Gradients of Matrices

- **Useful identities:**

$$\frac{\partial}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^\top = \left(\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right)^\top \tag{5.99}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \operatorname{tr}(\boldsymbol{f}(\boldsymbol{X})) = \operatorname{tr}\left(\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right) \tag{5.100}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \det(\boldsymbol{f}(\boldsymbol{X})) = \det(\boldsymbol{f}(\boldsymbol{X}))\operatorname{tr}\left(\boldsymbol{f}(\boldsymbol{X})^{-1}\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\right) \tag{5.101}$$

$$\frac{\partial}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^{-1} = -\boldsymbol{f}(\boldsymbol{X})^{-1}\frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}}\boldsymbol{f}(\boldsymbol{X})^{-1} \tag{5.102}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X}^{-1}\boldsymbol{b}}{\partial \boldsymbol{X}} = -(\boldsymbol{X}^{-1})^\top \boldsymbol{a}\boldsymbol{b}^\top (\boldsymbol{X}^{-1})^\top \tag{5.103}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{a}}{\partial \boldsymbol{x}} = \boldsymbol{a}^\top \tag{5.104}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}^\top \tag{5.105}$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X}\boldsymbol{b}}{\partial \boldsymbol{X}} = \boldsymbol{a}\boldsymbol{b}^\top \tag{5.106}$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{x}^\top(\boldsymbol{B}+\boldsymbol{B}^\top) \tag{5.107}$$

$$\frac{\partial}{\partial \boldsymbol{s}}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W}(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s}) = -2(\boldsymbol{x}-\boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W}\boldsymbol{A} \quad \text{for symmetric } \boldsymbol{W} \tag{5.108}$$

➔ **Note:**

Trace of a $D \times D \times E \times F$ tensor $\in \mathbb{R}^{E\times F}$

➔ $\mathbf{X} \in \mathbb{R}^{m\times n}$, $\mathbf{f}(\mathbf{X}) \in \mathbb{R}^{p\times p}$

Left term: $\frac{\partial}{\partial \mathbf{X}}\mathbf{f}(\mathbf{X})^{-1} \in \mathbb{R}^{p\times p\times m\times n}$

Right term:

$\mathbf{f}(\mathbf{X})^{-1} \in \mathbb{R}^{p\times p}$

$\frac{\partial \mathbf{f}(\mathbf{X})}{\partial \mathbf{X}} \in \mathbb{R}^{p\times p\times m\times n}$

For x_{ij}, $-\mathbf{f}(\mathbf{X})^{-1}\frac{\partial \mathbf{f}(\mathbf{X})}{\partial x_{ij}}\mathbf{f}(\mathbf{X})^{-1} \in \mathbb{R}^{p\times p}$

➔ $-\mathbf{f}(\mathbf{X})^{-1}\frac{\partial \mathbf{f}(\mathbf{X})}{\partial \mathbf{X}}\mathbf{f}(\mathbf{X})^{-1} \in \mathbb{R}^{p\times p\times m\times n}$

Gradients of Matrices

- **Clarification of some identities:**
  - $$\frac{\partial}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^{-1} = -\boldsymbol{f}(\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{f}(\boldsymbol{X})}{\partial \boldsymbol{X}} \boldsymbol{f}(\boldsymbol{X})^{-1} \qquad (5.102)$$
    - $$\mathbf{0} = \frac{\partial \mathbf{I}}{\partial \mathbf{X}} = \frac{\partial}{\partial \mathbf{X}} \left(\mathbf{f}(\mathbf{X})^{-1} \mathbf{f}(\mathbf{X}) \right) = \left(\frac{\partial}{\partial \mathbf{X}} \mathbf{f}(\mathbf{X})^{-1} \right) \mathbf{f}(\mathbf{X}) + \mathbf{f}(\mathbf{X})^{-1} \left(\frac{\partial}{\partial \mathbf{X}} \mathbf{f}(\mathbf{X}) \right)$$
  - $$\frac{\partial}{\partial \boldsymbol{s}} (\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W} (\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s}) = -2(\boldsymbol{x} - \boldsymbol{A}\boldsymbol{s})^\top \boldsymbol{W} \boldsymbol{A} \quad \text{for symmetric } \boldsymbol{W} \qquad (5.108)$$
    - Let $\mathbf{z} = \mathbf{x} - \mathbf{A}\mathbf{s}$, and use $\frac{\partial \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{x}^\top (\boldsymbol{B} + \boldsymbol{B}^\top)$:

    ➔ $\frac{\partial \mathbf{z}^\top \mathbf{W} \mathbf{z}}{\partial \mathbf{z}} = \mathbf{z}^\top (\mathbf{W} + \mathbf{W}^\top) = 2(\mathbf{x} - \mathbf{A}\mathbf{s})^\top \mathbf{W}$
    - Then compute $\frac{\partial \mathbf{z}^\top \mathbf{W} \mathbf{z}}{\partial \mathbf{z}} \frac{\partial \mathbf{z}}{\partial \mathbf{s}} = 2(\mathbf{x} - \mathbf{A}\mathbf{s})^\top \mathbf{W} \cdot (-\boldsymbol{A})$