Adjoint of a linear system - Lagrangian PoV

optimization problem

$$\min_{\underline{\theta}} \quad J(\underline{x}; \underline{\theta})$$
$$\text{s.t.} \quad \underline{\underline{A}}(\underline{\theta})\underline{x} = \underline{b}(\underline{\theta})$$

Want $\frac{dJ}{d\underline{\theta}}$ for gradient-based optimization, but how?

implicit relation (e.g. discrete PDE)

Similar to the sensitivity analysis of $\frac{dJ}{d\underline{\theta}}$ by implicit differentiation

$J \in \mathbb{R}$ $\quad \underline{\theta} \in \mathbb{R}^P$ $\quad \underline{x} \in \mathbb{R}^N$ $\quad \underline{\underline{A}} \in \mathbb{R}^{N \times N}$ $\quad \underline{b} \in \mathbb{R}^N$

$\to$ equality constrained optimization $\Rightarrow$ Lagrangian

$\Downarrow$

unconstrained optimization

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

total derivative

? not fully correct, it's ok!

$$\underbrace{\frac{d\mathcal{L}}{d\underline{\theta}}}_{\mathbb{R}^{1 \times P}} = \underbrace{\frac{\partial J}{\partial \underline{x}}}_{\mathbb{R}^{1 \times N}} \underbrace{\frac{d\underline{x}}{d\underline{\theta}}}_{\mathbb{R}^{N \times P}} + \underbrace{\frac{\partial J}{\partial \underline{\theta}}}_{\mathbb{R}^{1 \times P}} + \underbrace{\underline{\lambda}^T}_{1 \times N}\Big(\underbrace{\frac{d\underline{b}}{d\underline{\theta}}}_{\mathbb{R}^{N \times P}} - \Big(\underbrace{\frac{d\underline{\underline{A}}}{d\underline{\theta}}}_{\mathbb{R}^{N \times N \times P}} \underbrace{\underline{x}}_{\mathbb{R}^N} + \underbrace{\underline{\underline{A}}}_{\mathbb{R}^{N \times N}} \underbrace{\frac{d\underline{x}}{d\underline{\theta}}}_{\mathbb{R}^{N \times P}}\Big)\Big)$$

here: gradient is a row vector

difficult

rearrange

$$\frac{d\mathcal{L}}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T\Big(\frac{d\underline{b}}{d\underline{\theta}} - \frac{d\underline{\underline{A}}}{d\underline{\theta}}\underline{x}\Big) + \Big(\underbrace{\frac{\partial J}{\partial \underline{x}} - \underline{\lambda}^T \underline{\underline{A}}}\Big)\frac{d\underline{x}}{d\underline{\theta}}$$

make zero, because $\underline{\lambda}$ is arbitrary

$\hookrightarrow$ adjoint system

$$\frac{\partial J}{\partial \underline{x}} - \underline{\lambda}^T \underline{\underline{A}} = \underline{0}$$

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

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

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

adjoint of $\underline{\underline{A}}$

then

$$\frac{d\mathcal{L}}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T\Big(\frac{d\underline{b}}{d\underline{\theta}} - \frac{d\underline{\underline{A}}}{d\underline{\theta}}\underline{x}\Big)$$

But we wanted $\frac{dJ}{d\underline{\theta}}$, not $\frac{d\mathcal{L}}{d\underline{\theta}}$?

$\to$ they are identical

because $\mathcal{L} = J + \underline{\lambda}^T \underbrace{(\underline{b} - \underline{\underline{A}}\,\underline{x})}$ has to be $\underline{0}$

hence its derivative is zero

$$\Rightarrow \frac{d\mathcal{L}}{d\underline{\theta}} = \frac{dJ}{d\underline{\theta}}$$

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

1. Solve forward $\underline{\underline{A}}\,\underline{x} = \underline{b}$ for $\underline{x}$
2. Solve adjoint $\underline{\underline{A}}^T \underline{\lambda} = \left(\frac{\partial J}{\partial \underline{x}}\right)^T$ for $\underline{\lambda}$
3. Evaluate $\frac{dJ}{d\underline{\theta}} = \frac{\partial J}{\partial \underline{\theta}} + \underline{\lambda}^T\Big(\frac{d\underline{b}}{d\underline{\theta}} - \frac{d\underline{\underline{A}}}{d\underline{\theta}}\underline{x}\Big)$