Analysis III - 203(d)

Winter Semester 2024

Session Extra: December 29, 2025

Exercise 1 *Recall the ReLU function*

$$\mathrm{ReLU}(x) = \max(0, x). \tag{1}$$

A shallow neural network with one output neuron, m internal neurons, and n input neurons has the general form

$$f(x_1, \dots, x_n) = \sum_{k=1}^{m} \mathrm{ReLU}\left(\sum_{i=1}^{n} A_{ki} x_i + b_k\right) \tag{2}$$

where A_{ki} are weights and where b_k is a shift parameter.

- *Most training algorithms require the gradient of this network. Compute the derivatives $\partial_i f$*
- *There is interest in training algorithms that use the Hessian matrix. Compute the partial derivatives $\partial^2_{ij} f$.*

Solution 1 *We begin with the first derivatives. Obviously,*

$$\partial_i f(x_1, \dots, x_n) = \sum_{k=1}^{m} \partial_i \mathrm{ReLU}\left(\sum_{i=1}^{n} A_{ki} x_i + b_k\right) \tag{3}$$

The ReLU function can only be differentiated in the sense of distributions. Recalling the Heaviside function,

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}, \tag{4}$$

it is now possible to write the derivative as

$$\partial_i f(x_1, \dots, x_n) = \sum_{k=1}^{m} H\left(\sum_{i=1}^{n} A_{ki} x_i + b_k\right) A_{ki} \tag{5}$$

$$= \sum_{k=1}^{m} A_{ki} H\left(\sum_{i=1}^{n} A_{ki} x_i + b_k\right). \tag{6}$$

This computes the first derivatives.[1]

We continue with the second derivatives. Recall that the Heaviside function has a derivative in the sense of distributions, which is the Dirac Delta. Hence

$$\partial_{ji}^2 f(x_1, \ldots, x_n) = \sum_{k=1}^{m} A_{ki} \partial_j H \left(\sum_{i=1}^{n} A_{ki} x_i + b_k \right) \tag{7}$$

$$= \sum_{k=1}^{m} A_{ki} A_{kj} \delta_0 \left(\sum_{i=1}^{n} A_{ki} x_i + b_k \right). \tag{8}$$

This computes the second derivatives.[2]

Exercise 2 *In standard models of elasticity, a long straight beam of elastic material, such as wood or metal, can be modeled as a one-dimensional interval. When it is subject to an outside force f, such as gravity, than the deformation from the base is modeled by the beam equation*

$$u''''(x) = f(x) \tag{9}$$

Here, the fourth derivative u'''' can be interpreted as the curvature of a curvature and f describes the direction (upwards, downwards) and magnitude of the force.

So-called non-local interactions are modeled via a convolutional term $k \star u$, where k indicates how parts of a beam are influenced by neighboring parts. With that in mind, we consider a generalized beam equation

$$u''''(x) + cu(x) + (k \star u)(x) = f(x). \tag{10}$$

This is a so-called integro-differential equation.

Compute the Fourier transform of this differential equation for general source terms, and write it down for the particular example

$$k(x) = e^{-|y|}, \quad f(x) = e^{-y^2}. \tag{11}$$

You are not expected to solve this equation.

Solution 2 *The Fourier transform of this equation reads*

$$(i\alpha)^4 \hat{u}(\alpha) + c\hat{u}(\alpha) + \sqrt{2\pi}\hat{k}(\alpha)\hat{u}(\alpha) = \hat{f}(\alpha). \tag{12}$$

[1] Remark: the derivative does not have a meaningful value if one of the arguments of the Heaviside function is zero (or close to zero within the range of rounding errors). When training a neural network via gradient descent, this is "justified" by the assumption that these arguments being close to zero is very unlikely to happen in practice.

[2] Remark: the situation here is even worse than with the first derivatives. The Dirac Delta is zero everywhere except at the origin, and it equals a pointmass at the origin. Correspondingly, second-order training algorithms, such as, e.g., Newton's method, are not well-defined for such a network. This is a possible incentive to replace ReLU by other activation functions, such as $S(x) = \ln(1 + e^x)$.

We simplify this to:

$$\alpha^4\hat{u}(\alpha) + c\hat{u}(\alpha) + \sqrt{2\pi}\hat{k}(\alpha)\hat{u}(\alpha) = \hat{f}(\alpha). \tag{13}$$

We isolate $\hat{u}(\alpha)$*, giving us:*

$$\hat{u}(\alpha) = \frac{\hat{f}(\alpha)}{\alpha^4 + c + \sqrt{2\pi}\hat{k}(\alpha)}. \tag{14}$$

In the particular case of

$$k(x) = e^{-|y|}, \quad f(x) = e^{-y^2}, \tag{15}$$

one finds

$$\hat{k}(x) = \sqrt{\frac{2}{\pi}}\frac{1}{1+\alpha^2}, \quad \hat{f}(x) = \frac{1}{\sqrt{2}}e^{-\alpha^2/4}. \tag{16}$$

