

IDL

David Ponarovsky

April 2025

1 Theoretical Questions:

1.1 Composition of Linear question.

Let $f : D \rightarrow D'$ and $g : D' \rightarrow D''$ be linear functions. Then for any $x, y \in D'$ and coefficients a, b we have:

$$g \circ f(ax + by) = g(f(ax + by)) = g(af(x) + bf(y))$$

When in the last passages we used the linearity of f , Now since $f(x), f(y) \in D''$ we can use the linearity of g to get:

$$g \circ f(ax + by) = ag(f(x)) + bg(f(y)) = ag \circ f(x) + bg \circ f(y)$$

1.2 The Gradient Descent.

Denote by $f(x, y) = p^{-1}(x - 1)^2 + p(y + 1)^2$. Then:

$$\nabla f = \begin{bmatrix} 2p^{-1}(x - 1) \\ 2p(y + 1) \end{bmatrix}$$

So stepping at rate ε advance (x, y) as follows:

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} \leftarrow \begin{bmatrix} x_t \\ y_t \end{bmatrix} + \varepsilon \begin{bmatrix} 2p^{-1}(x_t - 1) \\ 2p(y_t + 1) \end{bmatrix}$$

So for requiring that $\Theta(\|\Delta x\|_2^2)$ will be small comparing to $\|\Delta x\|_1$ we have to ensure that $\|\Delta x\|_2 < 1$ (the two norm chosen arbitrary, any p - norm for $p \geq 1$ works). Thus:

$$\|\varepsilon \begin{bmatrix} 2p^{-1}(x_t - 1) \\ 2p(y_t + 1) \end{bmatrix}\|_2 \leq 1 \Rightarrow \varepsilon \leq \frac{1}{4p^{-2}(x - 1)^2 + 4p^2(y + 1)^2}$$

1.3 Prediction Loss.

$$L(\theta, \hat{\theta}) = |\text{Im}e^{i(\theta - \hat{\theta})}| \cdot \text{penalty}$$

1.4 Chain Rule.

1. Using the chain rule:

$$\begin{aligned}
 \frac{\partial}{\partial x} f(x+y, 2x, z) &= \frac{\partial}{\partial x} (x+y) \cdot \frac{\partial}{\partial w_1} f(w_1, w_2, w_3) \\
 &\quad + \frac{\partial}{\partial x} (2x) \cdot \frac{\partial}{\partial w_2} f(w_1, w_2, w_3) \\
 &\quad + \frac{\partial}{\partial x} (z) \cdot \frac{\partial}{\partial w_3} f(w_1, w_2, w_3) \\
 &= \frac{\partial}{\partial w_1} f(w_1, w_2, w_3) + 2 \cdot \frac{\partial}{\partial w_2} f(w_1, w_2, w_3) \Big|_{w_1=x+y, w_2=2x, w_3=z}
 \end{aligned}$$

2. Denote by g_n the concatenation of f_1, f_2, \dots, f_n with itself, and extend the notation to $n=0$ by defining $f_0(x) = x$. We will prove by induction that $\frac{d}{dx} f_n = \prod_{i=0}^{n-1} \frac{d}{dx} f_{i+1} \Big|_{f_i(x)}$:

(a) **Base.** $n=1$ and indeed $\prod_{i=0}^0 \frac{d}{dx} f_{i+1} \Big|_{f_i(x)} = \frac{d}{dx} f \Big|_x$.

(b) **Assumption.** Assumes correctness to $n' \leq n-1$.

