Bericht über ein Praktikum an der *Physikalisch-Technischen Bundesanstalt* zu Messunsicherheitspropagation in neuronalen Netzen

betreut von Dr. Sascha Eichstädt (E-Mail: sascha.eichstaedt@ptb.de)
geschrieben von Björn Ludwig (E-Mail: bjoern.ludwig@ptb.de)

Inhaltsverzeichnis

1 Die Physikalisch-Technische Bundesanstalt **1**

2 Meine Aufgaben **6**

3 Ergebnisse **6**

4 Fazit und Beurteilung **7**

A LPU for DNN **7**

B LPU for selected activations **11**

B.1 LPU for sigmoid 12

B.2 LPU for parametric softplus 13

B.3 LPU for QuadLU 16

1 Die Physikalisch-Technische Bundesanstalt

Die Inhalte dieses Abschnitts sind zum überwiegenden Teil aus unterschiedlichen Bereichen der Homepage der PTB unter ptb.de entnommen und hier zusammengetragen.

Die Physikalisch-Technische Bundesanstalt (PTB) ist das sogenannte nationale Metrologieinstitut (NMI) Deutschlands. Als solches ist sie verantwortlich für die Darstellung und die Weitergabe der physikalischen Einheiten, aber auch für metrologische Forschung und Dienstleistungen im Dienste der Wissenschaft, Politik, Wirtschaft und Gesellschaft. Die Metrologie, die wissenschaftliche Basis des Messens und aller daraus folgenden Anwendungen, ist essenziell für unsere moderne Welt, denn Messtechnik ist heutzutage in fast allen Bereichen des Lebens und Wirtschaftens integraler Bestandteil. Die PTB ist eine wissenschaftlich-technische Ressortforschungseinrichtung und Oberbehörde im Geschäftsbereich des Bundesministeriums für Wirtschaft und Klimaschutz. In allen Fragen des Messwesens und der Metrologie ist sie die höchste Beraterin der Bundesregierung. Als weltweit zweitgrösstes NMI, nach dem US-amerikanischen NIST, geniesst sie international hohes Ansehen, wenn es darum geht, Einheiten zu erforschen und präzise Messungen durchzuführen.

Der Großteil ihrer Arbeiten finden an den beiden Hauptstandorten in Berlin und Braunschweig statt:

- Bundesallee 100, 38116 Braunschweig: Der Hauptsitz der PTB befindet sich in Braunschweig, nordwestlich der Stadt gelegen. Seit 1947 ist die PTB dort ansässig. Auf dem $1\,\mathrm{km}^2$ großen Gelände sind etwa 1.700 Mitarbeitende in neun Abteilungen beschäftigt, darunter das Präsidium.

- Abbestraße 2–12, 10587 Berlin: Das Institut Berlin (IB) ist die historische Keimzelle der Physikalisch-Technischen Reichsanstalt (PTR), die 1887 auf Initiative von Werner von Siemens und Hermann von Helmholtz in Berlin-Charlottenburg gegründet wurde. In einer wechselvollen Geschichte wirkten herausragende Persönlichkeiten der Physik an der PTR. Sie wurde 1953 als IB in die zwischenzeitlich in Braunschweig gegründete PTB eingegliedert. Heute beherbergt das IB zwei der neun Fachabteilungen. Die Mitarbeiterinnen und Mitarbeiter arbeiten dort

in den Bereichen Temperatur und Physik mit Synchrotronstrahlung sowie Medizinphysik und Informationstechnik.

Daneben existieren zwei weitere Messstandorte für Metrologie mit Synchrotronstrahlung.

- am Speicherring BESSY II in Berlin-Adlershof beim Helmholtz-Zentrum Berlin in der Albert-Einstein-Str. 15, 12489 Berlin: Seit 1998 nutzt die PTB die Synchrotronstrahlung des Speicherrings BESSY II in Berlin-Adlershof und setzt damit die Arbeiten fort, die am Ende 1999 stillgelegten Speicherring BESSY I begonnen wurden. Die 12 Messplätze entwickeln sich zunehmend zu einem europäischen Zentrum für Metrologie mit Synchrotronstrahlung, das von Kooperationspartnern weltweit genutzt wird. Sie decken metrologische Dienstleistungen ab, von Routinekalibrierungen bis hin zu maßgeschneiderten metrologischen Lösungen und metrologieorientierter Forschung in kurzwelligen Spektralbereichen vom extremen Ultraviolett bis zum Röntgenbereich.

- Willy-Wien-Laboratorium, in der Magnusstr. 9, 12489 Berlin: In Ergänzung zum Speicherring BESSY II hat die PTB ihre Instrumentierung um einen eigenen Speicherring erweitert, die Metrology Light Source (MLS). Sie wurde speziell für die Anforderungen der Metrologie konzipiert und gebaut. Die MLS befindet sich in der Nähe des Speicherrings BESSY II in Berlin-Adlershof und wurde im Frühjahr 2008 in Betrieb genommen. Die PTB betreibt an der MLS ein Labor, in dem derzeit elf Messplätze zur Verfügung stehen. Dort werden Präzisionsmessungen und Kalibrierungen im Spektralbereich vom fernen Infrarot / THz bis zum extremen Ultraviolett angeboten.

Kein wissenschaftliches Experiment, kein industrieller Prozess, kein Waren- und Güterverkehr kommt ohne Quantifizierung aus. Die Messtechnik und ihr wissenschaftliches Rückgrat, die Metrologie, sind heute fast selbstverständlich geworden. Doch diese Selbstverständlichkeit präziser und zuverlässiger Messungen muss erarbeitet werden, und zwar nicht nur einmal, sondern immer wieder – synchron zu den ständig steigenden Genauigkeitsanforderungen der „metrologischen Kundinnen und Kunden“. Die Aufgabe eines NMI wie der PTB ist es daher, für die ständige Funktionsfähigkeit und damit für eine zuverlässige und fortschrittliche metrologische Infrastruktur zu sorgen, die den Anforderungen von Wissenschaft und Hightech-Industrie einerseits und den alltagsnahen Randbedingungen des gesetzlichen Messwesens andererseits gerecht wird.

Die Liste von derzeit insgesamt 22 Gesetzen, Verordnungen, Richtlinien und Abkommen, in denen Aufgaben und Ziele der Arbeit der PTB formuliert werden, ist lang und reicht von der Bekanntmachung der Meterkonvention am 20.05.1875 bis zur Bekanntmachung des Gesetzes über die Bereitstellung von Produkten auf dem Markt (Produktsicherheitsgesetz) vom 27.07.2021. In einem von der PTB geführten und öffentlich zugänglichen Dokument werden diese gesetzlichen Aufgaben gepflegt (Z.13, 2022). Konkret arbeitet die PTB unter anderem an den folgenden Aufgaben:

- Darstellung und Weitergabe der gesetzlichen Einheiten,

- Aufbewahrung der Prototypen der Bundesrepublik Deutschland sowie der Einheitenverkörperungen und Normale und Anschluss an die internationalen Prototypen oder Etalons nach der internationalen Meterkonvention,

- Wahrnehmung der Aufgaben einer Konformitätsbewertungsstelle,
- Metrologische Rückführung von Normalgeräten und Prüfungshilfsmitteln für Konformitätsbewertungsstellen, zuständige Behörden und staatlich anerkannte Prüfstellen,
- Erteilung von EG-Bauartzulassungen,
- Mitwirkung bei der Sicherstellung und Aufrechterhaltung eines bundesweit einheitlichen Sicherheitsniveaus für den Betrieb von zertifizierten Smart-MeterGateways,
- Zulassung der Bauart von Schussapparaten, Einsteckläufen und nicht der Beschusspflicht unterliegenden Feuerwaffen, Systemprüfungen von Schussapparaten und der in ihnen zu verwendenden Kartuschenmunition,
- Erstellung eines Leitfadens zu messtechnischen Kontrollen von Medizinprodukten mit Messfunktion, Bekanntmachung, Archivierung nicht mehr aktueller Leitfäden,
- Bereitstellung von Radioaktivitätsstandards,
- Bauartzulassungen von Spielgeräten mit Gewinnmöglichkeit,
- Prüfung der Bauart von Wahlgeräten,
- Bereitstellung von Aktivitätsnormalen für den Bereich Nuklearmedizin,
- internationale Zusammenarbeit zur Festlegung der physikalisch-technischen Einheiten.

