Buvdi Math w/ steps

↳ Discrete Adjoint Euler process

model problem: $y' = \theta y$ (θ is a param.)
$y(0) = 1 = y_0$

In this problem, $f(y) = \theta y$, $h = \frac{T}{N}$

time integrator: N steps of Euler's Method
↳ $y_{n+1} = y_n + h \cdot f(y_n)$

data: $\{(T, \hat{y})\}$ → only using this single data point for simplicity

---

A. Write down error fn: → for data $\{(T, \hat{y})\}$

$$E(\theta) = (y(T) - \hat{y})^2 = (y_N - \hat{y})^2$$

B. Write down the constraints

$$\left.\begin{aligned} y_N &= y_{N-1} + h \cdot f(y_{N-1}; \theta) \\ y_{N-1} &= y_{N-2} + h \cdot f(y_{N-2}; \theta) \\ &\vdots \\ y_1 &= y_0 + h\, f(y_0; \theta) \end{aligned}\right\} \text{N Euler steps}$$

↳ don't have analytic sol. for $y(t)$ to compute error, so we approx $y(t)$ w/ $y_1 \cdots y_N$ and init. cond. y_0.

C. Write down the Lagrangian

Function to optimize: $(y(T) - \hat{y})^2$

Constraints set $= 0$: $\sum_{i=1}^{N} y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta)$

$$L(y_1, \dots, y_N, \lambda_1, \dots, \lambda_N, \theta) = (y(T) - \hat{y})^2 + \sum_{i=1}^{N} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta))$$

D. Optimal θ (i.e. θ that minimizes our error fn $E(\theta)$) if the following conditions are satisfied.

$$\frac{\partial L}{\partial y_j} = 0;\ j = 1, \dots, N$$

} Gives us a formula for λ_j^*. The iteration is solved backward in time
↳ b/c y_i wiped out

$$\frac{\partial L}{\partial \lambda_j} = 0;\ j = 1, \dots, N$$

} Tells us to solve for y_i^* using Euler's Method
↳ b/c λ_i wiped out

↑ Our focus ↗

$$\frac{\partial L}{\partial \theta} = 0$$

↳ almost always too difficult to solve directly

i. Show that $\frac{\partial L}{\partial \lambda_j} = 0 \iff y_{j+1} = y_j + h \cdot f(y_j)$

$$L(y_1, \dots, y_N, \lambda_1, \dots, \lambda_N, \theta) = (y(T) - \hat{y})^2 + \sum_{i=1}^{N} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta))$$

$$L = (y(T) - \hat{y})^2 + \lambda_1 y_1 - \lambda_1 y_0 - \lambda_1 h\, f(y_0; \theta) + \sum_{i=2}^{N} \lambda_i (y_i - y_{i-1} - h\, f(y_{i-1}; \theta))$$

$$\frac{\partial L}{\partial \lambda_1} = 0 = 0 + y_1 - y_0 - h \cdot f(y_0; \theta) + 0$$
$$\iff y_1 = y_{1-1} + h \cdot f(y_{1-1}; \theta)$$

Similarly,

$$\frac{\partial L}{\partial \lambda_2} = 0 = 0 + y_2 - y_1 - h \cdot f(y_1; \theta)$$
$$\iff y_2 = y_{2-1} + h \cdot f(y_{2-1}; \theta)$$

$$\vdots$$

$$\frac{\partial L}{\partial \lambda_N} = 0 = 0 + y_N - y_{N-1} - h \cdot f(y_{N-1}; \theta)$$
$$\iff y_N = y_{N-1} + h \cdot f(y_{N-1}; \theta)$$

Thus, $\frac{\partial L}{\partial \lambda_j} = 0 \iff y_j = y_{j-1} + h \cdot f(y_{j-1}; \theta)$ (Euler's Method)
↳ $y_1^*, y_2^*, \dots, y_N^*$

ii. Show that $\frac{\partial L}{\partial y_i} = 0$ gives an iterative method for computing λ_j.

