Adjoint Sensitivities of a Nonlinear System

$$\begin{tikzcd}
\text{non linear system } \underline{f}(\underline{x}, \underline{\theta}) = \underline{0} \arrow[r, "\text{Newton Raphson}"] & \underline{x} \arrow[r] & \text{loss function } J(\underline{x}; \underline{\theta}) \arrow[d] \\
\text{Discretize non-linear PDEs} \arrow[u] & & \text{sensitivities } \frac{dJ}{d\underline{\theta}} \text{, but how?}
\end{tikzcd}$$

Discretize non-linear PDEs
e.g. large deformations in computational mechanics

(scalar-valued) loss function $J(\underline{x}; \underline{\theta})$ e.g. comparing against measurements

Dimensions:

$J \in \mathbb{R}$

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

$\underline{\theta} \in \mathbb{R}^P$ ← collection of parameter

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

total derivative

$$\underbrace{\frac{dJ}{d\underline{\theta}}}_{\mathbb{R}^{1\times P}} = \underbrace{\frac{\partial J}{\partial \underline{\theta}}}_{\mathbb{R}^{1\times P}} + \underbrace{\frac{\partial J}{\partial \underline{x}}}_{\mathbb{R}^{1\times N}} \underbrace{\frac{d\underline{x}}{d\underline{\theta}}}_{\in \mathbb{R}^{N\times P}}$$

— difficult, differentiate $\underline{f}$

here: gradient is a row vector

$$\frac{d\underline{f}}{d\underline{\theta}} = \underline{\underline{0}} = \frac{\partial \underline{f}}{\partial \underline{x}} \frac{d\underline{x}}{d\underline{\theta}} + \frac{\partial \underline{f}}{\partial \underline{\theta}}$$

$$\Leftrightarrow \underbrace{\frac{d\underline{x}}{d\underline{\theta}}}_{\mathbb{R}^{N\times P}} = - \underbrace{\left(\frac{\partial \underline{f}}{\partial \underline{x}}\right)^{-1}}_{\mathbb{R}^{N\times N}} \underbrace{\frac{\partial \underline{f}}{\partial \underline{\theta}}}_{\mathbb{R}^{N\times P}}$$

(inverse) Jacobian of $\underline{f}$ wrt $\underline{x}$
→ a matrix

plug in

$$\frac{dJ}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \Bigg(\overbrace{- \underbrace{\left(\frac{\partial J}{\partial \underline{x}} \left(\frac{\partial \underline{f}}{\partial \underline{x}} \right)^{-1} \right)}_{1 \text{ linear solve}}}^{=: \underline{\lambda}^T \; =: \underline{\lambda}^T} \underbrace{\left(\left(\frac{\partial \underline{f}}{\partial \underline{x}}\right)^{-1} \frac{\partial \underline{f}}{\partial \underline{\theta}} \right)}_{P \text{ linear solves}} \Bigg)$$

clever bracketing

yields the adjoint variable $\underline{\lambda} \in \mathbb{R}^N$

$$\frac{dJ}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{\theta}}$$

adjoint system

$$\underline{\lambda}^T := - \frac{\partial J}{\partial \underline{x}} \left(\frac{\partial \underline{f}}{\partial \underline{x}} \right)^{-1} \quad \Big| \, \frac{\partial \underline{f}}{\partial \underline{x}}$$

$$\underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{x}} = - \frac{\partial J}{\partial \underline{x}}$$

$$\left(\underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{x}} \right)^T = - \left(\frac{\partial J}{\partial \underline{x}} \right)^T$$

$$\left(\frac{\partial \underline{f}}{\partial \underline{x}} \right)^T \underline{\lambda} = - \left(\frac{\partial J}{\partial \underline{x}} \right)^T$$

strategy for $\frac{dJ}{d\underline{\theta}}$

① Solve forward $\underline{f}(\underline{x}, \underline{\theta}) = \underline{0}$ for $\underline{x}$
(e.g. by Newton-Raphson)

② Solve adjoint $\left(\frac{\partial \underline{f}}{\partial \underline{x}} \right)^T \underline{\lambda} = - \left(\frac{\partial J}{\partial \underline{x}} \right)^T$ for $\underline{\lambda}$
(e.g. by LU or Conjugate Gradient)

③ Evaluate $\frac{dJ}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{\theta}}$

Remarks:

→ forward solve is non-linear,
backward/adjoint solve is linear
↳ the adjoint problem is easier to solve

→ Scales constantly in the number of parameters $\underline{\theta} \in \mathbb{R}^P$

Relation to linear systems

$$\underline{\underline{A}}(\underline{\theta}) \, \underline{x} = \underline{b}(\underline{\theta})$$

then $\underline{f}(\underline{x}, \underline{\theta}) = \underline{b}(\underline{\theta}) - \underline{\underline{A}}(\underline{\theta}) \underline{x} = \underline{0}$

$$\frac{\partial \underline{f}}{\partial \underline{x}} = - \underline{\underline{A}}$$

$$\frac{\partial \underline{f}}{\partial \underline{\theta}} = \underline{\underline{0}} \qquad \text{hmm....}$$

we assumed $\underline{f}$ only had an explicit dependency on $\underline{\theta}$

correct would be

$$\underline{f} = \underline{f}\left(\underline{x}(\underline{\theta}), \underline{\underline{A}}(\underline{\theta}), \underline{b}(\underline{\theta})\right) = \underline{0}$$

then

$$\frac{d\underline{f}}{d\underline{\theta}} = \underbrace{\frac{\partial \underline{f}}{\partial \underline{x}}}_{-\underline{\underline{A}}} \frac{d\underline{x}}{d\underline{\theta}} + \underbrace{\frac{\partial \underline{f}}{\partial \underline{\underline{A}}} \frac{d\underline{\underline{A}}}{d\underline{\theta}}}_{-\frac{d\underline{\underline{A}}}{d\underline{\theta}} \underline{x}} + \underbrace{\frac{\partial \underline{f}}{\partial \underline{b}}}_{\underline{\underline{I}}} \frac{d\underline{b}}{d\underline{\theta}} = \underline{\underline{0}}$$

? not correct

rearrange for $\frac{d\underline{x}}{d\underline{\theta}}$

$$\frac{d\underline{x}}{d\underline{\theta}} = \underline{\underline{A}}^{-1} \left(\frac{d\underline{b}}{d\underline{\theta}} - \frac{d\underline{\underline{A}}}{d\underline{\theta}} \underline{x} \right)$$

Hence:

$$\frac{\partial \underline{f}}{\partial \underline{\theta}} \neq \underline{\underline{0}}$$

" means here $\frac{d\underline{f}}{d\underline{\theta}}$ without $\frac{d\underline{x}}{d\underline{\theta}}$ "