CME 108/MATH 114
Introduction to Scientific Computing
Summer 2020

Rootfinding
Bisection method

Guiding question How can I solve an equation? In particular, given a function f, how can I find x such that $f(x) = 0$?

Definition 1. A *zero* or a *root* of f is an element x in the domain of f such that $f(x) = 0$.

Remark 1. Note that the problem of solving an equation $f(x) = a$ reduces to finding a root of $g(x) = f(x) - a$. In the general case both x and f are vector-valued (systems of equations) but for now we'll focus on scalar-valued functions of a single var.

The most interesting case in our course is made un by functions which cannot be solved for analytically.

Example 2. *(Vander Waals equations) Recall from introductory chemistry the ideal gas law $PV = nRT$ used to model ideal gases. Real gases are in fact not fully compressible and there are attractive forces among their molecules, so a better model for their behavior is*

$$\big(P + \frac{n^2 a}{V^2}\big)\big(V - nb\big) = nRT, \tag{1}$$

where a and b are correction terms. In a lab, suppose 1 mol of chlorine gas has a pressure of 2 atm and a temperature of $313K$, for chlorine, $a = 6.29$ atm $L^2/mol^2, b = 0.0562 L/mol$. What is the volume?

We can't just "isolate" V. (Granted, in this simple case we obtain a low-degree polynomial in V and there are special methods for finding their roots. In this special case, the cubic formula will suffice.)

Bisection Method (Enclosure vs fixed point iteration schemes).

A basic example of enclosure methods: knowing f has a root p in $[a, b]$, we "trap" p in smaller and smaller intervals by halving the current interval at each step and choosing the half containing p.

Our method for determining which half of the current interval contains the root p is based on the *intermediate value theorem.*

Theorem 3 (IVT)**.** *Let f be a continuous function on $[a, b]$ and let k be any number between $f(a)$ and $f(b)$. Then there exists c in (a, b) such that $f(c) = k$.*

Informally, "A continuous function on an interval achieves all values between its values at the end points."

What does the IVT tell us about root finding?

Consider $\mathrm{sgn}(f(a)f(b))$. If $f(a)f(b) < 0$, what can you say?

Remark 2. Note that the same conclusion holds regardless of the magnitude of $f(a)$ and $f(b)$: only the sign of their product matters.

Example 4. $f(x) = x^3 + 2x^2 - 3x = 1$

$$\begin{aligned} &f(-3) = -1, \quad f(-1) = 3, \quad f(1) = -1 \\ &f(-2) = 5, \quad f(0) = -1, \quad f(2) = 9. \end{aligned}$$

Draw a picture! See Figure 1.

The Method Begin with an interval $[a, b]$ such that $f(a) \cdot f(b) < 0$. Find $p = (a + b)/2$. Test wether $f(a) \cdot f(p) < 0$. If so, then f has a root in $[a, p]$. Make $[a, p]$ the new interval and repeat the process. If not, then $f(a)$ and $f(p)$ have the same sign and therefore we are guaranteed that $f(p)f(b) < 0$ (since $f(a)$ and $f(b)$ have opposite signs), which then implies f has a root in $[p, b]$.

Make this the new interval and repeat the process. Generate a sequence of interval $[a_n, b_n]$ each guaranteed to contain a root of f and at each step use the approximation

$$p_n = \frac{a_n + b_n}{2} \tag{2}$$

of the enclosed root. Stop when p_n is "close enough" to a root of f in p.

The following example might suggest how to devise a stopping condition.

Example 5. *Square root via bisection.*

$f(x) = x^2 - 2$*, start with* $[0, 2]$

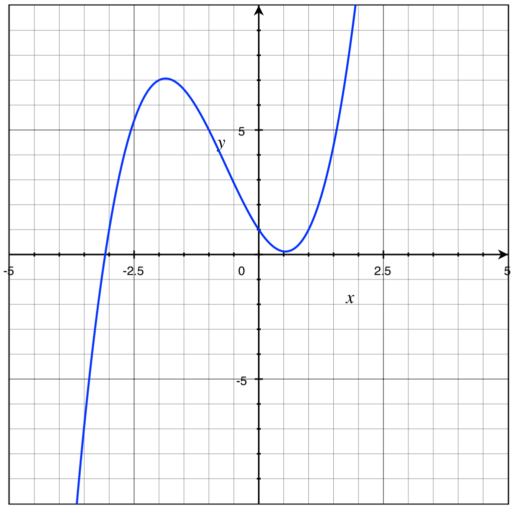

Figure 1: $f(x) = x^3 + 2x^2 - 3x - 1$

1. *Note* $f(0) \cdot f(2) = (-2)(2) = -4 < 0$*, so* f *indeed has a root in* $[0, 2]$ *by IVT. Set* $p_1 = 1$ *and notice* $f(0) \cdot f(1) = (-2)(-1) > 0$*, so choose* $[a_2, b_2] = [1, 2]$*.*

2. *Set* $p_2 = \frac{1+2}{2} = \frac{3}{2}$ *and notice* $f(a_2) \cdot f(p_2) = (-1)(\frac{9}{4} - 2) = \frac{-1}{4} < 0$ *so choose* $[a_3, b_3][1, \frac{3}{2}]$*.*

