Hessian of trace of some matrix functions

Xun Zheng
xzheng1@andrew.cmu.edu

1 Hessian of trace of some matrix functions

Let $F : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ be some matrix function and $\phi : \mathbb{R}^{n\times n} \to \mathbb{R}$ be its trace

$$\phi(X) = \operatorname{tr}\big(F(X)\big). \tag{1}$$

We are interested in the Hessian of ϕ:

$$H\phi(X) = \frac{\partial^2 \phi(X)}{\partial \operatorname{vec}(X) \partial \operatorname{vec}(X)^T}. \tag{2}$$

In the derivation, the following matrix serves as a handy tool to transform between $\operatorname{vec}(A)$ and $\operatorname{vec}(A^T)$.

Definition 1 (Commutation matrix) *A matrix K_{nm} is a commutation matrix if it satisfies*

$$K_{nm} \operatorname{vec}(A) = \operatorname{vec}(A^T) \tag{3}$$

for any $n \times m$ matrix A.

We will make use of matrix differential calculus to derive the Hessian.

Proposition 1 (Matrix power) *If $F(X) = X^k$ for some integer $k \geq 2$, the Hessian of ϕ is given by*

$$H\phi(X) = K_{nn}\big[k \sum_{r=0}^{k-2} \big(X^r\big)^T \otimes X^{k-2-r}\big]. \tag{4}$$

Proof: The first differential is

$$\mathrm{d}\phi(X) = \mathrm{d} \operatorname{tr}\big(X^k\big) = \operatorname{tr}\big(\mathrm{d}X^k\big) = \operatorname{tr}\big(kX^{k-1}\,\mathrm{d}X\big). \tag{5}$$

Then the second differential is

$$\begin{aligned}
\mathrm{d}^2\phi(X) &= \mathrm{d} \operatorname{tr}\big(kX^{k-1}\,\mathrm{d}X\big) && (6)\\
&= k \cdot \operatorname{tr}\big(\mathrm{d}\big(X^{k-1}\,\mathrm{d}X\big)\big) && (7)\\
&= k \cdot \operatorname{tr}\big((\mathrm{d}X)X^{k-2}(\mathrm{d}X) + X(\mathrm{d}X)X^{k-3}(\mathrm{d}X) + \cdots + X^{k-2}(\mathrm{d}X)(\mathrm{d}X)\big) && (8)\\
&= k \cdot \sum_{r=0}^{k-2} \operatorname{tr}\big(X^r(\mathrm{d}X)X^{k-2-r}(\mathrm{d}X)\big). && (9)
\end{aligned}$$

Using (Magnus and Neudecker, 2007, Theorem 10.1), we have

$$H\phi(X) = k \cdot \sum_{r=0}^{k-2} \frac{1}{2} K_{nn}\big[\big(X^r\big)^T \otimes X^{k-2-r} + \big(X^{k-2-r}\big)^T \otimes X^r\big]. \tag{10}$$

Result follows by noticing that the sum of two Kronecker products over $r = 0, \ldots, k-2$ are identical. ∎

Proposition 2 (Matrix exponential) *If $F(X) = e^X$, the Hessian of ϕ is given by*

$$H\phi(X) = K_{nn} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{1}{(p+q+1)!} \left[\left(X^p\right)^T \otimes X^q \right]. \tag{11}$$

Proof: The second differential is

$$\begin{aligned} \mathrm{d}^2 \phi(X) &= \sum_{k=2}^{\infty} \frac{1}{k!} \mathrm{d}^2 \operatorname{tr}(X^k) & (12) \\ &= \sum_{k=2}^{\infty} \frac{1}{(k-1)!} \sum_{r=0}^{k-2} \operatorname{tr} \left(X^r (\mathrm{d}X) X^{k-2-r} (\mathrm{d}X) \right) & (13) \\ &= \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{1}{(p+q+1)!} \operatorname{tr} \left(X^p (\mathrm{d}X) X^q (\mathrm{d}X) \right). & (14) \end{aligned}$$

Result follows by (Magnus and Neudecker, 2007, Theorem 10.1). ■

Now consider when element-wise product is involved.

