The Bisection Method

Motivations. In scientific inquiries, we often encounter the need to find the zeros of a function, i.e., $\{x \in \mathbb{R}^n \mid f(\boldsymbol{x}) = 0\}$ where $f : \mathbb{R}^n \to \mathbb{R}$. This type of problem is called **Root-Finding**. A typical example is from optimization. Consider a differentiable function $y = f(x)$ in 1D. Its local optima occur at critical points, which satisfy $f'(x) = 0$. In higher dimensions, we solve simultaneous equations as we set $\nabla f = \mathbf{0}$, the zero vector. The problem becomes increasingly hard not only as dimension increases, but also as the nonlinearity of f gets more complicated.

In this section, we focus on studying the elementary techniques to numerically find the root of a function $f(x)$, without relying on differentiability (something that can help us in writing, but not numerics).

Bisection: A Binary Search.

The idea of the **Bisection Method** originates from the conclusions of the **Intermediate Value Theorem**.

Theorem. *(Intermediate Value Theorem, IVT) Suppose f is continuous on $[a, b]$ with $f(a)$ and $f(b)$ taking different signs. Then, there exists $p \in (a, b)$ such that $f(p) = 0$.*

Remark. This procedure does NOT determine where p is. Nor does the theorem grant uniqueness, i.e., there may be multiple p's that give us $f(p) = 0$. Nonetheless, existence theorem gives us hope of eventually finding the p.

The Algorithm.

Begin with two candidates $x = a_1$ and $x = b_1$, such that $f(a_1)$ and $f(b_1)$ have different signs. Compute

$$p_1 = a_1 + \frac{b_1 - a_1}{2} = \frac{a_1 + b_1}{2}.$$

- If $f(p_1) = 0$, then we are done.
- If $f(p_1) \neq 0$, then $f(p_1)$ has the same sign as either $f(a_1)$ or $f(b_1)$.
  - If $f(p_1)$ and $f(a_1)$ share the same sign, then we know $p \in (p_1, b_1)$. Set $a_2 = p_1$ and $b_2 = b_1$.
  - If $f(p_1)$ and $f(a_1)$ have different signs, then we know $p \in (a_1, p_1)$. Set $a_2 = a_1$ and $b_2 = p_1$.
- Repeat this procedure for $[a_2, b_2]$ until the function value is sufficiently close to zero.

Example. Show that $f(x) = x^3 + 4x^2 - 10$ has a root in $[1, 2]$ and use the Bisection method to determine an approximation to the root.

Solution. We know $f(1) = -5$ and $f(2) = 14$, so IVT guarantees a root in this interval.

The first step is evaluate the half point of $[1, 2]$, namely, $p_1 = f(1.5) = 2.375 > 0$. This implies that we should search $[1, 1.5]$. Then, we go on with the midpoint of this new interval, and find

$$p_2 = f(1.25) = -1.796875 < 0,$$

which prompts the next search interval to be $[1.25, 1.5]$. We find

$$p_3 = f(1.375) = 0.16211 > 0.$$

Continuing this procedure, we eventually find that a_n and b_n become very close to each other. We find

$$p_{13} = 1.365112305$$

which is quite close to the actual root $p = 1.365230013$ (within 10^{-4} error).

Visually speaking, the **Bisection Method** simply halves the search interval until the interval collapses. To numerically carry out the **Bisection Method** is also not difficult (Algorithm 2.1). However, for any iterative algorithm, we want to study the conditions under which the algorithm converges, and if it does, at what rate?

From examples on problems with known solutions, we often observe that certain iterates already produce a good answer (since we know the true solution) but gets discarded because it is still too far from nearby iterates. Therefore, in practice, the **Bisection Method** may become excruciatingly slow since it is throwing out a lot of iterates that may be good enough. However, this method can be used a starter method to help us narrow down the search space, while a secondary method will be utilized for a finer search.

Nonetheless, no matter how slow **Bisection** goes, it **always** converges.

Theorem. *Suppose that $f \in C[a,b]$ and $f(a) \cdot f(b) < 0$. The Bisection method generates a sequence $\{p_n\}_{n=1}^{\infty}$ approximating a zero p of f with*

$$|p_n - p| \leq \frac{b-a}{2^n}, \quad n \geq 1.$$

Proof. For each $n \geq 1$, we have (halving the interval)

$$b_n - a_n = \frac{1}{2^{n-1}}(b-a), \quad p \in (a_n, b_n),$$

Since $p_n = \frac{1}{2}(a_n + b_n)$ for all $n \geq 1$, it follows that p_n and p should not differ by half the interval length, i.e.,

$$|p_n - p| \leq \frac{1}{2}(b_n - a_n) = \frac{1}{2^n}(b-a).$$

Sending $n \to \infty$, we achieve $p_n \to p$. □

Remark 1. The last line of the proof also informs us of the rate of convergence.

$$|p_n - p| \leq (b-a)\frac{1}{2^n}$$

implies that the rate of convergence is $O(2^{-n})$, that is,

$$p_n = p + O\left(2^{-n}\right).$$

Remark 2. The theorem only provides an **upper bound** for the error, which can be too conservative. In fact, the real error may be far smaller than the upper bound.

Example. Sketch the graphs of $y = x$ and $y = x^2 - 5$. Use the Bisection method to find the points of intersection.