Pushforward / Jvp rule for Neural ODEs / = Final time integration

$$q(\underline{\theta}, \underline{u}_0, T) = \Big\{ \text{solve } \frac{d\underline{u}}{dt} = f(\underline{u}, \underline{\theta}) \text{ for } \underline{u}(t) \text{ and evaluate it at } t = T \text{ with initial condition } \underline{u}(t=0) = \underline{u}_0 \Big\} =: \underline{u}(T)$$

e.g. RK 4/5

$\underline{\theta} \in \mathbb{R}^P$ ... parameter vector

$\underline{u}_0 \in \mathbb{R}^N$ ... initial condition

$T \in \mathbb{R}$ ... final time

$\underline{u}(T) \in \mathbb{R}^N$ ... final state

→ we are only interested in the final state and not in the full trajectory

task: forward-propagate tangent information on the input $\dot{\underline{\theta}} \in \mathbb{R}^P$, $\dot{\underline{u}}_0 \in \mathbb{R}^N$ & $\dot{T} \in \mathbb{R}$ to $\dot{\underline{u}}(T) \in \mathbb{R}^N$

↳ without unrolling the solver and applying forward-mode AD to its operations ($\hat{=}$ piggyback)

in general:

$$\dot{\underline{u}}(T) = \underbrace{\overset{\hat{=}\, \frac{\partial \underline{u}(T)}{\partial \underline{\theta}}}{\frac{\partial q}{\partial \underline{\theta}}} \dot{\underline{\theta}}}_{\dot{\underline{u}}(T)_\theta} + \underbrace{\frac{\partial q}{\partial \underline{u}_0} \dot{\underline{u}}_0}_{\dot{\underline{u}}(T)_{u_0}} + \underbrace{\frac{\partial q}{\partial T} \dot{T}}_{\dot{\underline{u}}(T)_T}$$

(1) (2) (3)

(1) Tangent contribution from the parameter vector $\underline{\theta}$

general solution to an ODE

$$\underline{u}(T) = \underline{u}_0 + \int_0^T f(\underline{u}(t), \underline{\theta})\, dt$$

total derivative wrt $\underline{\theta}$

$$\frac{\partial \underline{u}(T)}{\partial \underline{\theta}} = \frac{\partial \underline{u}_0}{\partial \underline{\theta}} + \int_0^T \frac{\partial f}{\partial \underline{u}} \frac{\partial \underline{u}}{\partial \underline{\theta}} + \frac{\partial f}{\partial \underline{\theta}}\, dt$$

multiply with $\dot{\underline{\theta}}$ from the right

$$\underbrace{\frac{\partial \underline{u}(T)}{\partial \underline{\theta}} \dot{\underline{\theta}}}_{\dot{\underline{u}}(T)_\theta} = \underbrace{\underbrace{\frac{\partial \underline{u}_0}{\partial \underline{\theta}}}_{\underline{0}} \dot{\underline{\theta}}}_{\underline{0}} + \int_0^T \frac{\partial f}{\partial \underline{u}} \underbrace{\frac{\partial \underline{u}}{\partial \underline{\theta}} \dot{\underline{\theta}}}_{\dot{\underline{u}}(t)_\theta} + \frac{\partial f}{\partial \underline{\theta}} \dot{\underline{\theta}}\, dt$$

$$\dot{\underline{u}}(T)_\theta = \underline{0} + \int_0^T \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}(t)_\theta + \frac{\partial f}{\partial \underline{\theta}} \dot{\underline{\theta}}\, dt$$

this is the general solution to the following ODE

$$\begin{cases} \frac{d\dot{\underline{u}}_\theta}{dt} = \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}_\theta + \frac{\partial f}{\partial \underline{\theta}} \dot{\underline{\theta}} \\ \dot{\underline{u}}_\theta(t=0) = \underline{0} \end{cases}$$

Jacobian-vector product
→ by autodiff
→ by knowing the Jacobian

$$\dot{\underline{u}}(T)_\theta = \Big\{ \text{solve } \frac{d\dot{\underline{u}}_\theta}{dt} = \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}_\theta + \frac{\partial f}{\partial \underline{\theta}} \dot{\underline{\theta}} \text{ for } \dot{\underline{u}}_\theta(t) \text{ and evaluate at } t=T \text{ with initial condition } \dot{\underline{u}}_\theta(t=0) = \underline{0} \Big\}$$

→ the tangent linear problem for the parameter tangent propagation is a linear ODE

→ it is forced (inhomogeneous in the ODE)

→ it has a zero IC (homogeneous in the IC)

(2) Tangent propagation from the initial condition $\underline{u}_0$

again take general solution

$$\underline{u}(T) = \underline{u}_0 + \int_0^T f(\underline{u}(t), \underline{\theta})\, dt$$

take total derivative wrt $\underline{u}_0$

$$\frac{\partial \underline{u}(T)}{\partial \underline{u}_0} = \frac{\partial \underline{u}_0}{\partial \underline{u}_0} + \int_0^T \frac{\partial f}{\partial \underline{u}} \frac{\partial \underline{u}}{\partial \underline{u}_0}\, dt$$

multiply with $\dot{\underline{u}}_0$ from the right

$$\underbrace{\frac{\partial \underline{u}(T)}{\partial \underline{u}_0} \dot{\underline{u}}_0}_{\dot{\underline{u}}(T)_{u_0}} = \underbrace{\frac{\partial \underline{u}_0}{\partial \underline{u}_0}}_{I} \dot{\underline{u}}_0 + \int_0^T \frac{\partial f}{\partial \underline{u}} \frac{\partial \underline{u}}{\partial \underline{u}_0} \dot{\underline{u}}_0\, dt$$

$$\dot{\underline{u}}(T)_{u_0} = \dot{\underline{u}}_0 + \int_0^T \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}(t)_{u_0}\, dt$$

