Lecture 2 notes

1 Introduction

We prove some assertions made in lecture 2 that were left as exercises.

2 Trace properties

Definition. *The trace of a square matrix A is defined as*

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii}. \tag{1}$$

Theorem. *Let $f(A) = \mathrm{tr}(AB)$, where A and B are square matrices. Then*

$$\nabla_A f(A) = B^T. \tag{2}$$

Proof. We have

$$\begin{aligned}
\nabla_A f(A) &= \nabla_A \mathrm{tr}(AB) && (3)\\
&= \nabla_A \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} && (4)\\
&= \begin{bmatrix} \frac{\partial}{\partial a_{11}} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} & \cdots & \frac{\partial}{\partial a_{1n}} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial a_{n1}} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} & \cdots & \frac{\partial}{\partial a_{nn}} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} \end{bmatrix} && (5)\\
&= \begin{bmatrix} \frac{\partial}{\partial a_{11}} a_{11}b_{11} + \cdots + \frac{\partial}{\partial a_{11}} a_{n1}b_{1n} & \cdots & \frac{\partial}{\partial a_{1n}} a_{11}b_{11} + \cdots + \frac{\partial}{\partial a_{1n}} a_{n1}b_{1n} \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial a_{n1}} a_{11}b_{11} + \cdots + \frac{\partial}{\partial a_{n1}} a_{n1}b_{1n} & \cdots & \frac{\partial}{\partial a_{nn}} a_{11}b_{11} + \cdots + \frac{\partial}{\partial a_{nn}} a_{n1}b_{1n} \end{bmatrix} && (6)\\
&= \begin{bmatrix} b_{11} & \cdots & b_{n1} \\ \vdots & \ddots & \vdots \\ b_{1n} & \cdots & b_{nn} \end{bmatrix} && (7)\\
&= B^T. && (8)
\end{aligned}$$

□

Remark. *Writing out proofs in this manner can become tedious. As a shorthand, we could have considered the matrix $\nabla_A f(A)$ at an individual element, say a_{mn}, and observe that $\nabla_A f(A)_{mn} = \frac{\partial}{\partial a_{mn}} f(A) = \frac{\partial}{\partial a_{mn}} \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} = b_{nm}$ and conclude that $\nabla_A f(A) = B^T$. This style of argumentation shall be used going forward.*

Theorem. *Let A and B be square matrices. Then*

$$\operatorname{tr}(AB) = \operatorname{tr}(BA). \tag{9}$$

Proof. We have

$$\begin{aligned}
\operatorname{tr}(AB) &= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}b_{ji} && (10)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ji}a_{ij} && (11)\\
&= \operatorname{tr}(BA). && (12)
\end{aligned}$$

All we had to do was swap the order of summation on line 11, which is valid because the order of summation for finite sums does not matter. □

Theorem. *Let A, B and C be square matrices. Then*

$$\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB). \tag{13}$$

Proof. We have

$$\begin{aligned}
\operatorname{tr}(ABC) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}b_{jk}c_{ki} && (14)\\
&= \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{i=1}^{n} b_{jk}c_{ki}a_{ij} && (15)\\
&= \operatorname{tr}(BCA) && (16)
\end{aligned}$$

and a similar argument shows that $\operatorname{tr}(CAB) = \operatorname{tr}(BCA)$. □

Theorem. *Let A and C be square matrices. Then*

$$\nabla_A \operatorname{tr}(AA^TC) = CA + C^TA. \tag{17}$$

Proof. Consider the space of $n \times n$ matricesand endow it with the inner product $\langle A, B\rangle = \operatorname{tr}(AB^T)$. This is indeed an inner product because, first and foremost it is symmetric:

$$\begin{aligned}
\langle A, B\rangle &= \operatorname{tr}(AB^T) && (18)\\
&= \operatorname{tr}(BA^T) && (19)\\
&= \langle B, A\rangle && (20)
\end{aligned}$$

It is also linear in the first argument:

$$\begin{aligned}
\langle \lambda A + \mu B, C\rangle &= \operatorname{tr}((\lambda A + \mu B)C^T) && (21)\\
&= \lambda \operatorname{tr}(AC^T) + \mu \operatorname{tr}(BC^T) && (22)\\
&= \lambda\langle A, C\rangle + \mu\langle B, C\rangle && (23)
\end{aligned}$$

Furthermore, it is positive definite:

$$\begin{aligned}
\langle A, A\rangle &= \operatorname{tr}(AA^T) && (24)\\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ij} && (25)\\
&\geq 0 && (26)
\end{aligned}$$

For equality to be achieved on line 26, we must have $a_{ij} = 0$ for all i and j, which is when A is the zero matrix. This completes the proof that $\langle A, B \rangle$ is an inner product on the space of $n \times n$ matrices.

Let $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ be defined by $f(A) = AA^T$ and $g : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ be defined by $g(A) = A^T$. Then $\nabla_A \mathrm{tr}(AA^T C) = \nabla_A \mathrm{tr}(CAA^T)$ by invariance of trace under cyclic permutations as previously established, and using the new notation $\nabla_A \mathrm{tr}(CAA^T) = \nabla_A \langle f(A), g(A) \rangle$.

The reason why we are talking about inner products is because the result being asked to be shown holds true in a more general setting. In particular, if f and g are two linear maps from a vector space V to itself, and $\langle \cdot, \cdot \rangle$ is an inner product on V, then $\nabla_v \langle f(v), g(v) \rangle = \langle \nabla_v f(v), g(v) \rangle + \langle f(v), \nabla_v g(v) \rangle$.

$\square$