STA 5106
Computational Methods in Statistics I

Department of Statistics
Florida State University

Class 10
September 26, 2019

Review: Simple Iterations

- We are given a real-valued function $f(x)$ where $x \in \mathbf{R}$ and the goal is to find $x^* \in \mathbf{R}$ such that $f(x^*) = 0$.

- We solve an equivalent problem of finding the fixed point of another function g. g is found in such a way that

$$f(x^*) = 0 \Leftrightarrow g(x^*) = x^*.$$

- One choice of g is

$$g(x) = x + f(x) .$$

- Let x_0 be some starting value. At the i-th stage of the algorithm, the next iterate is given by the formula:

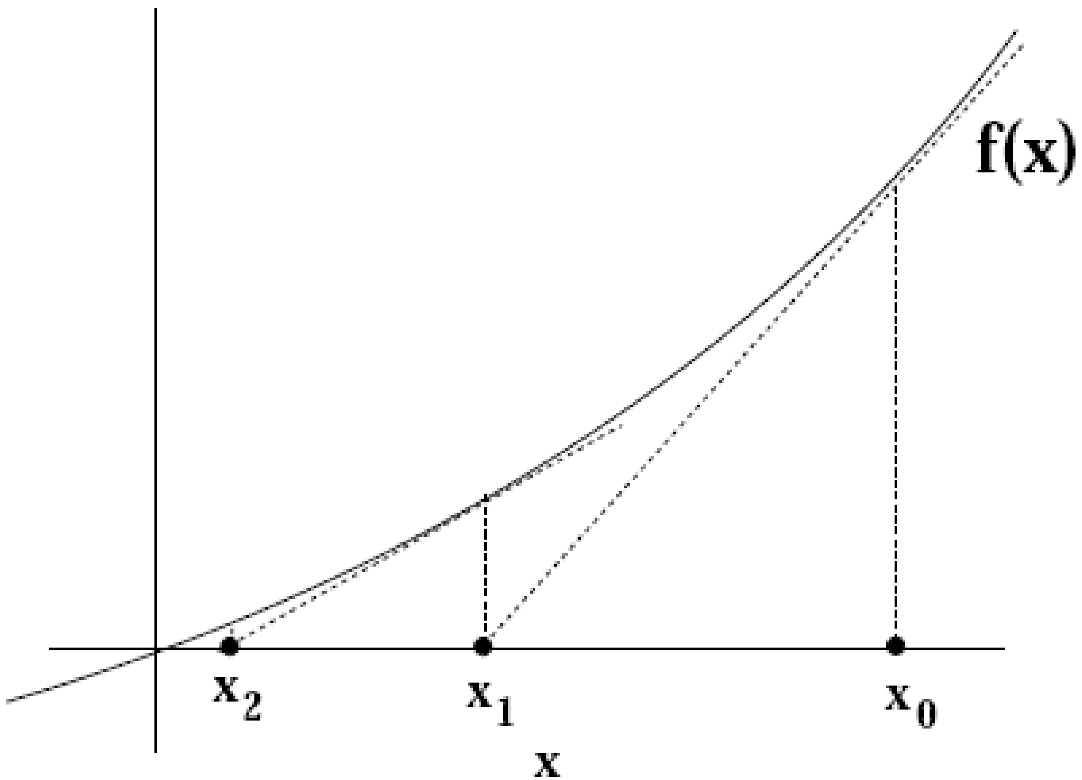
$$x_{i+1} = g(x_i) = x_i + f(x_i) .$$

Newton-Raphson's Method

- The slow convergence of the simple iteration is avoided by using the Newton-Raphson's Method.
- Newton-Raphson's Method is one of the most popular techniques used in numerical root finding.
- Stricter requirements: For Newton's method to apply, f should be continuously differentiable and
$$f'(x) \neq 0$$
near the solution x^*.
- We will derive an iterative procedure for computing x_{i+1} given x_i.

Basic Idea

- Approximate the function at a given point by a straight line. In this case, the line is given by the line tangent to *f* at that point.



