Last Time:

- Optimization Pt. 2
- Constraints

Today:

- Calculus of Variations

Calculus:

- Last time we looked at

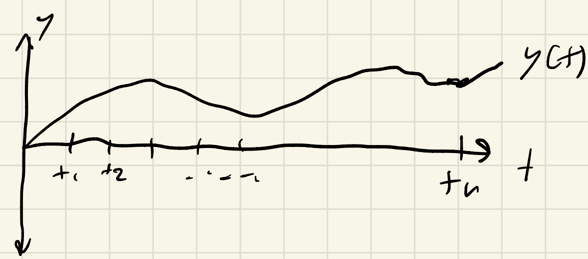
$$\min_x f(x) \quad , \quad f: \mathbb{R}^n \to \mathbb{R}$$

where $x \in \mathbb{R}^n$ is a vector and f is a mapping from $\mathbb{R}^n$ to a scalar

- We'll call this "finite dimensional" optimization

- In dynamics/physics/optimal control, we often want to solve "infinite dimensional" optimization problems.

- Imagine $x \in \mathbb{R}^n$ is a vector of samples from some continuous function $y(t)$:



$$x = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

- If we increase sample rate, n increases
- We can take $n \to \infty$ and recover the original continuous function $y(t)$
- Trajectories (functions of time or similar scalar) can be thought of as "infinite dimensional vectors".
- We can similarly define cost/objective functions that map infinite-dimensional vectors/trajectories to a scalar.

* Example:

$$f(x) = \frac{1}{N} x^T x = \sum_{n=1}^{N} \frac{1}{N} y_n^2$$

$$\Rightarrow \lim_{N \to \infty} f(x) = F(y(t)) = \int_0^1 y(t)^2 \, dt$$

- These "functions of functions" are often called "functionals"
- Let's look at minimizing a functional w.r.t. a function:

$$\min_x f(x) = \frac{1}{N} x^T x \Rightarrow \frac{\partial f}{\partial x} = \frac{2}{N} x^T = 0$$

$$\Rightarrow x^* = 0$$

- Let's look at this in components:

$$\min_x \sum_{n=1}^{N} \frac{1}{N} y_i^2$$

- $f(x)$ is a nonlinear function from $\mathbb{R}^n \to \mathbb{R}$ it's first derivative is the best linear approximation of $f(x)$ at a point.

- This generalizes to the infinite-dimensional case:

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

$$= \tfrac{1}{N}x^T x + \underbrace{\tfrac{2}{N}x^T \Delta x}_{\substack{\text{Think of this as a} \\ \text{linear operator/function} \\ \frac{\partial f}{\partial x}(\Delta x)}}$$

- Taking limit $N \to \infty$ we get:

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

$$= \int_0^1 y(t)^2\,dt + \underbrace{\int_0^1 2y(t)\,\Delta y(t)\,dt}_{\frac{\partial F}{\partial y}(\Delta y)}$$

- At a min, $F(y)$ is "locally flat", so doesn't change for small "variations" $\Delta y(t)$ of the function $y(t)$:

$$\Rightarrow \frac{\partial F}{\partial y}(\Delta y) = \int_0^1 2y(t)\,\Delta y(t)\,dt = 0 \;\forall\; \Delta y(t)$$

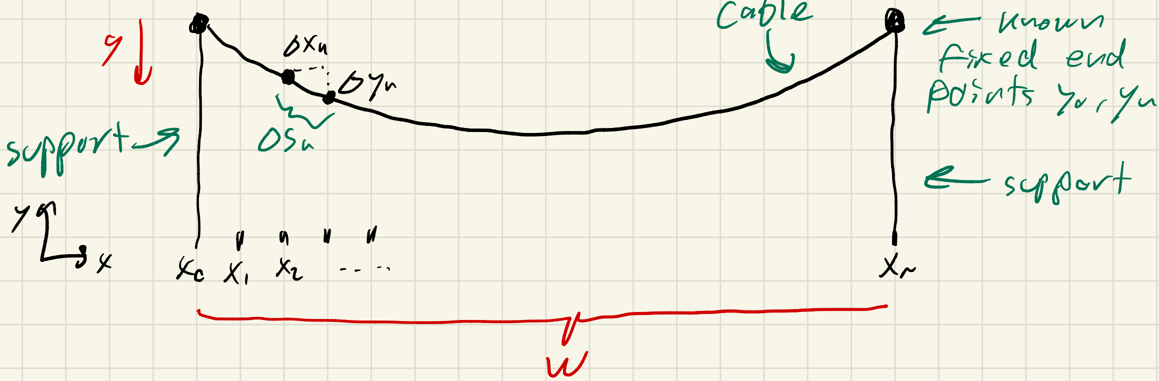
- The only way this can hold is if $y(t) = 0$

