Tutoraufgabe (Backpropagation für komplexwertige Netze)
In dieser Aufgabe sollen die Gleichungen (BP1) – (BP4) des Backpropagation-Algorithmus auf komplexwertige neuronale Netze verallgemeinert werden, d.h. w^ℓ, b^ℓ, z^ℓ, a^ℓ sind jetzt komplexe Matrizen bzw. Vektoren. Wir nehmen an, dass die Aktivierungsfunktion in einer offenen Menge $U \subset \mathbb{C}$ holomorph ist, was ja insbesondere auf die Sigmoid-Funktion $\sigma(z) = 1/(1+\mathrm{e}^{-z})$ zutrifft. Wie im reellen Fall verwenden wir die Hilfsgröße

$$\delta_j^\ell := \frac{\partial C}{\partial z_j^\ell},$$

wobei jetzt allerdings die Ableitungsoperatoren als *Wirtinger*- oder *Dolbeault*-Operatoren aufgefasst werden:

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

mit $z = x + iy$, $x, y \in \mathbb{R}$. Außerdem benötigen wir noch die folgende Kettenregel, die sich durch direktes Nachrechnen verifizieren lässt:

$$\frac{\partial}{\partial z}(g \circ f) = \left(\frac{\partial g}{\partial w} \circ f\right) \cdot \frac{\partial f}{\partial z} + \left(\frac{\partial g}{\partial \overline{w}} \circ f\right) \cdot \frac{\partial \overline{f}}{\partial z}.$$

Die Kostenfunktion C ist weiterhin reellwertig (so dass deren Minimierung überhaupt Sinn ergibt), und somit nicht komplex differenzierbar.

Hinweis: Die $\mathbb{C}$-lineare Multiplikation $w \mapsto z \cdot w$ mit einer vorgegebenen komplexen Zahl $z = x + iy$ lässt sich durch Identifikation $\mathbb{C} \cong \mathbb{R}^2$ auch als $\mathbb{R}$-lineare Abbildung auffassen. Anwendung auf $1 = (1, 0)$ und $i = (0, 1)$ ergibt die Matrixdarstellung

$$A_z = (z|iz) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}. \tag{1}$$

Umgekehrt ist eine reelle 2×2-Matrix $A = (z_1|z_2)$ mit Spaltenvektoren $z_1, z_2 \in \mathbb{C}$ also von der Form (1), wenn $z_2 = iz_1$ bzw. $z_1 + iz_2 = 0$.

Eine Funktion $f : U \subset \mathbb{C} \to \mathbb{C}$ ist bekanntlich genau dann in $z \in U$ komplex differenzierbar, wenn

$$f(w) = f(z) + f'(z) \cdot (w - z) + o(w - z) \qquad (w \to z). \tag{2}$$

Identifizieren wir $\mathbb{C} \cong \mathbb{R}^2$, dann ist f (als Funktion $U \subset \mathbb{R}^2 \to \mathbb{R}^2$) genau dann reell differenzierbar, wenn

$$f(w) = f(z) + Df(z) \cdot (w - z) + o(w - z) \qquad (w \to z).$$

Hierbei ist $Df(z) \in \mathbb{R}^{2\times 2}$ die Jakobi-Matrix. Die stärkere Bedingung “komplex differenzierbar” bedeutet also, dass das Matrix-Vektor-Produkt $Df(z) \cdot (w - z)$ als Multiplikation zweier komplexer Zahlen aufgefasst werden kann. Zerlegen wir $Df = (\partial_x f|\partial_y f)$, dann ist dies laut obiger Diskussion äquivalent zu den Cauchy-Riemann'schen Differentialgleichungen

$$\frac{\partial f}{\partial \overline{z}} \equiv \frac{1}{2}\left(\partial_x f + i\partial_y f\right) = 0,$$

wobei $\partial/\partial\overline{z}$ der (konjugierte) Wirtinger-Operator ist. Die komplexe Ableitung ist dann $f'(z) = \partial f(z)/\partial z$.

Als direkte Folgerung erhält man für f komplex differenzierbar in z:

$$\frac{\partial \overline{f}}{\partial z} = \frac{1}{2}\left(\partial_x \overline{f} - i\partial_y \overline{f}\right) = \overline{\frac{1}{2}\left(\partial_x f + i\partial_y f\right)} = \overline{\left(\frac{\partial f}{\partial \overline{z}}\right)} = 0. \tag{3}$$

(a) Zeigen Sie, dass (formal identisch zum reellwertigen Fall)

$$\delta_j^L = \frac{\partial C}{\partial a_j^L}\sigma'(z_j^L), \tag{BP1}$$

wobei $\partial/\partial a_j^L$ die Wirtinger-Ableitung bezüglich a_j^L ist und σ' die komplexe Ableitung der Aktivierungsfunktion. Wie lautet $\partial C/\partial a_j^L$ für die quadratische Kostenfunktion $C = \frac{1}{2}\|y - a^L\|^2$?

