Ableitungen von vektorwertigen Funktionen bzw. Matrizen

Im Folgenden sei $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ mit $\boldsymbol{x} \mapsto \begin{pmatrix} f_1(\boldsymbol{x}) & \cdots & f_m(\boldsymbol{x}) \end{pmatrix}^T$ eine vektorwertige Abbildung (Vektorfeld), definiere die Ableitung D von $\boldsymbol{f}$ (in beliebigem Punkt $\boldsymbol{x}$) als die Jacobimatrix:

$$(D\boldsymbol{f}(\boldsymbol{x})) := \left(\frac{\partial f_i(\boldsymbol{x})}{\partial x_j}\right)_{i=1,\dots,m;j=1,\dots,n} \in \mathbb{R}^{m\times n}$$

Übliche Ableitungsregeln sollen erhalten bleiben:
Der Übersicht halber schreiben wir jetzt $f := \boldsymbol{f}, g := \boldsymbol{g}$ und $x := \boldsymbol{x}$. Sei $f : \mathbb{R}^n \to \mathbb{R}^m$ und $\alpha, \beta \in \mathbb{R}$, betrachte folgende Anforderungen an die Ableitung:

1. **Linearität:**
$$D\left[\alpha \boldsymbol{f}(\boldsymbol{x}) + \beta \boldsymbol{g}(\boldsymbol{x})\right] = \alpha D\boldsymbol{f}(\boldsymbol{x}) + \beta D\boldsymbol{g}(\boldsymbol{x})$$
(nur falls $\boldsymbol{f}, \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$, da man dann zwei $m \times n$-Matrizen addiert).

2. **Kettenregel:**
$$D\left[\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x}))\right] = D\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \cdot D\boldsymbol{g}(\boldsymbol{x})$$
(nur falls $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ und $\boldsymbol{g} : \mathbb{R}^k \to \mathbb{R}^n$, da man dann $D\boldsymbol{f}(\boldsymbol{g}(\boldsymbol{x})) \in \mathbb{R}^{m\times n}$ mit $D\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^{n\times p}$ multipliziert; man kann beide Faktoren nicht vertauschen!).

3. **Produktregel:** (im Vektorfall)
$$D\left[\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{g}(\boldsymbol{x})\right] = \boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})^T D\boldsymbol{f}(\boldsymbol{x})$$
(nur falls $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$, da damit $\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}$ und $D(\boldsymbol{f}(\boldsymbol{x})^T\boldsymbol{g}(\boldsymbol{x})) \in \mathbb{R}^{1\times n}$ und $\boldsymbol{f}(\boldsymbol{x})^T D\boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^{1\times n}$)

4. **Produktregel:** (im Matrixfall)
Während die Ableitung von vektorwertigen Funktion nach einem Vektor intuitiv war, ist die Ableitung einer Matrixfunktion $\boldsymbol{A}(\boldsymbol{X})$ nach einer Matrix $\boldsymbol{X}$ etwas abstrakter. Um die Konsistenz zu wahren, liegt es nun nahe, dass man die Matrix $\boldsymbol{A}(\boldsymbol{X})$ mittels vec vektorisiert und dann nach $\text{vec}(\boldsymbol{X})$ ableitet:
$$D\left[\boldsymbol{A}(\boldsymbol{X})\right] := D\left[\text{vec}(\boldsymbol{A}(\boldsymbol{X})\right] := \frac{d\,\text{vec}(\boldsymbol{A}(\boldsymbol{X}))}{d\,\text{vec}(\boldsymbol{X})}$$
Seien also $\boldsymbol{A}(\boldsymbol{X}) \in \mathbb{R}^{m\times p}, \boldsymbol{B}(\boldsymbol{X}) \in \mathbb{R}^{p\times q}$ Matrizen abhängig von der Matrix $\boldsymbol{X}$, dann gilt: $\boldsymbol{A}(\boldsymbol{X}) \cdot \boldsymbol{B}(\boldsymbol{X}) \in \mathbb{R}^{m\times q}$. Seit weiter $I_n \in \mathbb{R}^{n\times n}$ die Einheitsmatrix dann gilt mit den Rechenregeln von vec und $\otimes$:
$$\text{vec}(\boldsymbol{I}_m \cdot \boldsymbol{A}(\boldsymbol{X}) \cdot \boldsymbol{B}(\boldsymbol{X}) \cdot \boldsymbol{I}_q) = (\boldsymbol{B}(\boldsymbol{X})^T \otimes \boldsymbol{I}_m) \cdot \text{vec}(\boldsymbol{A}(\boldsymbol{X})) = (\boldsymbol{I}_q \otimes \boldsymbol{A}(\boldsymbol{X})) \cdot \text{vec}(\boldsymbol{B}(\boldsymbol{X}))$$
Und somit wäre eine natürliche Produktregel:
$$D\left[\boldsymbol{A}(\boldsymbol{X}) \cdot \boldsymbol{B}(\boldsymbol{X})\right] = (\boldsymbol{B}(\boldsymbol{X})^T \otimes \boldsymbol{I}_m) \cdot D\boldsymbol{A}(\boldsymbol{X}) + (\boldsymbol{I}_q \otimes \boldsymbol{A}(\boldsymbol{X})) \cdot D\boldsymbol{B}(\boldsymbol{X})$$

Anwendungen der Matrixableitungen:
Sei wie oben wieder $\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{B} \in \mathbb{R}^{p\times q}, \boldsymbol{x} \in \mathbb{R}^n$ und $\boldsymbol{X}$ variabel, aber passend, dann gilt:

- $\frac{d\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}} = \boldsymbol{A}$
- $\frac{d\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{x}} = \boldsymbol{x}^T(\boldsymbol{A} + \boldsymbol{A}^T)$
- $\frac{d\,\text{vec}(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{d\,\text{vec}(\boldsymbol{A})} = \boldsymbol{x}^T \otimes \boldsymbol{x}^T$
- $\frac{d\boldsymbol{A}^T\boldsymbol{A}}{d\boldsymbol{A}} = (\boldsymbol{I}_{n^2} + \boldsymbol{T}_{n,n})(\boldsymbol{I}_n \otimes \boldsymbol{A}^T)$
- $\frac{d\boldsymbol{A}\boldsymbol{A}^T}{d\boldsymbol{A}} = (\boldsymbol{I}_{m^2} + \boldsymbol{T}_{m,m})(\boldsymbol{A} \otimes \boldsymbol{I}_m)$
- $\frac{d\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}}{d\boldsymbol{A}} = 2\boldsymbol{x}^T \otimes \boldsymbol{x}^T\boldsymbol{A}^T$
- $\frac{d\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{x}}{d\boldsymbol{A}} = 2\boldsymbol{x}^T\boldsymbol{A} \otimes \boldsymbol{x}^T$
- $\frac{d\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}}{d\boldsymbol{X}} = \boldsymbol{B}^T \otimes \boldsymbol{A}$
- $\frac{d\boldsymbol{A}^{-1}}{d\boldsymbol{A}} = -((\boldsymbol{A}^{-1})^T \otimes \boldsymbol{A}^{-1})$
- $\frac{d\log(\det(\boldsymbol{A})}{d\boldsymbol{A}} = \text{vec}((\boldsymbol{A}^{-1})^T)^T$
- $\frac{d\,\text{tr}(\boldsymbol{A}\boldsymbol{X})}{d\boldsymbol{X}} = \text{vec}(\boldsymbol{A}^T)^T$