$$\Rightarrow y^*(t) = 0$$

- The standard notation in calculus of variations is:

$$\underbrace{\delta F}_{\text{"variation of F"}} = \int_0^1 2y(t)\,\underbrace{\delta y(t)}_{\text{"variation of y"}}\,dt$$

- It is not common to write e.g. $\frac{\delta F}{\delta y}$
- Let's look at a more interesting problem



$$\underbrace{\Delta s_n^2}_{\text{incremental cable length}} = \Delta x_n^2 + \Delta y_n^2 \quad , \quad \Delta x = \frac{w}{N-1}$$

- Total cable length

$$l = \sum_{n=1}^{N-1} \Delta s_n = \sum_{n=1}^{N-1} \sqrt{\Delta x_n^2 + \Delta y_n^2} = \sum_{n=1}^{N-1} \Delta x \sqrt{1 + \left(\Delta y_n / \Delta x\right)^2}$$

- Total mass:

$$m = \sum_{n=1}^{N-1} \rho \Delta s_n \quad , \quad \rho = \frac{m}{l}$$

- What is the shape of the cable?

- Approach: minimize potential energy. If the system is at rest there, it can't move.

- We're going to think of shape $y(x)$ as a trajectory where x plays the role of time.

- Let's discretize x:

  Potential Energy:

$$U = \sum_{n=1}^{N-1} \underbrace{\rho \Delta s_n}_{m_n} g \left(\frac{y_{n+1} + y_n}{2} \right)$$

$$\sum_{n=1}^{N-1} \rho g \left(\frac{y_{n+1} + y_n}{2} \right) \Delta x \sqrt{1 + (\Delta y_n / \Delta x)^2}$$

- Optimization Problem:

$$\min_{y_{1:N-1}} \sum_{n=1}^{N-1} \rho g \left(\frac{y_{n+1} + y_n}{2} \right) \sqrt{1 + (\Delta y_n / \Delta x)^2} \, \Delta x$$

$$\text{s.t.} \sum_{n=1}^{N-1} \sqrt{1 + (\Delta y_n / \Delta x)^2} \, \Delta x - \ell = 0$$

- Note endpoints y_0 and y_N are given

- This problem has lots of structure that we can use.

- Let's look at a generic version:

$$\min_{x} \sum_{n=1}^{N-1} f(y_n, y_{n+1})$$

$$\text{s.t.} \sum_{n=1}^{N-1} c(y_n) = 0$$

$$\Rightarrow L(x,\lambda) = \sum_{n=1}^{N-1} f(y_n, y_{n+1}) + \lambda c(y_n)$$

- KKT Conditions:

$$\frac{\partial L}{\partial x}(\Delta x) = \sum_{n=0}^{N-1} D_1 f(y_n, y_{n+1})\, \Delta y_n$$

$$+ D_2 f(y_n, y_{n+1})\, \Delta y_{n+1}$$

$$+ \lambda\, D c(x_n)\, \Delta y_n = 0 \quad \forall \Delta x$$

"slot derivative"

$$\frac{\partial L}{\partial \lambda} = \sum_{n=0}^{N-1} c(x_n) = 0$$

- We can play an index trick knowing the end points are fixed:

$$\cancelto{0}{D_1 f(y_0, y_1)\, \Delta y_0} + D_2 f(y_0, y_1)\, \Delta y_1 + \cancelto{0}{\lambda\, D c(y_0)\, \Delta y_0}$$

$$+ \sum_{n=1}^{N-2} D_1 f(y_n, y_{n+1})\, \Delta y_n + D_2 f(y_n, y_{n+1})\, \Delta y_{n+1}$$

$$+ \lambda\, D c(y_n)\, \Delta y_n$$

$$D_1 f(y_{N-1}, y_N)\, \Delta y_{N-1} + \cancelto{0}{D_2 f(y_{N-1}, y_N)\, \Delta y_N}$$

$$+ \lambda\, D c(y_{N-1})\, \Delta y_{N-1}$$

$$= \sum_{n=1}^{N-1} D_2 f(y_{n-1}, y_n) \Delta y_n + D_1 f(y_n, y_{n+1}) \Delta y_n$$
$$+ \lambda DC(y_n) \Delta y_n$$

