Last Time:

- Newton-Euler Dynamics
- SE(3) group
- Quadrotors
- Airplanes

Today:

- Notation
- Root Finding
- Minimization

---

Some Notation:

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

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

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

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

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

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

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

- 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 can define:

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

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

---

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 point for discrete-time dynamics

* Fixed-point Itteration
  - simplest solution method
  - If fixed point is stable, just iterate the dynamics until you settle into the fixed point:

$$x_{n+1} \leftarrow f(x_n)$$

- Only works on stable fixed points
- 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 $\Delta x$:

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

  - Apply correction:

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

  - Repeat until convergence

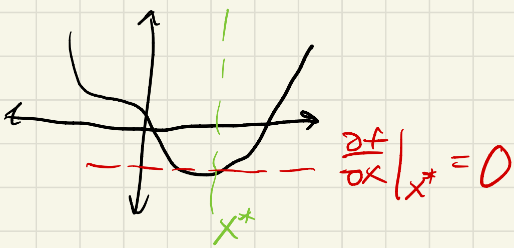
* Example: Backward Euler
  - Very fast convergence w/ Newton

* Take Aways for Newton:
  - 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:

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

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




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

$\Rightarrow$ Apply Newton's method

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

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

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

(repeate until convergence)

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

* Example:

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

  - 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 (max, min, saddle)

* Sufficient Condition
  - $\nabla f(x) = 0$ is "first-order necessary condition" for a min. Not sufficient.
  - Let's think about scalar case:

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

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

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

  $\Rightarrow$ descent

- If $\nabla^2 f \succ 0$ everywhere $\Leftrightarrow$ $f(x)$ is a "strongly convex" function

  $\Rightarrow$ Can always solve with Newton

- Usually not true for hard/nonlinear problems.

* Regularization

- Practical solution to make sure we always minimize:

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

$$\text{while } H \nsucc 0$$

$$H \leftarrow H + \underbrace{\beta}_{\text{scalar}} I \quad (\beta > 0)$$

$$\text{end}$$

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

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

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

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

* Line Search
  - Often $\Delta x$ from Newton is too big and overshoots the minimum
  - To fix this, check $f(x+\Delta x)$ and "back track" until we get a "good" reduction
  - Many strategies exist
  - A simple + effective one is "Armijo rule"

$$\alpha = 1 \quad \leftarrow \text{"step length"}$$

tolerance

$$\text{while } f(x+\alpha\Delta x) > f(x) + \underbrace{b\,\alpha\,\nabla f(x)^T \Delta x}_{\text{expected decrease from gradient}}$$

$$\alpha \leftarrow c\,\alpha$$

scalar $< 1$

end

* Intuition:
  - Make sure step agrees with linearization within some tolerance $b$

* Typical values:

$$c = 1/2\,, \quad b = 10^{-4} \sim 0.1$$

* Take-Away Message:

- Newton with simple + cheap modifications ("globalization strategies") is extremely effective at finding local optima.