Inverse and transpose of a matrix

If $\mathbf{A} \in \mathbb{R}^{n \times n}$ is invertible,

$$\left(\mathbf{A}^{-1}\right)^T = \left(\mathbf{A}^T\right)^{-1}$$

Determinant of a matrix

Absolute value of a matrix is nothing but the determinant of that matrix.

$$\det \mathbf{A} = |\mathbf{A}| = \prod_i \lambda_i \quad \lambda_i = \text{eig}(\mathbf{A})$$

Derivative of a scalar w.r.t. a vector

$$\nabla_{\mathbf{x}} y = \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}_n$$

$$\frac{\partial \mathbf{w}^T \mathbf{x}}{\partial \mathbf{w}} = \mathbf{x}$$

$$\frac{\partial \mathbf{y}^T \mathbf{y}}{\partial \mathbf{y}} = 2\mathbf{y}$$

$$\frac{\partial \left\|\mathbf{x} - \mathbf{h}\right\|^2}{\partial \mathbf{x}} = 2(\mathbf{x} - \mathbf{h})$$

Derivative of a scalar w.r.t. a matrix

$$\nabla_{\mathbf{A}} y = \begin{bmatrix} \frac{\partial y}{\partial a_{11}} & \frac{\partial y}{\partial a_{12}} & \cdots & \frac{\partial y}{\partial a_{1n}} \\ \frac{\partial y}{\partial a_{21}} & \frac{\partial y}{\partial a_{22}} & \cdots & \frac{\partial y}{\partial a_{2n}} \\ \cdots & \cdots & \ddots & \vdots \\ \frac{\partial y}{\partial a_{m1}} & \frac{\partial y}{\partial a_{m2}} & \cdots & \frac{\partial y}{\partial a_{mn}} \end{bmatrix}_{m \times n}$$

$$\frac{\partial y}{\partial \mathbf{W}^T} = (\frac{\partial y}{\partial \mathbf{W}})^T$$

$$\frac{\partial \left|\mathbf{X}\right|}{\partial \mathbf{X}} = \left|\mathbf{X}\right| (\mathbf{X}^{-1})^T = \left|\mathbf{X}\right| (\mathbf{X}^T)^{-1}$$

$$\frac{\partial \text{tr}(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}$$

Derivative of a vector w.r.t. a vector

$$\frac{\partial \mathbf{y}_n}{\partial \mathbf{x}_m} = \mathbf{J}^T = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_n}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_m} & \frac{\partial y_2}{\partial x_m} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix}_{m \times n}$$

$$\frac{\partial \mathbf{W}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{W}^T$$
$$\frac{\partial \mathbf{x}-\mathbf{h}}{\partial \mathbf{x}} = \mathbf{I}$$
$$\frac{\partial \mathbf{a}+\mathbf{c}}{\partial \mathbf{c}} = \mathbf{I}$$
$$\frac{\partial \mathrm{ReLU}(\mathbf{h})}{\partial \mathbf{h}} = \mathbb{I}(\mathbf{h} > 0)\odot$$

Scalar chain rule

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \quad z = g(y), y = f(x)$$
$$[f(g(x))]' = f'(g(x))g'(x)$$

Vector chain rule for vector valued functions

$$\frac{\partial \mathbf{z}}{\partial \mathbf{x}} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}}\frac{\partial \mathbf{z}}{\partial \mathbf{y}} \quad \mathbf{z} = g(\mathbf{y}), \mathbf{y} = f(\mathbf{x})$$

Derivative with respect to a tensor

If $\mathbf{X}\mathbf{Y} = \mathbf{Z} \longrightarrow \mathcal{L}$, and we have an upstream derivative $\partial\mathcal{L}/\partial\mathbf{Z}$:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}} = \frac{\partial \mathbf{Z}}{\partial \mathbf{X}}\frac{\partial \mathcal{L}}{\partial \mathbf{Z}} = \frac{\partial \mathcal{L}}{\partial \mathbf{Z}}\mathbf{Y}^T \quad (\mathbf{Y}^T\text{is on the right})$$
$$\frac{\partial \mathcal{L}}{\partial \mathbf{Y}} = \frac{\partial \mathbf{Z}}{\partial \mathbf{Y}}\frac{\partial \mathcal{L}}{\partial \mathbf{Z}} = \mathbf{X}^T\frac{\partial \mathcal{L}}{\partial \mathbf{Z}} \quad (\mathbf{X}^T\text{is on the left})$$

$\mathbf{W}\mathbf{x} = \mathbf{c} \longrightarrow \ell$, with $\mathbf{W} \in \mathcal{R}^{h\times m}, \mathbf{x} \in \mathcal{R}^m, \mathbf{c} \in \mathcal{R}^h$:

$$\underbrace{\frac{\partial \ell}{\partial \mathbf{x}}}_{m\times 1} = \underbrace{\frac{\partial \mathbf{c}}{\partial \mathbf{x}}}_{m\times h}\underbrace{\frac{\partial \ell}{\partial \mathbf{c}}}_{h\times 1} = \underbrace{\mathbf{W}^T}_{m\times h}\underbrace{\frac{\partial \ell}{\partial \mathbf{c}}}_{h\times 1} \quad (\mathbf{W}^T\text{is on the left})$$
$$\underbrace{\frac{\partial \ell}{\partial \mathbf{W}}}_{h\times m} = \underbrace{\frac{\partial \mathbf{c}}{\partial \mathbf{W}}}_{h\times m\times h}\underbrace{\frac{\partial \ell}{\partial \mathbf{c}}}_{h\times 1} = \underbrace{\frac{\partial \ell}{\partial \mathbf{c}}}_{h\times 1}\underbrace{\mathbf{x}^T}_{1\times m} \quad (\mathbf{x}^T\text{is on the right})$$

References

[1] Kaare Brandt Petersen; Michael Syskind Pedersen. The matrix cookbook. https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf.