Last Time!

- Discretization/Sim of dynamics
- Stability in discrete time
- Forward/Backward Euler, RK4
- Zero/First order hold controls

Today:

- Notation
- Root Finding
- Minimization

Some Notation:

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

$$\frac{\partial f}{\partial x} \in \mathbb{R}^{1 \times n}$$ is a <u>row</u> vector

- This is because $\frac{\partial f}{\partial x}$ is the linear opperator 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 $g(y+\Delta y) \approx g(y) + \frac{\partial g}{\partial y}\Delta y$

- This is important because it makes 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:

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

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

Root Finding:

- given $f(x)$, find x^* such that $f(x^*) = 0$
- Closely related to finding fixed points:

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

- Examples: continuous/discrete time equilibria

* Newton's Method:

$$f(x+\Delta x) \approx f(x) + \left.\frac{\partial f}{\partial x}\right|_x \Delta x = 0$$

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

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

repeate until convergence

* Example: Backward Euler

  - Very fast convergence vs. fixed-point iteration

* Take Away Messages:

  - Quadratic convergence rate
  - Easily 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

$\Rightarrow$ Apply Newton root finding to $\frac{\partial f}{\partial x} = 0$

$$\nabla f(x+\Delta x) \approx \nabla f(x) + \nabla^2 f(x)\, \Delta x = 0$$

$$\Delta x = -(\nabla^2 f)^{-1} \nabla f = 0$$

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

repeate until convergence

* Intuition:
  - Fit quadratic approximation to $f(x)$ using Taylor expansion at current guess solution
  - Exactly minimize quadratic approximation
* Example:
  - Minimize $f(x) = x^4 + x^3 - x^2 - x$
  - Start at 1.0, -1.5, $\underbrace{0.0}_{\text{maximizes!}}$
* Take Away Messages:
  - Newton is a <u>local</u> method. It will find the closest fixed point to the initial guess. (max, min, or saddle)

* Sufficient Conditions
  - How do we know if we're maximizing or minimizing?
  - Let's think about the 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 $\mathbb{R}^n$ case, $\nabla^2 f \succ 0$ (positive definite)
$\Rightarrow$ descent

- If $\nabla^2 f \succ 0$ everywhere $\Leftrightarrow$ $f(x)$ 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:

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

"not positive definite"

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

$$H \leftarrow H + \beta I$$

end

scalar hyper parameter

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

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

- Also called "damped Newton method"

- Guarantees descent + shrinks step size

* Back to Example:

  - Now minimizes starting at $x = 0$
  - What about overshoot?

* Line Search

  - Often Δx is too big and overshoots the minimum
  - To fix this, check $f(x+\Delta x)$ and "backtrack" until we get a "good" reduction
  - Many strategies exist.
  - A simple + effective one is the "Armijo rule":

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

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

$$\alpha \leftarrow \overset{\text{scalar} < 1}{c}\,\alpha$$

$$\text{end}$$

* Intuition:

  - Make sure step agrees with linearization within some tolerance.

* Typical Values:

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

*Take Away Messages:

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