Darüber hinaus ist die PTB im Rahmen der deutschen Entwicklungszusammenarbeit in zahlreichen Entwicklungs- und Schwellenländern aktiv, um Projekte zum Aufbau einer lokalen Qualitätsinfrastruktur zu fördern. Heutzutage geht es dabei vor allem um den bedarfsgerechten Auf- und Ausbau aller relevanten Institutionen der Qualitätsinfrastruktur (QI) in den Partnerländern, um deren Vernetzung untereinander sowie um die Nutzung der QI-Dienstleistungen durch Anwenderinnen und Anwender.

Der weitaus größte Teil der Arbeiten der PTB findet jedoch im Bereich der Wissenschaft statt. Mit rund 70 Prozent ist der Anteil an Forschungsaufgaben der PTB größer als der aller anderen gesetzlichen und Dienstleistungsaufgaben. Die Forschungsarbeiten in der PTB sind dabei stets metrologisch geprägt, prinzipiell anwendungsorientiert und damit deutlich unterschieden von reiner Grundlagenforschung. Mit ihren „Aufgaben und Ziele der Themenbereiche für 2023–2025“ (Physikalisch-Technische Bundesanstalt, n. d.) hat die PTB auch in diesem Bereich ein öffentlich zugängliches Dokument bereitgestellt, um einen umfassenden Überblick über ihre geplanten und aktuell laufenden Forschungsvorhaben zu erhalten. Die Arbeiten sind darin nach dem Vorbild der weiter oben schon erwähnten Meterkonvention in 14 sogenannte Themenbereiche gegliedert:

1. Akustik, Ultraschall, Beschleunigung: In diesem Themenbereich werden metrologische Fragestellungen bearbeitet, die die Darstellung und Weitergabe dynamischer mechanischer Einheiten zum Inhalt haben.

2. Durchfluss: Dieser Themenbereich befasst sich auf der Basis des Einheiten- und Zeitgesetzes (EinhZeitG) mit der Darstellung und Weitergabe der Einheiten für die Strömungsmessgrößen Menge, Durchfluss und thermische Energiemessung von Flüssigkeiten sowie Menge, Durchfluss, Strömungsgeschwindigkeit, thermische Energiemessung und kalorische Größen von Gasen.

3. Elektrizität und Magnetismus: Dieser Themenbereich befasst sich gemäß aus dem EinhZeitG und dem Mess- und Eichgesetz abgeleiteten Auftrag mit der Darstellung, Bewahrung und Weitergabe der elektrischen und magnetischen Einheiten.

4. Ionisierende Strahlung: In diesem Themenbereich werden gemäß EinhZeitG, Strahlenschutzgesetz, Strahlenschutzverordnung, Mess- und Eichgesetz, Mess- und Eichverordnung, IMIS-Zuständigkeitsverordnung, Atomgesetz sowie Medizinproduktegesetz und Medizinprodukte-Betreiberverordnung metrologische Aufgaben sowie Zulassungen und Konformitätsbewertungen wahrgenommen.

5. Länge, dimensionelle Metrologie: Dieser Themenbereich behandelt die sich aus dem gesetzlichen Auftrag gemäß dem EinhZeitG, dem Mess- und Eichgesetz und der Europäischen Messgeräterichtlinie (MID) 2014/32/EU zur Darstellung, Bewahrung und Weitergabe der SI-Einheit der Länge, Meter, sowie den daraus abgeleiteten dimensionellen Messgrößen ergebenden Aufgaben.

6. Masse und abgeleitete Größen: In diesem Themenbereich wird – basierend auf dem EinhZeitG, dem Mess- und Eichgesetz sowie der MID und der Richtlinie für nicht-selbsttätige Waagen (NAWID) – an der Darstellung und Weitergabe der mechanischen Einheiten für Masse, Festkörperdichte, Kraft, Drehmoment, Härte und Druck (auch Vakuumdruck) sowie an Dienstleistungen im Auftrag der Konformitätsbewertungsstelle der PTB gearbeitet.

7. Metrologie in der Chemie und Stoffeigenschaften: Dieser Themenbereich umfasst die metrologische Rückführung von chemisch-analytischen Messungen in Gasen, Flüssigkeiten und Feststoffen, einschließlich der Quantifizierung enthaltener (Nano-) Partikel.

8. Metrologie in der Medizin: Das Aufgabengebiet dieses Themenbereichs besteht in der Bearbeitung von biomedizinischen Messverfahren innerhalb der Metrologie mit dem Ziel, eine bestmögliche Qualitätssicherung zum Wohle der Patientinnen und Patienten zu ermöglichen und neue Lösungswege für metrologische Fragestellungen in der Medizin zu geben.

9. Photometrie und Radiometrie: Auf der Grundlage des EinhZeitG ist in diesem Themenbereich die zentrale Aufgabe die Darstellung und Weitergabe radiometrischer und photometrischer Einheiten.

10. Thermometrie: In der Thermometrie sorgt die Metrologie für die Darstellung, Weitergabe und Weiterentwicklung der internationalen Temperaturskalen, für damit zusammenhängende thermophysikalische Größen und für die Bereitstellung nationaler Normale für die thermische Energie.

11. Zeit und Frequenz: Die Forschungs- und Entwicklungsaufgaben dieses Themenbereichs leiten sich im Kernbereich aus dem EinhZeitG ab und umfassen die Weiterentwicklung des Feldes hin zur Nutzung optischer Uhren und optischer Referenzfrequenzen.

12. Mathematik und metrologische Informationstechnik: Das Gebiet Mathematik in der Metrologie beschäftigt sich mit mathematischen und statistischen Fragestellungen und unterstützt die experimentellen Gruppen der PTB. Das Aufgabengebiet der Fachdisziplin Metrologische Informationstechnik besteht einerseits in Querschnittsaufgaben auf dem Gebiet der Informationstechnik im gesetzlichen Messwesen und andererseits in der Bauartzulassung von Geldspielgeräten sowie den dazugehörigen speziellen Aufgaben.

13. Physikalische Sicherheitstechnik und Explosionsschutz: Dieser Themenbereich ist Teil der Kooperation mit der Bundesanstalt für Materialforschung und -prüfung (BAM) auf dem Gebiet „Chemisch-physikalische Sicherheitstechnik“.

14. Nanometrologie: In diesem Themenbereich werden Arbeiten zur messtechnischen Rückführung der Eigenschaften von Nanoobjekten sowie zum Verständnis von messtechnisch relevanten Prozessen auf der Nanoskala durchgeführt. Nanoobjekte sind durch Abmessungen im nanoskaligen Bereich definiert, das heißt von etwa 1 nm bis 100 nm in wenigstens einer Dimension.

