Improving on Newton?

How would we do "Newton + 1" (i.e. even faster, even better)?

- Use another term in the Taylor series
- even more locally convergent
- possibly, 3rd order convergent
- parabola might not have roots
- need 2 derivatives

Root Finding with Interpolants

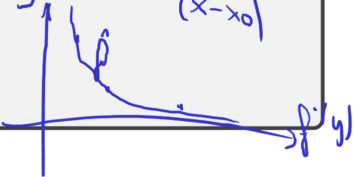
Secant method uses a linear interpolant based on points $f(x_k)$, $f(x_{k-1})$, could use more points and higher-order interpolant:

- Could fit polynomial to $(x_n, f(x_n))$, $(x_{n-1}, f(x_{n-1}))$, $(x_{n-2}, f(x_{n-2}))$
- Muller's method

What about existence of roots in that case?

If x_0 is a single root of f, $\hat{f}(x) = \dfrac{f(x)}{(x-x_0)}$

- Inverse quadratic interpolation
  Fit inverse to $(f(x_0), x_0)$, $(f(x_1), x_1)$, $(f(x_2), x_2)$



Next guess is $\hat{f}(0)$

Achieving Global Convergence

The linear approximations in Newton and Secant are only good locally. How could we use that?

- Hybrid: bisection + Newton
  - ↪ Revert to bisection if Newton leaves the bisection bracket
- "Trust region" method
- Limit step size

In-Class Activity: Nonlinear Equations

In-class activity: Nonlinear Equations

Fixed Point Iteration

$\|g(x)-g(y)\| \le \gamma \|x-y\|$

$\vec{x} = (x_1 \cdots x_n)$

$\vec{g} = (g_1 \cdots g_n)$

$$\begin{aligned} \boldsymbol{x}_0 &= \langle\text{starting guess}\rangle \\ \boldsymbol{x}_{k+1} &= \boldsymbol{g}(\boldsymbol{x}_k) \end{aligned}$$

$\hookrightarrow$ n-dimensional

When does this converge?

$$J_g(x^*) = \begin{bmatrix} \partial_{x_1} g_1 & \cdots & \partial_{x_n} g_1 \\ \vdots & & \\ \partial_{x_1} g_n & \cdots & \partial_{x_n} g_n \end{bmatrix}.$$

$$\|g(x)-g(x^*)\| \sim \|J_g(x^*)\cdot(x-x^*)\|$$

$$\le \|J_g(x^*)\|\cdot\|x-x^*\|.$$

$\underbrace{\|J_g(x^*)\|}_{=0.\ \text{quadratic.}} \le \rho(J_g(x^*))+\varepsilon < 1 \Rightarrow$ convergence.

Newton's Method

What does Newton's method look like in n dimensions?

$$f(x+s) \approx f(x) + J_f(x)\, s$$

$$0 = f(x) + J_f(x)\, s \Rightarrow s = -J_f(x)^{-1} f(x)$$

$$\begin{cases} x_0 = \text{initial guess} \\ x_{k+1} = x_k - J_f(x)^{-1} \cdot f(x). \end{cases}$$

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

Downsides of n-dim. Newton?

- locally convergent only
- computing / inverting J_f is expensive.

Demo: Newton's method in n dimensions [cleared]

Secant in n dimensions?

What would the secant method look like in n dimensions?

$$1D: \quad f'(x) \approx \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}$$

$$nD: \quad \underbrace{J_f(x)}_{n \times n} \approx \tilde{J}\left[\underbrace{f(x_{k+1})}_{n}, \underbrace{f(x_k)}_{n}, \underbrace{x_{k+1}}_{n}, \underbrace{x_k}_{n}\right]$$

$4n$

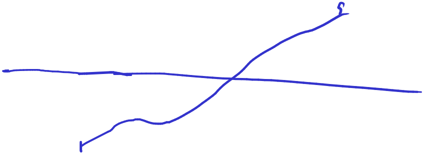
$$\tilde{J} \cdot \Delta x = \Delta f$$

Broyden's method etc.

$\|J_{k+1} - J_k\|_F$

Outline

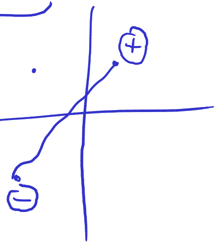
Optimization: Problem Statement



Have: Objective function $f : \mathbb{R}^n \to \mathbb{R}$
Want: Minimizer $\boldsymbol{x}^* \in \mathbb{R}^n$ so that

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}) = 0 \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) \leqslant 0.$$

- $\boldsymbol{g}(\boldsymbol{x}) = 0$ and $\boldsymbol{h}(\boldsymbol{x}) \leqslant 0$ are called constraints.
  They define the set of feasible points $\boldsymbol{x} \in S \subseteq \mathbb{R}^n$.
- If $\boldsymbol{g}$ or $\boldsymbol{h}$ are present, this is constrained optimization.
  Otherwise unconstrained optimization.
- If $\boldsymbol{f}$, $\boldsymbol{g}$, $\boldsymbol{h}$ are *linear*, this is called linear programming.
  Otherwise nonlinear programming.



Optimization: Observations

Q: What if we are looking for a *maximizer* not a minimizer?

Give some examples:

... lots

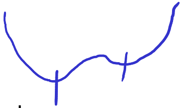
$\min \|F(x)\|$

What about multiple objectives?

Pareto optimality

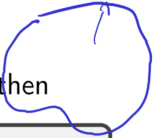
For us: combine into one

Existence/Uniqueness



Terminology: global minimum / local minimum

Under what conditions on f can we say something about existence/uniqueness?



If $f : S \to \mathbb{R}$ is continuous on a closed and bounded set $S \subseteq \mathbb{R}^n$, then

a minimum exists

$f : S \to \mathbb{R}$ is called *coercive* on $S \subseteq \mathbb{R}^n$ (which must be unbounded) if

$$\lim_{\|x\| \to \infty} f(x) = +\infty$$

If f is coercive, ...

a global minimum exists

Convexity

$S \subseteq \mathbb{R}^n$ is called convex if for all $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \leqslant \alpha \leqslant 1$

$f : S \to \mathbb{R}$ is called convex on $S \subseteq \mathbb{R}^n$ if for $\setminus$ $\boldsymbol{x}, \boldsymbol{y} \in S$ and all $0 \leqslant \alpha \leqslant 1$

Q: Give an example of a convex, but not strictly convex function.