Last Time:

- Discrete-time dynamics / sim
- Stability of discrete-time systems
- Forward/Backward Euler
- RK4

Today:

- Notation
- Root finding
- Minimization

Some Notation:

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

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

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

$$\frac{\partial g}{\partial y} \in \mathbb{R}^{n \times m} \quad \text{because:}$$

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

- 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 define:

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

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

$$f(x+\Delta x) \approx f(x) + \frac{\partial f}{\partial x}\Delta x + \frac{1}{2}\Delta x^T \frac{\partial^2 f}{\partial x^2}\Delta x$$

Root Finding:

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

* Example: equilibrium of a continuous-time dynamics

- Closely related: fixed point such that

$$f(x^*) = x^*$$

* Example: equilibrium of discrete-time dynamics

* Fixed-point Iteration
  - Simplest solution method
  - If fixed point is stable, just "iterate the dynamics" until it converges to x^*
  - Only works for stable x^* and has slow convergence

* Newton's Method

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

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

- 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 with Newton (Quadratic)
  - Can get machine precision
  - Most expensive part is solving a linear system $O(n^3)$
  - Can improve complexity by taking advantage of problem structure (more later)

Minimization:

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

- If f is smooth, $\frac{\partial f}{\partial x}\big|_{x^*} = 0$ at a local minimum
- Now we have a root-finding problem $\nabla f(x) = 0$

$\Rightarrow$ Apply Newton!

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

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

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

repeate until convergence

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

* Example

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

- start at: 1.0, -1.3, 0.0 maximizes!

* Take-Away Meessages

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

* Suffient Conditions

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

- Lets look at scalar case:

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

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

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

- In $\mathbb{R}^n$, $\nabla^2 f > 0$, $\nabla^2 f \in S^n_{++}$

(positive definite)

$\Rightarrow$ descent

- If $\nabla^2 f(x) > 0$ everywhere $(\forall x) \Leftarrow f(x)$ is strongly convex

$\Rightarrow$ Can always solve with Newton

- Usually not true for hard/nonlinear problems

* Regularization:

- Practical solution to make sure we're always minimizing

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

while $H \nsucc 0$ "not pos. def."

$$H = H + \beta I \qquad (\beta > 0)$$

scalar hyperparameter

end

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

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

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

* Example:

- Now we minimize
- What about overshoot?