Am 16.03.2021 wurde zu all diesen Aufgaben und Zielen der Arbeiten an der PTB ein sogenanntes Leitbild zunächst intern und dann auch in diversen Print- und Onlinemedien präsentiert. Das Leitbild selbst lautet:

> Die PTB ist das nationale Metrologieinstitut. Wir streben nach dem Maß der Dinge und entwickeln weltweit führende Standards für das Messen. Als Teil des gesetzlichen Messwesens und der Qualitätsinfrastruktur setzen wir zuverlässige und verbindliche Maßstäbe.
>
> Wir sorgen dafür, dass Menschen und Organisationen Messungen vertrauen können. Damit leisten wir einen grundlegenden Beitrag für unsere Wirtschaft, Wissenschaft, Gesellschaft und Umwelt. Mit international anerkannter Spitzenforschung erschließen wir immer neue Bereiche des Messbaren.
>
> Mit unseren vielfältigen Kompetenzen ermöglichen wir gemeinsam technologischen Fortschritt für die Herausforderungen der Zukunft.
>
> Heute, morgen und übermorgen. In allen Bereichen des Lebens.
>
> Gemeinsam mit unseren Kooperationspartnern erschließen wir neue Potentiale für Qualität, Innovation und Nachhaltigkeit. Für Deutschland, Europa und die Welt.

Die darunter liegenden Werte sind wie folgt formuliert:

> Größtmögliche Genauigkeit für zuverlässige und verbindliche Maßstäbe ist unser höchstes Ziel. Wir hinterfragen uns ständig neu und setzen auf modernste Technik, die wir stetig weiterentwickeln und für die wir weltweit anerkannt sind.
>
> Wir stehen für Objektivität, da wir unabhängig von finanziellen und politischen Interessen sind. Wir arbeiten ausschließlich nach wissenschaftlichen Prinzipien. Als Bundesanstalt sind wir neutral und nehmen eine wichtige Rolle bei der verlässlichen Regelsetzung ein.
>
> Seien die einzelnen Aufgabenbereiche an der PTB noch so verschieden: Was uns eint, ist die Leidenschaft fürs Messen.

2 Meine Aufgaben während des Praktikums

Während meines Praktikum war ich im Fachbereich 9.4 „Metrologie für die digitale Transformation“ beschäftigt. Dort durfte ich im Forschungsschwerpunkt „Metrologie für Sensornetzwerke“ im Umfeld erfahrener Wissenschaftler:innen mit vornehmlich mathematischem Hintergrund, aber auch mit Physiker:innen, Chemiker:innen, Maschinenbauer:innen und Informationswissenschaftler:innen arbeiten. Ich selbst setzt mich intensiv mit dem sogenannten „Guide to the expression of uncertainty in measurement“ (GUM) (JCGM, 2008a) auseinander. Dabei handelt es sich um das zentrale Dokument, welches international abgestimmt die Grundlage für das Verständnis von Messunsicherheit in der Metrologie bildet. Gemeinsam mit dem International Vocabulary of Metrology (JCGM, 2012) enthält der GUM alle wichtigen Begriffsbestimmungen und teils erläuternde Anmerkungen zur Interpretation und Anwendung dieser Grundsätze. Die enorme Bedeutung des GUM ist vor allem in der Formulierung des sogenannten „GUM uncertainty framework“ begründet. Im Kern handelt es sich dabei um die Formulierung des mathematischen Zusammenhangs, zwischen Messunsicherheiten im Eingang eines Messmodels und der/n Messunsicherheit/en der durch Messung realisierten Messgröße/n. Meine Auseinandersetzung bezog sich dabei auf die Übertragung des GUM uncertainty framework auf einfache neuronale Netze in Gestalt von sogenannten Mehrschicht Perzeptronen.

Das Ziel war insofern im ersten Schritt die rigorose Formulierung von mathematischen Voraussetzungen an die Netzkomponenten und gegebenenfalls die Architekturen, um die Anwendung des GUM uncertainty frameworks zu ermöglichen. Darauf aufbauend sollten dann erste Schlussfolgerungen gezogen werden und mathematische Aussagen über entsprechend ausgestattete, vortrainierte Netze getroffen werden.

Als Fernziel formulierten wir die exemplarische Simulation der Aussagen in geeigneten Beispielinstanzen solcher Netze, um die Ergebnisse quantitativ mit den Ergebnissen der Monte Carlo Methode zu vergleichen. Diese wird nämlich im Supplement 1 „Propagation of distributions“ des GUM (JCGM, 2008b) ausführlich diskutiert und als Referenzmethode vorgestellt.

3 Ergebnisse

Während der Dauer des Praktikums ist ein solides Rahmenwerk für die Formulierung von Voraussetzungen zur Unsicherheitspropagation in neuronalen Netzen entstanden. Darauf aufbauend habe ich mehrere mathematische Sätze und Propositionen formulieren können, welche die gezogenen Schlüsse exemplarisch zur Anwendung bringen. Darüber hinaus ist ein FAIRes Softwarepaket entstanden (Ludwig, 2023b), in welchem die sich ergebende Mathematik im Python Framework PyTorch umfassend und sehr gut getestet implementiert ist. Alle Ergebnisse des Praktikums konnten im Anschluss in einer Masterarbeit mit dem Titel „GUM-compliant neural network robustness verification“ (Ludwig, 2023a) zusammen mit dem Zuse-Institut Berlin (ZIB) und der PTB verarbeitet werden. Darüber hinaus werden die Ergebnisse auf dem *21st INTERNATIONAL METROLOGY CONGRESS* und voraussichtlich auch in einer Sonderausgabe der Zeitschrift *IOP Measurement Science and Technology* veröffentlicht. Die zentralen Ergebnisse finden sich in den Anhängen A und B.

4 Fazit und Beurteilung des Praktikums

Die wissenschaftliche Arbeit an der PTB war eine spannende Erfahrung für mich, obwohl ich als Forschungssoftwareentwickler und -projektleiter bereits mehrere Jahre Erfahrung im Umfeld wissenschaftlicher Fragestellungen sammeln konnte. Bisher drehten sich meine Aufgaben dabei eher um die Implementierung von fertigen Algorithmen und insofern die Anwendung insbesondere außerhalb der Forschungswelt etablierter Methoden und Werkzeuge zur Erstellung und Wartung von komplexen Softwareprodukten in Zusammenarbeit mit teils internationalen Teams. Während des Praktikums konnte ich also ein erstes Mal aus meiner Perspektive von der Implementierung neuer Ideen zur Findung solcher Ideen wechseln.

Der Austausch mit erfahrenen Wissenschaftlerinnen und Wissenschaftlern am ZIB und an der PTB in wöchentlichen und anderen regelmäßigen Austauschrunden sowie individuellen Gesprächen, hat dabei vielfach meinen Blick geschärft, welches die relevanten Fragen sind, mit denen ich mich als angehender Wissenschaftler im Feld der angewandten Mathematik in der Metrologie auseinandersetzen sollte.

Da mein Studium voraussichtlich bereits im April dieses Jahres abgeschlossen sein wird, habe ich die glückliche Gelegenheit ergriffen, mich auf eine Stelle an der PTB zu bewerben, in deren Rahmen ich voraussichtlich im Anschluss an den Abschluss an den Ergebnissen weiterarbeiten kann. Insofern freue ich mich, dass ich nicht nur erste Praxiserfahrung bei der Erarbeitung wissenschaftlicher Ergebnisse sammeln, sondern anscheinend auch einen relevanten Beitrag zu aktuellen Entwicklungen leisten konnte.

