Push forward / Jvp rule of nonlinear system solve

$$f(\underline{\theta}) = \left\{ \text{solve } g(\underline{x}, \underline{\theta}) = \underline{0} \text{ for } \underline{x} \right\} =: \underline{x}$$

e.g. Newton-Raphson scheme

$\underline{\theta} \in \mathbb{R}^P$

$\underline{x} \in \mathbb{R}^N$

$g: \mathbb{R}^N \times \mathbb{R}^P \to \mathbb{R}^N$

$f: \mathbb{R}^P \to \mathbb{R}^N$

→ assume it always converges

task: forward propagate tangent information on the input $\dot{\underline{\theta}} \in \mathbb{R}^P$ to the output $\dot{\underline{x}} \in \mathbb{R}^N$

without forward-mode AD through the solver

(= unrolling / piggy backing)

in general

$$\dot{\underline{x}} = \underbrace{\underbrace{\frac{\partial f}{\partial \underline{\theta}}}_{\text{Jacobian}} \dot{\underline{\theta}}}_{\text{Jacobian-vector product}}$$

more precisely $\left.\frac{\partial f}{\partial \underline{\theta}}\right|_{\underline{\theta} = \underline{\theta}^*}$

because we first solved the forward / primal problem $\underline{x}^* = f(\underline{\theta}^*)$

$\in \mathbb{R}^{N \times P}$

→ total derivative of the optimality criterion

$$\frac{dg}{d\underline{\theta}} = \frac{\partial g}{\partial \underline{x}} \underbrace{\frac{\partial \underline{x}}{\partial \underline{\theta}}}_{\hat{=} \frac{\partial f}{\partial \underline{\theta}}} + \frac{\partial g}{\partial \underline{\theta}} \overset{!}{=} \underline{\underline{0}}$$

$$\Leftrightarrow \quad \frac{\partial f}{\partial \underline{\theta}} \hat{=} \frac{\partial \underline{x}}{\partial \underline{\theta}} = - \left(\frac{\partial g}{\partial \underline{x}} \right)^{-1} \frac{\partial g}{\partial \underline{\theta}}$$

plug back in

$$\dot{\underline{x}} = - \left(\frac{\partial g}{\partial \underline{x}} \right)^{-1} \frac{\partial g}{\partial \underline{\theta}} \dot{\underline{\theta}}$$

$$\underline{d} = \frac{\partial g}{\partial \underline{\theta}} \dot{\underline{\theta}}$$

$$\frac{\partial g}{\partial \underline{x}} \dot{\underline{x}} = - \underline{d}$$

$$\Leftrightarrow \quad \dot{\underline{x}} = \left\{ \text{solve } \frac{\partial g}{\partial \underline{x}} \dot{\underline{x}} = -\underline{d} \text{ for } \dot{\underline{x}} \text{ with } \underline{d} = \frac{\partial g}{\partial \underline{\theta}} \dot{\underline{\theta}} \right\}$$

Full push forward rule

$$\boxed{\mathcal{F}\left(f, (\underline{\theta},), (\dot{\underline{\theta}},)\right) = \left((f(\underline{\theta}),), (\ldots) \right)}$$

given the optimality criterion is differentiable

$\underline{d} = \frac{\partial g}{\partial \underline{\theta}} \dot{\underline{\theta}}$ is just forward-mode AD

the solution to the linear system

$$\frac{\partial g}{\partial \underline{x}} \dot{\underline{x}} = -\underline{d}$$

can be solved with an iterative which just needs matrix-vector products

($\hat{=}$ matrix free solver)

→ these can be expressed using forward-mode AD as Jacobian-vector products