Newton-Raphson's Method

- Assume x_i is the current estimate. Then

$$f'(x_i) = \frac{0 - f(x_i)}{x_{i+1} - x_i}$$

- Therefore,

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}, \quad (f'(x_i) \neq 0)$$

- Newton-Raphson is one of the fastest known algorithms for root finding in general situations.

- Its order of convergence is $\beta = 2$.

Algorithm

- **Algorithm:** Given a function $f(x)$ find its roots using Newton-Raphson's method:

- **Algorithm 23 (Newton-Raphson Method)**

```
x(1) = x0;
i = 1;
gx = x(i) − f(x(i))/f′(x(i));
while (abs(x(i) − gx) > ε)
        gx = x(i);
        x(i + 1) = x(i) − f(x(i))/f′(x(i));
        i = i + 1;
end
```

Important Application

- Newton-Raphson method is particularly useful for minimizing (or maximizing) functions that are **convex** (or **concave**).

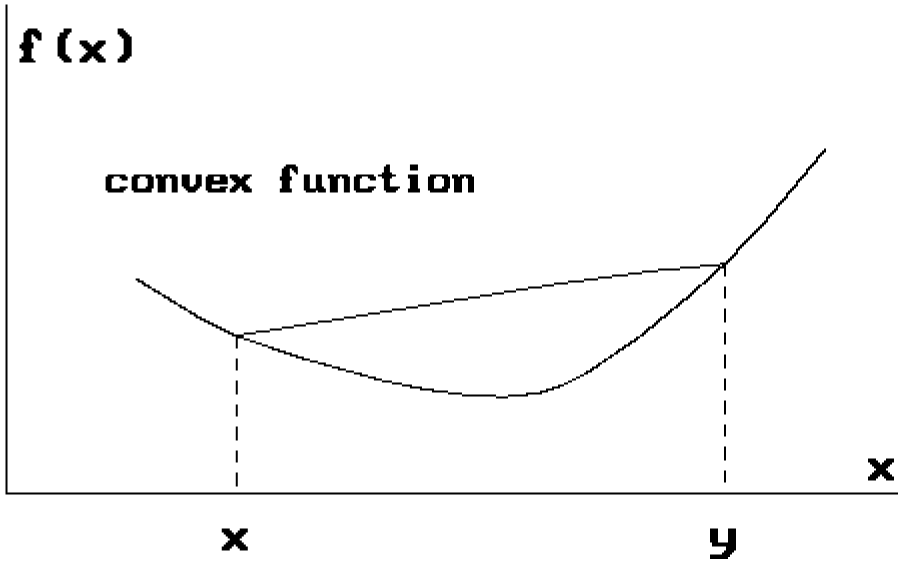
- **Definition:** A real-valued function $f(x)$ defined on an interval $[a, b]$ is called **convex,** if for any two points x_1 and x_2 in $[a, b]$ and any t in $[0, 1]$,

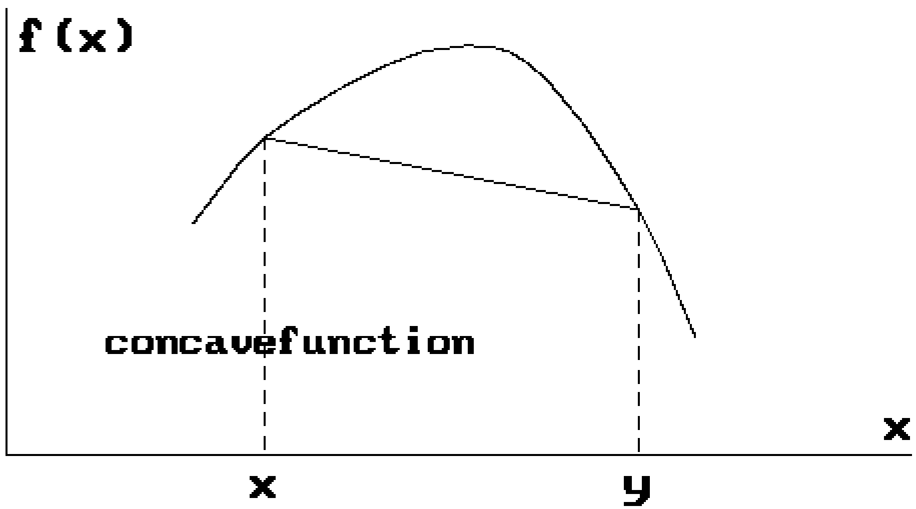
$$f(t\, x_1 + (1-t)\, x_2) \leq t\, f(x_1) + (1-t)\, f(x_2)$$