A Law of propagation of uncertainty for deep neural networks

In this section the foundation for the uncertainty propagation in compliance with the GUM and its supplements for deep neural networks is presented, which are defined for the rest of this work as follows:

Definition A.1. [fully connected, feedforward deep neural network (DNN)] Let $\ell \in \mathbb{N}$ be the number of layers not including the input layer, $n^{(0)}, \dots, n^{(\ell)} \in \mathbb{N}$ the number of neurons per layer, $x^{(0)} \in \mathbb{R}^{n^{(0)}}$ the input vector, $h^{(i)} \colon \mathbb{R}^{n^{(i)}} \to \mathbb{R}^{n^{(i)}}, i = 1, \dots, \ell$ the vector of the activation functions of the i-th layer and $x^{(i)} \colon \mathbb{R}^{n^{(i-1)}} \to \mathbb{R}^{n^{(i)}}, i = 1, \dots, \ell$ the hidden and output layers' activations defined as

$$x^{(i)} \colon x^{(i-1)} \mapsto x^{(i)}(x^{(i-1)}) := (h^{(i)} \circ z^{(i)})(x^{(i-1)}), \quad i = 1, \dots, \ell, \qquad (i\text{-th layer's activation})$$

where the i-th layer's affine transformation $z^{(i)} \colon \mathbb{R}^{n^{(i-1)}} \to \mathbb{R}^{n^{(i)}}$ is defined as

$$\begin{aligned} z^{(i)} \colon x^{(i-1)} \mapsto z^{(i)}(x^{(i-1)}) &:= W^{(i)} x^{(i-1)} + b^{(i)} \\ &= \Big(b_k^{(i)} + \sum_{j=1}^{n^{(i-1)}} w_{k,j}^{(i)} x_j^{(i-1)} \Big)_{k=1,\dots,n^{(i)}}, \quad i = 1, \dots, \ell, \end{aligned}$$

(i-th layer's affine transformation)

with weight matrices $W^{(i)} = (w^{(i)}_{k,j})_{\substack{k=1,\dots,n^{(i)}\\ j=1,\dots,n^{(i-1)}}} \in \mathbb{R}^{n^{(i)}\times n^{(i-1)}}$ and bias vectors $b^{(i)} \in \mathbb{R}^{n^{(i)}}, i = 1,\dots,\ell$. For the sake of readability, we will omit the indication of the argument $x^{(i-1)}$ in the following if it does not serve the purpose of comprehensibility. Then $f\colon \mathbb{R}^{n^{(0)}} \to \mathbb{R}^{n^{(\ell)}} : x^{(0)} \mapsto f(x^{(0)})$ defined as

$$\begin{aligned} f(x^{(0)}) &\coloneqq x^{(\ell)} \\ &= \underbrace{h^{(\ell)}\Big(W^{(\ell)}\, \underbrace{h^{(\ell-1)}\Big(W^{(\ell-1)} \cdots \underbrace{h^{(1)}\Big(W^{(1)}x^{(0)} + b^{(1)}\Big)}_{x^{(1)}} \cdots + b^{(\ell-1)}\Big)}_{\vdots\ x^{(\ell-1)}} + b^{(\ell)}\Big)}_{x^{(\ell)}} \end{aligned} \tag{DNN}$$

is called a *fully connected, feedforward deep neural network (DNN).*

A concrete instance of a DNN with $\ell = 3$ layers and $n^{(0)} = n^{(3)} = 2, n^{(1)} = n^{(2)} = 5$ neurons for the respective layers is exemplified in figure 1.

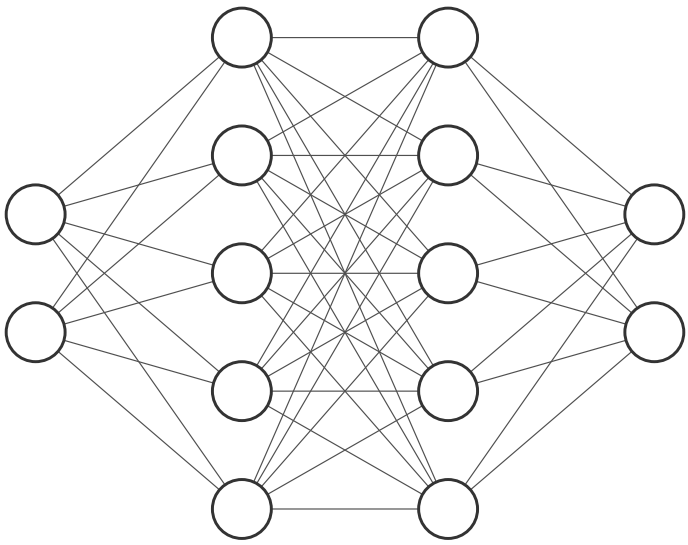

Abbildung 1: a schematic view on a DNN with each two in- and output neurons and two hidden layers, where each connecting line corresponds to one $x^{(i-1)}_k$ and each neuron to one mapping $h^{(i)}_k \circ z^{(i)}_k, i = 1,\dots,3$.

For the central definition of a law of propagation of uncertainty suited for neural network applications, first the notion of a *sensitivity matrix* needs to be defined.

Definition A.2 (Sensitivity matrix)**.** Let $N, M \in \mathbb{N}, \Omega \subseteq \mathbb{R}^N$ be an open set, $\mathbf{x} = (x_1,\dots,x_N) \in \Omega$ the best estimate of the vector input quantity $\mathbf{X} = (X_i)_{i=1,\dots,N}$ for a multivariate measurement function $f\colon \mathbb{R}^N \to \mathbb{R}^M : \mathbf{x} \mapsto f(\mathbf{x}) \coloneqq \mathbf{y}$. Let $f \in C^1(\Omega)$, i.e. continuously differentiable on Ω. Then the sensitivity matrix $\mathbf{C_x} \in \mathbb{R}^{M\times N}$ with respect to the input of this measurement is the Jacobian matrix of the measurement function

$$\mathbf{C_x} \coloneqq \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f_1}{\partial x_N}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1}(\mathbf{x}) & \dots & \frac{\partial f_M}{\partial x_N}(\mathbf{x}) \end{pmatrix} = J_f(\mathbf{x}). \tag{1}$$

Now a formal definition with explicit requirements for the law of propagation of uncertainties for real, multivariate measurement models can be presented.

Definition A.3 (multivariate law of propagation of uncertainty)**.** Let $N, M \in \mathbb{N}, \Omega \subseteq \mathbb{R}^N$ be an open set, $\mathbf{x} \in \Omega$ the best estimate of the vector input quantity $\mathbf{X} = (X_i)_{i=1,\dots,N}$ for a multivariate measurement function $f\colon \mathbb{R}^N \to \mathbb{R}^M\colon \mathbf{x} \mapsto f(\mathbf{x}) := \mathbf{y}$ and $\mathbf{U_x} \in \mathbb{R}^{N\times N}_{+,0}$ the associated covariance matrix. Let $f \in C^1(\Omega)$ and $\mathbf{C_x} \in \mathbb{R}^{M\times N}$ the sensitivity matrix with respect to the input as per Definition A.2. Then the covariance matrix $\mathbf{U_y} \in \mathbb{R}^{M\times M}_{+,0}$ associated with $\mathbf{y}$ is

$$\mathbf{U_y} = \mathbf{C_x}\mathbf{U_x}\mathbf{C_x}^T. \tag{2}$$

Remark A.4. A neural network as defined by A.1 from the perspective of the GUM is a real, multivariate, multistage measurement model with the measurement function f as in Definition A.1, best estimate $x^{(0)}$ for the real, vector valued input quantity $\mathbf{X}^{(0)}$, intermediate vector quantities $x^{(i)}, i = 1, \dots, \ell - 1$ and the best estimate $x^{(\ell)}$ for the vector measurand $\mathbf{X}^{(\ell)}$. As such the LPU A.3 requires continuous differentiability of each layer's mapping $x^{(i)}\colon \mathbb{R}^{n^{(i-1)}} \to \mathbb{R}^{n^{(i)}}\colon x^{(i-1)} \mapsto x^{(i)}(x^{(i-1)}) = (h^{(i)} \circ z^{(i)})(x^{(i-1)}), i = 1, \dots, \ell$ in a neighbourhood of the previous layer's activations $x^{(i-1)}$. As these in general are not known in advance, the continuous differentiability prerequisite of f implies the prerequisite of continuous differentiability of each of the $x^{(i)}$ on its entire domain $\mathbb{R}^{n^{(i-1)}}, i = 1, \dots, \ell$ to ensure the applicability of the 2.

