Last Time:

- Optimization Review

Today:

- Calculus of Variations

Calculus of Variations:

- Last time we looked at:

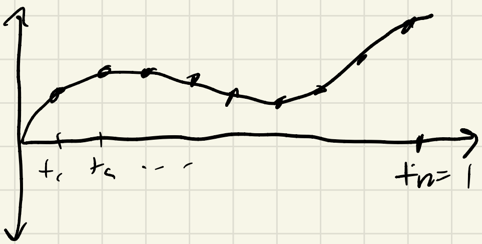
$$\min_x \; f(x) \quad , \quad f: \mathbb{R}^n \to \mathbb{R}$$

Where $x \in \mathbb{R}^n$ is a vector and f is a mapping from $\mathbb{R}^n$ to a scalar cost/objective.

- We'll call this "finite dimensional" optimization.

- In dynamics (and physics, optimal control, etc.) we often need 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 the sample rate, n increases

- We can take the limit $n \to \infty$ and recover the original continuous function $y(t)$

- Trajectories (functions of time or similar scalar) can be thought of as "infinit dimensional" vectors.

- We can similarly define cost/objective functions that map infinite-dimensional vectors/trajectories to a scalar. For example:

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

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

- 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:

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

- $f(x)$ 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:

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

$$= \frac{1}{N} x^T x + \underbrace{\frac{2}{N} x^T}_{} \Delta x$$

Think of this as a linear operator or function $\frac{\partial f}{\partial x}(\Delta x)$

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

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

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

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

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

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

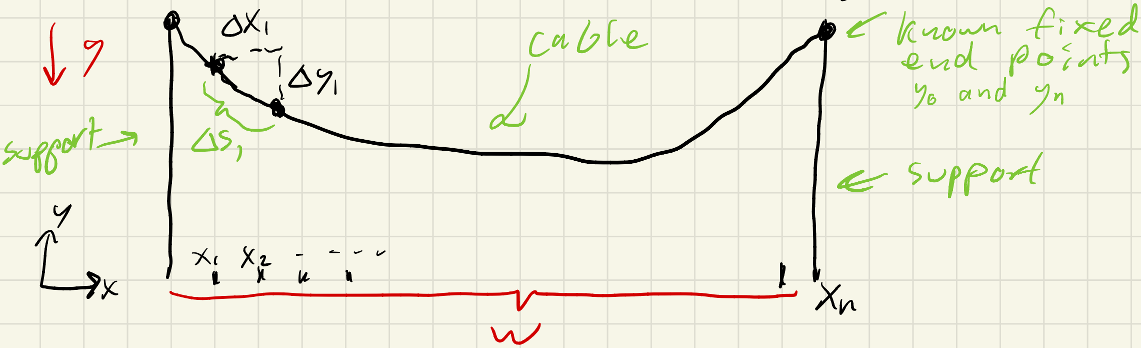
$$\Rightarrow y^*(t) = 0$$

- 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. $\frac{\delta F}{\delta y}$

- Lets look at a more interesting problem:



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

- Total cable length:

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

- Total mass:

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

density

- What is the shape of the cable?

- I could write down the dynamics $\dot{x} = f(x)$ and solve for an equilibrium $\dot{x} = f(x^*) = 0$

- Another approach: minimize potential energy. If the system starts there at rest, it can't move.

- We can think of the cable shape as a trajectory $y(x)$ where x plays the role of time.