Lösungsvorschlag Anwenden der Wirtinger-Kettenregel liefert zunächst

$$\delta_j^L = \frac{\partial C}{\partial z_j^L} = \sum_k \left(\frac{\partial C}{\partial a_k^L}\frac{\partial a_k^L}{\partial z_j^L} + \frac{\partial C}{\partial \overline{a_k^L}}\frac{\partial \overline{a_k^L}}{\partial z_j^L}\right) = \frac{\partial C}{\partial a_j^L}\frac{\partial a_j^L}{\partial z_j^L} + \frac{\partial C}{\partial \overline{a_j^L}}\frac{\partial \overline{a_j^L}}{\partial z_j^L},$$

wobei wir für das zweite Gleichheitszeichen ausgenutzt haben, dass $a_k^L = \sigma(z_k^L)$ nur von z_k^L abhängt. Laut Voraussetzung ist die Aktivierungsfunktion σ holomorph, und damit

$$\frac{\partial a_j^L}{\partial z_j^L} = \sigma'(z_j^L)$$

die komplexe Ableitung von σ in z_j^L. Außerdem gilt laut Gl. (3), dass $\partial\overline{a_j^L}/\partial z_j^L = 0$. Insgesamt erhält man

$$\delta_j^L = \frac{\partial C}{\partial a_j^L}\sigma'(z_j^L).$$

Zur Berechnung von $\partial C/\partial a_j^L$ für die quadratische Kostenfunktion betrachten wir zunächst die Abbildung $z \in \mathbb{C} \mapsto \frac{1}{2}|z|^2 = \frac{1}{2}(x^2+y^2)$, wobei $z = x + iy$. Deren Wirtinger-Ableitung beträgt

$$\frac{\partial}{\partial z}\frac{1}{2}|z|^2 = \frac{1}{2}\left(\partial_x - i\partial_y\right)\frac{1}{2}\left(x^2+y^2\right) = \frac{1}{2}\left(x - iy\right) = \frac{1}{2}\overline{z}.$$

Ebenso berechnet man

$$\frac{\partial}{\partial a_j^L}\frac{1}{2}\left\|y - a^L\right\|^2 = \frac{\partial}{\partial a_j^L}\sum_k \frac{1}{2}\left|y_k - a_k^L\right|^2 = \frac{1}{2}\overline{(a_j^L - y_j)}.$$

(b) Leiten Sie (wiederum formal identisch zum reellwertigen Fall) den Zusammenhang

$$\delta^\ell = \operatorname{diag}\bigl(\sigma'(z^\ell)\bigr)\cdot\left(w^{\ell+1}\right)^T\cdot\delta^{\ell+1} \tag{BP2}$$

für alle $\ell < L$ her, wobei die Ableitung σ' komponentenweise an den Einträgen von z^ℓ ausgewertet wird.

Lösungsvorschlag Laut Wirtinger-Kettenregel gilt

$$\delta_j^\ell = \frac{\partial C}{\partial z_j^\ell} = \sum_k\left(\frac{\partial C}{\partial z_k^{\ell+1}}\frac{\partial z_k^{\ell+1}}{\partial z_j^\ell} + \frac{\partial C}{\partial \overline{z_k^{\ell+1}}}\frac{\partial \overline{z_k^{\ell+1}}}{\partial z_j^\ell}\right).$$

$z_k^{\ell+1}$ als Funktion von z_j^ℓ ist eine Verkettung holomorpher Abbildungen und damit wiederum holomorph. Somit gilt $\partial\overline{z_k^{\ell+1}}/\partial z_j^\ell = 0$ gemäß Gl. (3), d.h. der zweite Term in der Klammer verschwindet. Für den ersten Term erhalten wir

$$\frac{\partial z_k^{\ell+1}}{\partial z_j^\ell} = \frac{\partial}{\partial z_j^\ell}\left(\sum_m w_{km}^{\ell+1}\sigma(z_m^\ell) + b_k^{\ell+1}\right) = w_{kj}^{\ell+1}\sigma'(z_j^\ell)$$

und somit insgesamt

$$\delta_j^\ell = \sum_k \delta_k^{\ell+1}w_{kj}^{\ell+1}\sigma'(z_j^\ell).$$

Dies ist (BP2) in Komponentenform.

(c) Verifizieren Sie ebenso

$$\frac{\partial C}{\partial b_j^\ell} = \delta_j^\ell \tag{BP3}$$

sowie

$$\frac{\partial C}{\partial w_{jk}^\ell} = a_k^{\ell-1}\delta_j^\ell. \tag{BP4}$$

Lösungsvorschlag Argumentation wie in (b) und im reellwertigen Fall.

(d) Wie kann man nun den Gradienten der Kostenfunktion bezüglich des Real- und Imaginärteils von w_{jk}^ℓ bzw. b_j^ℓ erhalten?

Lösungsvorschlag Direkt aus der Definition der Wirtinger-Ableitungsoperatoren erhält man für eine *reellwertige* Funktion $f : \mathbb{C} \to \mathbb{R}$

$$\frac{\partial f}{\partial x} = 2\operatorname{Re}\left(\frac{\partial f}{\partial z}\right), \quad \frac{\partial f}{\partial y} = -2\operatorname{Im}\left(\frac{\partial f}{\partial z}\right),$$

also interpretiert als Funktion $f : \mathbb{R}^2 \to \mathbb{R}$

$$\nabla f = 2\overline{\left(\frac{\partial f}{\partial z}\right)}.$$

Anwenden auf (BP3) und (BP4) liefert den gesuchten Gradienten der Kostenfunktion.