$$L(y_1, \dots, y_N, \lambda_1, \dots, \lambda_N, \theta) = (y(T) - \hat{y})^2 + \sum_{i=1}^{N} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta))$$

$$L = (y(T) - \hat{y})^2 + \sum_{i=1}^{N-2} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta))$$
$$+ \lambda_{N-1}(y_{N-1} - y_{N-2} - h \cdot f(y_{N-2}; \theta))$$
$$+ \lambda_N (y_N - y_{N-1} - h \cdot f(y_{N-1}; \theta))$$
↳ $\lambda_N y_N - \lambda_N y_{N-1} - \lambda_N \cdot h \cdot f(y_{N-1}; \theta)$

$$\frac{\partial L}{\partial y_N} = 0 = 0 + 0 + 0 + \lambda_N - 0 - 0$$
$$\iff \lambda_N = 0$$ → doesn't seem right, but no other y_N in L?

$$\frac{\partial L}{\partial y_{N-1}} = 0 = 0 + 0 + \lambda_{N-1} - 0 - 0 + 0 - \lambda_N - \lambda_N \cdot h \cdot \frac{\partial}{\partial_{N-1}} f(y_{N-1}; \theta)$$

B/c $f(y; \theta) = \theta y$ in this example, $f(y_{N-1}; \theta) = \theta y_{N-1}$.

So, $\frac{\partial}{\partial y_{N-1}} f(y_{N-1}; \theta) = \theta$.

Thus, $\frac{\partial L}{\partial y_{N-1}} = 0 = \lambda_{N-1} - \lambda_N - \lambda_N \cdot h \cdot \theta$

$$\lambda_{N-1} = \lambda_N + \lambda_N \cdot h \cdot \theta$$

Repeating we find $\lambda_j = \lambda_{j+1} + \lambda_{j+1} \cdot h \cdot \frac{\partial}{\partial y_j} f(y_j; \theta)$,
↳ $\lambda_1^*, \lambda_2^*, \dots, \lambda_N^*$ or $\lambda_j = \lambda_{j+1} + \lambda_{j+1} \cdot h \cdot \theta$ in our example problem where $y\ f(y; \theta) = \theta y$.

E. To use Gradient Descent to find the optimal θ, we require $\frac{\partial E}{\partial \theta}$

B/c our constraint fns are all $= 0$ and $E = (y(T) - \hat{y})^2$, then

$$L(y_1^*, \dots, y_N^*, \lambda_1^*, \dots, \lambda_N^*, \theta) = E(\theta).$$

So $\frac{\partial L}{\partial \theta} = \frac{\partial E}{\partial \theta}$. Find $\frac{\partial L}{\partial \theta}$.

$\lambda_i y_i - \lambda_i y_{i-1} - \lambda_i h\, f(y_{i-1}; \theta)$

$$L(y_1, \dots, y_N, \lambda_1, \dots, \lambda_N, \theta) = (y(T) - \hat{y})^2 + \sum_{i=1}^{N} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1}; \theta))$$

$$\frac{\partial L}{\partial \theta} = 0 + 0 - 0 - \sum_{i=1}^{N} \lambda_i^* \cdot h \cdot \frac{\partial}{\partial \theta} f(y_{i-1}^*; \theta),\quad f(y; \theta) = \theta y.$$

$\frac{\partial}{\partial \theta} f(y_{i-1}^*; \theta) = y_{i-1}^*$. Thus,

$$\frac{\partial L}{\partial \theta} = -\sum_{i=1}^{N} \lambda_i^* \cdot h \cdot y_{i-1}^*.$$

Therefore, $\frac{\partial E}{\partial \theta} = -\sum_{i=1}^{N} \lambda_i^* \cdot h \cdot y_{i-1}^*$, too.

Buvdi Math (Adjoint/Lagrangian/Euler) generalized

model problem: $y' = f(y;\theta)$, $y(t_0) = y_0$ (θ is a param.)

time integrator: N steps of Euler's Method

↳ $y_{n+1} = y_n + h \cdot f(y_n)$

In this problem, $f(y;\theta)$ is arb. $h = \frac{T}{N}$

data: $\{(T, \hat{y})\}$ → only using this single data point for simplicity

---

A. Error fn:

$$E(\theta) = (y(T) - \hat{y})^2 = (y_N - \hat{y})^2,$$

$y(T)$ is sol to $y' = f(y;\theta)$,

y_N is approx. sol to $y' = f(y;\theta)$.

