Pushforward / Jvp rule for Linear system solving

$$f(\underline{A}, \underline{b}) = \{ \text{solve } \underline{A}\underline{x} = \underline{b} \text{ for } \underline{x} \} =: \underline{x}$$

$\underline{A} \in \mathbb{R}^{N \times N}$
$\underline{b} \in \mathbb{R}^{N}$
$\underline{x} \in \mathbb{R}^{N}$

- Numerical solution to PDEs
- Optimization Methods
- ⋮

e.g.
- direct methods (e.g. LU)
- iterative methods (CG, Multigrid)

task: forward-propagate tangent information on the input $\dot{\underline{A}} \in \mathbb{R}^{N \times N}$ & $\dot{\underline{b}} \in \mathbb{R}^{N}$ to the output $\dot{\underline{x}} \in \mathbb{R}^{N}$

↳ without forward-mode AD through the solver

(unrolling / piggy backing)

in general

$$\dot{\underline{x}} = \frac{\partial f}{\partial \underline{A}} : \dot{\underline{A}} + \frac{\partial f}{\partial \underline{b}} \dot{\underline{b}}$$

index notation

primal / forward: $A_{ij} x_j = b_i$

pushforward: $\dot{x}_j = \frac{\partial f_j}{\partial A_{kl}} \dot{A}_{kl} + \frac{\partial f_j}{\partial b_k} \dot{b}_k$

$\dot{x}_j^{[A]}$ $\quad$ $\dot{x}_j^{[b]}$

② $\quad$ ①

① $\frac{\partial f_j}{\partial b_k}$

$$\Rightarrow \frac{\partial}{\partial b_k}(A_{ij} x_j) = \frac{\partial}{\partial b_k}(b_i)$$

$$A_{ij} \frac{\partial x_j}{\partial b_k} = \delta_{ik} \qquad | \cdot \dot{b}_k$$

$$\frac{\partial x_j}{\partial b_k} \hat{=} \frac{\partial f_j}{\partial b_k}$$

$$A_{ij} \frac{\partial x_j}{\partial b_k} \dot{b}_k = \delta_{ik} \dot{b}_k$$

$$A_{ij} \dot{x}_j^{[b]} = \dot{b}_i$$

"another linear system of equations"

② $\frac{\partial f_j}{\partial A_{kl}}$

$$\Rightarrow \frac{\partial}{\partial A_{kl}}(A_{ij} x_j) = \frac{\partial}{\partial A_{kl}}(b_i)$$

$$\delta_{ik} \delta_{jl} x_j + A_{ij} \frac{\partial x_j}{\partial A_{kl}} = 0 \qquad | \cdot \dot{A}_{kl}$$

$$\frac{\partial x_j}{\partial A_{kl}} \hat{=} \frac{\partial f_j}{\partial A_{kl}}$$

$$\delta_{ik} \delta_{jl} x_j \dot{A}_{kl} + A_{ij} \frac{\partial x_j}{\partial A_{kl}} \dot{A}_{kl} = 0$$

$$x_j \dot{A}_{ij} + A_{ij} \dot{x}_j^{[A]} = 0$$

$$\Leftrightarrow A_{ij} \dot{x}_j^{[A]} = - \dot{A}_{ij} x_j$$

"another linear system of equations"

back to symbolic notation

$$\underline{A} \dot{\underline{x}}^{[b]} = \dot{\underline{b}}$$

$$\underline{A} \dot{\underline{x}}^{[A]} = - \dot{\underline{A}} \underline{x}$$

fuse the two linear system solves into one

$$\dot{\underline{x}} = \dot{\underline{x}}^{[b]} + \dot{\underline{x}}^{[A]} = \underline{A}^{-1} \dot{\underline{b}} + \underline{A}^{-1}(-\dot{\underline{A}} \underline{x})$$

$$= \underline{A}^{-1} [\dot{\underline{b}} - \dot{\underline{A}} \underline{x}]$$

Hence: $\underline{A} \dot{\underline{x}} = \dot{\underline{b}} - \dot{\underline{A}} \underline{x}$

$$\dot{\underline{x}} = \{ \text{solve } \underline{A} \dot{\underline{x}} = \underline{d} \text{ for } \dot{\underline{x}} \text{ with } \underline{d} = \dot{\underline{b}} - \dot{\underline{A}} \underline{x} \}$$

Full pushforward rule

$$\boxed{\mathcal{F}(f, (\underline{A}, \underline{b}), (\dot{\underline{A}}, \dot{\underline{b}})) = ((\underbrace{\underline{A}^{-1} \underline{b}}_{\underline{x}}), (\underbrace{\underline{A}^{-1}(\dot{\underline{b}} - \dot{\underline{A}} \underline{x})}_{\dot{\underline{x}}}))}$$

- re-use the factorization in case of a direct solver
- re-use the preconditioner in case of an iterative solver