Jacobian-vector product
→ by autodiff
→ by knowing the Jacobian

$$\dot{\underline{u}}(T)_{u_0} = \Big\{ \text{solve } \frac{d\dot{\underline{u}}_{u_0}}{dt} = \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}_{u_0} \text{ for } \dot{\underline{u}}_{u_0}(t) \text{ and evaluate at } t=T \text{ with IC } \dot{\underline{u}}_{u_0}(t=0) = \dot{\underline{u}}_0 \Big\}$$ still linear

→ homogeneous in the ODE

→ inhomogeneous in the initial condition

(3) Tangent contribution from the final time $\dot{T}$

again general solution to the ODE

$$\underline{u}(T) = \underline{u}_0 + \int_0^T f(\underline{u}(t), \underline{\theta})\, dt$$

take total derivative wrt T

$$\frac{\partial \underline{u}(T)}{\partial T} = \frac{\partial \underline{u}_0}{\partial T} + f(\underline{u}(T), \underline{\theta})$$

multiply with $\dot{T}$ from the right

$$\underbrace{\frac{\partial \underline{u}(T)}{\partial T} \dot{T}}_{\dot{\underline{u}}(T)_T} = \underbrace{\cancel{\frac{\partial \underline{u}_0}{\partial T}}}_{=\underline{0}} \dot{T} + f(\underline{u}(T), \underline{\theta}) \dot{T}$$

$$\dot{\underline{u}}(T)_T = \underline{0} + f(\underline{u}(T), \underline{\theta}) \dot{T}$$

Summary

obtain components

(1) $\dot{\underline{u}}_\theta(T) = \Big\{ \text{integrate } \frac{d\dot{\underline{u}}_\theta}{dt} = \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}_\theta + \frac{\partial f}{\partial \underline{\theta}} \dot{\underline{\theta}} \text{ with } \dot{\underline{u}}_\theta(t=0) = \underline{0} \Big\}$

(2) $\dot{\underline{u}}_{u_0}(T) = \Big\{ \text{integrate } \frac{d\dot{\underline{u}}_{u_0}}{dt} = \frac{\partial f}{\partial \underline{u}} \dot{\underline{u}}_{u_0} \text{ with } \dot{\underline{u}}_{u_0}(t=0) = \dot{\underline{u}}_0 \Big\}$

both (1) & (2) similar to the primal problem we do not need the full trajectory, but only the solution at the final time $t=T$

(3) $\dot{\underline{u}}_T(T) = f(\underline{u}(T), \underline{\theta}) \dot{T}$

Full Pushforward rule

$$\boxed{\mathcal{F}\big(q, (\underline{\theta}, \underline{u}_0, T), (\dot{\underline{\theta}}, \dot{\underline{u}}_0, \dot{T})\big) = \Big(\underbrace{\big(q(\underline{\theta}, \underline{u}_0, T)\big)}_{\underline{u}(T)}, \underbrace{\big(\dot{\underline{u}}_\theta + \dot{\underline{u}}_{u_0} + \dot{\underline{u}}_T\big)}_{\dot{\underline{u}}(T)}\Big)}$$

Computational Considerations

→ instead of solving 2 additional problems both for $\mathbb{R}^N$ degrees of freedom, solve one large $\mathbb{R}^{3N}$ ODE problem of the tangent propagation together with the primal pass

(→ by this we also always have the primal values available for the Jacobian-vector product)

$$\frac{d}{dt} \begin{bmatrix} \underline{u} \\ \dot{\underline{u}}_\theta \\ \dot{\underline{u}}_{u_0} \end{bmatrix} = \begin{bmatrix} f(\underline{u}, \underline{\theta}) \\ \frac{\partial f}{\partial \underline{u}}\Big|_{\underline{u}, \underline{\theta}} \dot{\underline{u}}_\theta + \frac{\partial f}{\partial \underline{\theta}}\Big|_{\underline{u}, \underline{\theta}} \dot{\underline{\theta}} \\ \frac{\partial f}{\partial \underline{u}}\Big|_{\underline{u}, \underline{\theta}} \dot{\underline{u}}_{u_0} \end{bmatrix}$$

$$\text{with} \begin{bmatrix} \underline{u} \\ \dot{\underline{u}}_\theta \\ \dot{\underline{u}}_{u_0} \end{bmatrix} (t=0) = \begin{bmatrix} \underline{u}_0 \\ \underline{0} \\ \dot{\underline{u}}_0 \end{bmatrix}$$

Obtain the Jvp $\frac{\partial f}{\partial \underline{u}}$ sth. $\frac{\partial f}{\partial \underline{\theta}}$ sth. either:

→ forward-mode autodiff

→ knowing the Jacobian (e.g. by the sparsity structure of PDEs)

The auxiliary tangent linear problems for $\dot{\underline{u}}_\theta$ & $\dot{\underline{u}}_{u_0}$ will at max be as hard as the primal problem, since they are always linear