Deriving some Matrix Cookbook identities

Linear algebra without matrix notation

Proving identities in (Petersen & Pedersen, 2008) using standard matrix notation can be cumbersome. It can be helpful to employ explicit tensor notation with a basis of vectors $\{\boldsymbol{e}_i\}$ and dual vectors $\{\boldsymbol{e}^i\}$. Dual vectors act on vectors as $\boldsymbol{e}^j(\boldsymbol{e}_i) = \delta_i^j$.

To agree with the standard meaning of juxtaposition as matrix multiplication, juxtaposing (dual) vectors means either *application* $\boldsymbol{e}^i\boldsymbol{e}_j = \boldsymbol{e}_i(\boldsymbol{e}_j)$ or the *tensor product* $\boldsymbol{e}_i\boldsymbol{e}^j = \boldsymbol{e}_i \otimes \boldsymbol{e}^j$ depending on the order. Note that $\boldsymbol{e}_i\boldsymbol{e}_j$ and $\boldsymbol{e}^i\boldsymbol{e}^j$ are left undefined (in the same way that row-row or column-column multiplications are undefined).

In tensorial notation, we have

$$\begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix} \equiv u^i\boldsymbol{e}_i, \quad (v_1 \ \cdots \ v_m) \equiv v_j\boldsymbol{e}^j, \quad \begin{pmatrix} A^1_{\ 1} & \cdots & A^1_{\ m} \\ \vdots & \ddots & \vdots \\ A^n_{\ 1} & \cdots & A^n_{\ m} \end{pmatrix} \equiv A^i_{\ j}\boldsymbol{e}_i\boldsymbol{e}^j.$$

With this scheme, matrix multiplication looks like

$$Ax \equiv A^i_{\ j}x^k\boldsymbol{e}_i\boldsymbol{e}^j\boldsymbol{e}_k = A^i_{\ j}x^k\boldsymbol{e}_i\delta^j_k = A^i_{\ j}x^j\boldsymbol{e}_i$$

with implicit summation over i, j, k.

For transposition, define $(\boldsymbol{e}_i)^T = \boldsymbol{e}^i$. If $a = a^i\boldsymbol{e}_i$ and $a^T = a_i\boldsymbol{e}^i$, then $a^i = a_i$.

The derivative of a function from matrices to scalars

Suppose $f : \mathbb{K}^{n\times m} \to \mathbb{K}$ is a scalar-valued function of matrices. The derivative $\partial f(X)/\partial X$ is understood to be the matrix whose ij component is the derivative of $f(X)$ with respect to the ij component of the input matrix X.

This can be expressed concretely as

$$\frac{\partial}{\partial X}f(X) \equiv \sum_{ij} \frac{\mathrm{d}}{\mathrm{d}t} f(X + t\boldsymbol{e}_i\boldsymbol{e}^j)\bigg|_{t=0} \boldsymbol{e}_i\boldsymbol{e}^j$$

where the matrix form of $\boldsymbol{e}_i\boldsymbol{e}^j$ is the matrix with ij component one and others zero.

Identities from the Matrix Cookbook

$$\boxed{\frac{\partial}{\partial X}(a^TXb) = ab^T}$$

$$\begin{aligned}
\frac{\partial}{\partial X}(a^TXb) &= \sum_{ij}\frac{\mathrm{d}}{\mathrm{d}t}a^T(X + t\boldsymbol{e}_i\boldsymbol{e}^j)b\bigg|_{t=0}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \sum_{ij}(a^T\boldsymbol{e}_i\boldsymbol{e}^jb)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \sum_{ij}(\boldsymbol{e}^ia)^T(\boldsymbol{e}^jb)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \sum_{ij}(a^i)(b^j)\boldsymbol{e}_i\boldsymbol{e}^j && \text{since } \boldsymbol{e}^ia = a^j\boldsymbol{e}^i(\boldsymbol{e}_j) = a^j\delta^i_j \\
&= a^ib_j\boldsymbol{e}_i\boldsymbol{e}^j && \text{since } b^j = b_j \\
&= a^i\boldsymbol{e}_ib_j\boldsymbol{e}^j \\
&= ab^T
\end{aligned}$$

$$\boxed{\frac{\partial}{\partial X}\operatorname{tr}(AXB) = A^TB^T}$$

$$\begin{aligned}
\frac{\partial}{\partial X}\operatorname{tr}(AXB) &= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\big(A(X + t\boldsymbol{e}_i\boldsymbol{e}^j)B\big)\bigg|_{t=0}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \operatorname{tr}(A\boldsymbol{e}_i\boldsymbol{e}^jB)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \operatorname{tr}(A^a_{\ b}\boldsymbol{e}_a\boldsymbol{e}^b\boldsymbol{e}_i\boldsymbol{e}^jB^c_{\ d}\boldsymbol{e}_c\boldsymbol{e}^d)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= A^a_{\ b}B^c_{\ d}\operatorname{tr}(\boldsymbol{e}_a\boldsymbol{e}^b\boldsymbol{e}_i\boldsymbol{e}^j\boldsymbol{e}_c\boldsymbol{e}^d)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= A^a_{\ b}B^c_{\ d}\delta^b_i\delta^j_c\operatorname{tr}(\boldsymbol{e}_a\boldsymbol{e}^d)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= A^a_{\ i}B^j_{\ d}\delta^d_a\boldsymbol{e}_i\boldsymbol{e}^j \\
&= B^j_{\ a}A^a_{\ i}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= (BA)^j_{\ i}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= ((BA)^T)^i_{\ j}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= A^TB^T
\end{aligned}$$

$$\boxed{\frac{\partial}{\partial X}\operatorname{tr}(X^2) = 2X^T}$$

$$\begin{aligned}
\frac{\partial}{\partial X}\operatorname{tr}(X^2) &= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\Big((X + t\boldsymbol{e}_i\boldsymbol{e}^j)^2\Big)\bigg|_{t=0}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\big(X^2 + tX\boldsymbol{e}_i\boldsymbol{e}^j + t\boldsymbol{e}_i\boldsymbol{e}^jX + \mathcal{O}(t^2)\big)\bigg|_{t=0}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= \operatorname{tr}(X\boldsymbol{e}_i\boldsymbol{e}^j + \boldsymbol{e}_i\boldsymbol{e}^jX)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2X^a_{\ b}\operatorname{tr}(\boldsymbol{e}_a\boldsymbol{e}^b\boldsymbol{e}_i\boldsymbol{e}^j)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2X^a_{\ b}\delta^b_i\operatorname{tr}(\boldsymbol{e}_a\boldsymbol{e}^j)\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2X^a_{\ i}\delta^j_a\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2X^j_{\ i}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2(X^T)^i_{\ j}\boldsymbol{e}_i\boldsymbol{e}^j \\
&= 2X^T
\end{aligned}$$

References

Petersen, K. B., & Pedersen, M. S. (2008,). *The Matrix Cookbook*. Technical University of Denmark.