3. *Set* $p_3 = \frac{1+\frac{3}{2}}{2} = \frac{5}{4}$ *and notice* $f(a_3) \cdot f(p_3) = (-1)(\frac{25}{16} - 2) = \frac{7}{32} > 0$*, so choose* $[a_4, b_4] = [\frac{5}{4}, \frac{3}{2}]$*.*

4. *Now* $p_4 = \frac{\frac{5}{4}+\frac{3}{2}}{2} = \frac{11}{8} = 1.375\ldots$

Note that at iteration n our root approximation p_n is contained in an interval $[a_n, b_n]$ that is half the size of the previous interval $[a_{n-1}, b_{n-1}]$. Are we guaranteed convergence? If so in which cases?

Theorem 6. *Let* f *be a continuous function on* $[a, b]$ *and suppose that* $f(a) \cdot f(b) < 0$*. Then the bisection method generates a sequence of iterates* p_n *which converges to a root* $p \in (a, b)$ *with the property that*

$$|p_n - p| \leq \frac{b - a}{2^n}. \tag{3}$$

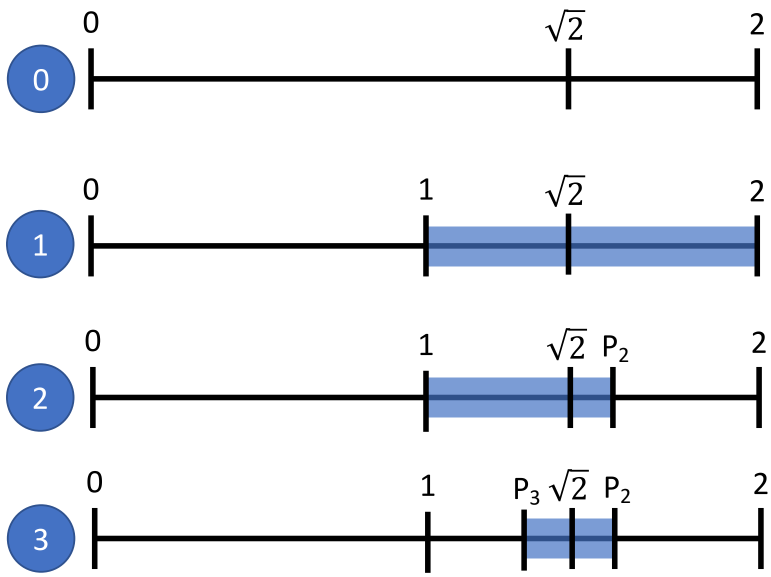

Figure 2: Bracketing $f(x) = x^2 - 2$ on $[0, 2]$

Remark 3.

1. Note conclusion: p_n converges to *a* root of f. If there are multiple roots in (a, b), we can't a priori know to which one p_n will converge.

2. We get a theoretical error bound, $|p_n - p|$, which can be helpful to eradicate bugs when coding.

3. Method cannot be used for locating roots of *even multiplicity*.

Definition 7. A root p of the equation $f(x) = 0$ is said to be of *multiplicity* m if f can be written as $f(x) = (x-p)^m q(x)$, with $\lim_{x \to p} q(x) \neq 0$.

Equivalently, p is of multiplicity m if $f(p) = f'(p) = \cdots = f^{(m-1)}(p) = 0$ and $f^{(m)}(p) \neq 0$ (assuming f is a smooth function).

We'll say more about multiplicity in the coming lectures.

Proof. We need only establish the error bound to prove convergence, since $\frac{b-a}{2^n} \to 0$ as $n \to \infty$.

By construction of the method, at each step we are guaranteed $p \in (a_n, b_n)$, and p_n is the midpoint of (a_n, b_n). This implies

$$|p_n - p| \leq \frac{b_n - a_n}{2}. \tag{4}$$

However,

$$\begin{aligned}
b_n - a_n &= \frac{1}{2}(b_{n-1} - a_{n-1}) \\
&= \frac{1}{2} \cdot \frac{1}{2}(b_{n-2} - a_{n-2}) \\
\cdots \\
&= \frac{1}{2^{n-1}}(b_1, a_1).
\end{aligned}$$

□

The convergence theorem suggests the stopping criterion $(b_n - a_n)/2 < \epsilon$. Since the absolute error $|p_n - p|$ is guaranteed to be no larger than $(b_n - a_n)/2$, this stopping criterion guarantees the root approximation p is no further than ϵ from p. That is, if we stop iterating when

$$\frac{b_n - a_n}{2} < \epsilon,$$

we are guaranteed that $|p_n - p| < \epsilon$.

Thus we obtain the following procedure.

Pseudocode

Algorithm 1 Bisection Method

```
Given f, [a, b], ϵ, N_max
sfa ← sign(f(a))
for i ← 1 to N_max do
    p ← (a + b)/2
    if (b − a)/2 < ϵ then
        return p
    end if
    sfp ← sign(f(p))
    if sfa · sfp < 0 then
        b ← p
    else
        a ← p
        sfa ← sfp
    end if
end for
```