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 $\{e_i\}$ and dual vectors $\{e^i\}$. Dual vectors act on vectors as $e^j(e_i) = \delta^j_i$.

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

In tensorial notation, we have

$$\begin{pmatrix} u^1 \\ \vdots \\ u^n \end{pmatrix} \equiv u^i e_i, \quad \begin{pmatrix} v_1 & \cdots & v_m \end{pmatrix} \equiv v_j e^j, \quad \begin{pmatrix} A^1{}_1 & \cdots & A^1{}_m \\ \vdots & \ddots & \vdots \\ A^n{}_1 & \cdots & A^n{}_m \end{pmatrix} \equiv A^i{}_j e_i e^j.$$

With this scheme, matrix multiplication looks like

$$Ax \equiv A^i{}_j x^k e_i e^j e_k = A^i{}_j x^k e_i \delta^j_k = A^i{}_j x^j e_i$$

with implicit summation over i, j, k.

For transposition, define $(e_i)^T = e^i$. If $a = a^i e_i$ and $a^T = a_i e^i$, 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

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

where the matrix form of $e_i e^j$ is the matrix with ij component one and others zero.

Identities from the Matrix Cookbook

$$\frac{\partial}{\partial X}(a^T X b) = a b^T$$

$$\begin{aligned}
\frac{\partial}{\partial X}(a^T X b) &= \sum_{ij} \frac{\mathrm{d}}{\mathrm{d}t} a^T (X + t e_i e^j) b\bigg|_{t=0} e_i e^j \\
&= \sum_{ij} (a^T e_i e^j b) e_i e^j \\
&= \sum_{ij} (e^i a)^T (e^j b) e_i e^j \\
&= \sum_{ij} (a^i)(b^j) e_i e^j && \text{since } e^i a = a^j e^i(e_j) = a^j \delta^i_j \\
&= a^i b_j e_i e^j && \text{since } b^j = b_j \\
&= a^i e_i b_j e^j \\
&= a b^T
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial}{\partial X}\operatorname{tr}(AXB) &= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\big(A(X + t e_i e^j)B\big)\bigg|_{t=0} e_i e^j \\
&= \operatorname{tr}(A e_i e^j B) e_i e^j \\
&= \operatorname{tr}(A^a{}_b e_a e^b e_i e^j B^c{}_d e_c e^d) e_i e^j \\
&= A^a{}_b B^c{}_d \operatorname{tr}(e_a e^b e_i e^j e_c e^d) e_i e^j \\
&= A^a{}_b B^c{}_d \delta^b_i \delta^j_c \operatorname{tr}(e_a e^d) e_i e^j \\
&= A^a{}_i B^j{}_d \delta^d_a e_i e^j \\
&= B^j{}_a A^a{}_i e_i e^j \\
&= (BA)^j{}_i e_i e^j \\
&= ((BA)^T)^i{}_j e_i e^j \\
&= A^T B^T
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial}{\partial X}\operatorname{tr}(X^2) &= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\Big((X + t e_i e^j)^2\Big)\bigg|_{t=0} e_i e^j \\
&= \frac{\mathrm{d}}{\mathrm{d}t}\operatorname{tr}\big(X^2 + tX e_i e^j + t e_i e^j X + \mathcal{O}(t^2)\big)\bigg|_{t=0} e_i e^j \\
&= \operatorname{tr}(X e_i e^j + e_i e^j X) e_i e^j \\
&= 2X^a{}_b \operatorname{tr}(e_a e^b e_i e^j) e_i e^j \\
&= 2X^a{}_b \delta^b_i \operatorname{tr}(e_a e^j) e_i e^j \\
&= 2X^a{}_i \delta^j_a e_i e^j \\
&= 2X^j{}_i e_i e^j \\
&= 2(X^T)^i{}_j e_i e^j \\
&= 2X^T
\end{aligned}$$

References

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