B. Constraints

$$y_N = y_{N-1} + h \cdot f(y_{N-1};\theta)$$
$$y_{N-1} = y_{N-2} + h \cdot f(y_{N-2};\theta)$$
$$\vdots$$
$$y_1 = y_0 + h \cdot f(y_0;\theta)$$

C. Lagrangian

Fn. to optimize: $(y(T) - \hat{y})^2$

Constraints set $= 0$: $\sum_{i=1}^{N} y_i - y_{i-1} - h \cdot f(y_{i-1};\theta)$

so:

$$L(y_1, \dots, y_N, \lambda_1, \dots, \lambda_N, \theta) = (y(T) - \hat{y})^2 + \sum_{i=1}^{N} \lambda_i (y_i - y_{i-1} - h \cdot f(y_{i-1};\theta))$$

D. Optimal θ if, for $j = 1, \dots, N$:

$\frac{\partial L}{\partial y_j} = 0$ → gives formula for λ_j^* (solved backward in time, so $N \to 1$)

$\frac{\partial L}{\partial \lambda_j} = 0$ → tells us to solve for y_j^* using Euler's Method (solved forward in time, so $1 \to T$)

$\frac{\partial L}{\partial \theta} = 0$

i. Show that $\frac{\partial L}{\partial \lambda_j} = 0 \iff y_{j+1} = y_j + h \cdot f(y_j;\theta)$

$$\frac{\partial L}{\partial \lambda_j} = 0 = 0 + y_N - y_{N-1} - h \cdot f(y_{N-1};\theta)$$

$\iff y_N = y_{N-1} + h \cdot f(y_{N-1};\theta)$.

Thus, $\frac{\partial L}{\partial \lambda_j} = 0 \iff y_{j+1}^* = y_j + h \cdot f(y_j;\theta)$

↳ $y_1^*, \dots, y_N^*$; y_0^* is init. cond.

Euler's Method.

ii. Show that $\frac{\partial L}{\partial y_j} = 0$ gives an iterative method for computing λ_j.

$\frac{\partial L}{\partial y_N} = 0 \iff \lambda_N^* = 0$ (b/c only one λ_N term in L?)

$$\frac{\partial L}{\partial y_{N-1}} = 0 = \lambda_{N-1} - \lambda_N - \lambda_N \cdot h \cdot \frac{\partial}{\partial y_{N-1}} f(y_{N-1};\theta)$$

$$\lambda_{N-1} = \lambda_N + \lambda_N \cdot h \cdot \frac{\partial}{\partial y_{N-1}} f(y_{N-1};\theta)$$

Repeating, we find

$$\frac{\partial L}{\partial y_j} = 0 \iff \lambda_j^* = \lambda_{j+1} + \lambda_{j+1} \cdot h \cdot \frac{\partial}{\partial y_j} f(y_j;\theta)$$

E. Need $\frac{\partial E}{\partial \theta}$ to perform Grad. Desc.

Know: $L(y_1^*, \dots, y_N^*, \lambda_1^*, \dots, \lambda_N^*, \theta) = E(\theta)$.

Thus, $\frac{\partial L}{\partial \theta} = \frac{\partial E}{\partial \theta}$. Find $\frac{\partial L}{\partial \theta}$.

$$\frac{\partial L}{\partial \theta} = 0 + 0 - 0 - \sum_{i=1}^{N} \lambda_i^* \cdot h \cdot \frac{\partial}{\partial \theta} f(y_{i-1}^*;\theta), \quad y' = f(y;\theta)$$

$$\iff \frac{\partial L}{\partial \theta} = -\sum_{i=1}^{N} \lambda_i^* \cdot h \cdot \frac{\partial}{\partial \theta} f(y_{i-1}^*;\theta).$$

Thus, $\frac{\partial E}{\partial \theta} = -\sum_{i=1}^{N} \lambda_i^* \cdot h \cdot \frac{\partial}{\partial \theta} f(y_{i-1}^*;\theta)$, too.