Lecture 2: Nonlinear Equations

CSC 338: Numerical Methods

Ray Wu

University of Toronto

January 18, 2023

Root-finding

- Root-finding is the algorithmic process of finding zeros of a function

$$f(x) = 0 \tag{1}$$

- We denote the root of interest to be x^*.
- Generally, x can be vectors, but in this section, x will be a scalar.
- Assume that f is continuous.

One-variable nonlinear equations

Why start with nonlinear equations?

- Linear equations are too simple

$$f(x) = ax + b = 0 \tag{2}$$

  solution is given by $x = -b/a$.
  - This kind of formula is known as a **direct solution**.
  - Generally, we have direct solutions for linear equations, but not for nonlinear equations.
- Systems of linear equations are more complicated, later topic.
- Specific algorithms covered:
  - Bisection method
  - Fixed point iteration
  - Newton's method
  - Secant method

Iterative methods

- Nonlinear equations can be arbitrarily complex
- Exact formulas such as $x = -b/a$ is not realistic.
- Hence, we need to use iterative methods.
- Iterative methods start from an initial guess x_0 and compute a sequence of iterates $x_1, x_2, x_3, \cdots$, eventually reaching an approximation of desired accuracy.

Terminating an iterative procedure

We do not expect the procedure to compute x^* exactly. Hence, we use "close enough" criteria.

- absolute error of iterate:

$$|x_n - x_{n-1}| < \text{atol} \tag{3}$$

- relative error of iterates:

$$|x_n - x_{n-1}| < x_n \text{rtol} \tag{4}$$

- function value:

$$|f(x_n)| < \text{ftol} \tag{5}$$

Desirable algorithm properties

How to compare algorithms?

- Efficient: the fewer function evaluations, the better
- Robust: fails rarely, if ever.
- Other information: function derivative, etc.
- Smoothness: the less requirements, the better.
- Generalization: Does the process generalize to many variables?

Rates of convergence

Linear, superlinear, quadratic.

- Linear convergence: There exists some constant $\rho < 1$ such that

$$|x_{k+1} - x^*| \leq \rho |x_k - x^*| \tag{6}$$

  for large enough k.

- superlinear convergence: There exists a sequence $\rho_k \to 0$ such that

$$|x_{k+1} - x^*| \leq \rho_k |x_k - x^*| \tag{7}$$

- quadratic convergence: There exists a constant M such that

$$|x_{k+1} - x^*| \leq M |x_k - x^*|^2 \tag{8}$$

You should understand these definitions.

A representative problem

Consider finding the root of

$$f(x) = \exp(-x) - x^3 \tag{9}$$

We know that the derivative is negative:

$$f'(x) = -\exp(-x) - 3x^2 < 0 \tag{10}$$

hence, there can be at most one root.

Bisection method

Assume we have a, $b > a$, and $f(a)$ and $f(b)$ have opposite signs.

- A continuous analogue to binary search.
- By intermmediate value theorem, there must be an x^* between a and b such that $f(x^*) = 0$.
- **Divide-and-conquer** strategy:
  - At each iteration, evaluate $f((a + b)/2)$.
  - Choose the bracket that ensures both ends have opposite signs.
  - repeat until bracket is small enough or function value is small enough.

Suppose we want to find a zero of $f(x) = \exp(-x) - x^3$.

- Bracketing:
  - Observe that $f(0) = \exp(0) - 0^3 = 1$, therefore, pick $a = 0$
  - Observe that $f(1) = \exp(-1) - 1 = 1/e - 1 < 0$, therefore, pick $b = 1$

Fixed-point iteration

Given $f(x)$, construct a function $g(x)$ such that when $f(x) = 0$, $g(x) = x$.

- Some examples:

$$g(x) = x - f(x) \tag{11}$$

$$g(x) = 2f(x) + x \tag{12}$$

$$g(x) = x - f(x)/f'(x) \tag{13}$$

After determining $g(x)$, we compute $x_{i+1} = g(x_i)$.
Many questions arise, such as:

- Is there a fixed point?
- If there is, is the fixed point unique?
- Does the sequence of iterates converge?
- If yes, at what rate?
- If it doesn't converge, does that mean no root exists?

Fixed point theorem

- Suppose we have for two values a and b:

$$g(a) > a \text{ and } g(b) < b. \tag{14}$$

- Apply the intermediate value theorem to show that a fixed point exists.
- If g is also differentiable and there is some $\rho < 1$ such that

$$|g'(x)| \leq \rho \quad \forall x \in [a, b], \tag{15}$$

  then the root x^* is unique.
- Assume that there exists another root y^*, then

$$|x^* - y^*| = |g(x^*) - g(y^*)| = |g'(z)(x^* - y^*)| \leq \rho|x^* - y^*| \tag{16}$$

- If $\rho < 1$, then y^* must equal x^*.

