Adjoint of a Non-Linear System – Lagrangian Perspective

① Solve $\underline{f}(\underline{x}, \underline{\theta}) = \underline{0}$ for $\underline{x}$ — non linear system — "classical problem"

② Solve $\left(\frac{\partial \underline{f}}{\partial \underline{x}}\right)^T \underline{\lambda} = -\left(\frac{\partial J}{\partial \underline{x}}\right)^T$ for $\underline{\lambda}$ — liner system — adjoint problem

③ Evaluate $\frac{dJ}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{\theta}}$ — explicit relation — gradient evaluation

$\underline{f}$ ... residuum function $\in \mathbb{R}^N$

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

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

J ... loss function $\in \mathbb{R}$

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

Why?

↳ Backpropagate through non-linear systems

e.g. discrete nonlinear PDEs

Derive using Lagrangian Perspective

$$\min_{\underline{\theta}} \; J(\underline{x}^{(\theta)}; \underline{\theta})$$

$$\text{s.t.} \quad \underline{f}(\underline{x}, \underline{\theta}) = \underline{0}$$

① Build Lagrangian

$$\mathcal{L}(\underline{x}, \underline{\lambda}; \underline{\theta}) = J + \underline{\lambda}^T \underline{f}$$

② Take total derivative wrt $\underline{\theta}$

$$\underbrace{\frac{d\mathcal{L}}{d\underline{\theta}}}_{\in \mathbb{R}^{1\times P}} = \frac{\partial J}{\partial \underline{\theta}} + \underbrace{\frac{\partial J}{\partial \underline{x}}}_{\in \mathbb{R}^{1\times N}} \underbrace{\frac{d\underline{x}}{d\underline{\theta}}}_{\in \mathbb{R}^{N\times P}} + \underline{\lambda}^T \left(\frac{\partial \underline{f}}{\partial \underline{\theta}} + \frac{\partial \underline{f}}{\partial \underline{x}} \frac{d\underline{x}}{d\underline{\theta}} \right)$$

"here: gradient is a row vector"

→ numerator layout

③ Isolate solution sensitivities $\frac{d\underline{x}}{d\underline{\theta}}$

$$\frac{d\mathcal{L}}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{\theta}} + \underbrace{\left(\frac{\partial J}{\partial \underline{x}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{x}} \right)}_{\text{if zero (i.e. } \underline{0}^T\text{) we do not need to calculate } \frac{d\underline{x}}{d\underline{\theta}}} \frac{d\underline{x}}{d\underline{\theta}}$$

④ Identify Adjoint problem

$$\frac{\partial J}{\partial \underline{x}} + \underline{\lambda}^T \frac{\partial \underline{f}}{\partial \underline{x}} = \underline{0}^T$$

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

a linear system of equations for $\underline{\lambda} \in \mathbb{R}^N$

⑤ Identify gradient evaluation

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