Theorem A.5 (law of propagation of uncertainty for deep neural networks)**.** Let f be a DNN according to Definition A.1, $\ell \in \mathbb{N}$ the number of layers in addition to the input layer, $n^{(0)}, \dots, n^{(\ell)} \in \mathbb{N}$ the number of neurons per layer, $x^{(0)} = \left(x_j^{(0)}\right)_{j=1}^{n^{(0)}} \in \mathbb{R}^{n^{(0)}}$ the input vector comprised of the best estimates of the input quantities $X_j, \mathbf{U}_{x^{(0)}}$ the associated covariance matrix and $W^{(i)}, i = 1, \dots, \ell$ the i-th layer's weight matrix. Further let each activation function $h_k^{(i)}\colon \mathbb{R} \to \mathbb{R}, i = 1, \dots, \ell, k = 1, \dots, n^{(i)}$ be continuously differentiable on $\mathbb{R}$. Then the LPU according to Definition A.3 can be applied to propagate the covariance matrix $\mathbf{U}_{x^{(0)}}$ associated with the inputs to compute the covariance matrix $\mathbf{U}_{x^{(\ell)}}$ associated with the outputs $x^{(\ell)}$ as

$$\mathbf{U}_{x^{(\ell)}} = \Big(\prod_{i=0}^{\ell-1} H^{(\ell-i)}W^{(\ell-i)}\Big)\mathbf{U}_{x^{(0)}}\Big(\prod_{i=1}^{\ell} {W^{(i)}}^T H^{(i)}\Big), \tag{LPU for DNN}$$

with the diagonal matrices $H^{(i)}$

$$H^{(i)} := \operatorname{diag}\left(\frac{\mathrm{d}h_k^{(i)}}{\mathrm{d}z}(z_k^{(i)})\right)_{k=1,\dots,n^{(i)}}, \quad i = 1, \dots, \ell, \tag{3}$$

where diag denotes the diag operator for vectors.

Remark A.6. In the notation of the equation (LPU for DNN) the matrix products must be expanded in the specified order as matrix multiplication does not commute.

Proof of Theorem A.5. It is to show that

- indeed this leads to the sensitivity matrices as stated by equations (LPU for DNN) and (3), and

- equation (LPU for DNN) then provides the outputs' covariance matrix.

Let f be a DNN as per Definition A.1 with input layer and additional $\ell \in \mathbb{N}$ layers. The applicability of the LPU A.3 and equation (LPU for DNN) are shown by means of complete induction on the number of layers, denoted by n.

For $n = 1$ it is

$$\begin{aligned} f(x^{(0)}) &= x^{(1)} \\ &= (h^{(1)} \circ z^{(1)})(x^{(0)}) \\ &= \begin{pmatrix} h_1^{(1)} \circ z_1^{(1)} \\ \vdots \\ h_{n^{(1)}}^{(1)} \circ z_{n^{(1)}}^{(1)} \end{pmatrix}. \end{aligned} \tag{4}$$

As affine map $z^{(1)}$ is differentiable with constant derivative, thus continuously differentiable on $\mathbb{R}^{n^{(0)}}$ with Jacobian

$$J_{z^{(1)}} = \begin{pmatrix} \frac{\partial z_1^{(1)}}{\partial x_1} & \cdots & \frac{\partial z_1^{(1)}}{\partial x_{n^{(0)}}} \\ \vdots & \ddots & \vdots \\ \frac{\partial z_{n^{(1)}}^{(1)}}{\partial x_1} & \cdots & \frac{\partial z_{n^{(1)}}^{(1)}}{\partial x_{n^{(0)}}} \end{pmatrix}. \tag{5}$$

From equation (i-th layer's affine transformation) it follows immediately

$$\frac{\partial z_k^{(1)}}{\partial x_j} = w_{k,j}^{(1)}, \quad , k = 1, \ldots, n^{(1)}, j = 1, \ldots, n^{(0)}. \tag{6}$$

Since $h_k^{(1)}, k = 1, \ldots, n^{(1)}$ is continuously differentiable on $\mathbb{R}$ the composition $h_k^{(1)} \circ z_k^{(1)}$ is continuously differentiable on $\mathbb{R}^{n^{(0)}}$ as well with its Jacobian being $J_{h_k^{(1)} \circ z_k^{(1)}} = J_{h_k^{(1)}} J_{z_k^{(1)}}$ with

$$J_{h_k^{(1)}} = \frac{\mathrm{d}h_k^{(1)}}{\mathrm{d}z}(z_k^{(1)}), \quad k = 1, \ldots, n^{(1)}. \tag{7}$$

As all thebe coordinate functions of f are continuously differentiable on $\mathbb{R}^{n^{(0)}}$ so is f and its derivative f', i.e. sensitivity matrix according to Definition A.2, is

$$\mathbf{C}_{x^{(0)}} = f' = \begin{pmatrix} h_1^{(1)} \circ z_1^{(1)} \\ \vdots \\ h_{n^{(1)}}^{(1)} \circ z_{n^{(1)}}^{(1)} \end{pmatrix}' = H^{(1)} W^{(1)}, \tag{8}$$

where $H^{(1)} := \operatorname{diag}\left(\frac{\mathrm{d}h_k^{(i)}}{\mathrm{d}z}(z_k^{(i)})\right)_{k=1,\ldots,n^{(1)}}$ with the diag operator for vectors (Horn & Johnson, 2012, Definition 0.9.1). Now the LPU for multivariate real measurement functions A.3 can be applied which provides for the covariance matrix $\mathbf{U}_{x^{(1)}}$ associated with the neural network's outputs $x^{(1)}$

$$\mathbf{U}_{x^{(1)}} = \mathbf{C}_{x^{(0)}} \mathbf{U}_{x^{(0)}} \mathbf{C}_{x^{(0)}}^T = H^{(1)} W^{(1)} \mathbf{U}_{x^{(0)}} \left(H^{(1)} W^{(1)}\right)^T = H^{(1)} W^{(1)} \mathbf{U}_{x^{(0)}} {W^{(1)}}^T H^{(1)}, \tag{9}$$

where $H^{(1)} = H^{(1)^T}$ was used. Assume now, that the LPU is applicable for each of the n first layers and $\mathbf{U}_{x^{(i)}}$ is the covariance matrix associated with the i-th layer's activations $x^{(i)}, i = 1, \dots, n$ given by

$$\mathbf{U}_{x^{(i)}} = \mathbf{C}_{x^{(i-1)}} \mathbf{U}_{x^{(i-1)}} \mathbf{C}_{x^{(i-1)}}^T = H^{(i)} W^{(i)} \mathbf{U}_{x^{(i-1)}} W^{(i)^T} H^{(i)}.$$

For the $(n+1)$-th layer it again is

$$x^{(n+1)} = (h^{(n+1)} \circ z^{(n+1)})(x^{(n)}) \tag{10}$$

$$= \begin{pmatrix} h_1^{(n+1)} \circ z_1^{(n+1)} \\ \vdots \\ h_{n^{(n+1)}}^{(n+1)} \circ z_{n^{(n+1)}}^{(n+1)} \end{pmatrix}. \tag{11}$$

Now by the same reasoning as above the LPU can be applied to $x^{(n+1)}$ and the resulting covariance matrix $\mathbf{U}_{x^{(n+1)}}$ associated with the output layer's activations $x^{(n+1)}$ is

$$\mathbf{U}_{x^{(n+1)}} = \mathbf{C}_{x^{(n)}} \mathbf{U}_{x^{(n)}} \mathbf{C}_{x^{(n)}}^T = H^{(n+1)} W^{(n+1)} \mathbf{U}_{x^{(n)}} W^{(n+1)^T} H^{(n+1)}.$$

Applying this result to f and replacing the covariance matrices recursively leads to

$$\mathbf{U}_{x^{(\ell)}} = \prod_{i=1}^{\ell} \mathbf{C}_{x^{(\ell-i)}} \mathbf{U}_{x^{(0)}} \prod_{i=0}^{\ell-1} \mathbf{C}_{x^{(i)}}^T \tag{12}$$

$$= \Big(\prod_{i=0}^{\ell-1} H^{(\ell-i)} W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} W^{(i)^T} H^{(i)}\Big). \tag{13}$$

□

B Law of propagation of uncertainty for selected activation functions

In each of the following subsections, networks with exclusively one distinct activation function are considered, although the same approach can be readily adopted for networks with varying activation functions per neuron. Due to the smoothness prerequisite on the activation functions employed (see Remark A.4), in the following we consider three continuously differentiable activation functions to exemplify the application of the LPU for DNN.