- Now I can factor ot Δy_n:

$$\frac{\partial L}{\partial x}(\Delta x) = \sum_{n=1}^{N-1} \left[D_2 f(y_{n-1}, y_n) + D_1 f(y_n, y_{n+1}) + \lambda D\, C(y_n) \right] \Delta y_n = 0 \quad \forall \Delta y_n$$

- For this to be true, we must have:

$$D_2 f(y_{n-1}, y_n) + D_1 f(y_n, y_{n+1}) + \lambda DC(y_n) = 0$$

- Now we have local optimality conditions that only depend on immediate neighbors
- We can use this to calculate the whole solution given 2 points as initial conditions.
- Let's look at limit $N \to \infty$

$$\min_{y(t)} \int_0^T f(y(t), \dot{y}(t))\, dt$$

$$\text{s.t.} \quad \int_0^T C(y(t), \dot{y}(t))\, dt = 0$$

- Lagrangian:

$$L(y(t), \lambda) = \int_0^T f(y(t), \dot{y}(t)) + \lambda C(y(t), \dot{y}(t))\, dt$$

- "KKT Conditions":

$$\frac{\partial L}{\partial y(t)}(\Delta y(t)) = \int_0^T D_1 f(y(t), \dot{y}(t))\Delta y(t) + D_2 f(y, \dot{y})\Delta \dot{y}(t)$$
$$+ \lambda D_1 C(y, \dot{y})\Delta y(t)$$
$$+ \lambda D_2 C(y, \dot{y})\Delta \dot{y}(t)\, dt = 0$$

$$\frac{\partial L}{\partial \lambda}(\Delta y) = \int_0^T C(y, \dot{y})\, dt = 0$$

- Similar to the discrete case, we want to factor out $\Delta y(t)$ to derive local optimality condition, but $\Delta \dot{y}(t)$ mess this up.
- The key trick is integration by parts:

$$\frac{d}{dt}(u(t)v(t)) = \dot{u}v + u\dot{v}$$

$$\Rightarrow u(t)v(t)\Big|_0^T = \int_0^T \dot{u}v\, dt + \int_0^T u\dot{v}\, dt$$

$$\Rightarrow \int_0^T u\dot{v}\, dt = \underbrace{uv\Big|_0^T}_{\text{boundary term}} - \int_0^T \underbrace{\dot{u}v}_{\text{"flip the dot"}}\, dt$$

- Apply to our KKT conditions:

$$\frac{\partial L}{\partial y(t)}(\Delta y(t)) = \int_0^T D_1 f(y,\dot{y})\Delta y - \frac{d}{dt}(D_2 f(y,\dot{y}))\Delta y + \lambda D_1 c(y,\dot{y})\Delta y - \frac{d}{dt}(D_2 c(y,\dot{y}))\Delta y \, dt$$

$$+ \cancelto{0}{D_2 f(y,\dot{y})\Delta y + D_2 c(y,\dot{y})\Delta y \Big|_0^T}$$

(fixed end points)

- Now we can pull out $\Delta y(t)$:

$$D_1 f(y,\dot{y}) - \frac{d}{dt}[D_2 f(y,\dot{y})] + \lambda\left[D_1 c(y,\dot{y}) - \frac{d}{dt} D_2 c(y,\dot{y})\right] = 0$$

$$\underbrace{\frac{\partial f}{\partial y} - \frac{d}{dt}\frac{\partial f}{\partial \dot{y}} + \lambda\left[\frac{\partial c}{\partial y} - \frac{d}{dt}\frac{\partial c}{\partial \dot{y}}\right]}_{\text{Euler-Lagrange Equation!}} = 0$$

- This is a 2^{nd}-order ODE for $y(t)$. Given initial conditions $y(0)$, $\dot{y}(0)$, we can numerically integrate to find $y(t)$.