One can simplify

$$\hat{u}(\alpha) = \frac{\hat{f}(\alpha)}{\alpha^4 + c + \sqrt{2\pi}\hat{k}(\alpha)} = \frac{\hat{f}(\alpha)}{\alpha^4 + c + \frac{2}{1+\alpha^2}} = \underbrace{\frac{1+\alpha^2}{(1+\alpha^2)(\alpha^4+c)+2}}_{=:\hat{h}(\alpha)}\hat{f}(\alpha). \tag{17}$$

The solution of the ODE can be found if we know the inverse Fourier transform $h(x)$ of the factor $\hat{h}(\alpha)$. Then the product of $\hat{h}(\alpha)\hat{f}(\alpha)$ can be transformed into the convolution $\sqrt{2\pi}h \star f$. Finding the inverse Fourier transform of such a factor will be possible with techniques of complex analysis, to be discussed next semester.

Exercise 3 *Solve the integro-differential equation*

$$9u(x) + 2\int_{-\infty}^{+\infty} \left(u''(t) - 4u(t)\right) e^{-2|x-t|}dt = \frac{1}{x^2+1}.$$

Solution 3 *We use the Fourier transform. First, we transform the equation. Beginning with the source term, we use the convolution theorem and transform the derivatives:*

$$\begin{aligned}\mathfrak{F}\left[\frac{1}{x^2+1}\right] &= 9\mathfrak{F}[u] + 2\mathfrak{F}\left[\int_{-\infty}^{+\infty} \left(u''(t) - 4u(t)\right) e^{-2|x-t|}dt\right] &(18)\\ &= 9\mathfrak{F}[u] + 2\sqrt{2\pi}\mathfrak{F}\left[u'' - 4u\right] \cdot \mathfrak{F}\left[e^{-2|x|}\right] &(19)\\ &= 9\mathfrak{F}[u] + 2\left(-\alpha^2\mathfrak{F}\left[u\right] - 4\mathfrak{F}\left[u\right]\right) \cdot \mathfrak{F}\left[e^{-2|x|}\right]. &(20)\end{aligned}$$

For convenience, we use the $\hat{u}$ notation for the Fourier transform and use the Fourier transform table:

$$\sqrt{\frac{\pi}{2}}e^{-|\alpha|} = 9\hat{u}(\alpha) - 2\sqrt{2\pi}\left(\alpha^2\hat{u}(\alpha) + 4\hat{u}(\alpha)\right)\cdot 2\sqrt{\frac{2}{\pi}}\frac{1}{4+\alpha^2} \tag{21}$$

$$= 9\hat{u}(\alpha) - 2\sqrt{2\pi}\hat{u}(\alpha)\left(\alpha^2+4\right)\cdot 2\sqrt{\frac{2}{\pi}}\frac{1}{4+\alpha^2} \tag{22}$$

$$= 9\hat{u}(\alpha) - 2\sqrt{2\pi}\hat{u}(\alpha)\cdot 2\sqrt{\frac{2}{\pi}} \tag{23}$$

$$= 9\hat{u}(\alpha) - 8\hat{u}(\alpha) = \hat{u}(\alpha). \tag{24}$$

We transform back and obtain the solution

$$u(x) = \frac{1}{1+x^2}. \tag{25}$$

Exercise 4 *Solve the integral equation*

$$u(t) + \lambda\int_0^{+\infty} e^{-|y|}u(t-y)\ dy = e^{-|t|}.$$

Solution 4 *We observe that this is a convolutional integral equation*

$$u(t) + \lambda(K\star u)(t) = f(t)$$

with

$$K(t) = \begin{cases} e^{-t} & t>0 \\ 0 & \textit{otherwise,} \end{cases} \qquad f(t) = e^{-|t|}.$$

We look up the Fourier transforms:

$$\hat{K}(\alpha) = \frac{1}{\sqrt{2\pi}}\frac{1}{1+i\alpha}, \quad \hat{f}(\alpha) = \sqrt{\frac{2}{\pi}}\frac{1}{1+\alpha^2}$$

We Fourier transform the equation,

$$\hat{u}(\alpha) + \lambda\sqrt{2\pi}\hat{K}(\alpha)\hat{u}(\alpha) = \hat{f}(\alpha),$$

and isolating $\hat{u}$ we find

$$\hat{u}(\alpha) = \frac{\hat{f}(\alpha)}{1+\lambda\sqrt{2\pi}\hat{K}(\alpha)},$$

Let us simplify this. Using $1+\alpha^2 = 1-i^2\alpha^2 = (1+i\alpha)(1-i\alpha)$, we obtain

$$\hat{u}(\alpha) = \frac{\hat{f}(\alpha)}{1+\lambda\frac{1}{1+i\alpha}}$$

$$= \frac{\hat{f}(\alpha)(1+i\alpha)}{(1+i\alpha)+\lambda} = \sqrt{\frac{2}{\pi}} \frac{\frac{1+i\alpha}{1+\alpha^2}}{(1+\lambda)+i\alpha} = \sqrt{\frac{2}{\pi}} \frac{1}{1-i\alpha} \cdot \frac{1}{(1+\lambda)+i\alpha}.$$