In particular, as already pointed out we cannot work with the *rectifier* or *ReLU* as it is usually referred to. Nevertheless, the definition will be given here, since in the further course at least a continuously differentiable approximation of ReLU will be worked with, namely the so-called (parametric) softplus activation function.

Definition B.1 (ReLU)**.** The mapping $\mathrm{ReLU}\colon \mathbb{R} \to \mathbb{R}_{+,0}$ defined as

$$\mathrm{ReLU}(x) \coloneqq \max\{0, x\}, \tag{ReLU}$$

is called the *ReLU activation function* (or just *ReLU*) because it transforms neurons in neural networks that are equipped with it into *rectified linear units*.

B.1 Uncertainty propagation for sigmoid activation

The sigmoid activation function is a common choice for DNNs when continuous differentiability is desired and the output of the neurons should lie in the open interval $(0, 1)$, as in the case of Bernoulli output probability distributions (Goodfellow et al., 2016, chapter 6.2.2.2).

Definition B.2 (sigmoid activation function)**.** The mapping $\sigma\colon \mathbb{R} \to (0, 1)$ defined as

$$\sigma(x) \coloneqq \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}}. \tag{sigmoid}$$

is called the *sigmoid activation function* (or just *sigmoid*).

Figure 2 shows the graph of the sigmoid activation function.

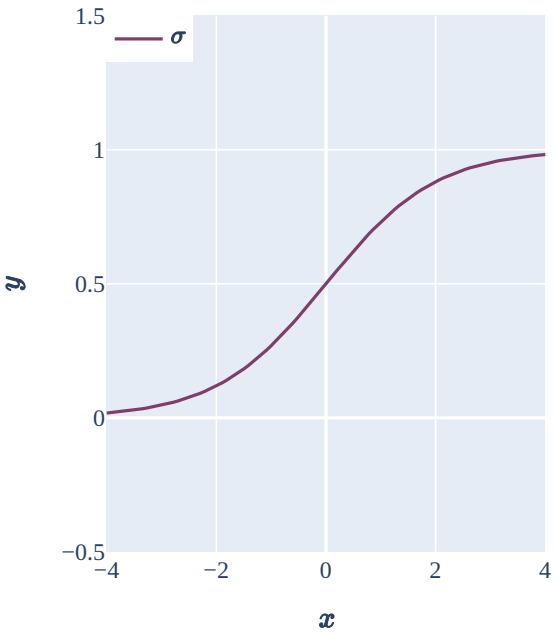

Abbildung 2: graph of sigmoid activation function for $x \in [-5, 5]$

Proposition B.3 (law of propagation of uncertainty for DNNs exclusively equipped with sigmoid activation function)**.** Let f be a DNN in the same situation as in Theorem A.5 exclusively equipped with the sigmoid activation function σ. Then the LPU according to Definition A.3 can be applied with $H^{(i)}$ equal to

$$H^{(i)} = \operatorname{diag} \sigma(z^{(i)}) \operatorname{diag}\big(\mathbf{1} - \sigma(z^{(i)})\big), \quad i = 1, \ldots, \ell,$$

with $\mathbf{1} \coloneqq (1, \ldots, 1) \in \mathbb{R}^{n^{(i)}}$, such that for the covariance matrix $\mathbf{U}_{x^{(\ell)}}$ associated with the output

layer's activations $x^{(\ell)}$ it is

$$\mathbf{U}_{x^{(\ell)}} = \Big(\prod_{i=0}^{\ell-1} \operatorname{diag}\sigma(z^{(\ell-i)}) \operatorname{diag}\big(\mathbf{1} - \sigma(z^{(\ell-i)})\big) W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} {W^{(i)}}^T \operatorname{diag}\sigma(z^{(i)}) \operatorname{diag}\big(\mathbf{1} - \sigma(z^{(i)})\big)\Big). \tag{LPU for sigmoid DNN}$$

Beweis. σ is a composition of the exponential $\exp\colon \mathbb{R} \to \mathbb{R}^+ \colon x \mapsto \exp(x) \coloneqq e^x$ and the affine function $1+x$, which are both continuously differentiable on $\mathbb{R}$, and $\frac{1}{x}$, which is continuously differentiable on $\mathbb{R}^+$. As such, σ is continuously differentiable on $\mathbb{R}$ with derivative

$$\frac{\mathrm{d}\sigma}{\mathrm{d}x}(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right) = \sigma(x)(1-\sigma(x)). \tag{14}$$

With $z_k^{(i)}, i = 1, \ldots, \ell, k = 1, \ldots, n^{(i)}$ as per (i-th layer's affine transformation) and equations (3) and (14) it is

$$H^{(i)} = \operatorname{diag}\frac{\mathrm{d}\sigma}{\mathrm{d}z}(z^{(i)}) = \operatorname{diag}\sigma(z^{(i)}) \operatorname{diag}(\mathbf{1} - \sigma(z^{(i)})), \quad i = 1, \ldots, n^{(i)}.$$

Theorem A.5 then yields

$$\begin{aligned}
\mathbf{U}_{x^{(\ell)}} &= \Big(\prod_{i=0}^{\ell-1} H^{(\ell-i)} W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} {W^{(i)}}^T H^{(i)}\Big) \\
&= \Big(\prod_{i=0}^{\ell-1} \operatorname{diag}\sigma(z^{(\ell-i)}) \operatorname{diag}(\mathbf{1} - \sigma(z^{(\ell-i)})) W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} {W^{(i)}}^T \operatorname{diag}\sigma(z^{(i)}) \operatorname{diag}(\mathbf{1} - \sigma(z^{(i)}))\Big).
\end{aligned}$$

□

B.2 Uncertainty propagation for parametric softplus activation

The second activation function that is considered is a continuously differentiable, parametric version of the so-called *rectifier* usually referred to as the *(parametric) softplus*. Although its deployment as an activation function is sometimes discouraged, e.g. by (Goodfellow et al., 2016, chapter 6.3.3) or (Szandała, 2021), there are also use cases where softplus performs better than ReLU regarding prediction accuracy (Bingham & Miikkulainen, 2022, chapters 5.2, 6.4), stabilization and error rates in speech recognition (Zheng et al., 2015) and during so-called *adversarial training* which aims to raise networks' resilience against adversarial attacks (Xie et al., 2021).

Definition B.4 (softplus activation function)**.** Let $\beta \in \mathbb{R}_+$. The mapping $\operatorname{softplus}_\beta \colon \mathbb{R} \to \mathbb{R}_+$ defined as

$$\operatorname{softplus}_\beta(x) = \frac{\ln\left(1 + e^{\beta x}\right)}{\beta}, \quad x \in \mathbb{R}, \tag{parametric softplus}$$

is called the *(parametric) softplus activation function* (or just *softplus*).

