Bonus Lecture: Solving Systems of Equations

Chris Conlon

September 7, 2020

Grad IO

Basic Setup

Often we are interested in solving a problem like this:

Root Finding $f(x) = 0$

Optimization $\arg\min_x f(x)$.

These problems are related because we find the minimum by setting: $f'(x) = 0$

Root Finding

Newton's Method for Root Finding

Consider the Taylor series for $f(x)$ approximated around $f(x_0)$:

$$f(x) \approx f(x_0) + f'(x_0) \cdot (x - x_0) + f''(x_0) \cdot (x - x_0)^2 + o_p(3)$$

Suppose we wanted to find a root of the equation where $f(x^*) = 0$ and solve for x:

$$\begin{aligned} 0 &= f(x_0) + f'(x_0) \cdot (x - x_0) \\ x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} \end{aligned}$$

This gives us an iterative scheme to find x^*:

1. Start with some x_k. Calculate $f(x_k), f'(x_k)$
2. Update using $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$
3. Stop when $|x_{k+1} - x_k| < \epsilon_{tol}$.

Halley's Method for Root Finding

Consider the Taylor series for $f(x)$ approximated around $f(x_0)$:

$$f(x) \approx f(x_0) + f'(x_0) \cdot (x - x_0) + f''(x_0) \cdot (x - x_0)^2 + o_p(3)$$

Now let's consider the second-order approximation:

$$\begin{aligned} x_{n+1} &= x_n - \frac{2f\left(x_n\right) f'\left(x_n\right)}{2\left[f'\left(x_n\right)\right]^2 - f\left(x_n\right) f''\left(x_n\right)} = x_n - \frac{f\left(x_n\right)}{f'\left(x_n\right) - \frac{f(x_n)}{f'(x_n)} \frac{f''(x_n)}{2}} \\ &= x_n - \frac{f\left(x_n\right)}{f'\left(x_n\right)} \left[1 - \frac{f\left(x_n\right)}{f'\left(x_n\right)} \cdot \frac{f''\left(x_n\right)}{2f'\left(x_n\right)}\right]^{-1} \end{aligned}$$

- Last equation is useful because we only need to know $f(x_n)/f'(x_n)$ and $f''(x_n)/f'(x_n)$
- If we are lucky $f''(x_n)/f'(x_n)$ is easy to compute or ≈ 0 (Newton's method).

Root Finding: Convergence

How many iterations do we need? This is a tough question to answer.

- However we can consider convergence where $f(a) = 0$:

$$|x_{n+1} - a| \leq K_d * |x_n - a|^d$$

  - $d = 2$ (Newton's Method) quadratic convergence (we need $f'(x)$)
  - $d = 3$ (Halley's Method) cubic convergence (but we need $f''(x)$)

Root Finding: Fixed Points

Some (not all) equations can be written as $f(x) = x$ or $g(x) = 0 : f(x) - x = 0$.

- In this case we can iterate on the fixed point directly

$$x_{n+1} = f(x_n)$$

- Advantage: we only need to calculate $f(x)$.
- There need not be a unique solution to $f(x) = x$.
- But... this may or may not actually work.

Contraction Mapping Theorem/ Banach Fixed Point

Consider a set $D \subset \mathbb{R}^n$ and a function $f : D \to \mathbb{R}^n$. Assume

1. D is closed (i.e., it contains all limit points of sequences in D)
2. $x \in D \Longrightarrow f(x) \in D$
3. The mapping g is a contraction on D : There exists $q < 1$ such that

$$\forall x, y \in D : \quad \|f(x) - f(y)\| \leq q\|x - y\|$$

Then

1. There exists a unique $x^* \in D$ with $f(x^*) = x^*$
2. For any $x^{(0)} \in D$ the fixed point iterates given by $x^{(k+1)} := f\left(x^{(k)}\right)$ converge to x^* as $k \to \infty$
3. $x^{(k)}$ satisfies the a-priori error estimate $\left\|x^{(k)} - x^*\right\| \leq \frac{q^k}{1-q} \left\|x^{(1)} - x^{(0)}\right\|$
4. $x^{(k)}$ satisfies the a-posteriori error estimate $\left\|x^{(k)} - x^*\right\| \leq \frac{q}{1-q} \left\|x^{(k)} - x^{(k-1)}\right\|$

Some notes

- Not every fixed point relationship is a contraction.
- Iterating on $x_{n+1} = f(x_n)$ will not always lead to $f(x) = x$ or $g(x) = 0$.
- Convergence rate of fixed point iteration is slow or $q-$linear.
- When q is small this will be faster.
- q is sometimes called modulus of contraction mapping.
- A key example of a contraction: value function iteration!

Accelerated Fixed Points: Secant Method

Start with Newton's method and use the finite difference approximation

$$f'(x_{n-1}) \approx \frac{f(x_{n-1}) - f(x_{n-2})}{x_{n-1} - x_{n-2}}$$

$$x_n = x_{n-1} - f(x_{n-1}) \frac{x_{n-1} - x_{n-2}}{f(x_{n-1}) - f(x_{n-2})}$$

- This doesn't have the actual $f'(x_n)$ so it isn't quadratically convergent
- Instead is is superlinear with rate $q = \frac{1+\sqrt{5}}{2} = 1.618 < 2$ (Golden Ratio)
- Faster than fixed-point iteration but doesn't require computing $f'(x_n)$.
- Idea: can use past iterations to approximate derivatives and accelerate fixed points.

Accelerated Fixed Points: Anderson (1965) Mixing

Define the residual $r(x_n) = f(x_n) - x_n$. Find weights on previous k residuals:

$$\widehat{\alpha^n} = \arg\min_{\alpha} \left\| \sum_{k=0}^{m} \alpha_k^n \cdot r_{n-k} \right\| \text{ subject to } \sum_{k=0}^{m} \alpha_k^n = 1$$

$$x_{n+1} = (1-\lambda) \sum_{j=0}^{m} \widehat{\alpha_k^n} \cdot x_{n-k} + \lambda \sum_{j=0}^{m} \widehat{\alpha_k^n} \cdot f\left(x_{n-k}\right)$$

- Convex combination of weighted average of: lagged x_{n-k} and lagged $f(x_{n-k})$.
- Variants on this are known as Anderson Mixing or Anderson Acceleration.

Accelerated Fixed Points: SQUAREM (Varadhan and Roland 2008)

Define the residual $r(x_n) = f(x_n) - x_n$ and $v(x_n) = f \circ r\,(x_n) = f \circ f\,(x_n) - f\,(x_n)$.

$$\begin{aligned} x_{n+1} =& x_n && -2s\left[f\,(x_n) - x_n\right] && +s^2\left[f \circ f\,(x_n) - 2f\,(x_n) + x_n\right] \\ =& x_n && -2sr && +s^2(v - r) \end{aligned}$$

Three versions of stepsize:

$$s_1 = \frac{r^t r}{r^t(v-r)}, \quad s_2 = \frac{r^t(v-r)}{(v-r)^t(v-r)}, \quad s_3 = -\sqrt{\frac{r^t r}{(v-r)^t(v-r)}}$$

Idea: use two iterations to construct something more like the quadratic/Halley method.

Note: I am hand-waving, don't try to derive this.