(c) **Step.**

$$\begin{aligned}
 \frac{d}{dx} g_n &= \frac{d}{dx} (f_n(g_{n-1}(x))) = \frac{d}{dx} g_{n-1} \cdot \left(\frac{d}{dx} f_n \right) \Big|_{f_{n-1}(x)} \\
 &= \prod_{i=0}^{n-2} \frac{d}{dx} f_{i+1} \Big|_{f_i(x)} \cdot \left(\frac{d}{dx} f_n \right) \Big|_{f_{n-1}(x)} = \prod_{i=0}^{n-1} \frac{d}{dx} f_{i+1} \Big|_{f_i(x)}
 \end{aligned}$$

3. Let's define $g_n(x)$ in recursive manner as follows: $g_0(x) = f(x)$ and $g_{n+1} = f_{n+1}(x, g_n(x))$. Then we will prove by induction that:

$$\frac{d}{dx} g_n(x) = \sum_i \frac{d}{dx} f_i(x, y) \Big|_{y=f_{i-1}(x)} \prod_{j=i-1}^n \frac{d}{dy} f_{j+1}(x, y) \Big|_{y=f_j(x)}$$

- (a) **Base.** For $n=2$ we have:

$$\frac{d}{dx} f_2(x, f_1(x)) = \frac{d}{dx} f_2(x, y) \Big|_{y=f_1(x)} + \frac{d}{dx} f_1(x) \cdot \frac{d}{dy} f_2(x, y) \Big|_{y=f_1(x)}$$

And that's exactly what we have in the formula.

- (b) **Assumption.** Assume the correctness of the claim for any $n' \leq n-1$

(c) **Step.**

$$\begin{aligned}
\frac{d}{dx} f_n(x, g_{n-1}(x)) &= \frac{d}{dx} f_n(x, y)|_{y=g_{n-1}(x)} + \frac{dy}{dx} \frac{d}{dy} f(x, y)|_{y=g_{n-1}(x)} \\
&= \frac{d}{dx} f_n(x, y)|_{y=g_{n-1}(x)} + \frac{d}{dx} g_{n-1} \frac{d}{dy} f_n(x, y)|_{y=g_{n-1}(x)} \\
&= \frac{d}{dx} f_n(x, y)|_{y=g_{n-1}(x)} + \sum_i^{n-1} \frac{d}{dx} f_i(x, y)|_{y=g_{i-1}(x)} \prod_{j=i}^{n-1} \frac{d}{dy} f_{j+1}(x, y)|_{y=g_j(x)} \\
&= \sum_i^n \frac{d}{dx} f_i(x, y)|_{y=g_{i-1}(x)} \prod_{j=i}^n \frac{d}{dy} f_j(x, y)|_{y=g_{j-1}(x)}
\end{aligned}$$

4.

$$f(x + g(x + h(x)))$$

First, notice that the derivative of $g(x + h(x))$ equals:

$$\begin{aligned}
\frac{d}{dx} g(x + h(x)) &= \frac{d}{dx} (x + h(x)) \cdot \frac{d}{dy} g(y)|_{y=x+h(x)} \\
&= \left(1 + \frac{d}{dx} h(x)\right) \cdot \frac{d}{dy} g(y)|_{y=x+h(x)}
\end{aligned}$$

So in overall we get:

$$\begin{aligned}
\frac{d}{dx} f(x + g(x + h(x))) &= \left(1 + \frac{d}{dx} g(x + h(x))\right) \cdot \frac{d}{dy} f(y)|_{y=x+g(x+h(x))} \\
&= \left(1 + \left(1 + \frac{d}{dx} h(x)\right) \cdot \frac{d}{dy} g(y)|_{y=x+h(x)}\right) \cdot \frac{d}{dy} f(y)|_{y=x+g(x+h(x))} \\
&= \frac{d}{dy} f(y)|_{y=x+g(x+h(x))} + \frac{d}{dy} f(y)|_{y=x+g(x+h(x))} \cdot \frac{d}{dy} g(y)|_{y=x+h(x)} \\
&\quad + \frac{d}{dy} f(y)|_{y=x+g(x+h(x))} \cdot \frac{d}{dy} g(y)|_{y=x+h(x)} \cdot \frac{d}{dx} h(x)
\end{aligned}$$