Figure 3 shows the graph of the softplus activation function with differing choice of β.

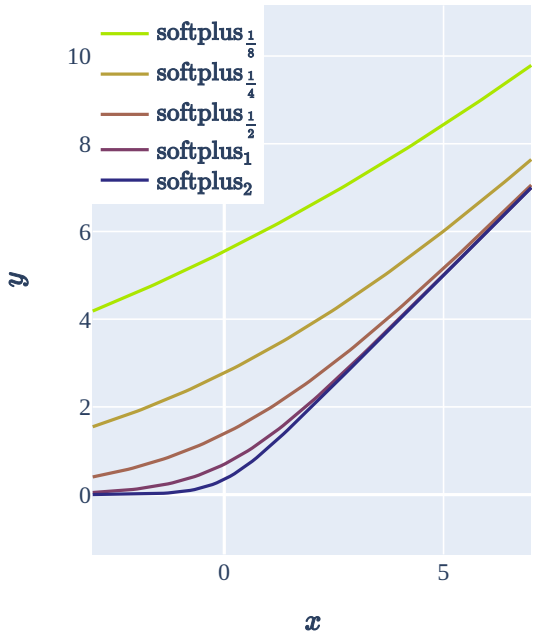

Abbildung 3: graph of softplus activation functions for $\beta \in \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1, 2\}$ and $x \in [-3, 7]$

Proposition B.5 (law of propagation of uncertainty for DNNs exclusively equipped with softplus activation function)**.** Let $\beta \in \mathbb{R}_+$ and f be a DNN in the same situation as in Theorem A.5 exclusively equipped with the softplus activation function $\text{softplus}_\beta \colon \mathbb{R} \to \mathbb{R}_+$. Then the LPU according to Definition A.3 can be applied with $H^{(i)}$ equal to

$$H^{(i)} = \operatorname{diag} \sigma(\beta z^{(i)}), \quad i = 1, \dots, \ell,$$

such that for the covariance matrix $\mathbf{U}_{x^{(\ell)}}$ associated with the output layer's activations $x^{(\ell)}$ it is

$$\mathbf{U}_{x^{(\ell)}} = \Big(\prod_{i=0}^{\ell-1} \operatorname{diag} \sigma(\beta z^{(\ell-i)}) W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} W^{(i)^T} \operatorname{diag} \sigma(\beta z^{(i)})\Big). \qquad \text{(LPU for softplus DNN)}$$

Beweis. With the same reasoning as in the proof for Proposition B.3 and the continuous differentiability of the logarithm on $\mathbb{R}_+$, softplus_β is continuously differentiable on $\mathbb{R}$ with derivative

$$\begin{aligned} \frac{\mathrm{d}\,\text{softplus}_\beta}{\mathrm{d}x}(x) &= \frac{\mathrm{d}}{\mathrm{d}x} \frac{\ln\left(1 + e^{\beta x}\right)}{\beta} \\ &= \frac{1}{\beta} \frac{\beta e^{\beta x}}{1 + e^{\beta x}} \\ &= \sigma(\beta x), \end{aligned} \tag{15}$$

where again the chain rule of differentiation was used and σ is the sigmoid activation function as per Definition B.2. With $z_k^{(i)}, i = 1, \dots, \ell, k = 1, \dots, n^{(i)}$ as per (i-th layer's affine transformation) and equations (3) and (15) it is

$$\begin{aligned} H^{(i)} &= \operatorname{diag} \frac{\mathrm{d}\,\text{softplus}_\beta}{\mathrm{d}z}(z^{(i)}) \\ &= \operatorname{diag} \sigma(\beta z^{(i)}), \quad i = 1, \dots, n^{(i)}. \end{aligned}$$

Theorem A.5 then yields

$$\begin{aligned}\mathbf{U}_{x^{(\ell)}} &= \Big(\prod_{i=0}^{\ell-1} H^{(\ell-i)} W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} W^{(i)^T} H^{(i)}\Big) \\ &= \Big(\prod_{i=0}^{\ell-1} \operatorname{diag} \sigma(\beta z^{(\ell-i)}) W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} W^{(i)^T} \operatorname{diag} \sigma(\beta z^{(i)})\Big).\end{aligned}$$

□

Further investigation of softplus$_\beta$ is justified by the fact that it is a „smooth approximation“ of the most widely spread activation function in the sense, that in the limit transition from $n \to \infty$ one obtains *ReLU*.

Proposition B.6 (ReLU is the limit case of softplus$_\beta$)**.**

$$\lim_{\beta\to\infty} \frac{\ln\left(1+e^{\beta x}\right)}{\beta} = \max\{0, x\}. \tag{16}$$

Beweis. First, it is observed that with the product rule of natural logarithm softplus$_\beta$ can be rewritten as follows

$$\begin{aligned}\operatorname{softplus}_\beta(x) &= \frac{\ln\left(1+e^{\beta x}\right)}{\beta} \\ &= \frac{\ln\left(e^{\beta x-\beta x}+e^{\beta x}\right)}{\beta} \\ &= \frac{\ln\left(e^{\beta x}(e^{-\beta x}+1)\right)}{\beta} \\ &= \frac{\ln\left(e^{-\beta x}+1\right)+\ln\left(e^{\beta x}\right)}{\beta}\end{aligned}$$

First let $x > 0$. Then it is

$$\begin{aligned}\lim_{\beta\to\infty} \frac{\ln\left(1+e^{\beta x}\right)}{\beta} &= \lim_{\beta\to\infty} \left(\frac{\ln\left(e^{-\beta x}+1\right)+\ln\left(e^{\beta x}\right)}{\beta}\right) \\ &= \lim_{\beta\to\infty} \frac{\ln\left(e^{-\beta x}+1\right)}{\beta} + \lim_{\beta\to\infty} \frac{\ln\left(e^{\beta x}\right)}{\beta} \\ &= 0 + \lim_{\beta\to\infty} \frac{\beta x}{\beta} \\ &= x.\end{aligned}$$

Now let $x < 0$. Then it is

$$\lim_{\beta\to\infty} \frac{\ln\left(1+e^{\beta x}\right)}{\beta} = 0,$$

as for $x < 0$ it is

$$\lim_{\beta\to\infty} \ln(1+e^{\beta x}) = \ln(1+0) = 0.$$

Finally for $x = 0$ it is

$$\lim_{\beta\to\infty} \frac{\ln\left(1 + e^{\beta x}\right)}{\beta} = \lim_{\beta\to\infty} \frac{\ln(2)}{\beta}$$
$$= 0.$$

□

B.3 Uncertainty propagation for parametric QuadLU activation

QuadLU emerged in the course of the work on this paper and is first introduced in this work as far as we know. For some time it appeared that for a GUM-compliant modification of existing methods the properties of QuadLU were necessary for an efficient operation. QuadLU is characterised by linearity over almost its entire domain, corresponds exactly to ReLU in this range, and fulfils the minimum requirements for smoothness resulting from the GUM. It is exactly once continuously differentiable, with little effort required to compute the respective derivatives.

Definition B.7 (QuadLU activation function)**.** Let $\alpha \in \mathbb{R}_+$. The mapping $\mathrm{QuadLU}_\alpha \colon \mathbb{R} \to \mathbb{R}_{+,0}$ defined as

$$\mathrm{QuadLU}_\alpha(x) := \begin{cases} 0, & \text{for } x \leq -\alpha \\ \frac{1}{4\alpha}(x+\alpha)^2, & \text{for } -\alpha < x < \alpha \\ x, & \text{for } x \geq \alpha \end{cases}. \tag{QuadLU}$$

is called the *(parametric) QuadLU activation function* (or just *QuadLU*) in reference to the widespread ReLU activation function B.1.

Figure 4 shows the graph of the QuadLU activation function with differing choice of α.