Proposition 3 (Matrix power of Hadamard square) *If $F(X) = (X \circ X)^k$, the Hessian of ϕ is*

$$\begin{aligned} H\phi(X) = K_{nn} \operatorname{diag}(\operatorname{vec} 2X^T) \big[k \sum_{r=0}^{k-2} \left[(X \circ X)^r \right]^T \otimes (X \circ X)^{k-2-r} \big] \operatorname{diag}(\operatorname{vec} 2X) & \quad (15) \\ + 2k \operatorname{diag} \left(\operatorname{vec} \left((X \circ X)^{k-1} \right)^T \right). & \quad (16) \end{aligned}$$

Proof: The first differential is

$$\begin{aligned} \mathrm{d}\phi(X) &= \operatorname{tr} \left(\mathrm{d}(X \circ X)^k \right) & (17) \\ &= \operatorname{tr} \left(k (X \circ X)^{k-1} \mathrm{d}(X \circ X) \right) & (18) \\ &= \operatorname{tr} \left(k (X \circ X)^{k-1} (2X \circ \mathrm{d}X) \right) & (19) \\ &= \operatorname{tr} \left(\left(k (X \circ X)^{k-1} \circ 2X^T \right) \mathrm{d}X \right), & (20) \end{aligned}$$

where last line is from (Magnus and Neudecker, 2007, Theorem 3.7).

The second differential is then

$$\begin{aligned} \mathrm{d}^2 \phi(X) &= \operatorname{tr} \left(\mathrm{d} \big(k (X \circ X)^{k-1} \circ 2X^T \big) \mathrm{d}X \right) & (21) \\ &= \operatorname{tr} \left(\left[\mathrm{d} \big(k (X \circ X)^{k-1} \big) \circ 2X^T + k (X \circ X)^{k-1} \circ \mathrm{d}(2X^T) \right] \mathrm{d}X \right). & (22) \end{aligned}$$

The first term in the trace:

$$\begin{aligned} \text{first} &= \operatorname{tr} \left(\left[\mathrm{d} \big(k (X \circ X)^{k-1} \big) \circ 2X^T \right] \mathrm{d}X \right) & (23) \\ &= \operatorname{tr} \left(\left[\big(k \sum_{r=0}^{k-2} (X \circ X)^r \, \mathrm{d}(X \circ X) (X \circ X)^{k-2-r} \big) \circ 2X^T \right] \mathrm{d}X \right) & (24) \\ &= \operatorname{tr} \left(\left[\big(k \sum_{r=0}^{k-2} (X \circ X)^r (2X \circ \mathrm{d}X) (X \circ X)^{k-2-r} \big) \circ 2X^T \right] \mathrm{d}X \right) & (25) \\ &= k \sum_{r=0}^{k-2} \operatorname{tr} \left(\left((X \circ X)^r (2X \circ \mathrm{d}X) (X \circ X)^{k-2-r} \circ 2X^T \right) \mathrm{d}X \right). & (26) \end{aligned}$$

Using (Magnus and Neudecker, 2007, Theorem 3.7),

$$\text{first} = k \sum_{r=0}^{k-2} \operatorname{tr}\big((X \circ X)^r (2X \circ \mathrm{d}X)(X \circ X)^{k-2-r}(2X \circ \mathrm{d}X)\big). \tag{27}$$

Using (Magnus and Neudecker, 2007, Theorem 2.3),

$$\text{first} = k \sum_{r=0}^{k-2} \big[\operatorname{vec}(2X \circ \mathrm{d}X)^T\big]^T \big[(X \circ X)^r\big]^T \otimes (X \circ X)^{k-2-r} \operatorname{vec}(2X \circ \mathrm{d}X). \tag{28}$$

Since $\operatorname{vec}(A \circ B) = \operatorname{diag}(\operatorname{vec} A) \operatorname{vec} B$,

$$\text{first} = k \sum_{r=0}^{k-2} (\mathrm{d}\operatorname{vec} X)^T K_{nn} \operatorname{diag}(\operatorname{vec} 2X^T)\big[(X \circ X)^r\big]^T \otimes (X \circ X)^{k-2-r} \operatorname{diag}(\operatorname{vec} 2X)(\mathrm{d}\operatorname{vec} X). \tag{29}$$

The second term in the trace:

$$\text{second} = \operatorname{tr}\big(\big[k(X \circ X)^{k-1} \circ \mathrm{d}(2X^T)\big]\, \mathrm{d}X\big) \tag{30}$$

$$= \Big(\operatorname{vec}\big(k(X \circ X)^{k-1} \circ \mathrm{d}(2X^T)\big)^T\Big)^T \mathrm{d}\operatorname{vec} X \tag{31}$$

$$= \Big(\operatorname{vec}\big(k\big((X \circ X)^{k-1}\big)^T \circ 2\,\mathrm{d}X\big)\Big)^T \mathrm{d}\operatorname{vec} X \tag{32}$$

$$= \Big(\operatorname{diag}\big(\operatorname{vec}\big[k\big((X \circ X)^{k-1}\big)^T\big]\big) 2\,\mathrm{d}\operatorname{vec} X\Big)^T \mathrm{d}\operatorname{vec} X \tag{33}$$

$$= (\mathrm{d}\operatorname{vec} X)^T 2k \operatorname{diag}\Big(\operatorname{vec}\big((X \circ X)^{k-1}\big)^T\Big)(\mathrm{d}\operatorname{vec} X). \tag{34}$$