- A function f is said to be **concave** if $-f$ is convex. That is, if for any two points x_1 and x_2 in $[a, b]$ and any t in $[0, 1]$,

$$f(t\, x_1 + (1-t)\, x_2) \geq t\, f(x_1) + (1-t)\, f(x_2)$$

Convex and Concave





Convex Condition

- **Theorem:** Assume f is twice continuously differentiable on $[a, b]$. If $f'' \geq 0$, then f is convex.

  Proof: For any two points x_1 and x_2 in $[a, b]$ and any t in $[0, 1]$, let $s = t\, x_1 + (1-t)\, x_2$. Then use Taylor series,

$$f(x_1) = f(s) + f'(s)(x_1-s) + f''(\xi_1)(x_1-s)^2/2,$$

$$f(x_2) = f(s) + f'(s)(x_2-s) + f''(\xi_2)(x_2-s)^2/2.$$

  Then

$$t\, f(x_1) + (1-t)\, f(x_2)$$

$$= f(s) + t\, f''(\xi_1)(x_1-s)^2/2 + (1-t)\, f''(\xi_2)(x_2-s)^2/2.$$

  As $f'' \geq 0$,

$$t\, f(x_1) + (1-t)\, f(x_2) \geq f(s) = f(t\, x_1 + (1-t)\, x_2).$$

  Therefore, f is convex.

Optimization

- Newton-Raphson's Method is a popular technique used in function optimization (or root-finding for derivative function)

- Iteration:

$$x_{i+1} = x_i - \frac{f'(x_i)}{f''(x_i)}, \qquad (f''(x_i) \neq 0)$$

- Newton-Raphson method is particularly useful for minimizing (or maximizing) functions that are **convex** (or **concave**).

- Approximation of a quadratic form (Taylor Expansion):

$$f(x) \approx f(x_i) + f'(x_i)(x - x_i) + \frac{1}{2} f''(x_i)(x - x_i)^2.$$

Minimizing this quadratic form results in the given iteration.

Multivariate Case

- Optimization of function $f(\mathbf{x}) = f(x_1, \ldots, x_n)$: $\mathbf{R^n} \rightarrow \mathbf{R}$.

- Let the gradient and Hessian matrix of f be

$$\nabla f = \begin{pmatrix} \partial f / \partial x_1 \\ \vdots \\ \partial f / \partial x_n \end{pmatrix} \qquad H_f = \begin{pmatrix} \partial^2 f / \partial x_1 \partial x_1 & & \partial^2 f / \partial x_1 \partial x_n \\ & \ddots & \\ \partial^2 f / \partial x_n \partial x_1 & & \partial^2 f / \partial x_n \partial x_n \end{pmatrix}$$

- Then the Newton-Raphson is updated as:

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - H_f^{-1}(\mathbf{x}^{(i)}) \nabla f(\mathbf{x}^{(i)})$$

- The method has optimal performance if
  - f is convex (when H_f is positive definite) -- for minimization
  - f is concave (when H_f is negative definite) -- for maximization

3.3 Starting Values and Stopping Criteria

Starting Values

- There is not much theory to finding starting values in general. It basically depends on the prior knowledge about the function.

- One way is to approximate the given function with another function with known roots.

- For example, a continuous function can be approximated by a polynomial of an appropriate order, and the roots of this polynomial can form the starting value for iterative methods.

Stopping Criteria

- We need to ascertain when the iteration has sufficiently converged to provide the desired accuracy in the final result.

- Two commonly-used methods:

  1: The absolute change in the values of the successive iterates:

$$| x_{i+1} - x_i | < \varepsilon$$

  2: The relative change in the values of the successive iterates:

$$\frac{| x_{i+1} - x_i |}{| x_i |} < \varepsilon$$