Proposition B.8 (law of propagation of uncertainty for DNNs exclusively equipped with QuadLU activation function)**.** Let $\alpha \in \mathbb{R}_+$ and f be a DNN in the same situation as in Theorem A.5 exclusively equipped with the QuadLU activation function $\mathrm{QuadLU}_\alpha \colon \mathbb{R} \to \mathbb{R}$. Then the LPU according to Definition A.3 can be applied with $H^{(i)}$ equal to

$$H^{(i)} = \operatorname{diag} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}z}(z^{(i)}), \quad i = 1, \ldots, \ell, \tag{17}$$

such that for the covariance matrix $\mathbf{U}_{x^{(\ell)}}$ associated with the output layer's activations $x^{(\ell)}$ it is

$$\mathbf{U}_{x^{(\ell)}} = \Big(\prod_{i=0}^{\ell-1} \operatorname{diag} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}z}(z^{(\ell-i)}) W^{(\ell-i)}\Big) \mathbf{U}_{x^{(0)}} \Big(\prod_{i=1}^{\ell} W^{(i)^T} \operatorname{diag} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}z}(z^{(i)})\Big),$$

(LPU for QuadLU DNN)

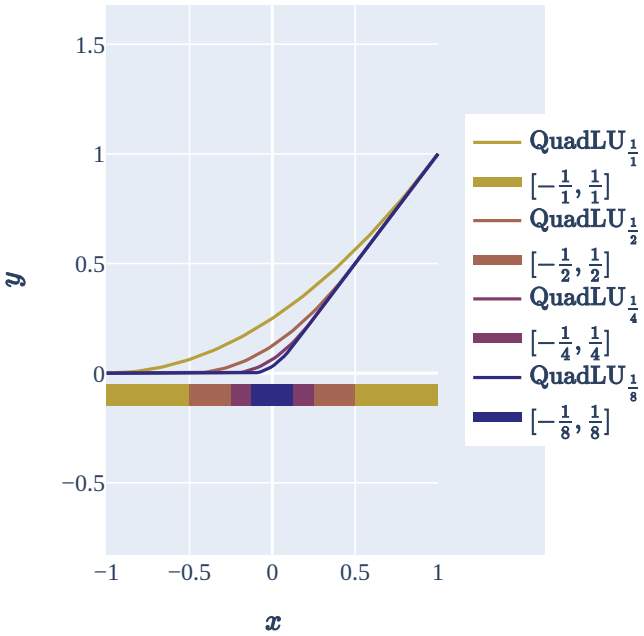

Abbildung 4: graph of QuadLU activation functions for $\alpha \in \{\frac{1}{8}, \frac{1}{4}, \frac{1}{2}, 1\}$ and $x \in [-1, 1]$ with corresponding regions of non-linearity

and the individual components of $\frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}z}$ given by

$$\frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}z}(z_k^{(i)}) = \begin{cases} 0, & \text{for } z_k^{(i)} \leq -\alpha \\ \frac{1}{2}\left(\frac{z_k^{(i)}}{\alpha} + 1\right), & \text{for } -\alpha < z_k^{(i)} < \alpha\,, \quad k = 1, \ldots, n^{(i)}. \\ 1, & \text{for } z_k^{(i)} \geq \alpha \end{cases} \tag{18}$$

Beweis. QuadLU_α is piecewise constant, linear or quadratic. Let $x \in \mathbb{R}$ and let $\alpha \in \mathbb{R}_+$ be fixed. First the differentiability in the corresponding open intervals is observed. At the transition points $\pm\alpha$ it is

$$\lim_{x \nearrow -\alpha} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}x}(x) = 0 = \frac{1}{2}\left(\frac{-\alpha}{\alpha} + 1\right) = \lim_{x \searrow -\alpha} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}x}(x),$$

and

$$\lim_{x \nearrow \alpha} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}x}(x) = \frac{1}{2}\left(\frac{\alpha}{\alpha} + 1\right) = 1 = \lim_{x \searrow \alpha} \frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}x}(x),$$

That shows that QuadLU_α is continuously differentiable on $\mathbb{R}$ with derivative

$$\frac{\mathrm{d}\,\mathrm{QuadLU}_\alpha}{\mathrm{d}x}(x) = \begin{cases} 0, & \text{for } x \leq -\alpha \\ \frac{1}{2}\left(\frac{x}{\alpha} + 1\right), & \text{for } -\alpha < x < \alpha\,, \quad i = 1, \ldots, n^{(i)}. \\ 1, & \text{for } x \geq \alpha \end{cases} \tag{19}$$

Applying Theorem A.5 concludes the proof. □

Literatur

Bingham, G., & Miikkulainen, R. (2022). Discovering Parametric Activation Functions [arXiv:2006.03179 [cs, stat]]. *Neural Networks*, *148*, 48–65. https://doi.org/10.1016/j.neunet.2022.01.001

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. https://www.deeplearningbook.org/

Horn, R. A., & Johnson, C. (2012). *Matrix analysis* (Second edition) [OCLC: 1151459555]. Cambridge University Press.

JCGM. (2008a). Evaluation of measurement data — Guide to the expression of uncertainty in measurement. https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6?version=1.7&download=true

JCGM. (2008b). Evaluation of measurement data — Supplement 1 to the "Guide to the expression of uncertainty in measurement" — Propagation of distributions using a Monte Carlo method. https://www.bipm.org/documents/20126/2071204/JCGM_101_2008_E.pdf/325dcaad-c15a-407c-1105-8b7f322d651c?version=1.5&download=true

JCGM. (2012). International Vocabulary of Metrology - Basic and general concepts and associated terms. Verfügbar 9. Oktober 2022 unter https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf/f0e1ad45-d337-bbeb-53a6-15fe649d0ff1?version=1.16&t=1659082802818&download=true

Ludwig, B. (2023a). *GUM-compliant neural network robustness verification* [Magisterarb., Technische Universität zu Berlin]. Berlin. https://gitlab1.ptb.de/ludwig10_masters_thesis/gum-compliant_neural_network_robustness_verification

Ludwig, B. (2023b). pytorch_gum_uncertainty_propagation. https://github.com/BjoernLudwigPTB/pytorch_gum_uncertainty_propagation

Physikalisch-Technische Bundesanstalt. (n. d.). Aufgaben und Ziele der Themenbereiche für 2023–2025. Verfügbar 8. Februar 2023 unter https://www.ptb.de/cms/fileadmin/internet/allgemeine_dokumente/Arbeits-_und_Forschungsprogramm/Aufgaben_und_Ziele_der_Themenbereiche_fuer_2023-2025.pdf

Szandała, T. (2021). Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks [Series Title: Studies in Computational Intelligence]. In A. K. Bhoi, P. K. Mallick, C.-M. Liu & V. E. Balas (Hrsg.), *Bio-inspired Neurocomputing* (S. 203–224, Bd. 903). Springer Singapore. https://doi.org/10.1007/978-981-15-5495-7_11

Xie, C., Tan, M., Gong, B., Yuille, A., & Le, Q. V. (2021). Smooth Adversarial Training [Number: arXiv:2006.14536 arXiv:2006.14536 [cs]]. Verfügbar 11. Juni 2022 unter http://arxiv.org/abs/2006.14536

Z.13. (2022). Gesetzliche Aufgaben der PTB. Verfügbar 8. Februar 2023 unter https://www.ptb.de/cms/fileadmin/internet/allgemeine_dokumente/Liste_der_gesetzlichen_Aufgaben_der_PTB.pdf

Zheng, H., Yang, Z., Liu, W., Liang, J., & Li, Y. (2015). Improving deep neural networks using softplus units. *2015 International Joint Conference on Neural Networks (IJCNN)*, 1–4. https://doi.org/10.1109/IJCNN.2015.7280459