Last Time:

- Discrete-time sim/dynamics
- Stability of discrete-time systems
- Forward/Backward Euler, RK4
- Zero/First-order hold on controls

Today:

- Notation
- Root Finding
- Minimization

Some Notation:

- Given $f(x) : \mathbb{R}^n \to \mathbb{R}$

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{1 \times n} \quad \text{is a \underline{row} vector}$$

- This is because $\frac{\partial f}{\partial x}$ is the linear operator mapping Δx into Δf:

$$f(x + \Delta x) \approx f(x) + \frac{\partial f}{\partial x} \Delta x$$

- Similarly given $g(y) : \mathbb{R}^m \to \mathbb{R}^n$

$$\frac{\partial g}{\partial y} \in \mathbb{R}^{n \times m}$$

because:

$$g(y + \Delta y) \approx g(y) + \frac{\partial g}{\partial y} \Delta y$$

- These conventions make the chain rule work:

$$f(g(y+\Delta y)) \approx f(g(y)) + \left.\frac{\partial f}{\partial x}\right|_{g(y)} \left.\frac{\partial g}{\partial y}\right|_{y} \Delta y$$

- For convenience, we will also define:

$$\nabla f(x) = \left(\frac{\partial f}{\partial x}\right)^T \in \mathbb{R}^{n \times 1} \quad \text{column vector}$$

$$\nabla^2 f(x) = \frac{\partial}{\partial x}\left(\nabla f(x)\right) = \frac{\partial^2 f}{\partial x^2} \in \mathbb{R}^{n \times n}$$

Root Finding:

- given $f(x)$, find x^* such that $f(x^*) = 0$

* Example: equilibrium point of continuous-time dynamics.

- Closely related: fixed point such that $f(x^*) = x^*$

* Example: equilibrium for discrete-time dynamics

* Fixed Point Iteration

  - Simplest solution method

  - If fixed point is stable, just iterate the "dynamics" until you settle into the fixed point.

$$x \leftarrow f(x)$$

- Only works on stable fixed points and it has slow convergence.

* Newton's Method:

  - Fit a linear approximation to $f(x)$:

$$f(x+\Delta x) \approx f(x) + \frac{\partial f}{\partial x}\Delta x$$

  - Set approximation to zero and solve for Δx:

$$f(x) + \frac{\partial f}{\partial x}\Delta x = 0 \Rightarrow \Delta x = -\left(\frac{\partial f}{\partial x}\right)^{-1} f(x)$$

  - Apply correction:

$$x \leftarrow x + \Delta x$$

  - Repeat until convergence

* Example: Backward Euler

  - Very fast convergence w/ Newton

* Take Away Messages:

  - Quadratic convergence
  - Can achieve machine precision
  - Most expensive part: solving linear system $O(n^3)$
  - Can improve complexity by taking advantage of problem structure (more later).

Minimization:

$$\min_x f(x) \quad , \quad f: \mathbb{R}^n \to \mathbb{R}$$

- If f is smooth, $\left.\frac{\partial f}{\partial x}\right|_{x^*} = 0$ at local min

- Now we have a root-finding problem $\nabla f(x^*) = 0$

$\Rightarrow$ Apply Newton root finding to $\nabla f(x) = 0$

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \underbrace{\frac{\partial}{\partial x}(\nabla f(x))}_{\nabla^2 f(x)} \Delta x = 0$$

$$\Rightarrow \Delta x = -(\nabla^2 f(x))^{-1} \nabla f(x)$$

$$x \leftarrow x + \Delta x$$

repeat until convergence

* Intuition:

- Fit a quadratic approximation to $f(x)$
- Exactly minimize quadratic approximation

* Example:

$$\min f(x) = x^4 + x^3 - x^2 - x$$

- Start at: 1.0, -1.5, 0.0 } maximizes!

* Take Away Message:

- Newton is a local root-finding method. Will converge to the closest fixed point to the initial guess (min, max, or saddle).

* Sufficient Conditions

- $\nabla f(x) = 0$ "first-order necessary condition" for a minimum. Not sufficient.

- Let's think about the scalar case:

$$\Delta x = \underset{\text{descent}}{-} \underbrace{(\nabla^2 f)^{-1}}_{\text{"learning rate"/"step length"}} \underbrace{\nabla f}_{\text{gradient}}$$

$\nabla^2 f > 0 \Rightarrow$ descent (minimization)

$\nabla^2 f < 0 \Rightarrow$ ascent (maximization)

- In $\mathbb{R}^n$, $\nabla^2 f \succ 0$, $\nabla^2 f \in S_{++}^n$ (positive definite)

$\Rightarrow$ descent

- If $\nabla^2 f > 0$ everywhere $\Leftrightarrow f(x)$ is strongly convex function.

$\Rightarrow$ Can always solve with Newton

- Usually not true for hard/nonlinear problems

* Regularization
  - Practical solution to make sure we're always minimizing:

$$H \leftarrow \nabla^2 f$$

"not pos. def."

$$\text{while } H \not\succ 0$$

$$H \leftarrow H + \beta I \quad (\beta > 0)$$

scalar hyper parameter

$$\text{end}$$

$$\Delta x = -H^{-1} \nabla f$$

$$x \leftarrow x + \Delta x$$

  - Also called "damped Newton"
  - Guarantees descent + shrinks step.

* Example:
  - Now we always minimize
  - What about overshoot?