Pullback / vJp rule of scalar root finding

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

e.g.: - bisection method
- Newton-Raphson
⋮

$x \in \mathbb{R}$
$\theta \in \mathbb{R}$
$g: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$
$f: \mathbb{R} \to \mathbb{R}$

task: backpropagate $\bar{x} \in \mathbb{R}$ to $\bar{\theta} \in \mathbb{R}$

without reverse-mode AD through the solver

(→ unrolling / piggy backing)

$$\bar{\theta} = \bar{x} \frac{\partial f}{\partial \theta} \quad \in \mathbb{R}$$

$$\left(\frac{\partial f}{\partial \theta} \stackrel{\wedge}{=} \frac{\partial x}{\partial \theta} \right)$$

total derivative of the optimality condition g wrt θ

$$\frac{dg}{d\theta} = \frac{\partial g}{\partial x} \frac{\partial x}{\partial \theta} + \frac{\partial g}{\partial \theta} \stackrel{!}{=} 0$$

$$\Leftrightarrow \quad \frac{\partial x}{\partial \theta} = - \frac{\overset{\in \mathbb{R}}{\frac{\partial g}{\partial \theta}}}{\underset{\in \mathbb{R}}{\frac{\partial g}{\partial x}}}$$

plug back into general definition

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

$$\boxed{\mathcal{B}(f, (\theta,), (\bar{x},)) = ((\underbrace{f(\theta)}_{x},), (\underbrace{-\bar{x} \left(\frac{\partial g}{\partial x} \right)^{-1} \frac{\partial g}{\partial \theta}}_{\bar{\theta}},))}$$