We already know that

$$\hat{v}(\alpha) = \frac{1}{(1+\lambda)+i\alpha} \qquad \Longrightarrow \qquad v(t) = \sqrt{2\pi} \begin{cases} e^{-(1+\lambda)t} & t > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Using the modulation identity, we find

$$\hat{w}(\alpha) = \frac{1}{1-i\alpha} \qquad \Longrightarrow \qquad w(t) = \sqrt{2\pi} \begin{cases} e^{t} & t < 0 \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$\hat{u}(\alpha) = \sqrt{\frac{2}{\pi}} \hat{v}(\alpha)\hat{w}(\alpha)$$

which means

$$u(t) = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2\pi}} (v \star w)(t)$$

We calculate

$$u(t) = \sqrt{\frac{2}{\pi}} \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} \sqrt{2\pi} e^{-(1+\lambda)y} w(t-y)\ dy = \sqrt{\frac{2}{\pi}} \int_0^{+\infty} e^{-(1+\lambda)y} w(t-y)\ dy.$$

The function w vanishes for $t - y > 0$, that is, $t > y$. Hence we only need to consider the case $t < y$. For negative $t < 0$, the condition $t < y$ is already implied by $0 < y$, and thus

$$u(t) = \sqrt{\frac{2}{\pi}} \sqrt{2\pi} \int_0^{\infty} e^{-(1+\lambda)y} e^{t-y}\ dy \tag{26}$$

$$= 2e^t \int_0^{\infty} e^{-(2+\lambda)y}\ dy = 2e^t \int_0^{\infty} e^{-(2+\lambda)y}\ dy = 2e^t \frac{1}{2+\lambda}. \tag{27}$$

For non-negative $t \geq 0$, we further simplify

$$\begin{aligned} u(t) &= \sqrt{\frac{2}{\pi}} \sqrt{2\pi} \int_t^{\infty} e^{-(1+\lambda)y} e^{t-y}\ dy \\ &= 2e^t \int_t^{\infty} e^{-(2+\lambda)y}\ dy = 2\frac{e^t}{-(2+\lambda)} \left(0 - e^{-(2+\lambda)t}\right) = 2\frac{e^t}{2+\lambda} e^{-(2+\lambda)t} = 2\frac{e^{-(1+\lambda)t}}{2+\lambda}. \end{aligned}$$

For completeness only, we check that this satisfies the equation. When $t < 0$, then

$$2\frac{e^t}{2+\lambda} + \lambda \int_0^{+\infty} e^{-y} 2\frac{e^{t-y}}{2+\lambda}\ dy \tag{28}$$

$$= \frac{2}{2+\lambda}e^t + \frac{2\lambda}{2+\lambda}e^t \int_0^{+\infty} e^{-2y}\ dy = \frac{2}{2+\lambda}e^t + \frac{\lambda}{2+\lambda}e^t = e^t = e^{-|t|}. \tag{29}$$

When $t > 0$, we split the integral to switch between the two different regimes of u. We find

$$\begin{aligned}
&u(t) + \lambda \int_0^t e^{-y}u(t-y)\ dy + \lambda \int_t^{\infty} e^{-y}u(t-y)\ dy \\
&= u(t) + \lambda \int_0^t e^{-y} 2 \frac{e^{-(1+\lambda)(t-y)}}{2+\lambda}\ dy + \lambda \int_t^{\infty} e^{-y} 2e^{t-y} \frac{1}{2+\lambda}\ dy \\
&= u(t) + \frac{2}{2+\lambda}e^{-(1+\lambda)t}\lambda \int_0^t e^{-y}e^{(1+\lambda)y}\ dy + \frac{2}{2+\lambda}e^t \lambda \int_t^{\infty} e^{-y}e^{-y}\ dy \\
&= u(t) + u(t)\lambda \int_0^t e^{\lambda y}\ dy + \frac{2}{2+\lambda}e^t \lambda \int_t^{\infty} e^{-2y}\ dy \\
&= u(t) + u(t)\left(e^{\lambda t} - 1\right) + \frac{2\lambda}{2+\lambda}e^t \frac{e^{-2t}}{2} \\
&= u(t) + u(t)\left(e^{\lambda t} - 1\right) + \frac{\lambda}{2+\lambda}e^{-t}.
\end{aligned}$$

We plug in the definition of $u(t)$ for $t \geq 0$ and observe

$$\begin{aligned}
&u(t) + u(t)\left(e^{\lambda t} - 1\right) + \frac{\lambda}{2+\lambda}e^{-t} \\
&= \frac{2}{2+\lambda}e^{-(1+\lambda)t} + \frac{2}{2+\lambda}e^{-(1+\lambda)t}\left(e^{\lambda t} - 1\right) + \frac{\lambda}{2+\lambda}e^{-t} \\
&= \frac{2}{2+\lambda}e^{-(1+\lambda)t} + \frac{2}{2+\lambda}\left(e^{-t} - e^{-(1+\lambda)t}\right) + \frac{\lambda}{2+\lambda}e^{-t} = e^{-t}.
\end{aligned}$$

This completes the verification of the solution.