Rate of convergence of fixed point iteration

- Since we know there is a unique solution, we can show that

$$|x_{k+1} - x^*| = |g(x_k) - g(x^*)| \leq \rho |x_k - x^*| \tag{17}$$

- Error reduces to at most ρ of error at previous iterate. Hence, convergence is established.
- Rate of convergence defined as

$$\text{rate} = -\log_{10} \rho. \tag{18}$$

  - For bisection, rate of convergence is $-\log_{10}(1/2) = 0.301$
  - For fixed point, depends on ρ and hence g.
- What kind of g should we choose?
  - The one that has a small value of ρ. (Ideally, close to zero!)

Newton's method

General idea: Linearize the equations locally, solve the linear equations, and repeat until convergence.

- Linearization: Using Taylor series,

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + f''(x_k)(x - x_k)^2/2 + \dots \tag{19}$$

- Drop the higher-order terms to get linearized equations:

$$f(x) = f(x_k) + f'(x_k)(x - x_k) \tag{20}$$

- Now, set $f(x) = 0$ and solve the linear equation to get

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} \tag{21}$$

- repeat until convergence.

Newton's Method and fixed-point iteration

- Newton's Method is a special case of fixed-point iteration.

$$g(x) = x - \frac{f(x)}{f'(x)}. \tag{22}$$

- What is ρ?

$$g'(x) = 1 - \frac{f'(x)f'(x) - f''(x)f(x)}{[f'(x)]^2} = \frac{f''(x)f(x)}{[f'(x)]^2} \tag{23}$$

- Sub in x^*:

$$g'(x^*) = \frac{f''(x^*)f(x^*)}{[f'(x^*)]^2} \tag{24}$$

- As $x \to x^*$, $g'(x^*) \to 0$.
- At least superlinear convergence, as long as $f'(x^*)$ is nonzero

Convergence of Newton's method

- Taylor expand f around x_k:

$$f(x^*) = f(x_k) + f'(x_k)(x^* - x_k) + \frac{1}{2} f''(\xi_n)(x^* - x_k)^2 \tag{25}$$

- Since x^* is the root, sub in $f(x^*) = 0$, divide by $f'(x_k)$, rearrange:

$$\frac{f(x_k)}{f'(x_k)} + (x^* - x_k) = -\frac{f''(\xi_n)(x^* - x_k)^2}{f'(x_k)} \tag{26}$$

- Recall that $x_{k+1} = x_k - f(x_k)/f'(x_k)$. Hence,

$$(x^* - x_{k+1}) = -\frac{f''(\xi_n)(x^* - x_k)^2}{f'(x_k)} \tag{27}$$

- **quadratic rate of convergence**, if $f'(x) \neq 0$, $f''(x)$ is continuous, and the iterates are close enough to the root.

Secant method

Newton's method relies on being able to evaluate $f'(x)$. Secant method avoids this.

- Instead of using derivative, start at two initial points x_0 and x_1, and compute the linearization.

$$f(x) = f(x_1) + \frac{f(x_1) - f(x_0)}{x_1 - x_0} x \tag{28}$$

- Solving $f(x) = 0$ gives the update

$$x_{k+1} = x_k - f(x_k) \frac{x_k - x_{k-1}}{f(x_k) - f(x_{k-1})} \tag{29}$$

- Relies on both previous estimates – **two-step** method.
- Convergence is superlinear.
- Intuition: approximation of derivative is more and more accurate.
- Proof: beyond the scope of this course

Multiple roots

A multiple root is when $f(x) = 0$ and $f'(x) = 0$.

- Newton and Secant method become linearly convergent.
- Example: $f(x) = x^m$, $m > 1$.
  - Writing down the Newton update, we get

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k}{m} = \frac{m-1}{m}x_k \tag{30}$$

  - Clearly, this is linear convergence to the root $x^* = 0$, with $\rho = \frac{m-1}{m}$.

Summary - bisection method vs Newton's method

Bisection Method:

- Efficient? no – convergence is only linear
- Robust? yes – never fails
- Other information? none required
- Smoothness requirements? Minimal
- Generalizes easily? No

Newton's Method:

- Efficient? yes – quadratic convergence
- Robust? no – sometimes fails
- Other information? need function derivative
- Smoothness requirements? Some requirements.
- Generalizes easily? Yes