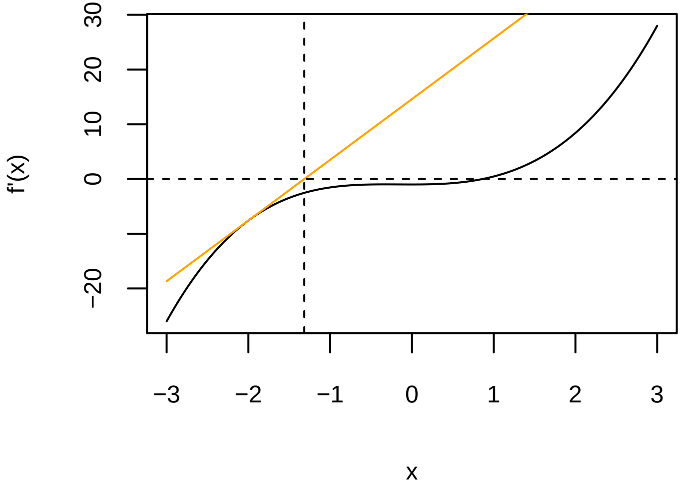
- $x^{(0)} = -2$

Illustration



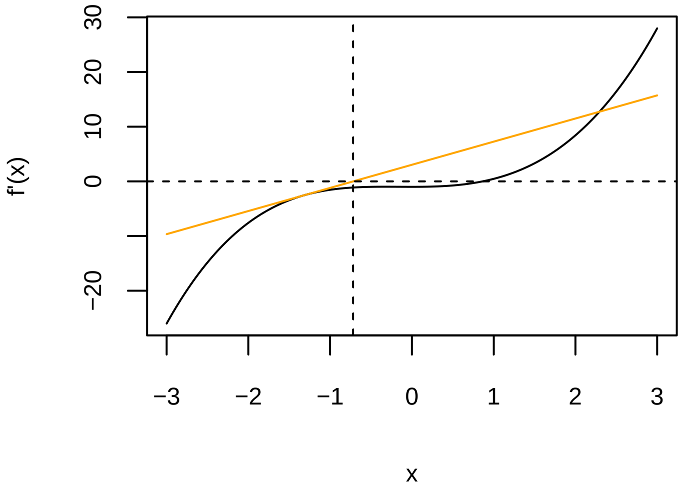
- $x^{(0)} = -2$
- $x^{(1)}$ solves $f'(x^{(0)}) + (x^{(1)} - x^{(0)})f''(x^{(0)}) = 0$

Illustration



- $x^{(1)}$ solves $f'(x^{(0)}) + (x^{(1)} - x^{(0)})f''(x^{(0)}) = 0$
- $x^{(1)} = -1.316$

Illustration



- $x^{(1)} = -1.316$
- $x^{(2)}$ solves $f'(x^{(1)}) + (x^{(2)} - x^{(1)})f''(x^{(1)}) = 0$

Extension to multiple variables

Example: $f(x, y) = \dfrac{x^4}{4} - \sin(x) + (x + y)^2$

$$\nabla f(x, y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} x^3 - \cos(x) + 2(x + y) \\ 2(x + y) \end{pmatrix}$$

Question: What is the *second* derivative of f?

Extension to multiple variables

Example: $f(x,y) = \dfrac{x^4}{4} - \sin(x) + (x+y)^2$

$$\nabla f(x,y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} \\ \dfrac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} x^3 - \cos(x) + 2(x+y) \\ 2(x+y) \end{pmatrix}$$

Hessian: $\mathbf{H}_f(x,y) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix}$

Newton's method in multiple dimensions

- $\mathbf{x} = (x_1, ..., x_d)^T \in \mathbb{R}^d$
- $f(\mathbf{x}) \in \mathbb{R}$

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_d} \end{pmatrix} \qquad \mathbf{H}_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_d} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \frac{\partial^2 f}{\partial x_d \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

Newton's method:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k (\mathbf{H}_f(\mathbf{x}^{(k)}))^{-1} \nabla f(\mathbf{x}^{(k)})$$

Your turn

- Newton's method practice questions on the course website:

https://sta379-s25.github.io/practice_questions/pq_17.html

- If done early, go back to momentum practice questions from Monday:

https://sta379-s25.github.io/practice_questions/pq_16.html