- Implicit Model

Instead of specifying how to compute the Layer's output from input
we specify the conditions that we want the Layers output to satisfy

explicit : $z = f(x)$.
implicit : find z. s.t. $g(x,z)=0$

- Implicit Function Theorem

let $f: \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}^n$ $a_0 \in \mathbb{R}^p$. $z_0 \in \mathbb{R}^n$ s.t.

$$\begin{cases} f(a_0, z_0) = 0 \\ f \text{ is continuously differentiable with non-singular Jacobian } \partial_1 f(a_0, z_0) \in \mathbb{R}^{n\times n} \end{cases}$$

then $\exists$ open set $S_{a_0} \subseteq \mathbb{R}^p$, $S_{z_0} \subseteq \mathbb{R}^n$ containing a_0, z_0

$\exists$ unique continuous function $z^*: S_{a_0} \to S_{z_0}$ s.t.

$$\begin{cases} z_0 = z^*(a_0) \\ f(a, z^*(a)) = 0 \quad \forall a \in S_{a_0} \\ z^* \text{ is differentiable on } S_{a_0} \end{cases}$$

$$\left.\begin{aligned} & f(a, z^*(a)) = 0 \\ & \partial f(a, z^*(a)) / \partial a = 0 \\ & \partial_0 f(a, z^*) + \partial_1 f(a, z^*)\, \partial z^*(a) = 0 \end{aligned}\right\} \quad \partial z^*(a) = -\partial_1 f(a, z^*)\, \partial_0 f(a, z^*)$$

Eg. Fixed Point Iteration

$z = \tanh(Wz + x)$

Forward Pass. (Newton's Method)

find z. s.t $g(z, x, W) = z - \tanh(Wz + x) = 0$

$$g(z + \Delta z, \cdots) \approx g(z, \cdots) + D_z g\, \Delta z$$
$$= g(z, \cdots) + \left[I - \mathrm{diag}\left(\tanh'(Wz + x) \right) \cdot W \right] \Delta z$$
$$:= 0$$

$$\Delta z = -\left[I - \mathrm{diag}\left(\tanh'(Wz + x) \right) \cdot W \right]^{-1} g(z, x, W)$$

Forward pass (unrolling)

$z := 0$

repeat until convergence

$z := \tanh(Wz + x)$

Backward Pass

$$g(x,w,z^*(x,w)) = z^*(x,w) - \tanh(w z^*(x,w) + x) \equiv 0$$

1° Backprop to x $\quad \dfrac{dz^*(w,x)}{dx}$

$$\frac{dg(x,w,z^*(x,w))}{dx} = \underbrace{\frac{dg(x,w,z^*)}{dx}}_{n\times n} + \underbrace{\frac{dg(x,w,z^*)}{dz^*}}_{n\times n} \cdot \underbrace{\frac{dz^*(x,w)}{dx}}_{n\times n} = 0$$

$$\frac{dz^*(x,w)}{dx} = -\left(\frac{dg(x,w,z^*)}{dz^*}\right)^{-1} \frac{dg(x,w,z^*)}{dx}$$

$$\underbrace{\frac{dL}{dx}}_{n} = \frac{dz^*(x,w)}{dx}^{T} \frac{dL}{dz^*}$$

$$= -\frac{dg(x,w,z^*)}{dx}^{T} \left(\frac{dg(x,w,z^*)}{dz^*}\right)^{-T} \frac{dL}{dz^*}$$

$$\frac{dL}{dx} = \mathrm{diag}(\tanh'(wz^*+x))\left(I - \mathrm{diag}(\tanh'(wz^*+x))w\right)^{-T} \frac{dL}{dz^*}$$

2° Backprop to w $\quad \dfrac{dz^*(x,w)}{dw}$

$$\frac{dg_i(x,w,z^*(x,w))}{dw} = \frac{dg_i(x,w,z^*)}{dw} + \sum_j \frac{dg_i(x,w,z^*)}{dz_j^*} \frac{dz_j^*(x,w)}{dw} = 0 \qquad \forall i$$

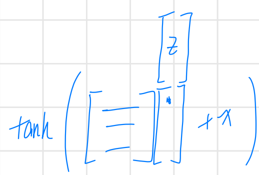
$$\underbrace{\Big[\; dg_i/dw \;\Big]}_{n \times n^2} + \underbrace{\Big[\; dg_i/dz_j^* \;\Big]}_{n\times n} \underbrace{\Big[\; dz_j^*/dw \;\Big]}_{n \times n^2} = 0$$

$$\frac{dg(x,w,z^*)}{dw} + \frac{dg(x,w,z^*)}{dz^*} \frac{dz^*(x,w)}{dw} = 0$$

$$\frac{dz^*(x,w)}{dw} = -\left(\frac{dg(x,w,z^*)}{dz^*}\right)^{-1} \frac{dg(x,w,z^*)}{dw}$$

$$\underbrace{\frac{dL}{dw}}_{n^2} = \frac{dz^*(x,w)}{dw}^{T} \frac{dL}{dz^*}$$

$$= -\frac{dg(x,w,z^*)}{dw}^{T} \left(\frac{dg(x,w,z^*)}{dz^*}\right)^{-T} \frac{dL}{dz^*}$$



$$= -\underbrace{\begin{bmatrix} dg_i/dw_1 & & 0 \\ \vdots & & \vdots \\ 0 & & dg_i/dw_n \end{bmatrix}}_{n^2 \times n} \left(\frac{dg(x,w,z^*)}{dz^*}\right)^{-T} \frac{dL}{dz^*}$$

$$\frac{\partial L}{\partial W} = \begin{bmatrix} \text{———} \\ \text{———} \\ \text{———} \end{bmatrix}$$

$$= -\,\mathrm{diag}\left(\frac{\partial g(x, W, z^*)}{\partial z^*}^{-T} \frac{\partial L}{\partial z^*} \right) \cdot -\begin{bmatrix} — \ \partial g_1 / \partial W_1 \ — \\ \vdots \\ — \ \partial g_n / \partial W_n \ — \end{bmatrix}$$

$$\frac{\partial L}{\partial W} = \mathrm{diag}\left(\underbrace{\frac{\partial g(x, W, z^*)}{\partial z^*}^{-T} \frac{\partial L}{\partial z^*}}_{\text{computed in } \frac{\partial L}{\partial x}} \right) \cdot \tanh'(Wz^* + x)\, {z^*}^T$$