- Let's discretize X:

Potential Energy:

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

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

- Optimization Problem:

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

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

- Note endpoints are given (y_0 and y_n)

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

- Let's look at a more generic version:

$$\min_{x} \sum_{n=1}^{N-1} g(x_n, x_{n+1})$$

$$\text{s.t.} \quad c(x_n) = 0 \quad \forall k$$

$$\Rightarrow L(x, \lambda) = \sum_{n=1}^{N-1} g(x_n, x_{n+1}) + \lambda_n^T c(x_n)$$

- KKT Conditions:

$$\frac{\partial L}{\partial x}(\Delta x) = \sum_{n=1}^{N-1} D_1 g(x_n, x_{n+1}) \Delta x_n$$

$$+ D_2 g(x_n, x_{n+1}) \Delta x_{n+1}$$

"slot derivative"

$$+ \lambda_n^T D c(x_n) \Delta x_n = 0 \quad \forall \Delta x$$

$$\frac{\partial L}{\partial \lambda} = c(x_n) = 0$$

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

$$D_1 g(x_1, x_2) \cancelto{0}{\Delta x_1} + D_2 g(x_1, x_2) \Delta x_2 + \lambda_1^T D c(x_1) \cancelto{0}{\Delta x_1}$$

$$+ \sum_{n=2}^{N-2} D_1 g(x_n, x_{n+1}) \Delta x_n + D_2 g(x_n, x_{n+1}) \Delta x_{n+1}$$

$$+ \lambda_n^T D c(x_n) \Delta x_n$$

$$+ D_1 g(x_{N-1}, x_N) \Delta x_{N-1} + D_2 g(x_{N-1}, x_N) \cancelto{0}{\Delta x_N}$$

$$+ \lambda_{N-1}^T D c(x_{N-1}) \Delta x_{N-1}$$

$$= \sum_{n=2}^{N-1} D_2 g(X_{n-1}, X_n)\Delta X_n + D_1 g(X_n, X_{n+1})\Delta X_n + \lambda_n^T DC(X_n)\Delta X_n$$

- Now I can factor out the ΔX_n:

$$\frac{\partial L}{\partial X}(\Delta X) = \sum_{n=2}^{N-1} \left[D_2 g(X_{n-1}, X_n) + D_1 g(X_n, X_{n+1}) + \lambda_n^T DC(X_n)\right]\Delta X_n = 0 \quad \forall \Delta X_n$$

- For this to be true for any ΔX, we must have:

$$D_2 g(X_{n-1}, X_n) + D_1 g(X_n, X_{n+1}) + \lambda_n^T DC(X_n) = 0$$

- 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 now look at limit $N \to \infty$:

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

- Lagrangian:

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

- KKT Conditions:

$$\frac{\partial L}{\partial x(t)}(\Delta x) = \int_0^T D_1 g(x,\dot{x})\Delta x + D_2 g(x,\dot{x})\Delta\dot{x}$$
$$+ \lambda^T D c(x)\Delta x \, dt = 0 \quad \forall \Delta x(t)$$

$$\frac{\partial L}{\partial \lambda}(\Delta\lambda) = \int_0^T \Delta\lambda^T c(x(t))\, dt = 0 \quad \forall \Delta\lambda$$

- Similar to the discrete case, we want to factor out $\Delta x(t)$ to derive local optimality condition, but $\Delta\dot{x}$ messes this up.

- Key trick is integration by parts

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

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

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

- Applying this to our KKT conditions:

$$\frac{\partial L}{\partial x(t)}(\Delta x) = \int_0^T D_1 g(x,\dot{x})\Delta x - \frac{d}{dt}\left(D_2 g(x,\dot{x})\right)\Delta x$$
$$+ \lambda^T D c(x)\Delta x\, dt + \cancel{D_2 g(x,\dot{x})\Big|_0^T}^{\,0}$$

(Fixed end points)

- Now we can pull out $\delta x(t)$:

$$D_1 g(x, \dot{x}) - \frac{d}{dt} D_2 g(x, \dot{x}) + \lambda^T D c(x) = 0$$

OR

$$\underbrace{\frac{\partial g}{\partial x} - \frac{d}{dt}\left(\frac{\partial g}{\partial \dot{x}}\right) + \lambda^T \frac{\partial c}{\partial x} = 0}_{\text{Euler-Lagrange Equation}}$$

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