Therefore the second differential is

$$\mathrm{d}^2\phi(X) = (\mathrm{d}\operatorname{vec} X)^T \Big\{ K_{nn} \operatorname{diag}(\operatorname{vec} 2X^T)\big[k \sum_{r=0}^{k-2} \big[(X \circ X)^r\big]^T \otimes (X \circ X)^{k-2-r}\big] \operatorname{diag}(\operatorname{vec} 2X) \tag{35}$$

$$+2k \operatorname{diag}\Big(\operatorname{vec}\big((X \circ X)^{k-1}\big)^T\Big)\Big\}(\mathrm{d}\operatorname{vec} X). \tag{36}$$

This is in the canonical form as in (Magnus and Neudecker, 2007, Table 10.1). The inner matrix is already symmetric, therefore it is the Hessian. ∎

Proposition 4 (Matrix exponential of Hadamard square) *If $F(X) = e^{X \circ X}$, the Hessian of ϕ is*

$$H\phi(X) = K_{nn} \operatorname{diag}(\operatorname{vec} 2X^T)\big[\sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \frac{1}{(p+q+1)!} \big[(X \circ X)^p\big]^T \otimes (X \circ X)^q\big] \operatorname{diag}(\operatorname{vec} 2X) \tag{37}$$

$$+\, 2 \operatorname{diag}\Big(\operatorname{vec}\big(e^{X \circ X}\big)^T\Big). \tag{38}$$

Proof: The second differential is

$$
\begin{aligned}
\mathrm{d}^2\phi(X) &= \sum_{k=2}^{\infty} \frac{1}{k!}\,\mathrm{d}^2\,\mathrm{tr}((X\circ X)^k) && (39)\\
&= \sum_{k=2}^{\infty} \frac{1}{(k-1)!}(\mathrm{d}\,\mathrm{vec}\,X)^T \times && (40)\\
&\quad \Big\{ K_{nn}\,\mathrm{diag}(\mathrm{vec}\,2X^T)\big[\sum_{r=0}^{k-2}\big[(X\circ X)^r\big]^T \otimes (X\circ X)^{k-2-r}\big]\,\mathrm{diag}(\mathrm{vec}\,2X)\Big\}(\mathrm{d}\,\mathrm{vec}\,X) && (41)\\
&\quad + \sum_{k=1}^{\infty}\frac{1}{(k-1)!}(\mathrm{d}\,\mathrm{vec}\,X)^T 2\,\mathrm{diag}\Big(\mathrm{vec}\big((X\circ X)^{k-1}\big)^T\Big)(\mathrm{d}\,\mathrm{vec}\,X) && (42)\\
&= (\mathrm{d}\,\mathrm{vec}\,X)^T \times && (43)\\
&\quad \Big\{ K_{nn}\,\mathrm{diag}(\mathrm{vec}\,2X^T)\big[\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{1}{(p+q+1)!}\big[(X\circ X)^p\big]^T\otimes(X\circ X)^q\big]\,\mathrm{diag}(\mathrm{vec}\,2X) && (44)\\
&\qquad + 2\,\mathrm{diag}\Big(\mathrm{vec}\big(\sum_{k=0}^{\infty}\frac{1}{k!}(X\circ X)^k\big)^T\Big)\Big\}(\mathrm{d}\,\mathrm{vec}\,X) && (45)\\
&= (\mathrm{d}\,\mathrm{vec}\,X)^T \times && (46)\\
&\quad \Big\{ K_{nn}\,\mathrm{diag}(\mathrm{vec}\,2X^T)\big[\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{1}{(p+q+1)!}\big[(X\circ X)^p\big]^T\otimes(X\circ X)^q\big]\,\mathrm{diag}(\mathrm{vec}\,2X) && (47)\\
&\qquad + 2\,\mathrm{diag}\Big(\mathrm{vec}\big(e^{X\circ X}\big)^T\Big)\Big\}(\mathrm{d}\,\mathrm{vec}\,X). && (48)
\end{aligned}
$$

This is in the canonical form as in (Magnus and Neudecker, 2007, Table 10.1). The inner matrix is already symmetric, therefore it is the Hessian. ■

References

J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. 2007.